UNIVERSITE DE LIEGE
Faculté d'Economie, de Gestion et de Sciences Sociales

et

UNIVERSITE PARIS XII VAL DE MARNE
Faculté de Sciences Economiques et de Gestion

Error components models and variable heterogeneity : modelisation, second order pseudo-maximum likelihood estimation and specification testing

Membres du jury :

L. Broze
Y. Langaskens
P. Sevestre
B. Thiry
A. Trognon
J-P. Urbain

Dissertation présentée par
Bernard LEJEUNE
pour l'obtention du grade de
Docteur en Science Economique

Année Académique 1997 - 1998

Remerciements

Je voudrais ici remercier toutes les personnes qui, de près ou de loin, par leur soutien, leurs commentaires et remarques, leurs encouragements ou plus simplement leur amitié, ont permis la réalisation du présent travail.

Plus particulièrement, je voudrais exprimer ma gratitude à Bernard Thiry pour m'avoir offert l'opportunité, la liberté intellectuelle et le temps de mener à bien cette dissertation. Je souhaiterais également exprimer ma reconnaissance à Patrick Sevestre pour avoir accepté de co-diriger cette thèse, ses conseils et commentaires toujours avisés, et son accueil au cours de mes divers séjours à l'ERUDITE.

Ce travail doit beaucoup au soutien, aux remarques et suggestions, et à la disponibilité de Laurence Broze. Nos fréquentes et fructueuses séances de travail au CORE auront été pour moi un stimulant irremplaçable, tant scientifiquement que moralement.

Je voudrais encore remercier Alain Trognon et Jean-Pierre Urbain pour leurs commentaires et encouragements répétés. Je souhaiterais aussi les remercier, ainsi que Yvan Langaskens, pour avoir accepté d'être membre du jury.

Diverses parties et versions préliminaires de ce travail ont également bénéficié des remarques et commentaires de Frédéric Jouneau, Kees Jan Van Garderen, James Lindsey, Olivier Donni, Philippe Lambert et Badi Baltagi. Barbara Sak m'a aidé dans la chasse aux coquilles typographiques. Je tiens à leur adresser à tous mes remerciements.

Résumé

Ce travail est consacré à l'étude de quelques problèmes économétriques associés à la modélisation de l'hétérogénéité des comportements individuels lorsque l'on travaille avec des données microéconomiques en panel. Plus précisément, il poursuit un double objectif : d'une part, proposer et discuter une extension du modèle à erreurs composées standard permettant de prendre en compte et de rendre compte de phénomènes d'hétérogénéité individuelle variables, et d'autre part, fournir pour l'estimation et la mise à l'épreuve de la spécification du modèle proposé un ensemble cohérent de procédures d'estimation et de tests prenant explicitement en compte une possible mauvaise spécification des moments d'ordre 2, c'est-à-dire de la forme d'hétérogénéité modélisée. Il est composé de quatre chapitres.

Dans un cadre qui dépasse largement — mais inclut comme cas particulier — les modèles à erreurs composées, le premier chapitre étudie la robustesse à une mauvaise spécification de la variance des estimateurs de type pseudo-maximum de vraisemblance au deuxième ordre, c'est-à-dire d'une classe d'estimateurs estimant conjointement, au travers de la maximisation d'une fonction de pseudo-vraisemblance, les paramètres de la moyenne et de la variance d'un modèle semi-paramétrique à l'ordre 2. On montre que des conditions nécessaires et suffisantes pour que ce type d'estimateur soit robuste à une mauvaise spécification de la variance sont (1) que les paramètres de la moyenne et de la variance varient indépendamment et (2) que les pseudo-densités utilisées pour former la fonction de pseudo-vraisemblance appartiennent à une famille particulière de distributions que nous avons appelée exponentielle quadratique restreinte. Les propriétés asymptotiques — convergence et distribution — de cette classe d'estimateurs robustes sont étudiées sous différentes hypothèses quant à l'importance de la mauvaise spécification présente dans le modèle.

Traité dans le même cadre général que le Chapitre 1, le second chapitre décrit comment, à partir d'un estimateur robuste tel que celui évoqué ci-dessus, tirer parti de l'approche 'm-test' / 'm-test' modifiée de Wooldridge pour tester, avec ou sans hypothèse alternative clairement définie, la spécification des modèles semi-paramétriques à l'ordre 2. On s'intéresse prioritairement aux hypothèses nulles de spécification correcte de la moyenne conditionnelle et de spécification correcte de la variance conditionnelle. Tant pour la moyenne que pour la variance, on montre comment mettre en oeuvre des tests de type 'Hausman', de type 'matrice d'information' ainsi que des tests contre des hypothèses alternatives auxiliaires emboîtées ou non-emboîtées. On s'intéresse également à des tests du caractère dynamiquement complet ou non des spécifications de la moyenne conditionnelle et de la variance conditionnelle. Dans tous les cas, les hypothèses maintenues des tests sont clairement précisées et réduites au minimum, de sorte que la validité des statistiques de tests proposées ne requiert généralement guère plus que l'hypothèse nulle testée.

Armé des outils statistiques généraux dérivés dans les deux premiers chapitres, le troisième chapitre revient à notre objectif initial : proposer et discuter une extension du modèle à erreurs composées standard permettant de prendre en compte et de rendre compte de phénomènes d'hétérogénéité individuelle variables. L'idée centrale en est simplement de "paramétriser" l'hétérogénéité, en d'autres termes, de faire dépendre d'un certain nombre de variables explicatives les paramètres représentatifs de l'hétérogénéité — variance du terme d'erreur général et variance du terme d'erreur individuel — dans le modèle standard. Cela signifie adopter pour les moments d'ordre deux une paramétrisation a priori assez flexible et intuitivement attractive, permettant à une hétérogénéité variable de se manifester tant dans la dimension 'intra' que 'inter'. Arguant de sa capacité à s'accommoder sans difficultés de données non-calibrées ("unbalanced panel"), de sa robustesse à une mauvaise spécification de l'hétérogénéité, de sa possible efficacité et de la commodité de son calcul, on plaide en faveur de l'estimation de ce modèle par un estimateur pseudo-maximum de vraisemblance gaussien à l'ordre 2. En conséquence, on fournit tous les ingrédients nécessaires à sa mise en oeuvre pratique. S'appuyant sur les Chapitres 1 et 2, on détaille les propriétés asymptotiques de l'estimateur et passe en revue les diverses façons de tester la bonne spécification du modèle. Finalement, on dérive un test simple permettant de se faire une idée de la pertinence du modèle hétéroscédastique proposé avant d'entreprendre son estimation.

Le quatrième chapitre propose une illustration empirique des procédures d'estimation et de tests exposées. Cette illustration, qui consiste en l'estimation et le test de la spécification de fonctions de production, est basée sur un échantillon — fortement non-calibré — de 824 entreprises françaises observées sur tout ou partie de la période 1979 - 1988 (5 201 observations).

Contents

Introduction and summary

Microeconomic theory typically thinks out in terms of 'representative agent'. In doing so, it provides "mean relationships" which should prevail, if the theory were true, between given sets of economic variables.

However, when econometricians try to assess the relevance of the models proposed by the theory, they rapidly face a fundamental problem: the strong heterogeneity of individual behaviors (see for example Mairesse (1988)). As such, the empirical observation of some heterogeneity across individual behaviors does not necessarily invalidate a theoretical hypothesis derived from a 'representative agent' economic model. But it makes at best incomplete, and at worst incorrect — sometimes in terms of consistency, most of the time in terms of inference — any econometric model which does not take it into account.

The problem of heterogeneity is obviously not specific to microeconometric panel data analysis. However, by the very nature of the data and economic models at hand, it is particularly prevalent in this context.

Microeconomic panel dataset usually contains a lot of individuals and a few periods of observation. Moreover, the observed dispersion typically appears to be much stronger in the individual dimension, i.e., across individuals, than in the time dimension, i.e., over time for a given individual. Accordingly, most of the literature in this area has been concerned with modelling individual heterogeneity, as opposed to time heterogeneity. In such panel data models, individual heterogeneity is typically captured through time-invariant individual-specific effects, allowing for intercept variation and/or (in linear models) slope variation across individuals. Virtually all panel data models, either static or dynamic, with exogenous or endogenous regressors, conform to this basic scheme.

Among this very large spectrum of available models (see for example Mátyás-Sevestre (1996)), the most popular and widely used in application is undoubtedly the one-way error components model. In this model, individual heterogeneity is captured through random time-invariant individual-specific effects allowing for intercept variation across individuals.

According to Hsiao's (1986) view, the one-way error components model is intended for making inference about some dependent variable Y conditionally to some given set of explanatory variables X, but unconditionally to the observed individuals, i.e., unconditionally to the individual effects. In this perspective, it just appears as a (multivariate, possibly non-linear) regression model allowing for intercept variation across individuals where the individual effect has been "purged" from the conditional mean of Y, given X and the individual effect, by integrating it out. Viewed in this way, the time-invariant individual-specific error term appearing in the model simply corresponds to the part of the individual effect which is (first order) unpredictable given the chosen set of explanatory variables X. Obviously, the conditional and the unconditional inference about the impact of X on Y will usually be different, and will vary for different choices of the conditioning variables X. For given X, they will be the same in some situations, including the well-known linear case where the individual effects are uncorrelated with the regressors. It is ultimately up to the researcher to choose whether he wants to make conditional or unconditional inference and eventually revise his judgment if his original choice appears empirically "unfeasible".

As a result of its focus on making inference unconditionally to the individual effects, the one-way error components model confines the modelling of heterogeneity to the second order conditional moments of Y given X. In its standard formulation, it assumes that both the individual-specific error term and the general error term are identically and independently distributed — and thus have constant variance —, as well as mutually independent. In other words, following the heuristic interpretation of the model, the individual heterogeneity in behavior across individuals, as well as the heterogeneity of the repeated observations of an individual through time, are assumed to be constant, unrelated to the individuals's characteristics. This is quite unrealistic. From an empirical point of view, heteroscedasticity, which in the present framework is equivalent to variable heterogeneity, is indeed largely acknowledged as an endemic problem when working with cross-sectional microeconomic data. There is of course no reason to believe that the problems encountered in cross-section "disappear" when considering panel data. On the other hand, from a conceptual point of view, in many situations it appears reasonable to expect that the degree to which an economic relationship may describe the actual behavior of individuals is in one way or another related to their characteristics. Nonetheless, this issue seems somewhat to have been ignored in both the theoretical and empirical literature.

Then, there is some need for generalized versions of the standard model which take into account and account for phenomenons of variable heterogeneity. To propose and discuss such an extension of the standard model is one of the two basic purposes of this thesis. The second one, which is its natural complement, is to provide a comprehensive and integrated inferential framework for its estimation and specification testing. We shall argue that second order pseudo-maximum likelihood methods in association with the m-testing / Wooldridge's modified m-testing framework offer an attractive statistical tool-box for the job.

As suggested above, one-way error components models (either standard or generalized) are nothing more than static multivariate second order semi-parametric models, i.e., models which jointly specify the mean and the variance of some T-variate (T being the number of observations over time) dependent variable Y conditional on some information set, or conditioning variables X. The problem of estimation and specification testing of error components models is then just a particular case of the one of estimating and testing multivariate second order semi-parametric models. This is the subject of our two first chapters.

An important issue when dealing with the estimation of second order semi-parametric models is the question of the robustness of the estimation procedure to a possible conditional variance misspecification. This is the question of whether or not, while yielding consistent estimators of both the conditional mean and the conditional variance parameters when they are both correctly specified, it continues to provide a consistent estimator of the mean parameters when the conditional mean is correctly specified but the conditional variance is misspecified. From the one-way error components models perspective, this is the question of whether assuming a wrong specification for the heterogeneity is innocuous or not for the estimation of the mean parameters, which usually are of primary interest.

In the first chapter, we study, in a somewhat abstract but widely applicable multivariate non-linear dynamic framework, this robustness question for arbitrary second order pseudo-maximum likelihood estimators, i.e., a class of estimators which jointly estimate, through the maximisation of a pseudo log-likelihood function, the mean and variance parameters of a second order semi-parametric model. We show that sufficient and essentially necessary conditions for such an estimator to be robust to conditional variance misspecification are (1) that the mean and variance parameters vary independently and (2) that the pseudo-likelihood used as criterion function belongs to a family of distributions that we call restricted quadratic exponential families and whose prominent example is just the (multivariate) normal density. We name RPML2 a second order pseudo-maximum likelihood estimator which satisfies these conditions, the 'R' standing for either robust or restricted. Furthermore, we provide the limiting distribution properties of this class of estimators under different assumptions regarding the degree of misspecification present in the model.

In Chapter 2, we deal with specification testing, in the same general framework than in Chapter 1. According to the results of Chapter 1, the gaussian RPML2 estimator, i.e., RPML2 implemented using the gaussian density as pseudo-densities, appears, because of its robustness, as a very convenient go-between estimator. Indeed, it simultaneously allows to get efficiency gains from approximately taking into account the scedastic structure of the data when, in a first step, concentrating on the conditional mean specification, and, once this first step completed, to further explore the conditional variance specification. The purpose of this chapter is to describe how, from this nice go-between estimator, to take advantage of the very powerful m-testing / Wooldridge's modified m-testing framework for testing, either with or without clear alternatives in mind, the specification of second order semi-parametric models. We sequentially consider nested, non-nested, Hausman-type and information matrix-type testing of the prominent hypotheses of first order correct specification and second order correct specification. We also cover the testing of first order and second order dynamic completeness. In all cases, maintained hypotheses of the tests are precisely stated and reduced to their minimum so that the validity of the tests usually requires no more than just the hypothesis of interest under the null. Although much of the material of this chapter is built from a collection of published works, some of the proposed test statistics are new.

Armed with the quite comprehensive statistical tool-box provided in the first two chapters, we go back to our first purpose in Chapter 3. We propose and discuss an extension of the standard one-way error components model which allows to take into account and to account for phenomenons of variable heterogeneity. The basic idea underlying this extension is very simple. It amounts to letting both the individual-specific and the general error terms variances change by parametrically specifying these variances as functions of the individual's characteristics X. Doing this means adopting for the conditional variance of Y given X a quite flexible parametrization allowing for variable heterogeneity both in the between and within dimensions. This specification obviously contains the standard model as a particular case. Given that the model contains no functional links between mean and variance parameters, we argue for estimating this model by gaussian pseudo-maximum likelihood of order 2, on the grounds of its ability to straightforwardly handle incomplete (unbalanced) panel, its robustness to distributional misspecification and possible misspecification of the heterogeneity, its computational convenience and its potential efficiency. Consequently, we provide all the required ingredients needed for its implementation. Further, as an application of the general results derived in Chapters 1 and 2, we outline its limiting distribution properties, survey the different ways in which its specification may be tested, and finally, derive a convenient joint test statistic for checking the potential relevance of the heteroscedastic model before undertaking the estimation procedure. This chapter emphasizes issues of practical interest.

When proposing an extension of a well-established model, some questions naturally come out : how does its estimation and testing work in practice ? what is its empirical significance ? Chapter 4 exemplifies the potential usefulness of the proposed model and statistical tools through an empirical illustration consisting in production functions estimation and specification testing. This illustration is based on a strongly unbalanced panel dataset of 824 french firms observed over the period 1979 - 1988 (5 201 observations). It suggests (a) that heteroscedasticity-related problems are indeed likely to be present when estimating production models using (cross-section or) panel data, (b) that the proposed full heteroscedastic one-way error components model and its accompanying robust inferential methods may offer a sensible way to deal with it, and finally (c) that the set of proposed specification tests allows to get interesting insights about the empirical correctness of the different estimated models. In this latter respect, it shows in particular that more detailed models do not necessarily turn out to be the most appropriate. On balance, it then provides some support to the key points which motivate the theoretical developments undertaken in this dissertation.

Chapter 1

Second order pseudo-maximum likelihood estimation and conditional variance misspecification

1.1. Introduction

Several econometric models are interested in modelling the expectation of some dependent variable conditional on some information set or conditioning variables. For efficiency reasons and/or because it is of interest of its own, this basic specification is sometimes completed by jointly specifying the conditional variance of the dependent variable. Examples of such second order semi-parametric models are numerous : cross-section models with parametrized heteroscedasticity (e.g. Harvey (1976) or Amemiya (1973)), SURE models (e.g. Magnus (1982)), panel data error components models and their various extensions (random coefficients, heteroscedastic or autocorrelated errors, see Mátyás-Sevestre (1996)), ARCH-type models in a dynamic framework (see Bollerslev-Engle-Nelson (1994)), etc.

Since the seminal works of White (1982) and Gourieroux-Monfort-Trognon (1984a,b), because of the relative simplicity of their implementation — in practice, all which is required is a maximum likelihood optimization routine, a feature provided by most statistical software —, their close relationship with the foundational standard maximum likelihood theory, and the fact that they do not rely on any distributional assumption, pseudo-maximum likelihood methods have become increasingly popular. For the estimation of second order semi-parametric models, the methods proposed by the pseudo-maximum likelihood theory are twofold : quasi-generalized pseudo-maximum likelihood of order one (hereafter denoted QGPML1) and pseudo-maximum likelihood of order two (hereafter denoted PML2). The first one (QGPML1) is based on the properties of the so-called generalized linear exponential families. It is a three-step method whose first step consists in a preliminary estimation — typically by pseudo-maximum likelihood of order one (hereafter denoted PML1) — of the conditional mean parameters, the second step, based on the previous estimator, consists in the estimation of the conditional variance parameters, while the third step is just a generalized PML1 re-estimation of the mean parameters incorporating the conditional variance estimates, as well as possibly the first step conditional mean estimates, as auxiliary parameters. It contains as a special case the well-known feasible generalized (nonlinear) least squares estimator. The second method (PML2) is based on the properties of quadratic exponential families. It is a one-step method in which the mean and variance parameters are jointly estimated. It contains as a special (and prominent) case, multivariate maximum likelihood estimation under normality.

QGPML1 is chiefly intended to tackle with situations where primary interest lies in the conditional mean estimation, i.e., when the conditional variance has mainly been specified for efficiency reasons. On the other hand, PML2 is primarily intended to deal with situations where both the conditional mean and the conditional variance are of interest and/or when do exist functional links between mean and variance parameters. However, these typical roles are not exclusive. In particular, depending on the case at hand, it may be wise to resort to PML2 even when the conditional variance has mainly been specified for efficiency reasons.

Although at first sight more complex, compared to QGPML1, PML2 indeed presents some potential attractive features, both from a computational and a statistical point of view. From a computational point of view, because of the non-negative nature of the variance — positive definiteness of covariance matrix —, the conditional variance specification may be (should often be in order to prevent troubles) nonlinear, implying that variance parameters cannot always be obtained in a simple way, i.e., in avoiding nonlinear optimization. On the order hand, PML2 also requires nonlinear optimization but simultaneously provides mean and variance parameters. This argument is of course strengthened if the conditional mean is also nonlinear. In this case, multiple nonlinear optimizations involved by QGPML1 are replaced by a single nonlinear program. From a statistical point of view, under second order correct (dynamic) specification, i.e., when the model is jointly correctly specified (dynamically complete) for the conditional mean and the conditional variance, not only PML2 may (almost) always be implemented in a way such that it provides an estimator of the mean parameters which is at least as efficient as the one obtained from QGPML1 — PML2 will usually be more efficient if possible structural links between mean and variance parameters are taken into account, a possibility ruled out by QGPML1 —, but it also has additional by-product properties for the variance parameters. Among them, the asymptotic precision of the variance estimator is always easily obtained, and under favorable circumstances, it may be asymptotically efficient. In fact, most of the time, it may be expected to be more efficient than the PML1-like estimator usually computed in the second step of QGPML1. This is of course important when the conditional variance has not only been specified for efficiency reasons but also because it is of interest of its own.

If the quest for first order correct specification, i.e., a model correctly specified for the conditional mean, is already a thorny assignment, the complementary search for second order correct specification may be viewed as a heroic mission. This is particularly true in a multivariate framework where not only variances but also covariances have to be modeled. As a matter of fact, economic theory generally offers less guidelines for specifying the conditional variance than for specifying the conditional mean. Of course, intuition and/or empirical regularities often suggest plausible specifications. For example, in panel data models, the observations of each individual over time may be expected to be serially correlated. Also, if based on microeconomic data — heteroscedasticity seems to be endemic in this kind of data —, some heteroscedasticity-related phenomenon may be anticipated. Based on these stylized facts, a heteroscedastic one-way error components disturbance structure such as the one proposed in Chapter 3 appears plausible but alternative specifications, for example based on a heteroscedastic autoregresssive disturbance structure, are equally appealing. Likewise, to give another example, for modelling the prominent empirical regularities pertaining to the temporal variation in financial market volatility, ARCH-type models offer a quite large spectrum of — non necessarily nested — alternative plausible specifications. Clearly, the problem is not in nature radically different for the specification of the conditional mean. Actually, the point we want to stress here is that specification uncertainty, and thus possible misspecification, is likely to be more severe for the second order conditional moments than for the first order conditional moments.

The recognition of this fact naturally leads to the question of robustness to conditional variance misspecification of the pseudo-maximum likelihood methods outlined above, i.e., to the question of whether or not they continue to provide a consistent estimator of the mean parameters when the conditional mean is correctly specified but the conditional variance is not jointly correctly specified. QGPML1 has been shown to be robust to conditional variance misspecification under weak conditions in a very general multivariate dynamic framework. An extensive discussion and proof of this result may be found in White (1994) (see also Wooldridge (1994)). This point has not been investigated for general PML2 estimators. As far as we know, Pagan-Sabau (1991) is the only available paper related to this problem. In this paper, the authors examine, for univariate linear heteroscedastic regression models such as Poisson and ARCH models, the robustness of the gaussian maximum likelihood estimator to conditional variance misspecification. The present chapter offers a general treatment of the robustness question for arbitrary second order pseudo-maximum likelihood estimators without neither relying on distributional assumptions nor restricting to specific forms the conditional variance misspecification. Further, it provides limiting distribution results.

Derived in a somewhat abstract but widely applicable multivariate nonlinear dynamic framework, the fundamental consistency results of this chapter are twofold. First, sufficient and essentially necessary conditions for a second order pseudo-maximum likelihood estimator to be, regardless of the conditional variance (mis)specification, consistent for the mean parameters when the conditional mean is correctly specified are (1) that mean and variance parameters vary independently and (2) that the pseudo-likelihood used as criterion function belongs to a sub-family of generalized linear exponential families, a sub-family that we entitle restricted generalized linear exponential families. These conditions imply that, as it stands and even if mean and variance parameters vary independently, PML2 is generally not robust to conditional variance misspecification. In other words, when the conditional mean is correctly specified but the conditional variance is not jointly correctly specified, the use of quadratic exponential families as pseudo-likelihood is no longer a sufficient condition for consistent estimation of the conditional mean parameters. Second, sufficient and essentially necessary conditions for a second order pseudo-maximum likelihood estimator to be consistent for both mean and variance parameters when the conditional mean and the conditional variance are jointly correctly specified, and to remain consistent for the mean parameters when the conditional mean is correctly specified but the conditional variance is not jointly correctly specified are (1) again that the mean and variance parameters vary independently and (2) that the pseudo-likelihood used as criterion function belongs to a sub-family of both quadratic exponential families and restricted generalized linear exponential families, a sub-family that we call restricted quadratic exponential families.

We entitle a second order pseudo-maximum likelihood estimator which satisfies the latter conditions RPML2, the 'R' standing for either robust or restricted. As it could be expected, the (multivariate) normal density is a member of restricted quadratic exponential families, implying that (provided of course that mean and variance parameters vary independently) pseudo-maximum likelihood estimation under normality is a particular case — and undoubtedly the prominent one — of RPML2. The requirement that mean and variance parameters have to vary independently is no more surprising. It is also imposed by QGPML1. It is important to note that this requirement does not signify that mean and variance parameters have to be functionally unrelated in the structural model but simply that they have to be treated as if they were not functionally related. In other words, for gaining robustness, eventual structural cross-constraints between mean and variance parameters (here considered in their reduced-form through parameters common to the conditional mean and the conditional variance) have to be discarded.

Besides these consistency results, we investigate the limiting distribution of RPML2 under different assumptions regarding the degree of misspecification present in the model. So, along with possible dynamic misspecification, are covered the cases where the model is only correctly specified for the conditional mean, the model is jointly correctly specified for the conditional mean and the conditional variance, the model is in addition jointly correctly specified for the third or the third and the fourth order conditional moments and, finally, the model is correctly specified for the entire conditional distribution. Further, we provide lower bounds for its asymptotic covariance matrix and compare these bounds with the semi-parametric efficiency bounds based on conditional moments restrictions. Also, we treat the special case where the observations are independent, as in cross-section or panel data, and briefly discuss the possible efficiency price to pay for robustness it may entail. Finally, we compare its asymptotic distribution and relative merits with those of QGPML1.

We concentrate on pseudo-maximum likelihood methods. This is by no means the only way to handle second order semi-parametric models. As prominent alternatives, the generalized method of moments framework (hereafter denoted GMM) mainly offers two methods which may be viewed, in the present context, as GMM analogues of QGPML1 and PML2 (see Newey (1993) and Wooldridge (1994)). The first one consists in instrumental variable estimation of the first order conditional moments parameters using optimal instruments. The optimal instruments are, among other things, functions of the second order conditional moments. They have to be estimated — possibly non-parametrically, which is more tricky but, as the first order over-identified optimally weighted GMM estimator suggested by Cragg (1983), allows handling heteroscedasticity of unknown form — in a first step. In the parametric case, this estimator has essentially the same properties than QGPML1. In particular, it is robust to conditional variance misspecification. The second GMM technique consists in joint instrumental variable estimation of the first and second order conditional moments parameters, possibly using optimal instruments. If optimal instruments are used, this estimator, which achieves a semi-parametric efficiency bound, will usually be more efficient than PML2. However this would require non-parametric estimation of (dynamic) conditional third and fourth moments, as well as numerous cross-product moments in the multivariate case. Needless to say, except in very special cases, this estimator is not robust to conditional variance misspecification.

The chapter proceeds as follows. Section 1.2 describes the general set-up and notations. Section 1.3 defines second order pseudo-maximum likelihood estimation. As a preliminary, Section 1.4 provides a generalized version of the standard consistency properties of PML2 and outlines a first result suggesting its general inconsistency under conditional variance misspecification. Section 1.5 provides sufficient and essentially necessary conditions for a second order pseudo-maximum likelihood estimator to be robust to conditional variance misspecification. Section 1.6 deals with the limiting distribution of RPML2, mentionning in passing its close connection with those of QGPML1. Finally, concluding comments are proposed in Section 1.7.

1.2. Set-up and Notation

We adopt a general multivariate dynamic framework essentially similar to those of White (1994) and Wooldridge (1994). Throughout the chapter, matrix calculus notational conventions are in accordance with those of Magnus-Neudecker (1986,1988).

We assume that the observed data are a realization of a stochastic process $W \equiv \{W_t : \Omega \to \mathbb{R}^\nu,\ \nu \in \mathbb{N},\ t = 1, 2, ...\}$ on a complete probability space $(\Omega, \mathcal{F}, P_o)$. We will refer to P_o as the "true data generating process" (true DGP). Unless otherwise explicitly stated, all expectations and conditional expectations are taken with respect to this true DGP.

Let W_t be partitioned as $W_t = (Y_t', Z_t')'$, where Y_t is a $G \times 1$ vector and Z_t is a $(\nu - G) \times 1$ vector. Further, let X_t stand for some subset of the information set $(Z_t, \tilde{W}_{t-1})$, where $\tilde{W}_{t-1} \equiv (Y_{t-1}, Z_{t-1}, ..., Y_1, Z_1)$ is the information available on Y and Z at time $t-1$. Let also $\mathcal{X}_t \subset \mathbb{R}^{k_{x_t}}$ denote the range of X_t. The dimension k_{x_t} of $\mathcal{X}_t$ is allowed to depend on t, in particular to grow with t. Finally, let $Y^n \equiv (Y_1, Y_2, ..., Y_n)$ and $X^n \equiv (X_1, X_2, ..., X_n)$ be finite random samples of size n.

Y_t denotes the vector of dependent or endogenous variables. We suppose that interest lies in explaining Y_t in terms of the explanatory or conditioning variables X_t. This setting puts up with most of usual practical situations. In a pure time-series context, X_t will only contain some (possibly growing) number of lags of the dependent variable Y_t. In more general time-series, X_t will usually contain (some sub-vector of) Z_t and some (possibly growing) number of lags of both Y_t and (some sub-vector of) Z_t. In a cross-section or panel data framework, X_t is by definition restricted to (some sub-vector of) Z_t and the observations are assumed to be independently distributed across t. Note that X_t may be a constant vector for all t, in which case interest lies in the unconditional distribution of Y_t.

Let $D_t(.|X_t)$ denote the conditional distribution of Y_t given X_t and $p_t(.|X_t)$ stand for the conditional density of Y_t given X_t. $D_t(.|X_t)$ always exists while $p_t(.|X_t)$ exists under weak conditions on $D_t(.|X_t)$ (see White (1994)). We will assume that these conditions hold whenever needed.

In what follows, we suppose that the researcher's primary interest lies in particular aspects of $D_t(.|X_t)$, namely the conditional expectation of Y_t given X_t and, either for efficiency reasons or because it is of interest of its own, the conditional variance of Y_t given X_t. Accordingly, we assume that the following semi-parametric model $\mathcal{S}$ is jointly specified for respectively $E(Y_t|X_t)$ and $V(Y_t|X_t)$

$$\mathcal{S} \equiv \begin{cases} \{m_t(X_t, \theta) : X_t \in \mathcal{X}_t,\ \theta \in \Theta \subset \mathbb{R}^{k_\theta}\} \\ \{\Omega_t(X_t, \theta) : X_t \in \mathcal{X}_t,\ \theta \in \Theta \subset \mathbb{R}^{k_\theta}\} \end{cases},\ t = 1, 2, ...$$

where θ is a $k_\theta \times 1$ vector of parameters, the functions m_t are known $G \times 1$ vector functions which may depend on t, and the functions Ω_t are $G \times G$ known matrix functions which may also depend on t and are symmetric positive definite, $\forall\, \theta \in \Theta$, $\forall\ X_t \in \mathcal{X}_t$, $t = 1, 2, ...$

It is worth noting that the choice of the conditioning variables X_t is entirely free. It only depends on what is of interest to the researcher. For example, in a cross-section or panel data context, one might be more interested in the "reduced-form relationship" which may exist between Y_t and a small subset of Z_t rather than in the "structural relationship" which presumably exists between Y_t and the entire information set Z_t. Likewise, in a time-series framework, one might only be interested in the contemporaneous relationship between Y_t and Z_t. Also, note that in the present framework, there is no presumption that some sort of strict exogeneity (Granger noncausality) of the process $\{Z_t : t = 1, 2, ...\}$ holds (for a discussion of these points, see Wooldridge (1994)).

Actually, from a statistical point of view — but of course not from an economical interpretation point of view —, only the following definitions of correct specification matter.

Definition 1 The semi-parametric model $\mathcal{S}$ is said (a) *first order correctly specified* (correctly specified for the conditional mean) if there exists a true value θ^o in Θ such that

$$m_t(X_t, \theta^o) = E(Y_t|X_t), \qquad a.s. - P_o, \quad t = 1, 2, ...$$

and (b) *second order correctly specified* (jointly correctly specified for the conditional mean and the conditional variance) if there exists a true value θ^o in Θ such that

$$\begin{cases} m_t(X_t, \theta^o) = E(Y_t|X_t) \\ \Omega_t(X_t, \theta^o) = V(Y_t|X_t) \end{cases}, \qquad a.s. - P_o, \quad t = 1, 2, ...$$

First (resp. second) order correct specification basically means that, for a given choice of the conditional variables X_t, we have been able to (resp. jointly) correctly specify, up to an unknown vector of parameters, the functional forms of the first (resp. the two first) conditional moment(s) of Y_t given X_t. When first order correct specification holds, second order misspecification may follow either from the fact that the conditional variance is misspecified, i.e., there is no θ^o in Θ such that $\Omega_t(X_t, \theta^o) = V(Y_t|X_t)$, or from the fact that, although in itself correctly specified, the correct conditional variance specification holds at a true value θ^o of θ different from the one at which the correct conditional mean specification holds. Obviously, the latter situations are tautologically ruled out whenever mean and variance parameters vary independently.

As it may be seen, the above definitions have nothing to do with the fact that the semi-parametric model $\mathcal{S}$ captures or not all the dynamics of Y_t, i.e., the entire dependence of Y_t on the past. Such dynamic incompleteness does not matter for the issue of consistency. However, it has important consequences for inference. The concept of dynamic misspecification crucially hinges on the choice of the conditioning variables X_t.

Definition 2 The semi-parametric model $\mathcal{S}$ is said (a) *first order dynamically complete* (dynamically complete for the conditional mean) if

$$E(Y_t|X_t) = E(Y_t|X_t, \Psi_{t-1}), \qquad a.s. - P_o, \quad t = 1, 2, ...$$

and (b) *second order dynamically complete* (dynamically complete for the conditional mean and the conditional variance) if it is first order dynamically complete and in addition

$$V(Y_t|X_t) = V(Y_t|X_t, \Psi_{t-1}), \qquad a.s. - P_o, \quad t = 1, 2, ...$$

where $\Psi_{t-1} \equiv (Y_{t-1}, X_{t-1}, ..., Y_1, X_1)$ is the information available at time $t-1$.

Note that Definition 2(a) and 2(b) allow X_t and Ψ_{t-1} to overlap, as happens if X_t contains lags of Y_t or lags of some other variables Z_t. For example, if $X_t \equiv (Z_t, Y_{t-1}, Z_{t-1})$, then $\Psi_{t-1} \equiv (Y_{t-1}, Z_{t-1}, ..., Y_1, Z_1)$.

Note also that the concept of dynamic misspecification is only related to the choice of the conditioning variables X_t, and not to the functional forms of the conditional moments. So, the semi-parametric model $\mathcal{S}$ could be second order dynamically complete although misspecified for the conditional mean and/or the conditional variance. Following Wooldridge (1994), correct dynamic specification basically means that if interest lies in explaining Y_t in terms of past Y and possibly current and past values of some other sequence $\{Z_t\}$, then enough lags of Y and Z have been included in the conditioning variables X_t to capture the entire dependence of Y_t on the past. Clearly, the whole concept of dynamic misspecification is irrelevant when dealing with independent observations as in cross-section or panel data.

Finally, in order to prevent misunderstandings, a last remark. In the definition of the semi-parametric model $\mathcal{S}$, as well as in the various definitions of correct specification, the assumed set of conditioning variables X_t may be seen to be the same in the conditional mean and in the conditional variance. Just as for the vector of parameters θ , this does not mean that both the conditional mean and the conditional variance actually depend on all the conditioning variables X_t. It simply means that the information set X_t is defined as including all the variables which appear either in the conditional mean or in the conditional variance. For example, if the conditional mean only depends on say X_t^1 and the conditional variance only depends on say X_t^2, then X_t must be defined as $X_t \equiv (X_t^1, X_t^2)$. This point is important since it implies that for judging conditional mean or conditional variance (dynamic) specification, we must take into account the variables which appear in both moments and not only in the one under scrutiny. In other words, continuing the above example, first order correct specification requires that there exists θ^o in Θ such that $E(Y_t|X_t^1, X_t^2) = m_t(X_t^1, \theta^o)$ and not only such that $E(Y_t|X_t^1) = m_t(X_t^1, \theta^o)$. The same reasoning applies to every definition.

Throughout the chapter, it is assumed that first order correct specification holds. Our primary concern is the consequences of the violation of second order correct specification when performing second order pseudo-maximum likelihood estimation. The consequences for inference of dynamic misspecification — as well as higher order misspecification — is also examined.

1.3. Second order pseudo-maximum likelihood estimation

We concentrate on second order pseudo-maximum likelihood estimators, i.e., on a class of estimators which jointly estimate, through the maximization of a pseudo log-likelihood function, the mean and variance parameters of the semi-parametric model $\mathcal{S}$. The following definition makes this statement more precise.

Definition 3 A second order pseudo-maximum likelihood estimator $\hat{\theta}_n$ of the semi-parametric model $\mathcal{S}$ is defined as a solution of

$$\text{Max}_{\theta \in \Theta}\, L_n(Y^n, X^n, \theta) \equiv \frac{1}{n} \sum_{t=1}^{n} \ln f_t\left(Y_t, m_t(X_t, \theta), \Omega_t(X_t, \theta)\right)$$

where the p.d.f. $f_t\left(Y, m, \Sigma\right)$ are indexed by their mean $m \in \mathcal{M}_t \subset \mathbb{R}^G$ and by their covariance matrix $\Sigma \in \mathcal{E}_t$, $\mathcal{E}_t$ being a subset of the $G \times G$ positive definite matrices, $t = 1, 2, ...$, and where, $\forall\ \theta \in \Theta$ and $\forall\ X_t \in \mathcal{X}_t$, $m_t(X_t, \theta) \in \mathcal{M}_t$ and $\Omega_t(X_t, \theta) \in \mathcal{E}_t$, $t = 1, 2, ...$

According to Definition 3, a second order pseudo-maximum likelihood estimator is based on a sequence $\{f_t\}$ of probability distribution functions adapted for the first and second order moments and "compatible" with the semi-parametric model $\mathcal{S}$. Note that f_t may be different for all t. Note also that the above definition contains no explicit "compatibility" assumption in terms of the range of Y. This is simply because such assumption is not always necessary (see below). If necessary, it will be implicitly assumed to hold.

Finding p.d.f. f_t adapted for the first and second order moments, i.e., such that $E(Y) = m$ and $V(Y) = \Sigma$, is not very complicated[1]. It suffices to start with a G-variate p.d.f. $g_t(Y^*)$ such that $E(Y^*) = 0$ and $V(Y^*) = I_G$ (for example, take the product of the p.d.f. of suitably scaled univariate random variables), and then define

[1] I owe this way of looking at the point to Alain Trognon.

Y through the linear transformation $Y = m + \Sigma^{1/2} Y^*$, where $\Sigma^{1/2}$ is a symmetric positive definite matrix. The desired first and second order moments adapted p.d.f. f_t is given by

$$f_t(Y, m, \Sigma) = g_t\left(\Sigma^{-1/2}(Y - m)\right) \det \Sigma^{-1/2}$$

Since f_t is a p.d.f. adapted for the first and second order moments and "compatible" with $\mathcal{S}$, $f_t(., m_t(X_t, \theta), \Omega_t(X_t, \theta)) = \lambda_t(., X_t, \theta)$ characterizes a well-defined conditional density for Y_t given X_t, for which the two first conditional moments are by definition respectively equal to $m_t(X_t, \theta)$ and $\Omega_t(X_t, \theta)$, and the higher conditional moments (if they exist) depend on the choice of the adapted p.d.f. f_t. In other words, for each choice of the sequence $\{f_t\}$, a second order pseudo-maximum likelihood estimator $\hat{\theta}_n$ corresponds to a well-defined standard[2] (conditional) maximum likelihood estimator, namely a standard maximum likelihood estimator of the following possibly misspecified parametric model $\mathcal{P}$ implicitly defined by the semi-parametric model $\mathcal{S}$ and the chosen sequence $\{f_t\}$

$$\mathcal{P} \equiv \left\{\lambda_t(., X_t, \theta) = f_t(., m_t(X_t, \theta), \Omega_t(X_t, \theta)) : X_t \in \mathcal{X}_t,\ \theta \in \Theta \subset \mathbb{R}^{k_\theta}\right\}, \ t = 1, 2, ...$$

Obviously, when $\mathcal{S}$ is first order correctly specified, $\mathcal{P}$, i.e., the collection of conditional densities for Y_t given X_t underlying $\hat{\theta}_n$, is correct for the first conditional moments of Y_t given X_t, while possibly misspecified for higher conditional moments. Likewise, if $\mathcal{S}$ is second order correctly specified, $\mathcal{P}$ is correct for the two first conditional moments of Y_t given X_t, while again possibly misspecified for higher conditional moments.

In some circumstances, i.e., in particular, under second order correct specification, for a suitable choice of the sequence $\{f_t\}$, $\mathcal{P}$ may be correctly specified for some higher conditional moments or further for the entire conditional densities of Y_t given X_t. In this respect, the following definitions of correct specification, as well as further definitions of correct dynamic specification, will be useful in the sequence.

Definition 4 The parametric model $\mathcal{P}$ arising from the semi-parametric model $\mathcal{S}$ and a given sequence $\{f_t\}$ as defined in Definition 3 is said (a) *third order correctly specified* (jointly correctly specified for the three first conditional moments) if the semi-parametric model $\mathcal{S}$ is second order correctly specified and in addition

$$Cov_{\lambda_t^o}\left[\left(\text{vec}\left(Y_t Y_t'\right), Y_t\right) | X_t\right] = Cov\left[\left(\text{vec}\left(Y_t Y_t'\right), Y_t\right) | X_t\right], \qquad a.s. - P_o, \quad t = 1, 2, ...$$

(b) *fourth order correctly specified* (jointly correctly specified for the four first conditional moments) if it is third order correctly specified and in addition

$$V_{\lambda_t^o}\left[\text{vec}\left(Y_t Y_t'\right) | X_t\right] = V\left[\text{vec}\left(Y_t Y_t'\right) | X_t\right], \qquad a.s. - P_o, \quad t = 1, 2, ...$$

where $Cov_{\lambda_t^o}[.|X_t]$ and $V_{\lambda_t^o}[.|X_t]$ respectively denote covariance and variance taken with respect to $\lambda_t(Y_t, X_t, \theta^o)$.

and (c) *correctly specified for the conditional density* if there exists a true value θ^o

[2] But possibly different from a classical (conditional) maximum likelihood estimator which requires the specification of the joint density of $Y^n \equiv (Y_1, Y_2, ..., Y_n)$ conditional on $Z^n \equiv (Z_1, Z_2, ..., Z_n)$. See Wooldridge (1994) for a discussion of this point.

in Θ such that

$$\lambda_t\left(.,X_t,\theta^o\right)=p_t(.|X_t),\qquad \forall X_t\in\mathcal{X}_t,\quad t=1,2,...$$

Definition 5 The semi-parametric model $\mathcal{S}$ is said (a) *third order dynamically complete* (dynamically complete for the three first conditional moments) if it is second order dynamically complete and in addition

$$Cov\left[\left(\text{vec}\left(Y_tY_t'\right),Y_t\right)\middle|\, X_t\right]=Cov\left[\left(\text{vec}\left(Y_tY_t'\right),Y_t\right)\middle|\, X_t,\Psi_{t-1}\right],\qquad a.s.{-}P_o,\quad t=1,2,...$$

(b) *fourth order dynamically complete* (dynamically complete for the four first conditional moments) if it is third order dynamically complete and in addition

$$V\left[\text{vec}\left(Y_tY_t'\right)\middle|\, X_t\right]=V\left[\text{vec}\left(Y_tY_t'\right)\middle|\, X_t,\Psi_{t-1}\right],\qquad a.s.-P_o,\quad t=1,2,...$$

and (c) *dynamically complete for the conditional distribution* if

$$D_t(.|X_t)=\tilde{D}_t(.|X_t,\Psi_{t-1}),\qquad t=1,2,...$$

where $\tilde{D}_t(.|X_t,\Psi_{t-1})$ denotes the conditional distribution of Y_t given X_t and Ψ_{t-1}.

These definitions of correct (dynamic) specification may be interpreted exactly in the same way than Definition 1 and Definition 2. Definition 4(a), 4(b) and 4(c) basically mean that, for a given choice of the conditioning variables X_t, we have been able to jointly correctly specify, up to an unknown vector of parameters and only through the parametrization of the two first conditional moments and a suitable choice of the sequence $\{f_t\}$, the functional forms of, respectively, the three first conditional moments, the four first conditional moments and the entire conditional densities of Y_t given X_t, i.e., all conditional moments of Y_t given X_t. In the latter case, $\hat{\theta}_n$ is just a standard maximum likelihood estimator, and thus, under usual regularity conditions, is consistent for θ^o (see for example Wooldridge (1994)). On the other hand, Definition 5(a), 5(b) and 5(c) allow to state to which order the semi-parametric model $\mathcal{S}$ captures or not all the dynamics of Y_t, i.e., the entire dependence of Y_t on the past, a feature which only hinges on the choice of the conditioning variables X_t. As outlined above, such dynamic completeness does typically not matter for the issue of consistency but may have important consequences for inference. As a matter of fact, under conditional density correct specification, i.e., when $\hat{\theta}_n$ is a just standard maximum likelihood estimator, conditional distribution correct dynamic specification is usually needed for the traditional information matrix equality to hold (again, see for example Wooldridge (1994)). Obviously, conditional density correct specification implies first, second, third and fourth order correct specification and conditional distribution correct dynamic specification implies first, second, third and fourth order correct dynamic specification.

Hereafter, the fact that the implicit parametric model $\mathcal{P}$ corresponding to $\hat{\theta}_n$ (resp. the semi-parametric model $\mathcal{S}$) might be correctly specified (resp. dynamically complete) up to orders higher than the two first will only be used as benchmark when discussing the limiting distribution of RPML2.

1.4. Pseudo-maximum likelihood of order 2 (PML2)

As mentioned in the introduction, PML2, which is a particular sub-class of second order pseudo-maximum likelihood estimators, relies on the properties of quadratic exponential families. We first outline the definition and essential characteristics of quadratic exponential families. Then we provide a generalized version of the standard consistency properties of PML2 under second order correct specification. Finally, we give a first result suggesting its general inconsistency under first order correct specification but second order misspecification.

1.4.1. Quadratic exponential families

According to Gourieroux-Monfort-Trognon (1984a), quadratic exponential families may be defined as follows.

Definition 6 A family of probability measures on $\mathbb{R}^G$ indexed by $m \in \mathcal{M} \subset \mathbb{R}^G$ and $\Sigma \in \mathcal{E}$, where $\mathcal{E}$ is a subset of the $G \times G$ positive definite matrices, is called quadratic exponential if (a) every element of the family has a density function with respect to a given measure $\upsilon(dY)$ which may be written as

$$l(Y, m, \Sigma) = \exp\left(A(m, \Sigma) + B(Y) + C(m, \Sigma)'Y + Y'D(m, \Sigma)Y\right)$$

where $A(m, \Sigma)$ and $B(Y)$ are scalar, $C(m, \Sigma)$ is a $G \times 1$ vector and $D(m, \Sigma)$ is a $G \times G$ matrix, and (b) m is the mean and Σ is the covariance matrix of the distribution $l(Y, m, \Sigma)$.

The prominent member of quadratic exponential families is undoubtedly the normal density. For the normal density, we simply have

$$\begin{array}{c} A(m, \Sigma) = -\frac{G}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}m'\Sigma^{-1}m, \\ B(Y) = 0,\ C(m, \Sigma) = \Sigma^{-1}m \ \text{ and } \ D(m, \Sigma) = -\frac{1}{2}\Sigma^{-1} \end{array} \tag{1.1}$$

Quadratic exponential families have some important properties. Three of them will be particularly useful in the sequence.

Property 1 If $l(Y, m, \Sigma)$ is a quadratic exponential family, then $\forall\ m,\ m_o \in \mathcal{M}$, $\forall\ \Sigma,\ \Sigma_o \in \mathcal{E}$, we have

$$\begin{aligned} & A(m_o, \Sigma_o) + C(m_o, \Sigma_o)'m_o + \operatorname{tr}\left(D(m_o, \Sigma_o)(\Sigma_o + m_o m_o')\right) \\ & \geq A(m, \Sigma) + C(m, \Sigma)'m_o + \operatorname{tr}\left(D(m, \Sigma)(\Sigma_o + m_o m_o')\right) \end{aligned}$$

where the equality holds if and only if $m = m_o$ and $\Sigma = \Sigma_o$.

Proof. See Appendix B. ■

Property 2 If $l(Y, m, \Sigma)$ is a quadratic exponential family, then $\forall\ m_o \in \mathcal{M}$, $\forall\ \Sigma$,

$\Sigma_o \in \mathcal{E}$ such that $\Sigma \neq \Sigma_o$, it may exist $m \in \mathcal{M}$ such that $m \neq m_o$ and that we have

$$\begin{aligned} &A(m_o, \Sigma) + C(m_o, \Sigma)' m_o + \operatorname{tr}\left(D(m_o, \Sigma)(\Sigma_o + m_o m_o')\right) \\ &< A(m, \Sigma) + C(m, \Sigma)' m_o + \operatorname{tr}\left(D(m, \Sigma)(\Sigma_o + m_o m_o')\right) \end{aligned}$$

Proof. See Appendix B. ■

Property 3 If $l(Y, m, \Sigma)$ is a quadratic exponential family and if the functions $A(m, \Sigma)$, $C(m, \Sigma)$ and $D(m, \Sigma)$ are continuously differentiable with respect to m and Σ on respectively $int\,\mathcal{M}$ and $int\,\mathcal{E}$, then $\forall\ m \in int\,\mathcal{M}$, $\forall\ \Sigma \in int\,\mathcal{E}$, we have

$$\begin{aligned} \frac{\partial A(m, \Sigma)}{\partial m} + \frac{\partial C(m, \Sigma)'}{\partial m} m + \frac{\partial\left(\operatorname{vec} D(m, \Sigma)\right)'}{\partial m} \operatorname{vec}(\Sigma + mm') = 0 \\ \frac{\partial A(m, \Sigma)}{\partial \operatorname{vec} \Sigma} + \frac{\partial C(m, \Sigma)'}{\partial \operatorname{vec} \Sigma} m + \frac{\partial\left(\operatorname{vec} D(m, \Sigma)\right)'}{\partial \operatorname{vec} \Sigma} \operatorname{vec}(\Sigma + mm') = 0 \end{aligned}$$

Proof. See Appendix B. ■

The consistency of PML2 under second order correct specification basically relies on Property 1. On the other hand, the general inconsistency of PML2 under first order correct specification but second order misspecification essentially derives from Property 2. Property 3 will be used later for outlining a property of restricted quadratic exponential families.

1.4.2. Consistency of PML2 under second order correct specification

In this section, we focus on conditions ensuring consistent estimation of θ^o when the semi-parametric model $\mathcal{S}$ is jointly correctly specified for the conditional mean and the conditional variance.

In general, i.e., for an arbitrary choice of the sequence $\{f_t\}$, a second order pseudo-maximum likelihood estimator $\hat{\theta}_n$ as given in Definition 3 will not provide a consistent estimator of the true value θ^o. However, both sufficient and essentially necessary conditions for $\hat{\theta}_n$ to be a consistent estimator of θ^o may be derived. Sufficient conditions for consistency are given in Proposition 1.

Proposition 1 *Suppose that $\hat{\theta}_n$ is as given in Definition 3 and that regularity conditions R1-R6 in Appendix A hold. If the semi-parametric model $\mathcal{S}$ is second order correctly specified, and if, for all $t = 1, 2, ...,$ f_t belongs to the quadratic exponential family, then $\hat{\theta}_n \to \theta^o$ as $n \to \infty$ a.s. $- P_o$.*

Proof. Given regularity conditions R1-R6, from Theorem 3.5 of White (1994), we have $\hat{\theta}_n - \theta_n^* \to 0$ as $n \to \infty$ $a.s. - P_o$, where $\theta_n^* = \text{Argmax}_{\theta \in \Theta} E(L_n(Y^n, X^n, \theta))$. Thus, it is enough to show that $\theta_n^* = \theta^o$ for all $n = 1, 2, ...$ Since, for all $t = 1, 2, ...,$ f_t belongs to the quadratic exponential family, letting m_t stand for $m_t(X_t, \theta)$ and Ω_t stand for $\Omega_t(X_t, \theta)$, we have that, for all $t = 1, 2, ...,$ $E(\ln f_t(Y_t, m_t(X_t, \theta), \Omega_t(X_t, \theta)))$

is equal to

$$E\left[A_t(m_t,\Omega_t)+B_t(Y_t)+C_t(m_t,\Omega_t)'Y_t+\operatorname{tr}\left(D_t(m_t,\Omega_t)Y_tY_t'\right)\right] \tag{1.2}$$

Given second order correct specification, $E\left(Y_t|X_t\right)$ and $E(Y_tY_t'|X_t)=V\left(Y_t|X_t\right)+E\left(Y_t|X_t\right)E\left(Y_t|X_t\right)'$ by definition exist and the law of iterated expectations applies such that (1.2) may be written

$$E\left[A_t(m_t,\Omega_t)+B_t(Y_t)+C_t(m_t,\Omega_t)'E\left(Y_t|X_t\right)+\operatorname{tr}\left(D_t(m_t,\Omega_t)E(Y_tY_t'|X_t)\right)\right] \tag{1.3}$$

Since $B_t(Y_t)$ does not depend on m_t and Ω_t, from Property 1, (1.3) has a unique maximum when $m_t(X_t,\theta)=E\left(Y_t|X_t\right)$ and $\Omega_t(X_t,\theta)=V\left(Y_t|X_t\right)$, or, given second order correct specification, a maximum at $\theta=\theta^o$. The identifiable uniqueness of $\{\theta_n^*\}$ ensures that θ^o is the unique maximum of $E\left(L_n(Y^n,X^n,\theta)\right)$, i.e., that $\theta_n^*=\theta^o$ for all $n=1,2,...$ ■

In other words, a second order pseudo-maximum likelihood estimator obtained by specifying the pseudo-densities f_t as members of the quadratic exponential family, i.e., PML2, provides a consistent estimator of the true value of a second order correctly specified semi-parametric model $\mathcal{S}$ regardless of the true DGP P_o, i.e., regardless of whether or not the implicit parametric model $\mathcal{P}$ corresponding to $\hat{\theta}_n$ is correctly specified for other aspects of the "true conditional densities" of Y_t given X_t, and thus in particular whether or not these "true underlying densities" are in the quadratic exponential family. This seminal result has been first brought out, in a more restrictive framework, by Gourieroux-Monfort-Trognon (1984a). It has been shown to hold for dynamic models under the assumption that the pseudo log-likelihood is specified as a normal density by Bollerslev-Wooldridge (1992). Proposition 1 is a straightforward generalization of these previous results.

Note that the identifiability of θ^o is ensured by the regularity condition R6. Such an identifiability condition typically holds under the more primitive — but also more restrictive — assumption that $\mathcal{S}$ is second order identifiable, i.e., that $\forall\ \theta,\ \theta^o\in\Theta$

$$\begin{cases} m_t(X_t,\theta)=m_t(X_t,\theta^o) \\ \Omega_t(X_t,\theta)=\Omega_t(X_t,\theta^o) \end{cases} \Rightarrow \theta=\theta^o \qquad a.s.-P_o, \quad t=1,2,...$$

By the way, note also that, as pointed out by Gourieroux-Monfort-Trognon (1984a) for PML1, because in the definition of the quadratic exponential family $B(Y)$ does not depend on parameters, it makes no differences to maximize $\sum_{t=1}^n[A_t(m_t,\Omega_t)+B_t(Y_t)+C_t(m_t,\Omega_t)'Y_t+Y_t'D_t(m_t,\Omega_t)Y_t]$ or $\sum_{t=1}^n[A_t(m_t,\Omega_t)+C_t(m_t,\Omega_t)'Y_t+Y_t'D_t(m_t,\Omega_t)Y_t]$. Therefore, it is not necessary to impose on Y_t the "compatibility" constraints which may be implied by the definition of $B_t(.)$.

Remarkably, at least for $G=1$, there exists a reciprocal to Proposition 1 which shows that, for a second order pseudo-maximum likelihood estimator to be consistent for θ^o, a specification of f_t based on quadratic exponential families is not only sufficient but also necessary in some sense. This necessary condition is outlined in Proposition 2.

Proposition 2 *($G=1$) Suppose that $\hat{\theta}_n$ is as given in Definition 3, where $\mathcal{M}_t$ and $\mathcal{E}_t$ are closures of open connected sets. If for any probability measure P_o such* *that regularity conditions R1-R5, R6′, R7-R9 in Appendix A hold, when the semi-parametric model $\mathcal{S}$ is second order correctly specified, we have that* $\theta_n^* = Argmax_{\theta\in\Theta} E\left(L_n(Y^n, X^n, \theta)\right) = \theta^o$ *for all* $n = 1, 2, ...,$ *then, for all* $t = 1, 2, ...,$ f_t *belongs to the quadratic exponential family.*

Proof. See Appendix C. ■

Again, Proposition 2 is a straightforward generalization of a seminal result of Gourieroux-Monfort-Trognon (1984a). Note that Proposition 2 is not entirely a converse of Proposition 1. Besides the fact that $G = 1$, as pointed out by White (1994) for a similar result holding for PML1, Proposition 2 imposes additional regularity conditions, in particular differentiability (R7 and R8) as well as interiority of θ_n^* (R6′). Moreover, for consistent estimation of θ^o, it is only necessary that $\theta_n^* \to \theta^o$ rather than $\theta_n^* = \theta^o$ for all $n = 1, 2, ...$ as assumed. For this reason, it is possible to consistently estimate θ^o even when f_t is not a member of the quadratic exponential family for some (necessary asymptotically negligeable) indices t. Bearing this minor qualification in mind, Proposition 2 basically says that the quadratic exponential family is essentially the only family that provides a consistent second order pseudo-maximum likelihood estimator of the true value of a second order correctly specified semi-parametric model $\mathcal{S}$ regardless of the true DGP P_o. In other words, only a very limited subset of the class of second order pseudo-maximum likelihood estimators is insensitive to distributional misspecification. If P_o was further restricted, it would presumably be possible to find others families which would also yield such estimators. For an interesting result closely related to this point, see (Theorem 1 of) Newey-Steigerwald (1997).

1.4.3. PML2 and first order correct specification but second order misspecification

We now turn our attention to the consistency properties of PML2 when the semi-parametric model $\mathcal{S}$ is first order correctly specified but second order misspecified.

As intuitively appealing and according to the special results obtained by Pagan-Sabau (1991) for the univariate gaussian maximum likelihood estimator, if $\mathcal{S}$ is such that the conditional mean and the conditional variance depend on common parameters, PML2 may be expected to be generally inconsistent for the mean parameters true value whenever the conditional variance is not jointly correctly specified. It is not so obvious if mean and variance parameters vary independently.

In this section, we assume that the vector of parameters θ is partitioned as $\theta = (\theta_1', \theta_2')'$, where θ_1 and θ_2 denote respectively mean-specific and variance-specific parameters, such that the semi-parametric model $\mathcal{S}$ is as described in the following assumption.

Assumption 1 The semi-parametric model $\mathcal{S}$ is such that for θ partitioned as $\theta = (\theta_1', \theta_2')'$ and Θ accordingly defined as $\Theta = \Theta_1 \times \Theta_2$, we have

$$\mathcal{S} \equiv \left\{ \begin{array}{l} \left\{m_t(X_t, \theta) = m_t(X_t, \theta_1) : X_t \in \mathcal{X}_t,\ \theta_1 \in \Theta_1 \subset \mathbb{R}^{k_{\theta_1}}\right\} \\ \left\{\Omega_t(X_t, \theta) = \Omega_t(X_t, \theta_2) : X_t \in \mathcal{X}_t,\ \theta_2 \in \Theta_2 \subset \mathbb{R}^{k_{\theta_2}}\right\} \end{array} \right. ,\ t = 1, 2, ...$$

where the $k_{\theta_1} \times 1$ vector of parameters θ_1 and the $k_{\theta_2} \times 1$ vector of parameters θ_2 $(k_{\theta_1} + k_{\theta_2} = k_\theta)$ vary independently on respectively Θ_1, a compact subset of $\mathbb{R}^{k_{\theta_1}}$, and Θ_2, a compact subset of $\mathbb{R}^{k_{\theta_2}}$.

The semi-parametric model $\mathcal{S}$ as defined in Assumption 1 is just a special case of the one defined in Section 1.2. Proposition 3 outlines the fact that, even under the a priori favorable circumstances where mean and variance parameters vary independently, specifying the pseudo-densities f_t underlying a second order pseudo-maximum likelihood estimator as members of the quadratic exponential family does no longer appear as a sufficient condition for getting a consistent estimator of the mean-specific parameters when the model is correctly specified for the conditional mean but not jointly correctly specified for the conditional variance.

Proposition 3 *Suppose that $\hat{\theta}_n$ is as given in Definition 3 and that regularity conditions R1-R6 in Appendix A hold. Suppose further that the semi-parametric model $\mathcal{S}$ is as described in Assumption 1 and is first order correctly specified but second order misspecified, and that, for all $t = 1, 2, ...,$ f_t belongs to the quadratic exponential family. Then we may have that $\theta_n^* = Argmax_{\theta\in\Theta} E\left(L_n(Y^n, X^n, \theta)\right) \neq \left(\theta_1^{o\prime}, \theta_{2_n}^{*\prime}\right)'$ for all $n = 1, 2, ...,$ where θ_1^o is the true value of θ_1.*

Proof. Given regularity conditions R1-R6, from Theorem 3.5 of White (1994), we have $\hat{\theta}_n - \theta_n^* \to 0$ as $n \to \infty$ $a.s. - P_o$, where, given $\mathcal{S}$ as defined in Assumption 1, $\theta_n^* = \left(\theta_{1_n}^{*\prime}, \theta_{2_n}^{*\prime}\right)' = \text{Argmax}_{\theta\in\Theta} E\left(L_n(Y^n, X^n, \theta)\right)$. It is enough to show that for a given choice of P_o, $\mathcal{S}$ and $\{f_t\}$ which satisfy the assumptions of the Proposition, we may already have $\theta_{1_n}^{*\prime} \neq \theta_1^o$ for all $n = 1, 2, ...$ Suppose that P_o is such that Y_t is i.i.d. with $E\left(Y_t|X_t\right) = \theta_1^o$ and $V\left(Y_t|X_t\right) = \Sigma^o$, that $\mathcal{S}$ is specified as $\{m_t(X_t, \theta_1) = \theta_1\}$ and $\{\Omega_t(X_t, \theta_2) = \Omega\left(\theta_2\right)\}$, where $\Omega\left(.\right)$ is one-to-one and such that, $\forall\, \theta_2 \in \Theta_2$, $\Omega\left(\theta_2\right) \neq \Sigma^o$, and that, for all $t = 1, 2, ...,$ $f_t = f$, where f belongs to the quadratic exponential family. Then, we have that $E\left(L_n(Y^n, X^n, \theta)\right) = E\left(\ln f\left(Y_t, \theta_1, \Omega\left(\theta_2\right)\right)\right)$ is equal to

$$
\begin{aligned}
& E\left[A(\theta_1, \Omega\left(\theta_2\right)) + B(Y_t) + C(\theta_1, \Omega\left(\theta_2\right))'Y_t + \text{tr}\left(D(\theta_1, \Omega\left(\theta_2\right))Y_tY_t'\right)\right] \\
= {} & A(\theta_1, \Omega\left(\theta_2\right)) + E\left(B(Y_t)\right) + C(\theta_1, \Omega\left(\theta_2\right))'E\left(Y_t\right) + \text{tr}\left(D(\theta_1, \Omega\left(\theta_2\right))E(Y_tY_t')\right) \\
= {} & A(\theta_1, \Omega\left(\theta_2\right)) + E\left(B(Y_t)\right) + C(\theta_1, \Omega\left(\theta_2\right))'\theta_1^o + \text{tr}\left(D(\theta_1, \Omega\left(\theta_2\right))(\Sigma^o + \theta_1^o\theta_1^{o\prime})\right)
\end{aligned}
$$

From Property 2, since $\forall\, \theta_2 \in \Theta_2$, $\Omega\left(\theta_2\right) \neq \Sigma^o$, $\forall\, \theta_1^o$, it may exist $\theta_1^* \neq \theta_1^o$ ($\theta_{n_1}^* = \theta_1^*$, $n = 1, 2, ...$ because Y_t is i.i.d. and, f_t, m_t and Ω_t do not depend on t) such that we have

$$
\begin{aligned}
& A(\theta_1^o, \Omega\left(\theta_2\right)) + C(\theta_1^o, \Omega\left(\theta_2\right))'\theta_1^o + \text{tr}\left(D(\theta_1^o, \Omega\left(\theta_2\right))(\Sigma^o + \theta_1^o\theta_1^{o\prime})\right) \\
< {} & A(\theta_1^*, \Omega\left(\theta_2\right)) + C(\theta_1^*, \Omega\left(\theta_2\right))'\theta_1^o + \text{tr}\left(D(\theta_1^*, \Omega\left(\theta_2\right))(\Sigma^o + \theta_1^o\theta_1^{o\prime})\right)
\end{aligned}
$$

or, in other words, $\theta_n^* = \left(\theta_{1_n}^{*\prime}, \theta_{2_n}^{*\prime}\right)' = \text{Argmax}_{\theta\in\Theta} E\left(L_n(Y^n, X^n, \theta)\right) \neq \left(\theta_1^{o\prime}, \theta_{2_n}^{*\prime}\right)'$ for all $n = 1, 2, ...$ ■

The result of Proposition 3 is not in itself very strong. It does not say that PML2 will never be consistent for the mean-specific parameters when the semi-parametric model $\mathcal{S}$ is correctly specified for the conditional mean but not jointly correctly specified for the conditional variance. It does not even show that such inconsistency occurs in a particular case. It simply means that PML2, i.e., specifying the pseudo-densities f_t underlying a second order pseudo-maximum likelihood estimator as members of the quadratic exponential family, offers no guaranties for consistent estimation of the correctly specified part of the model. Clearly, what is true for models where mean and variance parameters vary independently is all the more so true for more general models, such that the following more general form of Proposition 3 obviously holds.

Corollary 4 *Suppose that $\hat{\theta}_n$ is as given in Definition 3 and that regularity conditions R1-R6 in Appendix A hold. Suppose further that the semi-parametric model $\mathcal{S}$ is first order correctly specified but second order misspecified, and that, for all $t = 1, 2, ...,$ f_t belongs to the quadratic exponential family. Then we may have that $\theta_n^* = Argmax_{\theta \in \Theta} E\left(L_n(Y^n, X^n, \theta)\right) \neq \theta^o$ for all $n = 1, 2, ...$*

Proof. This directly follows from Proposition 3 by considering the special case where mean and variance parameters vary independently. ■

1.5. Robust pseudo-maximum likelihood of order 2 (R1PML2 and RPML2)

We are ultimately looking for sufficient and essentially necessary conditions for a second order pseudo-maximum likelihood estimator to be, regardless of the true DGP P_o, consistent for both mean and variance parameters when the conditional mean and the conditional variance are jointly correctly specified, and to remain consistent for the mean parameters when the conditional mean is correctly specified but the conditional variance is not jointly correctly specified. Although they do not formally prove it, the results of Section 1.4.3 strongly suggest that the conditions underlying PML2 do not ensure such consistency properties.

In order to find out sufficient and essentially necessary conditions for a second order pseudo-maximum likelihood estimator to behave as just outlined, namely to be robust to conditional variance misspecification, it seems logical to first looking at sufficient and essentially necessary conditions for a second order pseudo-maximum likelihood estimator to be, regardless of the conditional variance (mis)specification, consistent for the mean parameters when the conditional mean is correctly specified. At this stage, further consistency for the variance parameters when second order correct specification also holds is not required. We entitle this intermediary class of second order pseudo-maximum likelihood estimators R1PML2. These conditions now formally imply the general inconsistency of PML2 under conditional variance misspecification. Then, by “mixing” the conditions underlying PML2 and R1PML2, we deduce sufficient and essentially necessary conditions for a second order pseudo-maximum likelihood estimator to be robust to conditional variance misspecification, i.e., to be what we call a RPML2 estimator.

1.5.1. R1PML2

Sufficient and essentially necessary conditions for a second order pseudo-maximum likelihood estimator to be, regardless of the conditional variance (mis)specification, consistent for the mean parameters when the conditional mean is correctly specified are based on a sub-family of generalized linear exponential families. We entitle this sub-family restricted generalized linear exponential families. We first define it and outline its main properties. Then, we investigate the consistency properties of the class of second order pseudo-maximum likelihood estimators based on it, i.e., R1PML2 estimators.

1.5.1.1. Restricted generalized linear exponential families

Restricted generalized linear exponential families may be defined as follows.

Definition 7 A family of probability measures on $\mathbb{R}^G$ indexed by $m \in \mathcal{M} \subset \mathbb{R}^G$ and $\Sigma \in \mathcal{E}$, where $\mathcal{E}$ is a subset of the $G \times G$ positive definite matrices, is called restricted generalized linear exponential if (a) every element of the family has a density function with respect to a given measure $\upsilon(dY)$ which may be written as

$$l(Y, m, \Sigma) = \exp\left(A(m, \Sigma) + B(\Sigma, Y) + C(m, \Sigma)'Y\right)$$

where $A(m, \Sigma)$ and $B(\Sigma, Y)$ are scalar, $C(m, \Sigma)$ is a $G \times 1$ vector, and (b) m is the mean and Σ is the covariance matrix of the distribution $l(Y, m, \Sigma)$.

The restricted generalized linear exponential family is just a special case of generalized linear exponential families, the family of density functions which underlies QGPML1. To see this, just recall that the "generic form" of generalized linear exponential families is $l(Y, m, \eta) = \exp\left(A(m, \eta) + B(\eta, Y) + C(m, \eta)'Y\right)$, where m is the mean of the distribution $l(Y, m, \eta)$ and the extra parameter η is, for any given m, one-to-one related with the "built into" covariance matrix Σ of $l(Y, m, \eta)$ through the function $\eta = \Gamma(m, \Sigma)$ (see Gourieroux-Monfort-Trognon (1984) or White (1994)). The result follows by letting $\eta = \Gamma(m, \Sigma) = \Sigma$. Note also that the restricted generalized linear exponential family does not contain the quadratic exponential family. For this to be true, $B(\Sigma, Y)$ should be allowed to depend on m, a feature ruled out in the above definition. Finally, remark that, as it may be readily seen from (1.1), the normal density is a member — and undoubtedly again the prominent one — of this sub-family of the generalized linear exponential family.

Restricted generalized linear exponential families have essentially the same properties than generalized linear exponential families.

Property 4 If $l(Y, m, \Sigma)$ is a restricted generalized linear exponential family, then $\forall\ m,\ m_o \in \mathcal{M}$, $\forall\ \Sigma,\ \Sigma_o \in \mathcal{E}$, we have (a)

$$\begin{aligned} &A(m_o, \Sigma_o) + E_{l_o}\left[B(\Sigma_o, Y)\right] + C(m_o, \Sigma_o)'m_o \\ &\geq A(m, \Sigma) + E_{l_o}\left[B(\Sigma, Y)\right] + C(m, \Sigma)'m_o \end{aligned}$$

where $E_{l_o}[.]$ denotes expectation taken with respect to $l(Y, m_o, \Sigma_o)$ and the equality holds if and only if $m = m_o$ and $\Sigma = \Sigma_o$, and (b)

$$\begin{aligned} &A(m_o, \Sigma_o) + C(m_o, \Sigma_o)'m_o \\ &\geq A(m, \Sigma_o) + C(m, \Sigma_o)'m_o \end{aligned}$$

where the equality holds, $\forall\ \Sigma_o$, if and only if $m = m_o$.

Proof. See Appendix B. ■

Property 5 If $l(Y, m, \Sigma)$ is a restricted generalized linear exponential family, and if the functions $A(m, \Sigma)$, $B(\Sigma, Y)$ and $C(m, \Sigma)$ are continuously differentiable with respect to m and Σ on respectively $int\, \mathcal{M}$ and $int\, \mathcal{E}$, then $\forall\ m \in int\, \mathcal{M}$, $\forall\ \Sigma \in int\, \mathcal{E}$, we have (a)

$$\frac{\partial A(m, \Sigma)}{\partial m} + \frac{\partial C(m, \Sigma)'}{\partial m} m = 0$$

$$\frac{\partial A(m, \Sigma)}{\partial \operatorname{vec} \Sigma} + E_l \left[\frac{\partial B(\Sigma, Y)}{\partial \operatorname{vec} \Sigma} \right] + \frac{\partial C(m, \Sigma)'}{\partial \operatorname{vec} \Sigma} m = 0$$

where $E_l\,[.]$ denotes expectation taken with respect to $l(Y, m, \Sigma)$, and (b)

$$\frac{\partial C(m, \Sigma)'}{\partial m} = \Sigma^{-1}$$

Proof. See Appendix B. ■

Property 6 If $l(Y, m, \Sigma)$ is a restricted generalized linear exponential family and if the functions $A(m, \Sigma)$, $B(\Sigma, Y)$ and $C(m, \Sigma)$ are continuously differentiable with respect to m and Σ on respectively $int\, \mathcal{M}$ and $int\, \mathcal{E}$, then we cannot have, $\forall\ Y \in \mathcal{Y}$, where $\mathcal{Y}$ denotes the support of Y, $\forall\ m \in int\, \mathcal{M}$, $\forall\ \Sigma \in int\, \mathcal{E}$,

$$\frac{\partial B(\Sigma, Y)}{\partial \operatorname{vec} \Sigma} = 0$$

Proof. See Appendix B. ■

The crucial property of the restricted generalized linear exponential family is Property 4(b). The insensitivity to the conditional variance (mis)specification of R1PML2 basically relies on this property. Note that it is a similar property which underlies the robustness to conditional variance misspecification of QGPML1. On the other hand, the general inconsistency for variance parameters of R1PML2 despite second order correct specification essentially stems from Property 4(a). Property 6 outlines the fact that for $l(Y, m, \Sigma)$ to be a proper restricted generalized linear exponential family, the term $B(.)$ has to depend on Σ. This fact, as well as Property 5(a), will serve when demonstrating Proposition 6 in the next section. Finally, Property 5(b) will be used later for outlining a property of restricted quadratic exponential families.

1.5.1.2. Consistency of R1PML2 under first order correct specification but possible second order misspecification

Throughout this section, we assume that the semi-parametric model $\mathcal{S}$ is at least correctly specified for the conditional mean, while possibly second order misspecified. Further, as in section 1.4.3, we assume for now that the vector of parameters θ is partitioned as $\theta = (\theta_1', \theta_2')'$, where θ_1 and θ_2 respectively denote mean-specific and variance-specific parameters, such that the semi-parametric model $\mathcal{S}$ is as described in Assumption 1.

Sufficient conditions for a second order pseudo-maximum likelihood estimator to be, regardless of the conditional variance (mis)specification, consistent for the assumed correctly specified part of $\mathcal{S}$ are given in the following proposition.

Proposition 5 *Suppose that $\hat{\theta}_n$ is as given in Definition 3 and that regularity conditions R1-R6 in Appendix A hold. If the semi-parametric model $\mathcal{S}$ is as described in Assumption 1 and is first order correctly specified, and if, for all $t = 1, 2, ...,$ f_t belongs to the restricted generalized linear exponential family, then $\hat{\theta}_n - \theta_n^* \to 0$ as $n \to \infty$ a.s. $- P_o$, where $\theta_n^* = \left(\theta_1^{o\prime}, \theta_{2_n}^{*\prime}\right)'$ and θ_1^o is the true value of θ_1.*

Proof. Given regularity conditions R1-R6, from Theorem 3.5 of White (1994), we have $\hat{\theta}_n - \theta_n^* \to 0$ as $n \to \infty$ *a.s.* $- P_o$, where, given $\mathcal{S}$ as defined in Assumption 1, $\theta_n^* = \left(\theta_{1_n}^{*\prime}, \theta_{2_n}^{*\prime}\right)' = \text{Argmax}_{\theta \in \Theta} E\left(L_n(Y^n, X^n, \theta)\right)$. Thus, it is enough to show that $\theta_{1_n}^* = \theta_1^o$ for all $n = 1, 2, ...$ Since, for all $t = 1, 2, ...,$ f_t belongs to the restricted generalized linear exponential family, letting m_t stand for $m_t(X_t, \theta_1)$ and Ω_t stand for $\Omega_t(X_t, \theta_2)$, we have that, for all $t = 1, 2, ...,$ $E\left(\ln f_t\left(Y_t, m_t(X_t, \theta_1), \Omega_t(X_t, \theta_2)\right)\right)$ is equal to

$$E\left[A_t(m_t, \Omega_t) + B_t(\Omega_t, Y_t) + C_t(m_t, \Omega_t)' Y_t\right] \tag{1.4}$$

Given first order correct specification, $E\left(Y_t | X_t\right)$ by definition exists and the law of iterated expectations applies such that (1.4) may be written

$$E\left[A_t(m_t, \Omega_t) + B_t(\Omega_t, Y_t) + C_t(m_t, \Omega_t)' E\left(Y_t | X_t\right)\right] \tag{1.5}$$

Since $B_t(\Omega_t, Y_t)$ does not depend on m_t, from Property 4(b), $\forall\ \Omega_t \in \mathcal{E}_t$, (1.5) has a unique maximum in m_t when $m_t(X_t, \theta_1) = E\left(Y_t | X_t\right)$, or, given that θ_1 and θ_2 vary independently and first order correct specification, $\forall\ \theta_2 \in \Theta_2$, a maximum in θ_1 at $\theta_1 = \theta_1^o$. The identifiable uniqueness of $\{\theta_n^*\}$ ensures that, $\forall\ \theta_{2_n}^* \in \Theta_2$, θ_1^o is the unique maximum in θ_1 of $E\left(L_n(Y^n, X^n, \theta)\right)$, i.e., that $\theta_{1_n}^* = \theta_1^o$ for all $n = 1, 2, ...$ ■

In other words, provided that mean and variance parameters vary independently, a second order pseudo-maximum likelihood estimator obtained by specifying the pseudo-densities f_t as members of the restricted generalized linear exponential family, i.e., R1PML2, yields a consistent estimator of the true mean parameters value of a first order correctly specified semi-parametric model $\mathcal{S}$ regardless of the true DPG P_o and the conditional variance (mis)specification, i.e., regardless of whether or not the implicit parametric model $\mathcal{P}$ corresponding to $\hat{\theta}_n$ is correctly specified for other aspects of the "true conditional densities" of Y_t given X_t, and thus in particular whether or not these "true underlying densities" are in the restricted generalized linear exponential family. This result contains as a particular case — but under much less restrictive assumptions — Theorem 1 of Pagan-Sabau (1991).

Note that in Proposition 5, the forms of the conditional variance (mis)specification allowed are only restricted through the "compatibility" assumption contained in Definition 3 and the regularity conditions. In this latter respect, it is worth noting that Assumption R6 is actually unnecessary restrictive. This assumption requires that $\{E\left(L_n(Y^n, X^n, \theta)\right)\}$ has identifiably unique maximizers $\left\{\theta_n^* = \left(\theta_{1_n}^{*\prime}, \theta_{2_n}^{*\prime}\right)'\right\}$ on

$\Theta = \Theta_1 \times \Theta_2$. However, for the most important part of Proposition 5 to hold, i.e., $\hat{\theta}_{1_n} \to \theta_1^o$ as $n \to \infty$ $a.s.{-}P_o$, it is not necessary that $\hat{\theta}_{2_n}$ converges to the well-defined unique quantity $\theta_{2_n}^* = \text{Argmax}_{\theta_2 \in \Theta_2} \frac{1}{n} \sum_{t=1}^n E(\ln f_t(Y_t, m_t(X_t, \theta_1^o), \Omega_t(X_t, \theta_2)))$. For example, multiple maximums could be allowed. The identifiability of θ_1^o is of course required. As above, such an identifiability condition typically holds under the more primitive — but also more restrictive — assumption that $\mathcal{S}$ is first order identifiable, i.e., that $\forall\ \theta_1,\ \theta_1^o \in \Theta_1$

$$m_t(X_t, \theta_1) = m_t(X_t, \theta_1^o) \Rightarrow \theta_1 = \theta_1^o, \qquad a.s. - P_o, \quad t = 1, 2, ...$$

By the way, it is also worth recalling the remark made at the end of Section 1.2: for judging conditional mean or conditional variance specification, we must take into account the variables which appear in both moments. In other words, specifying the conditional variance as functions of variables which do not enter in the conditional mean may dismantle first order correct specification.

Noticeably, if $\mathcal{S}$ is in addition second order correctly specified, according to Proposition 2, R1PML2 will usually not further provide a consistent estimator of the variance parameters. It simply follows from the fact that members of the restricted generalized linear exponential family are not necessarily — although some are, see below — members of the quadratic exponential family. Another way to see this is to remember Property 4(a). From this property, it is easily seen that unless the implicit parametric model $\mathcal{P}$ is correctly specified for the conditional density, in which case $\hat{\theta}_n$ is just a genuine maximum likelihood estimator, expectations taken with respect to the true DGP P_o and expectations taken with respect to the pseudo-densities $\{\lambda_t(Y_t, X_t, \theta^o)\}$ — in terms of Property 4(a), $E_{l_o}(.)$ — will usually differ and thus, although always maximized in θ_1 at θ_1^o, $E(L_n(Y^n, X^n, \theta))$ will usually not be maximized in θ_2 at the true value θ_2^o.

Just as for Proposition 1, there exists a reciprocal to Proposition 5 which shows that the outlined conditions for consistent estimation of the assumed correctly specified part of $\mathcal{S}$ are not only sufficient but also necessary in a sense.

To show that, it is convenient to again reparametrize the semi-parametric model $\mathcal{S}$. We assume for now that the vector of parameters θ is partitioned as $\theta = (\theta_1', \theta_2')'$, where θ_1 and θ_2 respectively denote mean-and-variance-common and variance-specific parameters, such that the semi-parametric model $\mathcal{S}$ is as described in the following assumption.

Assumption 2 The semi-parametric model $\mathcal{S}$ is such that for θ partitioned as $\theta = (\theta_1', \theta_2')'$ and Θ accordingly defined as $\Theta = \Theta_1 \times \Theta_2$, we have

$$\mathcal{S} \equiv \begin{cases} \{m_t(X_t, \theta) = m_t(X_t, \theta_1) : X_t \in \mathcal{X}_t,\ \theta_1 \in \Theta_1 \subset \mathbb{R}^{k_{\theta_1}}\} \\ \{\Omega_t(X_t, \theta) = \Omega_t(X_t, \theta_1, \theta_2) : X_t \in \mathcal{X}_t,\ \theta_1 \in \Theta_1 \subset \mathbb{R}^{k_{\theta_1}}, \theta_2 \in \Theta_2 \subset \mathbb{R}^{k_{\theta_2}}\} \end{cases},$$

$t = 1, 2, ...$ where the $k_{\theta_1} \times 1$ vector of parameters θ_1 and the $k_{\theta_2} \times 1$ vector of parameters θ_2 ($k_{\theta_1} + k_{\theta_2} = k_\theta$) vary independently on respectively Θ_1, a compact subset of $\mathbb{R}^{k_{\theta_1}}$, and Θ_2, a compact subset of $\mathbb{R}^{k_{\theta_2}}$.

As defined in Assumption 2, $\mathcal{S}$ may either be taken literally or be viewed as the "reduced-form" of a more general model incorporating cross-constraints between mean and variance parameters. We are now ready to state our converse of Proposition 5.

Proposition 6 *Suppose that $\hat{\theta}_n$ is as given in Definition 3, where $\mathcal{M}_t$ is the closure of an open connected set, and that the semi-parametric model $\mathcal{S}$ is as described in Assumption 2. If for any probability measure P_o such that regularity conditions R1-R5, R6′, R7-R9 in Appendix A hold, when the semi-parametric model $\mathcal{S}$ is first order correctly specified, we have that $\theta_n^* = Argmax_{\theta \in \Theta} E\left(L_n(Y^n, X^n, \theta)\right) = \left(\theta_1^{o\prime}, \theta_{2_n}^{*\prime}\right)'$ for all $n = 1, 2, ...,$ where θ_1^o is the true value of θ_1, then, for all $t = 1, 2, ...,$ f_t belongs to the restricted generalized linear exponential family and $\Omega_t(X_t, \theta_1, \theta_{2_n}^*) = \Omega_t(X_t, \theta_{2_n}^*)$, $\forall\ \theta_1 \in int\,\Theta_1$, $\forall\ \theta_{2_n}^* \in int\,\Theta_2$ and $\forall\ X_t \in \mathcal{X}_t$.*

Proof. See Appendix C. ■

As Proposition 2, Proposition 6 imposes additional regularity conditions, in particular differentiability (R7 and R8) as well as interiority of θ_n^* (R6′), and thus is not entirely a converse of Proposition 5. Keeping this minor qualification in mind, Proposition 6 basically says that a specification of f_t belonging to restricted generalized linear exponential families and functional independence between mean and variance parameters are essentially necessary conditions for a second order pseudo-maximum likelihood estimator to yield a consistent estimator of the mean parameters true value of a first order correctly specified semi-parametric model $\mathcal{S}$ regardless of the true DGP P_o and the conditional variance (mis)specification. In other words, again only a very limited subset of the class of second order pseudo-maximum likelihood estimators is jointly insensitive to distributional misspecification and conditional variance (mis)specification.

If P_o and the kind of allowed conditional variance (mis)specification were further restricted, these necessary conditions would no longer hold. As a matter of fact, Pagan-Sabau (1991) gives conditions under which the univariate gaussian maximum likelihood estimator of certain conditional variance misspecified ARCH regression models provides a consistent estimator of the mean parameters despite functional links between mean and variance parameters.

An important corollary of Proposition 6 is that, even under the a priori favorable circumstances where mean and variance parameters vary independently, PML2 is generally not robust to conditional variance misspecification. Again, it follows from the fact that members of the quadratic exponential family are not necessarily members of the restricted generalized linear exponential family.

1.5.2. RPML2

Let us summarize the results that we already obtained. From Proposition 1 and 2, we know that a sufficient and essentially necessary condition for a second order pseudo-maximum likelihood estimator $\hat{\theta}_n$ to be, regardless of the true DGP P_o, consistent for the true parameters value of a second order correctly specified semi-parametric model $\mathcal{S}$ is to specify the pseudo-densities f_t as members of the quadratic exponential family. On the other hand, from Proposition 5 and 6, we know that sufficient and essentially necessary conditions for such an estimator $\hat{\theta}_n$ to be, regardless of the true DGP P_o and the conditional variance (mis)specification, consistent for the mean parameters true value of a first order correctly specified semi-parametric model $\mathcal{S}$ are that the pseudo-densities f_t belong to the restricted generalized linear exponential family and that mean and variance parameters vary independently.

Mixing the sufficient conditions of Proposition 1 and Proposition 5 in order to get sufficient conditions for a second order pseudo-maximum likelihood estimator $\hat{\theta}_n$ to be robust to conditional variance misspecification, namely to be, regardless of the true DGP P_o, consistent for both mean and variance parameters when the conditional mean and the conditional variance are jointly correctly specified, and to remain consistent for the mean parameters when the conditional mean is correctly specified but the conditional variance is not jointly correctly specified, basically means finding a family of density functions which jointly belongs to both quadratic exponential families and restricted generalized linear exponential families. On the other hand, mixing the essentially necessary conditions of Proposition 2 and Proposition 6 in order to get essentially necessary conditions for such an estimator $\hat{\theta}_n$ to be robust to conditional variance misspecification basically means finding the largest family of density functions which jointly belongs to both quadratic exponential families and restricted generalized linear exponential families. We entitle this family of density functions restricted quadratic exponential families. We first define it and sketch out its main properties. Then, we outline the consistency properties of the class of second order pseudo-maximum likelihood estimators based on it, i.e., RPML2 estimators.

1.5.2.1. Restricted quadratic exponential families

Restricted quadratic exponential families may be defined as follows.

Definition 8 A family of probability measures on $\mathbb{R}^G$ indexed by $m \in \mathcal{M} \subset \mathbb{R}^G$ and $\Sigma \in \mathcal{E}$, where $\mathcal{E}$ is a subset of the $G \times G$ positive definite matrices, is called restricted quadratic exponential if (a) every element of the family has a density function with respect to a given measure $\upsilon(dY)$ which may be written as

$$l(Y, m, \Sigma) = \exp\left(A(m, \Sigma) + B(Y) + C(m, \Sigma)'Y + Y'D(\Sigma)Y\right)$$

where $A(m, \Sigma)$ and $B(Y)$ are scalar, $C(m, \Sigma)$ is a $G \times 1$ vector and $D(\Sigma)$ is a $G \times G$ matrix, and (b) m is the mean and Σ is the covariance matrix of the distribution $l(Y, m, \Sigma)$.

The only difference between quadratic exponential families and restricted quadratic exponential families is that in the expression of the latter, the $G \times G$ matrix $D(.)$ does no longer depend on the mean m. While not preventing it from still being a member of quadratic exponential families, this small change makes the restricted quadratic exponential family a special case of restricted generalized linear exponential families, a special case where the term $B(\Sigma, Y)$ appearing in the expression of the latter is simply given by $B(Y) + Y'D(\Sigma)Y$, i.e., as required, does not depend on m. Further, it is readily seen that the restricted quadratic exponential family is indeed the largest class of density functions which jointly belongs to both quadratic exponential families and restricted generalized linear exponential families. By the way, note that, as it may be easily checked from (1.1), the normal density is still a member — and undoubtedly again the prominent one — of this family.

Since the restricted quadratic exponential family is a sub-family of both quadratic exponential families and restricted generalized linear exponential families, it obviously inherits all their properties. It also has some additional properties.

Property 7 If $l(Y, m, \Sigma)$ is a restricted quadratic exponential family, then $\forall$ m, $m_o \in \mathcal{M}$, $\forall$ Σ, $\Sigma_o \in \mathcal{E}$, we have (a)

$$\begin{aligned} & A(m_o, \Sigma_o) + C(m_o, \Sigma_o)' m_o + \operatorname{tr}(D(\Sigma_o)(\Sigma_o + m_o m_o')) \\ & \geq A(m, \Sigma) + C(m, \Sigma)' m_o + \operatorname{tr}(D(\Sigma)(\Sigma_o + m_o m_o')) \end{aligned}$$

where the equality holds if and only if $m = m_o$ and $\Sigma = \Sigma_o$, and (b)

$$\begin{aligned} & A(m_o, \Sigma_o) + C(m_o, \Sigma_o)' m_o \\ & \geq A(m, \Sigma_o) + C(m, \Sigma_o)' m_o \end{aligned}$$

where the equality holds, $\forall$ Σ_o, if and only if $m = m_o$.

Proof. See Appendix B. ■

Property 8 If $l(Y, m, \Sigma)$ is a restricted quadratic exponential family and if the functions $A(m, \Sigma)$, $C(m, \Sigma)$ and $D(\Sigma)$ are continuously differentiable with respect to m and Σ on respectively $int\, \mathcal{M}$ and $int\, \mathcal{E}$, then $\forall$ $m \in int\, \mathcal{M}$, $\forall$ $\Sigma \in int\, \mathcal{E}$, we have (a)

$$\frac{\partial A(m, \Sigma)}{\partial m} + \frac{\partial C(m, \Sigma)'}{\partial m} m = 0$$

$$\frac{\partial A(m, \Sigma)}{\partial \operatorname{vec} \Sigma} + \frac{\partial C(m, \Sigma)'}{\partial \operatorname{vec} \Sigma} m + \frac{\partial (\operatorname{vec} D(\Sigma))'}{\partial \operatorname{vec} \Sigma} \operatorname{vec}(\Sigma + mm') = 0$$

(b)

$$\frac{\partial C(m, \Sigma)'}{\partial m} = \Sigma^{-1}$$

(c)

$$\frac{\partial C(m, \Sigma)'}{\partial \operatorname{vec} \Sigma} = -\frac{\partial (\operatorname{vec} D(\Sigma))'}{\partial \operatorname{vec} \Sigma} Cov_l \left[(\operatorname{vec}(YY'), Y) \right] \Sigma^{-1}$$

$$\frac{\partial C(m, \Sigma)'}{\partial \operatorname{vec} \Sigma} Cov_l \left[(Y, \operatorname{vec}(YY')) \right] = I_{G^2} - \frac{\partial (\operatorname{vec} D(\Sigma))'}{\partial \operatorname{vec} \Sigma} V_l \left[\operatorname{vec}(YY') \right]$$

and (d)

$$\frac{\partial (\operatorname{vec} D(\Sigma))'}{\partial \operatorname{vec} \Sigma} N_G = \left(D_{uG}^{+} \right)' \left(\begin{array}{c} V_l \left[\operatorname{vech}(YY') \right] - Cov_l \left[(\operatorname{vech}(YY'), Y) \right] \\ \Sigma^{-1} Cov_l \left[(Y, \operatorname{vech}(YY')) \right] \end{array} \right)^{-1} D_{uG}^{+}$$

where[3] $Cov_l[.]$ and $V_l[.]$ denote respectively covariance and variance taken with respect to $l(Y, m, \Sigma)$, $D^+_{uG} = (D'_{uG}D_{uG})^{-1} D'_{uG}$, $N_G = D_{uG}D^+_{uG} = \frac{1}{2}(I_{G^2} + K_{GG})$, D_{uG} is the $G^2 \times \frac{1}{2}G(G+1)$ duplication matrix, i.e., a matrix such that, for any symmetric $G \times G$ matrix A, $D_{uG} \operatorname{vech} A = \operatorname{vec} A$, and K_{GG} is the $G^2 \times G^2$ commutation matrix, i.e., a matrix such that, for any $G \times G$ matrix A, $K_{GG} \operatorname{vec} A = \operatorname{vec} A'$.

Proof. See Appendix B. ■

The robustness to conditional variance misspecification of RPML2 basically relies on Property 7. Just reported for the sake of clarity, Property 7 outlines the way in which restricted quadratic exponential families cumulate the nice properties of quadratic exponential families and restricted generalized linear exponential families. Property 8 will be used later for deriving the limiting distribution of RPML2.

1.5.2.2. Consistency of RPML2 under first order correct specification but possible second order misspecification

As in Section 1.5.1.2, we assume throughout this section that the semi-parametric model $\mathcal{S}$ is at least correctly specified for the conditional mean, while possibly second order misspecified. Further, we assume for now that the vector of parameter θ is partitioned as $\theta = (\theta_1', \theta_2')'$, where θ_1 and θ_2 denote respectively mean-specific and variance-specific parameters, such that the semi-parametric model $\mathcal{S}$ is as described in Assumption 1.

Obtained by mixing the conditions of Proposition 1 and Proposition 5, sufficient conditions for a second order pseudo-maximum likelihood estimator to be robust to conditional variance misspecification are given in the following proposition.

Proposition 7 *Suppose that $\hat{\theta}_n$ is as given in Definition 3 and that regularity conditions R1-R6 in Appendix A hold. If the semi-parametric model $\mathcal{S}$ is as described in Assumption 1 and is first order correctly specified, and if, for all $t = 1, 2, ...,$ f_t belongs to the restricted quadratic exponential family, then $\hat{\theta}_n - \theta^*_n \to 0$ as $n \to \infty$ a.s. $- P_o$, where $\theta^*_n = \left(\theta_1^{o\prime}, \theta_{2_n}^{*\prime}\right)'$ and θ_1^o is the true value of θ_1. If, in addition, the semi-parametric model $\mathcal{S}$ is also second order correctly specified, then $\hat{\theta}_n \to \theta^o$ as $n \to \infty$ a.s. $- P_o$, where $\theta^o = (\theta_1^{o\prime}, \theta_2^{o\prime})'$ and θ_2^o is the true value of θ_2.*

Proof. Since restricted quadratic exponential families are jointly members of both quadratic exponential families and restricted generalized linear exponential families, it directly follows from Proposition 1 and Proposition 5. ■

In other words, provided that mean and variance parameters vary independently, a second order pseudo-maximum likelihood estimator obtained by specifying the pseudo-densities f_t as members of the restricted quadratic exponential family yields, regardless of the true DGP P_o — i.e., regardless of whether or not the implicit parametric model $\mathcal{P}$ corresponding to $\hat{\theta}_n$ is correctly specified for other aspects of the "true conditional densities" of Y_t given X_t, and thus in particular whether or not these "true underlying densities" are in the restricted quadratic exponential family —, a consistent estimator of both mean and variance parameters when the

[3] For details on the properties of D_{uG}, K_{GG} and N_G, see Magnus-Neudecker (1986).

semi-parametric model $\mathcal{S}$ is jointly correctly specified for the conditional mean and the conditional variance, and a consistent estimator of the mean-specific parameters when the semi-parametric model $\mathcal{S}$ is correctly specified for the conditional mean but not jointly correctly specified for the conditional variance.

As outlined for PML2, because in the definition of the restricted quadratic exponential family $B(Y)$ does not depend on parameters, note that the terms $B_t(Y_t)$ may be dropped from the pseudo log-likelihood such that it is not necessary to impose on Y_t the "compatibility" constraints which may be implied by the definition of $B_t(.)$. Recall also that, as outlined after Proposition 5, the form of the conditional variance misspecification allowed is only restricted through the "compatibility" assumption contained in Definition 3 and the regularity conditions.

Again by mixing the conditions of Proposition 2 and Proposition 6, at least for $G = 1$, a converse of Proposition 7 may be obtained. So, Proposition 8 shows that the outlined conditions for robust to conditional variance misspecification estimation of $\mathcal{S}$ are not only sufficient but also necessary in a sense.

As above, for convenience, we assume for now that the vector of parameter θ is partitioned as $\theta = (\theta_1', \theta_2')'$, where θ_1 and θ_2 denote respectively mean-and-variance-common and variance-specific parameters, such that the semi-parametric model $\mathcal{S}$ is as described in Assumption 2.

Proposition 8 $(G = 1)$ *Suppose that $\hat{\theta}_n$ is as given in Definition 3, where $\mathcal{M}_t$ and $\mathcal{E}_t$ are closures of open connected sets, and that the semi-parametric model $\mathcal{S}$ is as described in Assumption 2. If for any probability measure P_o such that conditions R1-R5, R6$'$, R7-R9 in Appendix A hold, when the semi-parametric model $\mathcal{S}$ is first order correctly specified, we have that $\theta_n^* = Argmax_{\theta\in\Theta} E\left(L_n(Y^n, X^n, \theta)\right) = \left(\theta_1^{o\prime}, \theta_{2_n}^{*\prime}\right)'$ for all $n = 1, 2, ...$, where θ_1^o is the true value of θ_1, and when, in addition, the semi-parametric model $\mathcal{S}$ is also second order correctly specified, we have that $\theta_n^* = Argmax_{\theta\in\Theta} E\left(L_n(Y^n, X^n, \theta)\right) = (\theta_1^{o\prime}, \theta_2^{o\prime})'$ for all $n = 1, 2, ...$, where θ_2^o is the true value of θ_2, then, for all $t = 1, 2, ...$, f_t belongs to the restricted quadratic exponential family and $\Omega_t(X_t, \theta_1, \theta_{2_n}^*) = \Omega_t(X_t, \theta_{2_n}^*)$, $\forall\ \theta_1 \in int\,\Theta_1$, $\forall\ \theta_{2_n}^* \in int\,\Theta_2$ and $\forall\ X_t \in \mathcal{X}_t$.*

Proof. $(G = 1)$ Since the restricted quadratic exponential family is the largest family of density functions which jointly belongs to both quadratic exponential families and restricted generalized linear exponential families, it directly follows from Proposition 2 and Proposition 6. ■

As Proposition 2 and 6, Proposition 8 is not entirely a converse of Proposition 7. Keeping this qualifying statement in mind, Proposition 8 basically says that both a specification of f_t belonging to restricted quadratic exponential families and functional independence between mean and variance parameters are essentially necessary conditions for a second order pseudo-maximum likelihood estimator to yield, regardless of the true DGP P_o, a consistent estimator of both mean and variance parameters when the semi-parametric model $\mathcal{S}$ is second order correctly specified, and a consistent estimator of the mean parameters when the semi-parametric model $\mathcal{S}$ is first order correctly specified but second order misspecified. In other words, again only a very narrow subset of the class of second order pseudo-maximum likelihood estimators, and in particular a narrow subset of PML2 estimators, is robust to conditional variance misspecification.

Before concluding this section, let us stress that the requirement that mean and variance parameters have to vary independently for robust estimation does not signify that mean and variance parameters have to be functionally unrelated in the structural model, but simply that they have to be treated as if they were not functionally related. In other words, eventual structural cross-constraints between mean and variance parameters have to be discarded. For example, if the structural semi-parametric model $\mathcal{S}$ of interest is as given in Assumption 2, RPML2 means that, for robustness, it has to be treated as if it were given by

$$\mathcal{S} \equiv \begin{cases} \left\{m_t(X_t, \theta) = m_t(X_t, \theta_1) : X_t \in \mathcal{X}_t,\ \theta_1 \in \Theta_1 \subset \mathbb{R}^{k_{\theta_1}}\right\} \\ \left\{\Omega_t(X_t, \theta) = \Omega_t(X_t, \theta_{12}, \theta_2) : X_t \in \mathcal{X}_t,\ \theta_{12} \in \Theta_{\theta_{12}} \subset \mathbb{R}^{k_{\theta_1}},\ \theta_2 \in \Theta_2 \subset \mathbb{R}^{k_{\theta_2}}\right\} \end{cases},$$

$t = 1, 2, \ldots$ where the vector of parameters θ_1 and θ_{12} are now assumed to vary independently. Note that QGPML1 means exactly the same thing, but is implemented in three steps rather than in one step.

Although in most cases it should not entail any problems — it is for example true for GARCH models —, in certain cases, discarding cross-constraints between mean and variance parameters while continuing to jointly estimate them could lead to identification problems. If, as it may be expected, the identification problem arises in the variance, as an alternative to RPML2 and QGPML1, a quasi-generalized robust pseudo-maximum likelihood of order 2 estimation procedure (hereafter denoted QGRPML2) may be proposed. It is a two-step procedure which basically amounts to first estimating the conditional mean parameters — presumably by PML1 — and then applying a RPML2-like procedure where the mean parameters which appear in the conditional variance, in the above example θ_{12}, are replaced by their first step estimates. This allows to bypass one of the three steps implied by QGPML1, namely the estimation of the variance-specific parameters which, as outlined above, is potentially the hardiest, in particular in multivariate cases. Clearly, if the structural model does not contain any variance-specific parameters, i.e., if all variance parameters appear in the conditional mean, then QGRPML2 collapses to QGPML1. Using Property 7 and the general results of White (1994) for quasi-generalized pseudo-maximum likelihood, it is not very complicated to show that this estimator is indeed consistent for the mean parameters true value of a first order correctly specified semi-parametric model $\mathcal{S}$ regardless of the true DGP P_o and possible second order misspecification, while consistent for the mean and variance parameters true value under second order correct specification, again regardless of the true DGP P_o.

1.6. Limiting distribution of RPML2

We now examine the limiting distribution of RPML2 estimators under different assumptions regarding the degree of misspecification present in the model. We first state a classical limiting distribution result for M-estimators.

Proposition 9 *Suppose that $\hat{\theta}_n$ is as given in Definition 3 and that regularity conditions R1-R5, R6′-R8′, R9-R12 in Appendix A hold. Then $\hat{\theta}_n - \theta_n^* \to 0$ as $n \to \infty$*

a.s. $- P_o$,

$$\sqrt{n}(\hat{\theta}_n - \theta_n^*) = -A_n^{*^{-1}}\sqrt{n}\frac{\partial}{\partial\theta}L_n(Y^n, X^n, \theta_n^*) + o_{P_o}(1)$$

and

$$B_n^{*-1/2}A_n^*\sqrt{n}(\hat{\theta}_n - \theta_n^*) \xrightarrow{d} N(0, I_{k_\theta})$$

where $\theta_n^* = Argmax_{\theta\in\Theta}E\left(L_n(Y^n, X^n, \theta)\right)$, $A_n^* = E\left(\frac{\partial^2}{\partial\theta\partial\theta'}L_n(Y^n, X^n, \theta_n^*)\right)$, $B_n^* = V\left(n^{1/2}\frac{\partial}{\partial\theta}L_n(Y^n, X^n, \theta_n^*)\right)$, *so that avar* $\hat{\theta}_n = C_n^* = A_n^{*^{-1}}B_n^*A_n^{*^{-1}}$.

Proof. It follows from Theorem 3.5 and Theorem 6.4 of White (1994). ■

Proposition 9 describes the asymptotic behavior and the asymptotic covariance matrix of an arbitrary second order pseudo-maximum likelihood estimator $\hat{\theta}_n$ under arbitrary misspecification. All subsequent results are particular cases of this result.

According to the conditions underlying RPML2 estimation, throughout this section, it is assumed both that the pseudo-densities f_t used to form $\hat{\theta}_n$ all belong to restricted quadratic exponential families and that the semi-parametric model $\mathcal{S}$ is as given in Assumption 1, i.e., that mean and variance parameters vary independently, either as a consequence of the structural model or because, as required for robustness, structural cross-constraints has been discarded. Further, we will always maintain the hypothesis that the model is first order correctly specified. As outlined by Proposition 10, this minimal assumption already ensures a nice structure for the asymptotic covariance matrix C_n^* of RPML2 estimators, a nice structure which basically follows from the block-diagonality of A_n^* between mean and variance parameters. By the way, note that this nice structure does generally not hold for PML2 estimators, even if mean and variance parameters vary independently.

Proposition 10 *Suppose that all the assumptions of Proposition 9 hold. If, for all* $t = 1, 2, ...,$ f_t *belongs to the restricted quadratic exponential family, and if the semi-parametric model* $\mathcal{S}$ *is as described in Assumption 1 and is first order correctly specified, then, for all* $n = 1, 2, ...,$ $\theta_n^* = \left(\theta_1^{o\prime}, \theta_{2_n}^{*\prime}\right)'$ *and we have*

$$A_{n_{12}}^* = A_{n_{21}}^{*\prime} = E\left[\frac{\partial^2}{\partial\theta_1\partial\theta_2'}L_n(Y^n, X^n, \theta_n^*)\right] = 0$$

such that

$$C_n^* = \begin{bmatrix} C_{n_{11}}^* & C_{n_{12}}^* \\ C_{n_{12}}^{*\prime} & C_{n_{22}}^* \end{bmatrix} = \begin{bmatrix} A_{n_{11}}^{*^{-1}}B_{n_{11}}^*A_{n_{11}}^{*^{-1}} & A_{n_{11}}^{*^{-1}}B_{n_{12}}^*A_{n_{22}}^{*^{-1}} \\ A_{n_{22}}^{*^{-1}}B_{n_{12}}^{*\prime}A_{n_{11}}^{*^{-1}} & A_{n_{22}}^{*^{-1}}B_{n_{22}}^*A_{n_{22}}^{*^{-1}} \end{bmatrix}$$

where

$$A_{n_{11}}^* = -\frac{1}{n}\sum_{t=1}^{n}E\left[\frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{*^{-1}}\frac{\partial m_t^o}{\partial\theta_1'}\right]$$

$$A_{n_{22}}^* = -\frac{1}{n}\sum_{t=1}^{n}E\left[\frac{\partial\left(\operatorname{vec}\Omega_t^*\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}D_t^*\right)'}{\partial\operatorname{vec}\Omega_t}\frac{\partial\operatorname{vec}\Omega_t^*}{\partial\theta_2'} - \Delta_t^*\right]$$

$$
\begin{aligned}
&= -\frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial\left(\operatorname{vech}\Omega_t^*\right)'}{\partial\theta_2}\left(M_t^{4*}-M_t^{3*}\Omega_t^{*^{-1}}M_t^{3*\prime}\right)^{-1}\frac{\partial\operatorname{vech}\Omega_t^*}{\partial\theta_2'}-\Delta_t^*\right] \\
B_{n_{ij}}^* &= E\left[\left(n^{-1/2}\sum_{t=1}^{n}s_t^{i*}\right)\left(n^{-1/2}\sum_{t=1}^{n}s_t^{j*}\right)'\right], \qquad i=1,2;\; j=1,2 \\
&= \frac{1}{n}\sum_{t=1}^{n}E\left[s_t^{i*}s_t^{j*\prime}\right]+\frac{1}{n}\sum_{\tau=1}^{n-1}\sum_{t=\tau+1}^{n}\left(E\left[s_t^{i*}s_{t-\tau}^{j*\prime}\right]+E\left[s_{t-\tau}^{i*}s_t^{j*\prime}\right]\right)
\end{aligned}
$$

and

$$
\begin{aligned}
s_t^{1*} &= \frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{*^{-1}}\left(Y_t-m_t^o\right), \quad m_t^o=m_t\left(X_t,\theta_1^o\right), \quad \Omega_t^*=\Omega_t\left(X_t,\theta_{2_n}^*\right) \\
s_t^{2*} &= \frac{\partial\left(\operatorname{vec}\Omega_t^*\right)'}{\partial\theta_2}\left(\frac{\partial C_t^{*\prime}}{\partial\operatorname{vec}\Omega_t}\left(Y_t-m_t^o\right)+\frac{\partial\left(\operatorname{vec}D_t^*\right)'}{\partial\operatorname{vec}\Omega_t}\operatorname{vec}\left(Y_tY_t'-\Omega_t^*-m_t^om_t^{o\prime}\right)\right) \\
&= \frac{\partial\left(\operatorname{vech}\Omega_t^*\right)'}{\partial\theta_2}\left(M_t^{4*}-M_t^{3*}\Omega_t^{*^{-1}}M_t^{3*\prime}\right)^{-1} \\
&\quad \left(\operatorname{vech}\left(Y_tY_t'-\Omega_t^*-m_t^om_t^{o\prime}\right)-M_t^{3*}\Omega_t^{*^{-1}}\left(Y_t-m_t^o\right)\right) \\
\Delta_t^* &= \left(\left(\operatorname{vec}\left(Y_tY_t'-\Omega_t^*-m_t^om_t^{o\prime}\right)\right)'\otimes I_{k_{\theta_2}}\right)\frac{\partial}{\partial\theta_2'}\left[\operatorname{vec}\left(\frac{\partial\left(\operatorname{vec}\Omega_t^*\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}D_t^*\right)'}{\partial\operatorname{vec}\Omega_t}\right)\right] \\
M_t^{3*} &= Cov_{\lambda_t^*}\left[\left(\operatorname{vech}\left(Y_tY_t'\right),Y_t\right)\middle|X_t\right], \qquad M_t^{4*}=V_{\lambda_t^*}\left[\operatorname{vech}\left(Y_tY_t'\right)\middle|X_t\right]
\end{aligned}
$$

with $Cov_{\lambda_t^*}[.|X_t]$ *and* $V_{\lambda_t^*}[.|X_t]$ *denoting respectively covariance and variance taken with respect to* $\lambda_t\left(Y_t,X_t,\theta_1^o,\theta_{2_n}^*\right)$.

Proof. See Appendix C. ■

Proposition 10 sketches out two important things. First, regardless of second order misspecification and dynamic incompleteness, first order correct specification ensures a sort of "independence" between mean and variance parameter estimators: the fact that $\theta_{2_n}^*$ is estimated has no effect on the asymptotic distribution of $\hat\theta_{1_n}$, and conversely. Note that a similar property holds for QGPML1, at least for the asymptotic distribution of $\hat\theta_{1_n}$. Second, under the same conditions and contrary to C_{n22}^* or $C_{n_{12}}^*$, the analytical expression of $C_{n_{11}}^*$ — but not its value since, under second order misspecification, $\theta_{2_n}^*$ depends on $\{f_t\}$ — is unchanged whatever the choice of f_t, $t=1,2,...$ Again, a similar property holds for QGPML1. This double similarity with QGPML1 simply follows from the fact that the restricted quadratic exponential family is a sub-family of the generalized linear exponential family. Actually, the RPML2 mean parameters estimator $\hat\theta_{1_n}$ is a just a QGPML1-like estimator where the auxiliary (or nuisance) parameters are jointly estimated with the parameters of interest θ_1 (see below for more details).

According to the above results, when only first order correct specification is assumed, the asymptotic covariance matrix $C_{n_{11}}^*$ of $\hat\theta_{1_n}$ basically depends on the chosen (misspecified) conditional variance specification and on the chosen sequence $\{f_t\}$ — through the pseudo-true value $\theta_{2_n}^*$ implied by $\{f_t\}$ (as well as, of course, by the variance specification itself) —, but not on the fact that $\theta^*_{2_n}$ has been estimated. The dependence of $C^*_{n_{11}}$ on the sequence $\{f_t\}$ only through the pseudo-true value $\theta^*_{2_n}$ is important since it implies that if second order correct specification is in addition assumed, in which case $\theta^*_{2_n} = \theta^o_2$, $C^o_{n_{11}}$ will no longer depend on which members $\{f_t\}$ of the restricted quadratic exponential family are used to form $\hat{\theta}_n$. On the other hand, the asymptotic covariance matrix $C^*_{n_{22}}$of $\hat{\theta}_{2_n}$ and the asymptotic covariance matrix $C^*_{n_{12}}$between $\hat{\theta}_{1_n}$ and $\hat{\theta}_{2_n}$ similarly depend on the chosen (misspecified) conditional variance specification and on the chosen sequence $\{f_t\}$ — in particular through the quantities $\partial C(m,\Sigma)'/\,\partial \operatorname{vec}\Sigma$ and $\partial\left(\operatorname{vec} D(\Sigma)\right)'/\,\partial \operatorname{vec}\Sigma$, which, as shown by Property 8(c)-(d) and made explicit in the second expressions of $A^*_{n_{22}}$ and s^{2*}_t, are related to the third and fourth order moments "built into" the restricted quadratic exponential specification —, but not on the fact that θ^o_1 has been estimated. In contrast with $C^*_{n_{11}}$, $C^*_{n_{22}}$ and $C^*_{n_{12}}$ directly and strongly depend on the chosen sequence $\{f_t\}$ and this dependence does not disappear if second order correct specification is in addition assumed.

If, in addition to first order correct specification, only second order correct specification is assumed, the basic structure of C^*_n is only marginally affected. Proposition 11 portrays this minor change.

Proposition 11 *Suppose that all the assumptions of Proposition 9 hold. If, for all $t = 1, 2, ...,$ f_t belongs to the restricted quadratic exponential family, and if the semi-parametric model $\mathcal{S}$ is as described in Assumption 1 and is second order correctly specified, then, for all $n = 1, 2, ...,$ $\theta^*_n = (\theta^{o\prime}_1, \theta^{o\prime}_2)'$ and, for all $t = 1, 2, ...,$ $E\left(\Delta^o_t\right) = 0$ such that $A^*_{n_{22}}$ collapses to*

$$
\begin{aligned}
A^o_{n_{22}} &= -\frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial\left(\operatorname{vec}\Omega^o_t\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}D^o_t\right)'}{\partial\operatorname{vec}\Omega_t}\frac{\partial\operatorname{vec}\Omega^o_t}{\partial\theta'_2}\right] \\
&= -\frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial\left(\operatorname{vech}\Omega^o_t\right)'}{\partial\theta_2}\left(M^{4o}_t - M^{3o}_t\Omega^{o^{-1}}_t M^{3o\prime}_t\right)^{-1}\frac{\partial\operatorname{vech}\Omega^o_t}{\partial\theta'_2}\right]
\end{aligned}
$$

Proof. See Appendix C. ■

Besides the fact that all quantities are now defined at the true value $\theta^o = (\theta^{o\prime}_1, \theta^{o\prime}_2)'$ rather than at a mixture $\theta^*_n = \left(\theta^{o\prime}_1, \theta^{*\prime}_{2_n}\right)'$ of the true value θ^o_1 and of the pseudo-true value $\theta^*_{2_n}$, the only noticeable change is thus that, as $A^*_{n_{11}}$, $A^o_{n_{22}}$ now only depends on first order derivatives, a feature which is very convenient for their estimation.

Without relying on additional assumptions to either first or second order correct specification, a consistent estimator of C^*_n may not be easy to obtain. Indeed, if obvious consistent — under usual regularity conditions — estimators of $A^*_{n_{11}}$ and $A^*_{n_{22}}$ are given by their empirical counterparts, i.e., under only first order correct specification,

$$
\hat{A}_{n_{11}} = -\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\hat{m}'_t}{\partial\theta_1}\hat{\Omega}^{-1}_t\frac{\partial\hat{m}_t}{\partial\theta'_1}
$$

$$\tilde{A}_{n_{22}} = -\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial\left(\operatorname{vec}\hat{\Omega}_t\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}\hat{D}_t\right)'}{\partial\operatorname{vec}\Omega_t}\frac{\partial\operatorname{vec}\hat{\Omega}_t}{\partial\theta_2'}-\hat{\Delta}_t\right],$$

or, if in addition second order correct specification is assumed,

$$\hat{A}_{n_{22}} = -\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial\left(\operatorname{vec}\hat{\Omega}_t\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}\hat{D}_t\right)'}{\partial\operatorname{vec}\Omega_t}\frac{\partial\operatorname{vec}\hat{\Omega}_t}{\partial\theta_2'}\right],$$

the story is considerably more complicated for $B^*_{n_{11}}$, $B^*_{n_{12}}$ and $B^*_{n_{22}}$. In this respect, the basic problem arises from the second term of $B^*_{n_{ij}}$. This term contains $n-1$ quantities which have each to be adequately estimated. With only n observations, this is not possible unless relying on additional specific assumptions. Different consistent — under such additional specific assumptions — estimators of quantities like $B^*_{n_{ij}}$ have been proposed in the literature. We shall not discuss them here. We refer the reader to White (1994), Wooldridge (1994) or Pötscher-Prucha (1997) for both a general discussion and references. We will only outline one point: generally speaking, under arbitrary misspecification, such a consistent estimator need not exist. In the present context, because of the maintained hypothesis of first order correct specification, such a consistent estimator of $B^*_{n_{11}}$ will usually be available. Likewise, if second order correct specification is in addition assumed, so it will usually be for $B^*_{n_{12}}$ and $B^*_{n_{22}}$. However, under (first order correct specification but) arbitrary second order misspecification, i.e., not only dynamic misspecification, it is not necessarily the case.

As suggested by the above discussion and as already mentioned, dynamic completeness or incompleteness has important consequences for inference. Indeed, if it does not influence the expression of A^*_n, it crucially governs the form of B^*_n, and thus the expression of C^*_n. Proposition 12 details this crucial influence as well as the effect of second order correct specification and possible correct specification up to higher order conditional moments on the limiting distribution of RPML2 estimators.

Proposition 12 *Suppose that all the assumptions of Proposition 9 hold. If, for all $t = 1, 2, ...$, f_t belongs to the restricted quadratic exponential family, and if the semi-parametric model $\mathcal{S}$ is as described in Assumption 1 and is first order correctly specified and first order dynamically complete, then, for all $n = 1, 2, ...$, $\theta^*_n = \left(\theta^{o\prime}_1, \theta^{*\prime}_{2_n}\right)'$ and $B^*_{n_{11}}$ collapses to*

$$\overline{B}^*_{n_{11}} = \frac{1}{n}\sum_{t=1}^{n}E\left[s^{1*}_t s^{1*\prime}_t\right] = \frac{1}{n}\sum_{t=1}^{n}E\left[\frac{\partial m^{o\prime}_t}{\partial\theta_1}\Omega^{*^{-1}}_t\Sigma^o_t\Omega^{*^{-1}}_t\frac{\partial m^o_t}{\partial\theta_1'}\right]$$

where $\Sigma^o_t = V(Y_t|X_t)$ is the actual conditional covariance matrix of Y_t given X_t, such that $C^*_{n_{11}}$ collapses to

$$\overline{C}^*_{n_{11}} = A^{*^{-1}}_{n_{11}}\overline{B}^*_{n_{11}}A^{*^{-1}}_{n_{11}}$$

If, in addition, the semi-parametric model $\mathcal{S}$ is also second order correctly specified, *then, for all $n = 1, 2, ...,$ $\theta_n^* = (\theta_1^{o\prime}, \theta_2^{o\prime})'$ and $B_{n_{11}}^*$ further collapses to*

$$\overline{B}_{n_{11}}^o = \frac{1}{n}\sum_{t=1}^n E\left[s_t^{1o} s_t^{1o\prime}\right] = \overline{\overline{B}}_{n_{11}}^o = \frac{1}{n}\sum_{t=1}^n E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{o^{-1}}\frac{\partial m_t^o}{\partial \theta_1'}\right] = -A_{n_{11}}^o$$

such that $C_{n_{11}}^*$ further collapses to

$$\overline{C}_{n_{11}}^o = \overline{\overline{C}}_{n_{11}}^o = -A_{n_{11}}^{o^{-1}} = \overline{\overline{B}}_{n_{11}}^{o^{-1}} = \overline{B}_{n_{11}}^{o^{-1}}$$

and we have

$$\overline{C}_{n_{11}}^* - \overline{\overline{C}}_{n_{11}}^o \gg 0$$

i.e, $\overline{\overline{C}}_{n_{11}}^o$ is the minimum asymptotic covariance matrix of a RPML2 mean parameters estimator of a semi-parametric model $\mathcal{S}$ first order correctly specified and first order dynamically complete.

If, in addition, the semi-parametric model $\mathcal{S}$ is also second order dynamically complete, then, for all $n = 1, 2, ...,$ $B_{n_{12}}^*$ and $B_{n_{22}}^*$ collapse to

$$\begin{aligned} \overline{B}_{n_{12}}^o &= \frac{1}{n}\sum_{t=1}^n E\left[s_t^{1o} s_t^{2o\prime}\right] \\ \overline{B}_{n_{22}}^o &= \frac{1}{n}\sum_{t=1}^n E\left[s_t^{2o} s_t^{2o\prime}\right] \end{aligned}$$

such that $C_{n_{12}}^*$ and $C_{n_{22}}^*$ collapse to

$$\begin{aligned} \overline{C}_{n_{12}}^o &= A_{n_{11}}^{o^{-1}} \overline{B}_{n_{12}}^o A_{n_{22}}^{o^{-1}} \\ \overline{C}_{n_{22}}^o &= A_{n_{22}}^{o^{-1}} \overline{B}_{n_{22}}^o A_{n_{22}}^{o^{-1}} \end{aligned}$$

If, in addition, the implicit parametric model $\mathcal{P}$ arising from the semi-parametric model $\mathcal{S}$ and the sequence $\{f_t\}$ is also third order correctly specified, then, for all $n = 1, 2, ...,$ $B_{n_{12}}^*$ further collapses to

$$\overline{\overline{B}}_{n_{12}}^o = A_{n_{12}}^o = 0$$

such that $C_{n_{12}}^*$ further collapses to

$$\overline{\overline{C}}_{n_{12}}^o = 0$$

Finally, if, in addition, the implicit parametric model $\mathcal{P}$ arising from the semi-parametric model $\mathcal{S}$ and the sequence $\{f_t\}$ is also fourth order correctly specified, then, for all $n = 1, 2, ...,$ $B_{n_{22}}^*$ further collapses to

$$\begin{aligned} \overline{\overline{B}}_{n_{22}}^o &= \frac{1}{n}\sum_{t=1}^n E\left[\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial \theta_2}\frac{\partial\left(\operatorname{vec}D_t^o\right)'}{\partial \operatorname{vec}\Omega_t}\frac{\partial \operatorname{vec}\Omega_t^o}{\partial \theta_2'}\right] = -A_{n_{22}}^o \\ &= \frac{1}{n}\sum_{t=1}^n E\left[\frac{\partial\left(\operatorname{vech}\Omega_t^o\right)'}{\partial \theta_2}\left(\overline{\overline{M}}_t^{4o} - \overline{\overline{M}}_t^{3o}\Omega_t^{o^{-1}}\overline{\overline{M}}_t^{3o\prime}\right)^{-1}\frac{\partial \operatorname{vech}\Omega_t^o}{\partial \theta_2'}\right] \end{aligned}$$

where $\overline{\overline{M}}_t^{3o} = Cov\left[\left(\text{vech}\left(Y_tY_t'\right), Y_t\right)\middle| X_t\right]$ *and* $\overline{\overline{M}}_t^{4o} = V\left[\text{vech}\left(Y_tY_t'\right)\middle| X_t\right]$ *are respectively the actual third and fourth order conditional moments of* Y_t *given* X_t*, such that* C_{n22}^* *further collapses to*

$$\overline{\overline{C}}_{n22}^{o} = -A_{n22}^{o^{-1}} = \overline{\overline{B}}_{n22}^{o^{-1}}$$

and we have

$$\overline{C}_n^o - \overline{\overline{C}}_n^o \gg 0$$

i.e., $\overline{\overline{C}}_n^o$ *is the minimum asymptotic covariance matrix of a RPML2 estimator of a semi-parametric model* $\mathcal{S}$ *second order correctly specified and second order dynamically complete.*

Proof. See Appendix C. ■

As it may be seen from Proposition 12, dynamic completeness is of great practical importance for inference: when associated with the assumption of first (resp. second) order correct specification, first (resp. second) order correct dynamic specification indeed drastically simplifies the correlative expression of $B_{n_{ij}}^*$, both making it possible to be consistently — under usual regularity conditions — estimated by the common outer-product gradient estimator

$$\hat{B}_{n_{ij}} = \frac{1}{n}\sum_{t=1}^{n} \hat{s}_t^i \hat{s}_t^{j\prime}$$

and allowing, under additional assumptions, to retrieve traditional information matrix equalities and to outline efficiency bounds.

So, when $\mathcal{S}$ is assumed correctly specified and dynamically complete for the conditional mean, the asymptotic covariance matrix $C_{n_{11}}^*$ of $\hat{\theta}_{1_n}$ is equal to $\overline{C}_{n_{11}}^* = A_{n_{11}}^{*-1}\overline{B}_{n_{11}}^* A_{n_{11}}^{*-1}$, and an appropriate consistent — under usual regularity conditions — estimator of it is simply $\hat{C}_{n_{11}} = \hat{A}_{n_{11}}^{-1}\hat{B}_{n_{11}}\hat{A}_{n_{11}}^{-1}$, i.e., a generalized form of the seminal White's (1980c) heteroscedasticity-consistent covariance matrix estimator. As it may be check from White (1994)[4], the only difference between the asymptotic covariance matrix $\overline{C}_{n_{11}}^*$ of $\hat{\theta}_{1_n}$ and those of a QGPML1 estimator under the same conditions (i.e., first order correct specification and first order correct dynamic specification but second order misspecification) stems from the fact the pseudo-true values $\theta_{2_n}^*$, and thus the misspecified covariance matrices $\Omega_t^* = \Omega_t\left(X_t, \theta_{2_n}^*\right)$, associated with each estimator are generally different: in the RPML2 case, $\theta_{2_n}^*$ is "endogenously" determined through the choice of $\{f_t\}$, while in the QGPML1 case, it "exogenously" depends on the way the misspecified conditional variance specification $\{\Omega_t\left(X_t, \theta_2\right)\}$ is estimated. Note that, as suggested above, the same similarity between $\hat{\theta}_{1_n}$ and QGPML1 already holds when only first order correct specification is assumed.

Further simplifications arise if, in addition to first order correct specification and

[4] This result is not readily apparent from the results reported in White (1994) since his results are expressed in terms of a direct parametrization $\eta = \kappa_t\left(X_t, \alpha\right)$ of the extra parameter η of the generalized linear exponential family, rather than in terms of a parametrization $\Sigma = \Omega_t(X_t, \theta_2)$ of the covariance matrix (uniquely for given m) associated with η through the function $\eta = \Gamma(m, \Sigma)$.

first order correct dynamic specification, it is also assumed that $\mathcal{S}$ is jointly correctly specified for the conditional variance. In this case, the asymptotic covariance matrix $C^*_{n_{11}}$ of $\hat{\theta}_{1_n}$ is equal to $\overline{\overline{C}}^o_{n_{11}} = -A^{o^{-1}}_{n_{11}} = \overline{\overline{B}}^{o^{-1}}_{n_{11}}$, i.e., the traditional information equality $B^o_{n_{11}} = -A^o_{n_{11}}$ holds for the mean parameters, and appropriate consistent — under usual regularity conditions — estimators of it are $\hat{C}_{n_{11}} = -\hat{A}^{-1}_{n_{11}}$ or $\check{C}_{n_{11}} = \hat{B}^{-1}_{n_{11}}$ ($\hat{C}_{n_{11}}$ typically has better finite sample properties than $\check{C}_{n_{11}}$). $\overline{\overline{C}}^o_{n_{11}}$ does no longer depend in any way on the choice of $\{f_t\}$. Interestingly, note that all this holds regardless of whether or not $\mathcal{S}$ is dynamically complete for the conditional variance: only first order correct dynamic specification is required. According to the similarity between QGPML1 and RPML2 just described, the asymptotic covariance matrix $\overline{\overline{C}}^o_{n_{11}}$ of $\hat{\theta}_{1_n}$ now exactly corresponds to the asymptotic covariance matrix of a QGPML1 estimator under the same conditions (i.e., second order correct specification and first order correct dynamic specification). It simply follows from the fact that we now have $\theta^*_{2_n} = \theta^o_2$ for both estimators.

$\overline{\overline{C}}^o_{n_{11}}$ is the minimum asymptotic covariance matrix of a RPML2 mean parameters estimator of a semi-parametric model $\mathcal{S}$ correctly specified and dynamically complete for the conditional mean. For reaching this lower bound, it suffices to be able to jointly correctly specify the conditional variance. It may be easily checked that $\overline{\overline{C}}^o_{n_{11}}$ is identical to the common lower bound for the asymptotic covariance matrix of PML1 and QGPML1 estimators outlined by Gourieroux-Monfort-Trognon (1984a) and White (1994). In other words, as QGPML1, RPML2 will never yield a mean parameters estimator $\hat{\theta}_{1_n}$ (of a first order dynamically complete model) with an asymptotic covariance matrix smaller than $\overline{\overline{C}}^o_{n_{11}}$. Likewise, a mean parameters estimator $\hat{\theta}_{1_n}$ as efficient as (asymptotically equivalent to) a genuine maximum likelihood estimator of a (first order dynamically complete) parametric model whose "true conditional densities" of Y_t given X_t belong to the linear exponential family $\tilde{\lambda}_t(Y_t, X_t, \theta_1) = \exp(A_t(m_t) + B_t(Y_t) + C_t(m_t)'Y_t)$, where $m_t = m_t(X_t, \theta_1)$, may always be obtained by RPML2 or QGPML1[5]. Note that this lower bound is identical to the well-known semi-parametric efficiency bound (e.g., Chamberlain (1987), Newey (1990,1993), Wooldridge (1994)) associated with optimal GMM estimation based on the first order conditional moments restrictions $E[(Y_t - m_t(X_t, \theta_1))\,|X_t] = 0$.

When second order correct dynamic specification is added to the previous assumptions of second order correct specification and first order correct dynamic specification, the asymptotic covariance matrix $C^*_{n_{22}}$ of $\hat{\theta}_{2_n}$ and the asymptotic covariance matrix $C^*_{n_{12}}$between $\hat{\theta}_{1_n}$ and $\hat{\theta}_{2_n}$ are respectively equal to $\overline{C}^o_{n_{22}} = A^{o^{-1}}_{n_{22}}\overline{B}^o_{n_{22}}A^{o^{-1}}_{n_{22}}$ and $\overline{C}^o_{n_{12}} = A^{o^{-1}}_{n_{11}}\overline{B}^o_{n_{12}}A^{o^{-1}}_{n_{22}}$, and appropriate consistent — under usual regularity conditions — estimators of them are $\hat{C}_{n_{22}} = \hat{A}^{-1}_{n_{22}}\hat{B}_{n_{22}}\hat{A}^{-1}_{n_{22}}$ and $\hat{C}_{n_{12}} = \hat{A}^{-1}_{n_{11}}\hat{B}_{n_{12}}\hat{A}^{-1}_{n_{22}}$, i.e., again generalized forms of the seminal White's (1980c) heteroscedasticity-consistent

[5]For an example, see Gourieroux-Monfort-Trognon (1984a). Note that a stronger result is actually available. According to Theorem 5 of Gourieroux-Monfort-Trognon (1984a), it may be showed that a mean parameters estimator $\hat{\theta}_{1_n}$ as efficient as a genuine maximum likelihood estimator of a (first order dynamically complete) parametric model whose "true conditional densities" of Y_t given X_t belong to the generalized linear exponential family $\check{\lambda}_t(Y_t, X_t, \theta_1, \alpha) = \exp(A_t(m_t, \eta_t) + B_t(\eta_t, Y_t) + C_t(m_t, \eta_t)'Y_t)$, where $m_t = m_t(X_t, \theta_1)$, $\eta_t = \kappa_t(X_t, \alpha)$ and θ_1 and α vary independently, may always be obtained by RPML2 or QGPML1.

covariance matrix estimator. In this case, the entire covariance matrix C_n^* may thus be readily estimated. Needless to say, both $\overline{C}_{n_{12}}^o$ and $\overline{C}_{n_{22}}^o$ strongly depend on which members $\{f_t\}$ — through their "built into" third and fourth order moments — of the restricted quadratic exponential family are used to form $\hat{\theta}_n$.

The two last situations considered in Proposition 12 are those where, in addition to the previous assumptions, we are lucky enough for the implicit parametric model $\mathcal{P}$ arising from $\mathcal{S}$ and the sequence $\{f_t\}$ to be also jointly correctly specified for the third or for the third and the fourth order conditional moments. Since it already attained its lower bound, nothing changes for the asymptotic covariance matrix $C_{n_{11}}^*$ of $\hat{\theta}_{1_n}$, which is still as given above by $\overline{\overline{C}}_{n_{11}}^o = -A_{n_{11}}^{o^{-1}} = \overline{\overline{B}}_{n_{11}}^{o^{-1}}$. Under third order correct specification, the asymptotic covariance matrix $C_{n_{12}}^*$ between $\hat{\theta}_{1_n}$ and $\hat{\theta}_{2_n}$ is simply equal to $\overline{\overline{C}}_{n_{12}}^o = A_{n_{12}}^o = \overline{\overline{B}}_{n_{12}}^o = 0$: $\hat{\theta}_{1_n}$ and $\hat{\theta}_{2_n}$ are now truly asymptotically independent. If, in addition, fourth order correct specification is also assumed, the asymptotic covariance matrix $C_{n_{22}}^*$ of $\hat{\theta}_{2_n}$is equal to $\overline{\overline{C}}_{n_{22}}^o = -A_{n_{22}}^{o^{-1}} = \overline{\overline{B}}_{n_{22}}^{o^{-1}}$ and appropriate consistent — under usual regularity conditions — estimators of it are $\hat{C}_{n_{22}} = -\hat{A}_{n_{22}}^{-1}$ or $\check{C}_{n_{22}} = \hat{B}_{n_{22}}^{-1}$ ($\hat{C}_{n_{22}}$ typically has better finite sample properties than $\check{C}_{n_{22}}$). $\overline{\overline{C}}_n^o$ does no longer depend in any way on the choice of $\{f_t\}$ and the traditional information matrix equality $B_n^o = -A_n^o$ now fully applies. Interestingly, note that all this holds without requiring $\mathcal{S}$ to be dynamically complete for the third and the fourth order conditional moments, or, a fortiori, dynamically complete for the conditional distribution: only second order correct dynamic specification is required. This follows from the very special form of the score associated with (restricted) quadratic exponential families.

$\overline{\overline{C}}_n^o$ is the minimum asymptotic covariance matrix of a RPML2 estimator of a semi-parametric model $\mathcal{S}$ jointly correctly specified and dynamically complete for the conditional mean and the conditional variance. In other words, RPML2 will never yield a estimator $\hat{\theta}_n$ (of a second order dynamically complete model) with an asymptotic covariance matrix smaller than $\overline{\overline{C}}_n^o$. For fully reaching this lower bound — according to the results outlined above, its sub-block $\overline{\overline{C}}_{n_{11}}^o$ corresponding to the mean parameters is reached under much less restrictive assumptions — it is necessary to be able to pick up a sequence of pseudo-densities $\{f_t\}$ belonging to restricted quadratic exponential families and such that the implicit parametric model $\mathcal{P}$ is jointly correctly specified for the third and the fourth order conditional moments. Contrary to $\overline{\overline{C}}_{n_{11}}^o$, the lower bound $\overline{\overline{C}}_n^o$ is mainly of theoretical interest since it is usually unfeasible both because third and fourth order conditional moments are typically unknown and, if it were known, because nothing guarantees that such a choice for $\{f_t\}$ always exists. If it exists, as shown in Appendix D, it is worth noting that this bound is again identical to the semi-parametric efficiency bound $\overline{\overline{C}}_n^{oGMM}$ associated with optimal GMM estimation jointly based on the first and second order conditional moments restrictions $E\left[\left(Y_t - m_t(X_t, \theta_1)\right)|X_t\right] = 0$ and $E\left[\mathrm{vech}\left(Y_t Y_t' - \Omega_t\left(X_t, \theta_2\right) - m_t(X_t, \theta_1) m_t(X_t, \theta_1)'\right)|X_t\right] = 0$.

Obviously, since it already attained its lower bound, if instead of fourth order correctly specified, $\mathcal{P}$ were assumed correctly specified for the conditional density, i.e., for all conditional moments, the asymptotic covariance matrix C_n^* of $\hat{\theta}_n$ would still be given by $\overline{\overline{C}}_n^o$. However, $\hat{\theta}_n$ would now reach, or at least would be closer to, asymptotic efficiency. Although $\hat{\theta}_n$ would be now a standard maximum likelihood estimator, it is worth emphasizing that it would not be necessarily asymptotically efficient, and this for at least two reasons. First, if we actually had functional links between mean and variance parameters in the structural model, taking them into account — a possibility ruled out for robustness by RPML2 — would usually yield a more efficient estimator. See for example Gourieroux (1992) for a discussion of this efficiency loss in a univariate conditionally gaussian ARCH(p) model. Further, even in absence of such structural cross-constraints, as extensively discussed in White (1994), conditional density correct specification and conditional distribution correct dynamic specification (a condition which is not assumed here) are not sufficient for assuring asymptotic efficiency (see White (1994) for details). Be that as it may, the fact is that RPML2 may get efficiency gains from an eventual proximity between the "true conditional densities" and the chosen sequence $\{f_t\}$. It is a practical important point since, as already outlined, the prominent member of restricted quadratic exponential families is just the normal density[6], i.e., a distribution which is often presented as a plausible approximation of the true underlying distribution in a lot of empirical works.

To follow, a special result which is of importance when dealing with independent observations as in cross-section or panel data. When the observations are independent, second order correct dynamic specification tautologically holds. Proposition 13 outlines the fact that, in such situations and contrary to the general dynamic case, whenever first order correct specification holds, the covariance matrix $C^*_{n_{12}}$ between $\hat{\theta}_{1_n}$ and $\hat{\theta}_{2_n}$ and an upper bound for the covariance matrix $C^*_{n_{22}}$ of $\hat{\theta}_{2_n}$ can always be easily obtained.

Proposition 13 *Suppose that all the assumptions of Proposition 9 hold. If, for all $t = 1, 2, ...,$ f_t belongs to the restricted quadratic exponential family, if the semi-parametric model $\mathcal{S}$ is as described in Assumption 1 and is first order correctly specified, and if the observations are independent across t, then, for all $n = 1, 2, ...,$ $\theta^*_n = \left(\theta^{o\prime}_1, \theta^{*\prime}_{2_n}\right)'$, $B^*_{n_{12}}$ and $B^*_{n_{22}}$ collapse to*

$$\begin{aligned} \overline{B}^*_{n_{12}} &= \frac{1}{n}\sum_{t=1}^{n} E\left[s^{1*}_t s^{2*\prime}_t\right] \\ \ddot{B}^*_{n_{22}} &= \frac{1}{n}\sum_{t=1}^{n} E\left[s^{2*}_t s^{2*\prime}_t\right] - \frac{1}{n}\sum_{t=1}^{n} E\left(s^{2*}_t\right) E\left(s^{2*}_t\right)' \end{aligned}$$

such that $C^*_{n_{12}}$ collapses to

$$\overline{C}^*_{n_{12}} = A^{*^{-1}}_{n_{11}} \overline{B}^*_{n_{12}} A^{*^{-1}}_{n_{22}}$$

and

$$C^*_{n_{22}} \ll \overline{Q}^*_{n_{22}} = A^{*^{-1}}_{n_{22}} \overline{B}^*_{n_{22}} A^{*^{-1}}_{n_{22}}$$

[6] Simplified forms, for the normal density, of the general expressions appearing in Proposition 10-13 are given in Appendix E.

where

$$\overline{B}^*_{n22} = \frac{1}{n}\sum_{t=1}^{n} E\left[s_t^{2*} s_t^{2*\prime}\right] \gg \ddot{B}^*_{n22}$$

Proof. See Appendix C. ■

Thus, when the observations are independent and $\mathcal{S}$ is first order correctly specified but second order misspecified, the covariance matrix $C^*_{n_{12}}$ between $\hat{\theta}_{1_n}$ and $\hat{\theta}_{2_n}$ is simply equal to $\overline{C}^*_{n_{12}} = A^{*-1}_{n_{11}} \overline{B}^*_{n_{12}} A^{*-1}_{n_{22}}$ and an appropriate consistent — under usual regularity conditions — estimator of it is $\tilde{C}_{n_{12}} = \hat{A}^{-1}_{n_{11}} \hat{B}_{n_{12}} \tilde{A}^{-1}_{n_{22}}$ (note that $\hat{A}_{n_{22}}$cannot be used). On the other hand, an upper bound $\overline{Q}^*_{n_{22}}$ of $C^*_{n_{22}}$ is given by $A^{*-1}_{n_{22}} \overline{B}^*_{n_{22}} A^{*-1}_{n_{22}}$ and an appropriate consistent — under usual regularity conditions — estimator of it is $\hat{Q}_{n_{22}} = \tilde{A}^{-1}_{n_{22}} \hat{B}_{n_{22}} \tilde{A}^{-1}_{n_{22}}$ (again, note that $\hat{A}_{n_{22}}$cannot be used). Needless to say, the covariance matrix $C^*_{n_{11}}$of $\hat{\theta}_{1_n}$ is, and may be estimated, as outlined above when assuming first order correct specification and first order correct dynamic specification. This ability to easily obtain an upper bound of $C^*_{n_{22}}$ allows to perform conservative tests, i.e., tests with true asymptotic size necessarily inferior to their specified nominal size, on the pseudo-true value $\theta^*_{2_n}$. This may for example be useful for readily checking through Wald tests some possibly meaningful restrictions despite conditional variance misspecification. Note that a similar result, i.e., the ability to readily obtain an upper bound $\overline{Q}^*_n$ of the true asymptotic covariance matrix C^*_n of $\hat{\theta}_n$ when the observations are independent, holds for the case where both the conditional mean and the conditional variance are misspecified and a consistent — under usual regularity conditions — estimator $\hat{Q}_n$ of it may simply be obtained through the empirical hessian and the empirical outer-product gradient: $\hat{Q}_n = \breve{A}^{-1}_n \breve{B}_n \breve{A}^{-1}_n$ where $\breve{A}_n = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta'} \ln \hat{f}_t$, $\breve{B}_n = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial}{\partial\theta}\ln \hat{f}_t \frac{\partial}{\partial\theta'}\ln \hat{f}_t\right]$, $\ln \hat{f}_t = \ln f_t\left(Y_t, m_t(X_t, \hat{\theta}_{1_n}), \Omega_t(X_t, \hat{\theta}_{2_n})\right)$ and $\theta = (\theta_1', \theta_2')'$ (note that no sub-blocks of $\breve{A}_n$ is equal to the already outlined estimators $\hat{A}_{n_{11}}$, $\hat{A}_{n_{22}}$ or $\tilde{A}_{n_{22}}$). This crude estimator, or some of its components (sub-blocks), may of course also be used — although it may have poor finite sample properties — for computing in a purely numerical way the true asymptotic covariance matrix C^*_n under correct (dynamic) specification whenever analytical derivatives are difficult to derive and/or to compute.

To conclude this section, a few words about the possible efficiency price to pay for robustness entailed by RPML2. Besides using pseudo-densities belonging to restricted quadratic exponential families, a feature which, compared to PML2, at least theoretically reduces the possible choices for the sequence $\{f_t\}$ and thus also the possibility to get efficiency benefits from an eventual proximity between of the "true conditional densities" and the chosen sequence $\{f_t\}$, RPML2 means discarding eventual structural cross-constraints between mean and variance parameters. We already mentioned that, when the implicit parametric model $\mathcal{P}$ is correctly specified for the conditional density and second order correct dynamic specification holds, if we actually had functional links between mean and variance parameters in the structural model, taking them into account would usually yield a more efficient estimator. The point we want to emphasize here is that, in most other situations, this efficiency price is not at all enforced. Indeed, generally speaking, the difference between the asymptotic covariance matrix of an unconstrained consistent asymptotically normal (CAN) extremum estimator and a constrained version of it may be shown to be necessarily positive semidefinite only when the unconstrained CAN estimator is such that the information matrix equality $B_n^o = -A_n^o$ holds and, of course, the constraints are correct (see for example Gourieroux-Monfort (1989)). In all other cases, the constrained estimator may be either more efficient or less efficient than the unconstrained one. In other words, unless the information matrix equality holds — in the present context, when $\mathcal{P}$ is jointly correctly specified for the first four order conditional moments and second order correct dynamic specification holds — taking into account structural cross-constraints between mean and variance parameters while continuing to use the same restricted quadratic exponential pseudo-densities $\{f_t\}$ to form $\hat{\theta}_n$ does not necessarily, although it likely will whenever the distributional misspecification is not too severe, improve efficiency. On the contrary, it may entail efficiency losses.

The above discussion concerns second order pseudo-maximum likelihood estimators. From a GMM perspective, according to the intuitively appealing inequality $\overline{C}_n^o - \overline{\overline{C}}_n^{oGMM} \gg 0$ shown in Appendix D, unless the implicit parametric model $\mathcal{P}$ is fourth order correctly specified, optimal GMM estimation jointly based on the first and second order conditional moments restrictions $E[(Y_t - m_t(X_t, \theta_1))|X_t] = 0$ and $E[\text{vech}(Y_tY_t' - \Omega_t\,(X_t, \theta_2) - m_t(X_t, \theta_1)m_t(X_t, \theta_1)')|X_t] = 0$ will usually yield, at least for the variance parameters, a more efficient — but of course not robust to conditional variance misspecification — estimator than RPML2, and, because for such an estimator the information matrix equality holds, it is all the more so true if there are functional links between mean and variance parameters in the structural model. From this perspective, at least under severe distributional misspecification and when mean and variance parameters are functionally related, the possible efficiency price to pay for robustness entailed by RPML2 might be more substantial. However, it is worth recalling that such an "ideal" optimal GMM estimator requires non-parametric estimation of (dynamic) third and fourth order conditional moments. As a practical matter, it then have limitations, especially in the multivariate case and when the dimension of X_t is large.

1.7. Concluding comments

In view of the limiting distribution results outlined in the preceding section, just as PML2 but now as a robust to conditional variance misspecification alternative, compared to QGPML1, RPML2 presents some potential attractive features. Under second order correct specification (and dynamic completeness of the conditional mean), we saw that the asymptotic distributions of the mean parameters estimator of RPML2 and QGPML1 are identical. Both of them attain the semiparametric efficiency bound based on first order conditional moments restrictions. This signifies that most of what we said about the potential advantages of PML2 over QGPML1 remains valid for RPML2: one-step estimation procedure and additional by-product properties for the variance parameters including (under dynamic completeness) asymptotic precision always easily obtained and, under favorable circumstances, asymptotic efficiency. Moreover, at least when there are no functional links between mean and variance parameters in the structural model and because of its ability to get efficiency benefits from an eventual proximity between of the "true conditional densities" and the used pseudo-densities, it may be conjectured that in a lot of practical cases, the RPML2 variance parameters estimator will compete favorably with the PML1-like estimator usually computed in the second step of QGPML1. On the other hand, under first order correct specification (and dynamic completeness of the conditional mean) but second order misspecification, we saw that the asymptotic distributions of the mean parameters estimator of RPML2 and QGPML1 are generally different. They can generally not be compared. Roughly speaking, the most efficient will be the one which approximates the best the true second order conditional moments of the observations. In this respect, RPML2 could again compete favorably with QGPML1. This is in particular true when the structural second order misspecified model which is thought to be a good approximation of the true one contains functional links between mean and variance parameters. Indeed, in this case, the way in which RPML2 is computed (recall that such structural links must be discarded) will generally allow more flexibility for the second order conditional moments approximate adjustment than the standard second step of QGPML1 which usually uses the first step conditional mean estimates as auxiliary parameters. Be that as it may, RPML2 continues to preserve the following attractive features: one-step estimation procedure and, in the case of independent observations (cross-section and panel data), an upper bound for the asymptotic precision of the estimated variance parameters pseudo-true value always easily obtained.

From a practical point of view, the easiest way — and probably the only manageable one — to implement RPML2 is to use the normal density as pseudo-densities. Because of the relative simplicity of its implementation and its potential efficiency, we believe that this estimator should be useful in a variety of situations. In particular as an alternative to QGPML1, it constitutes an attractive tool for implementing the natural sequential "bottom-up" model construction/specification testing strategy advocated by Wooldridge (1991a). As such, it appears as a very convenient go-between estimator which simultaneously allows to get efficiency gains from approximately taking into account the scedastic structure of the data when, in a first step, concentrating on the conditional mean specification, and, once this first step completed, to further explore, for efficiency reasons and/or because it is of interest of its own, the conditional variance specification. Accordingly, it should encourage researchers to use second order semi-parametric models whenever both good reasons suggest that second order conditional moments are not trivial (proportional to an identity matrix) and a plausible, even approximative, specification is available for them. Likewise, besides its natural role as an estimator of second order semi-parametric models where there is no link between mean and variance parameters, it should be considered as an attractive first step estimator when dealing with second order semi-parametric models containing structural links between mean and variance parameters such as ARCH-type models.

Appendix A

This appendix contains the set of regularity conditions to which it is referred to in the various propositions.

Assumption R1 Θ is compact.

Assumption R2 The functions $m_t : \mathbb{R}^{k_{x_t}} \times \Theta \to \mathbb{R}^G$ are such that for each θ in Θ, a compact subset of $\mathbb{R}^{k_\theta}$, $m_t(.,\theta)$ is measurable -$\mathcal{B}^{k_{x_t}}$, and $m_t(X_t,.)$ is continuous on Θ, $a.s. - P_o$, $t = 1, 2, ...$

Assumption R3 The functions $\Omega_t : \mathbb{R}^{k_{x_t}} \times \Theta \to \mathbb{R}^{G\times G}$ are such that for each θ in Θ, a compact subset of $\mathbb{R}^{k_\theta}$, $\Omega_t(.,\theta)$ is measurable -$\mathcal{B}^{k_{x_t}}$, and $\Omega_t(X_t,.)$ is continuous on Θ, $a.s. - P_o$, $t = 1, 2, ...$

Assumption R4 The functions $f_t : \mathbb{R}^G \times \mathbb{R}^G \times \mathbb{R}^{G\times G} \to \mathbb{R}^+$ are such that for each $m \in \mathcal{M}_t \subset \mathbb{R}^G$and each $\Sigma \in \mathcal{E}_t \subset \mathbb{R}^{G\times G}$, $f_t\,(.,m,\Sigma)$ is measurable -$\mathcal{B}^G$, and $f_t\,(Y_t,.,.)$ is continuous on $\mathcal{M}_t \times \mathcal{E}_t$, $a.s. - P_o$, $t = 1, 2, ...$

Assumption R5 (a) $E\left(\ln f_t\left(Y_t, m_t(X_t,\theta), \Omega_t(X_t,\theta)\right)\right) < \infty$, for each θ in Θ, $t = 1, 2, ...$; (b) $E\left(\ln f_t\left(Y_t, m_t(X_t,\theta), \Omega_t(X_t,\theta)\right)\right)$ is continuous on Θ, $t = 1, 2, ...$; and (c) $\{\ln f_t\left(Y_t, m_t(X_t,\theta), \Omega_t(X_t,\theta)\right)\}$ obeys a strong ULLN (White (1994), Definition 3.1).

Assumption R6 $\{E\left(L_n(Y^n, X^n, \theta)\right)\}$ has identifiably unique maximizers $\{\theta_n^*\}$ on Θ (White (1994), Definition 3.3).

Assumption R7 The functions $m_t(X_t,.)$ and $\Omega_t(X_t,.)$ are continuously differentiable on Θ, $a.s. - P_o$, $t = 1, 2, ...$

Assumption R8 The functions $f_t\,(Y_t,.,.)$ are continuously differentiable on $\mathcal{M}_t \times \mathcal{E}_t$, $a.s. - P_o$, $t = 1, 2, ...$

Assumption R6′ $\{E\left(L_n(Y^n, X^n, \theta)\right)\}$ has identifiably unique maximizers $\{\theta_n^*\}$ on $int\,\Theta$, the interior of Θ, uniformly in n.

Assumption R7′ The functions $m_t(X_t,.)$ and $\Omega_t(X_t,.)$ are twice continuously differentiable on Θ, $a.s. - P_o$, $t = 1, 2, ...$

Assumption R8′ The functions $f_t\,(Y_t,.,.)$ are twice continuously differentiable on $\mathcal{M}_t \times \mathcal{E}_t$, $a.s. - P_o$, $t = 1, 2, ...$

Assumption R9 $\frac{\partial}{\partial\theta} E\left(L_n(Y^n, X^n, \theta)\right) = E\left(\frac{\partial}{\partial\theta} L_n(Y^n, X^n, \theta)\right) < \infty$, for each θ in Θ, $n = 1, 2, ...$

Assumption R10 (a) $\frac{\partial^2}{\partial\theta\partial\theta'} E\left(L_n(Y^n, X^n, \theta)\right) = E\left(\frac{\partial^2}{\partial\theta\partial\theta'} L_n(Y^n, X^n, \theta)\right) < \infty$, for each θ in Θ, $n = 1, 2, ...$; (b) $E\left(\frac{\partial^2}{\partial\theta\partial\theta'} L_n(Y^n, X^n, \theta)\right)$ is continuous on Θ uniformly in $n = 1, 2, ...$; and (c) $\left\{\frac{\partial^2}{\partial\theta\partial\theta'} \ln f_t\left(Y_t, m_t(X_t,\theta), \Omega_t(X_t,\theta)\right)\right\}$ obeys a weak ULLN (White (1994), Definition 3.2).

Assumption R11 $\{A_n^*\}$ is $O(1)$ and negative definite uniformly in n.

Assumption R12 The double array $\left\{n^{-1/2} \frac{\partial}{\partial\theta} \ln f_t\left(Y_t, m_t(X_t,\theta_n^*), \Omega_t(X_t,\theta_n^*)\right)\right\}$ obeys a central limit theorem with covariance matrix $\{B_n^*\}$ (White (1994), Definition 6.3), where $\{B_n^*\}$ is $O(1)$ and positive definite uniformly in n.

Appendix B

This appendix contains the proofs of Property 1-8.

Proof of Property 1 From the Kullback inequality (e.g. White (1994), theorem 2.3), we have

$$\mathbb{I}\left[l(Y, m_o, \Sigma_o) : l(Y, m, \Sigma)\right] = \int \ln\left(\frac{l(Y, m_o, \Sigma_o)}{l(Y, m, \Sigma)}\right) l(Y, m_o, \Sigma_o) \upsilon(dY)$$

$$= \int \ln l(Y, m_o, \Sigma_o) l(Y, m_o, \Sigma_o) \upsilon(dY) - \int \ln l(Y, m, \Sigma) l(Y, m_o, \Sigma_o) \upsilon(dY)$$

$$\begin{aligned} &= A(m_o, \Sigma_o) + C(m_o, \Sigma_o)' E_{l_o}(Y) + \operatorname{tr}\left(D(m_o, \Sigma_o) E_{l_o}(YY')\right) \\ &\quad - \left[A(m, \Sigma) + C(m, \Sigma)' E_{l_o}(Y) + \operatorname{tr}\left(D(m, \Sigma) E_{l_o}(YY')\right)\right] \\ &= A(m_o, \Sigma_o) + C(m_o, \Sigma_o)' m_o + \operatorname{tr}\left(D(m_o, \Sigma_o)(\Sigma_o + m_o m_o')\right) \\ &\quad - \left[A(m, \Sigma) + C(m, \Sigma)' m_o + \operatorname{tr}\left(D(m, \Sigma)(\Sigma_o + m_o m_o')\right)\right] \geq 0 \end{aligned}$$

or

$$\begin{aligned} &A(m_o, \Sigma_o) + C(m_o, \Sigma_o)' m_o + \operatorname{tr}\left(D(m_o, \Sigma_o)(\Sigma_o + m_o m_o')\right) \\ &\quad \geq A(m, \Sigma) + C(m, \Sigma)' m_o + \operatorname{tr}\left(D(m, \Sigma)(\Sigma_o + m_o m_o')\right) \end{aligned}$$

where the equality holds if and only if $l(Y, m, \Sigma) = l(Y, m_o, \Sigma_o)$ $a.s. - \upsilon$, or, given identifiability, if and only if $m = m_o$ and $\Sigma = \Sigma_o$. ■

Proof of Property 2 From Property 1, $\forall\ m,\ m_o \in \mathcal{M}$, $\forall\ \Sigma,\ \Sigma_o \in \mathcal{E}$ such that $\Sigma \neq \Sigma_o$, the following strict inequality holds

$$\begin{aligned} &A(m_o, \Sigma_o) + C(m_o, \Sigma_o)' m_o + \operatorname{tr}\left(D(m_o, \Sigma_o)(\Sigma_o + m_o m_o')\right) \\ &\quad > A(m, \Sigma) + C(m, \Sigma)' m_o + \operatorname{tr}\left(D(m, \Sigma)(\Sigma_o + m_o m_o')\right) \end{aligned}$$

or equivalently, subtracting the same quantity from both sides,

$$\begin{aligned} &A(m_o, \Sigma_o) + C(m_o, \Sigma_o)' m_o + \operatorname{tr}\left(D(m_o, \Sigma_o)(\Sigma_o + m_o m_o')\right) \\ &- \left[A(m_o, \Sigma) + C(m_o, \Sigma)' m_o + \operatorname{tr}\left(D(m_o, \Sigma)(\Sigma_o + m_o m_o')\right)\right] \\ > \quad &A(m, \Sigma) + C(m, \Sigma)' m_o + \operatorname{tr}\left(D(m, \Sigma)(\Sigma_o + m_o m_o')\right) \\ &- \left[A(m_o, \Sigma) + C(m_o, \Sigma)' m_o + \operatorname{tr}\left(D(m_o, \Sigma)(\Sigma_o + m_o m_o')\right)\right] \end{aligned}$$

or, multiplying both sides by minus 1,

$$\begin{aligned} &A(m_o, \Sigma) + C(m_o, \Sigma)' m_o + \operatorname{tr}\left(D(m_o, \Sigma)(\Sigma_o + m_o m_o')\right) \\ &- \left[A(m, \Sigma) + C(m, \Sigma)' m_o + \operatorname{tr}\left(D(m, \Sigma)(\Sigma_o + m_o m_o')\right)\right] \\ > \quad &A(m_o, \Sigma) + C(m_o, \Sigma)' m_o + \operatorname{tr}\left(D(m_o, \Sigma)(\Sigma_o + m_o m_o')\right) \\ &- \left[A(m_o, \Sigma_o) + C(m_o, \Sigma_o)' m_o + \operatorname{tr}\left(D(m_o, \Sigma_o)(\Sigma_o + m_o m_o')\right)\right] \end{aligned} \tag{B-1}$$

Again from property 1, since $\forall\, m_o \in \mathcal{M}$, $\forall\, \Sigma,\, \Sigma_o \in \mathcal{E}$ such that $\Sigma \neq \Sigma_o$, we have

$$\begin{aligned} &A(m_o,\Sigma_o) + C(m_o,\Sigma_o)'m_o + \mathrm{tr}\,(D(m_o,\Sigma_o)(\Sigma_o + m_o m_o')) \\ &> A(m_o,\Sigma) + C(m_o,\Sigma)'m_o + \mathrm{tr}\,(D(m_o,\Sigma)(\Sigma_o + m_o m_o')) \end{aligned}$$

the right-hand side of (B-1) is strictly inferior to zero. Thus, the left-hand side of (B-1) is only necessary superior to a strictly negative quantity, or equivalently, it may exist $m \in \mathcal{M}$ such that $m \neq m_o$ and that we have

$$\begin{aligned} &A(m_o,\Sigma) + C(m_o,\Sigma)'m_o + \mathrm{tr}\,(D(m_o,\Sigma)(\Sigma_o + m_o m_o')) \\ &< A(m,\Sigma) + C(m,\Sigma)'m_o + \mathrm{tr}\,(D(m,\Sigma)(\Sigma_o + m_o m_o')) \end{aligned}$$

■

Proof of Property 3 Differentiating

$$\int l(Y,m,\Sigma)\upsilon(dY) = 1$$

with respect to m and $\mathrm{vec}\,\Sigma$, we get respectively

$$\begin{aligned} \int \frac{\partial \ln l(Y,m,\Sigma)}{\partial m} l(Y,m,\Sigma)\upsilon(dY) = 0 \\ \int \frac{\partial \ln l(Y,m,\Sigma)}{\partial\, \mathrm{vec}\,\Sigma} l(Y,m,\Sigma)\upsilon(dY) = 0 \end{aligned}$$

or, noting that $Y'D(m,\Sigma)Y = (\mathrm{vec}\,D(m,\Sigma))'\,\mathrm{vec}\,(YY')$,

$$\int \left(\frac{\partial A(m,\Sigma)}{\partial m} + \frac{\partial C(m,\Sigma)'}{\partial m} Y + \frac{\partial\,(\mathrm{vec}\,D(m,\Sigma))'}{\partial m}\,\mathrm{vec}\,(YY') \right) l(Y,m,\Sigma)\upsilon(dY) = 0$$

$$\int \left(\frac{\partial A(m,\Sigma)}{\partial\, \mathrm{vec}\,\Sigma} + \frac{\partial C(m,\Sigma)'}{\partial\, \mathrm{vec}\,\Sigma} Y + \frac{\partial\,(\mathrm{vec}\,D(m,\Sigma))'}{\partial\, \mathrm{vec}\,\Sigma}\,\mathrm{vec}\,(YY') \right) l(Y,m,\Sigma)\upsilon(dY) = 0$$

which is equivalent to

$$\begin{aligned} \frac{\partial A(m,\Sigma)}{\partial m} + \frac{\partial C(m,\Sigma)'}{\partial m} E_l(Y) + \frac{\partial\,(\mathrm{vec}\,D(m,\Sigma))'}{\partial m}\,\mathrm{vec}(E_l(YY')) = 0 \\ \frac{\partial A(m,\Sigma)}{\partial\, \mathrm{vec}\,\Sigma} + \frac{\partial C(m,\Sigma)'}{\partial\, \mathrm{vec}\,\Sigma} E_l(Y) + \frac{\partial\,(\mathrm{vec}\,D(m,\Sigma))'}{\partial\, \mathrm{vec}\,\Sigma}\,\mathrm{vec}(E_l(YY')) = 0 \end{aligned}$$

and to

$$\begin{aligned} \frac{\partial A(m,\Sigma)}{\partial m} + \frac{\partial C(m,\Sigma)'}{\partial m} m + \frac{\partial\,(\mathrm{vec}\,D(m,\Sigma))'}{\partial m}\,\mathrm{vec}(\Sigma + mm') = 0 \\ \frac{\partial A(m,\Sigma)}{\partial\, \mathrm{vec}\,\Sigma} + \frac{\partial C(m,\Sigma)'}{\partial\, \mathrm{vec}\,\Sigma} m + \frac{\partial\,(\mathrm{vec}\,D(m,\Sigma))'}{\partial\, \mathrm{vec}\,\Sigma}\,\mathrm{vec}(\Sigma + mm') = 0 \end{aligned}$$

■

Proof of Property 4 (a) As Property 1, it directly follows from the Kullback inequality

$$\mathbb{I}\left[l(Y,m_o,\Sigma_o):l(Y,m,\Sigma)\right]=\int \ln\left(\frac{l(Y,m_o,\Sigma_o)}{l(Y,m,\Sigma)}\right) l(Y,m_o,\Sigma_o)\upsilon(dY)$$

$$\begin{aligned}
&= A(m_o,\Sigma_o)+E_{l_o}\left[B(\Sigma_o,Y)\right]+C(m_o,\Sigma_o)'E_{l_o}(Y)\\
&\quad -\left[A(m,\Sigma)+E_{l_o}\left[B(\Sigma,Y)\right]+C(m,\Sigma)'E_{l_o}(Y)\right]\\
&= A(m_o,\Sigma_o)+E_{l_o}\left[B(\Sigma_o,Y)\right]+C(m_o,\Sigma_o)'m_o\\
&\quad -\left[A(m,\Sigma)+E_{l_o}\left[B(\Sigma,Y)\right]+C(m,\Sigma)'m_o\right]\geq 0
\end{aligned}$$

or

$$\begin{aligned}
&A(m_o,\Sigma_o)+E_{l_o}\left[B(\Sigma_o,Y)\right]+C(m_o,\Sigma_o)'m_o\\
&\geq A(m,\Sigma)+E_{l_o}\left[B(\Sigma,Y)\right]+C(m,\Sigma)'m_o
\end{aligned}$$

where the equality holds if and only if $l(Y,m,\Sigma)=l(Y,m_o,\Sigma_o)$ $a.s.-\upsilon$, or, given identifiability, if and only if $m=m_o$ and $\Sigma=\Sigma_o$. (b) It follows from (a) by taking $\Sigma=\Sigma_o$ and then subtracting equal terms from both sides of the inequality. ■

Proof of Property 5 (a) As for Property 3, differentiating

$$\int l(Y,m,\Sigma)\upsilon(dY)=1$$

with respect to m and $\operatorname{vec}\Sigma$, we get respectively

$$\int \frac{\partial \ln l(Y,m,\Sigma)}{\partial m} l(Y,m,\Sigma)\upsilon(dY)=0$$

$$\int \frac{\partial \ln l(Y,m,\Sigma)}{\partial \operatorname{vec}\Sigma} l(Y,m,\Sigma)\upsilon(dY)=0$$

or

$$\int\left(\frac{\partial A(m,\Sigma)}{\partial m}+\frac{\partial C(m,\Sigma)'}{\partial m}Y\right) l(Y,m,\Sigma)\upsilon(dY)=0$$

$$\int\left(\frac{\partial A(m,\Sigma)}{\partial \operatorname{vec}\Sigma}+\frac{\partial B(\Sigma,Y)}{\partial \operatorname{vec}\Sigma}+\frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma}Y\right) l(Y,m,\Sigma)\upsilon(dY)=0$$

which is equivalent to

$$\frac{\partial A(m,\Sigma)}{\partial m}+\frac{\partial C(m,\Sigma)'}{\partial m}m=0$$

$$\frac{\partial A(m,\Sigma)}{\partial \operatorname{vec}\Sigma}+E_l\left[\frac{\partial B(\Sigma,Y)}{\partial \operatorname{vec}\Sigma}\right]+\frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma}m=0$$

(b) Since m is the mean of the distribution $l(Y,m,\Sigma)$, we have

$$\int Y'l(Y,m,\Sigma)\upsilon(dY)=m'$$

According to the analycity property of the Laplace transform, differentiating this equality with respect to m, we get

$$\int \frac{\partial \ln l(Y,m,\Sigma)}{\partial m} Y' l(Y,m,\Sigma) \upsilon(dY) = I_G$$

or

$$\int \left(\frac{\partial A(m,\Sigma)}{\partial m} + \frac{\partial C(m,\Sigma)'}{\partial m} Y \right) Y' l(Y,m,\Sigma) \upsilon(dY) = I_G$$

Further, using the first equality of (a), we obtain

$$\int \left(\frac{\partial C(m,\Sigma)'}{\partial m} Y - \frac{\partial C(m,\Sigma)'}{\partial m} m \right) Y' l(Y,m,\Sigma) \upsilon(dY) = I_G$$

or

$$\int \frac{\partial C(m,\Sigma)'}{\partial m} (Y - m) Y' l(Y,m,\Sigma) \upsilon(dY) = I_G$$

which, since $E_l\left((Y-m)Y'\right) = \Sigma$ and Σ is positive definite, is equivalent to

$$I_G = \frac{\partial C(m,\Sigma)'}{\partial m} \Sigma \Leftrightarrow \frac{\partial C(m,\Sigma)'}{\partial m} = \Sigma^{-1}$$

■

Proof of Property 6 Since m is the mean of the distribution $l(Y,m,\Sigma)$, we have

$$\int Y' l(Y,m,\Sigma) \upsilon(dY) = m'$$

As above, differentiating this equality with respect to $\operatorname{vec}\Sigma$, we get

$$\int \frac{\partial \ln l(Y,m,\Sigma)}{\partial \operatorname{vec}\Sigma} Y' l(Y,m,\Sigma) \upsilon(dY) = 0$$

or

$$\int \left(\frac{\partial A(m,\Sigma)}{\partial \operatorname{vec}\Sigma} + \frac{\partial B(\Sigma,Y)}{\partial \operatorname{vec}\Sigma} + \frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma} Y \right) Y' l(Y,m,\Sigma) \upsilon(dY) = 0$$

which is equivalent to

$$\frac{\partial A(m,\Sigma)}{\partial \operatorname{vec}\Sigma} m' + E_l\left[\frac{\partial B(\Sigma,Y)}{\partial \operatorname{vec}\Sigma} Y'\right] + \frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma} (\Sigma + mm') = 0 \qquad \text{(B-2)}$$

If $\forall\, Y \in \mathcal{Y}$, $\forall\, m \in int\,\mathcal{M}$, $\forall\, \Sigma \in int\,\mathcal{E}$, we have $\partial B(\Sigma,Y)/\partial \operatorname{vec}\Sigma = 0$, using Property 5(a) which in this case collapses to the equality

$$\frac{\partial A(m,\Sigma)}{\partial \operatorname{vec}\Sigma} = -\frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma} m$$

(B-2) implies that, $\forall\, Y \in \mathcal{Y}$, $\forall\, m \in int\,\mathcal{M}$, $\forall\, \Sigma \in int\,\mathcal{E}$, we have

$$\frac{\partial A(m,\Sigma)}{\partial \operatorname{vec}\Sigma} m' + \frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma} (\Sigma + mm') = \frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma} \Sigma = 0$$

or, since Σ is positive definite,

$$\frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma} = 0 \quad \text{and} \quad \frac{\partial A(m,\Sigma)}{\partial \operatorname{vec}\Sigma} = 0$$

which would mean that $l(Y,m,\Sigma)$ does actually not depend on Σ, i.e., that $l(Y,m,\Sigma)$ is not a restricted generalized linear exponential family. ■

Proof of Property 7 (a) As Property 1 and Property 4(a), whose it is a special case, it directly follows from the Kullback inequality. (b) As Property 4(b), whose it is a special case, it follows from (a) by taking $\Sigma = \Sigma_o$ and then subtracting equal terms from both sides of the inequality. ■

Proof of Property 8 (a) It follows from Property 3 since $D(m,\Sigma) = D(\Sigma) \Leftrightarrow \frac{\partial(\operatorname{vec} D(m,\Sigma))'}{\partial m} = 0, \forall\ m \in int\,\mathcal{M}, \forall\ \Sigma \in int\,\mathcal{E}$ (Obviously, it is also a special case of Property 5(b)). (b) It follows from Property 5(b) since the restricted quadratic exponential family is a special case of restricted generalized linear exponential families. (c) Since m is the mean of the distribution $l(Y,m,\Sigma)$, we have

$$\int Y' l(Y,m,\Sigma)\upsilon(dY) = m'$$

As for Property 6, differentiating this equality with respect to $\operatorname{vec}\Sigma$, we get

$$\int \frac{\partial \ln l(Y,m,\Sigma)}{\partial \operatorname{vec}\Sigma} Y' l(Y,m,\Sigma)\upsilon(dY) = 0$$

or, noting that $Y'D(\Sigma)Y = (\operatorname{vec} D(\Sigma))' \operatorname{vec}(YY')$,

$$\int \left(\frac{\partial A(m,\Sigma)}{\partial \operatorname{vec}\Sigma} + \frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma} Y + \frac{\partial (\operatorname{vec} D(\Sigma))'}{\partial \operatorname{vec}\Sigma} \operatorname{vec}(YY') \right) Y' l(Y,m,\Sigma)\upsilon(dY) = 0$$

which is equivalent to

$$\frac{\partial A(m,\Sigma)}{\partial \operatorname{vec}\Sigma} m' + \frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma} (\Sigma + mm') + \frac{\partial (\operatorname{vec} D(\Sigma))'}{\partial \operatorname{vec}\Sigma} E_l \left[\operatorname{vec}(YY')\, Y'\right] = 0$$

Since $E_l \left[\operatorname{vec}(YY')\, Y'\right] = Cov_l \left[(\operatorname{vec}(YY'), Y)\right] + (\operatorname{vec}(\Sigma + mm'))\, m'$, using Property 8(a) which may be written as

$$\frac{\partial A(m,\Sigma)}{\partial \operatorname{vec}\Sigma} m' + \frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma} mm' + \frac{\partial (\operatorname{vec} D(\Sigma))'}{\partial \operatorname{vec}\Sigma} (\operatorname{vec}(\Sigma + mm'))\, m' = 0$$

by subtraction, we get

$$\frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma} \Sigma + \frac{\partial (\operatorname{vec} D(\Sigma))'}{\partial \operatorname{vec}\Sigma} Cov_l \left[(\operatorname{vec}(YY'), Y)\right] = 0$$

or, since Σ is positive definite,

$$\frac{\partial C(m,\Sigma)'}{\partial \operatorname{vec}\Sigma} = -\frac{\partial (\operatorname{vec} D(\Sigma))'}{\partial \operatorname{vec}\Sigma} Cov_l \left[(\operatorname{vec}(YY'), Y)\right] \Sigma^{-1} \qquad \text{(B-3)}$$

Similarly, by definition, we have

$$\int (\mathrm{vec}\,(YY'))'\, l(Y,m,\Sigma)\upsilon(dY) = (\mathrm{vec}(\Sigma + mm'))'$$

As above, differentiating this equality with respect to $\mathrm{vec}\,\Sigma$, we get

$$\int \frac{\partial \ln l(Y,m,\Sigma)}{\partial\,\mathrm{vec}\,\Sigma} (\mathrm{vec}\,(YY'))'\, l(Y,m,\Sigma)\upsilon(dY) = I_{G^2}$$

or

$$\int \left(\frac{\partial A(m,\Sigma)}{\partial\,\mathrm{vec}\,\Sigma} + \frac{\partial C(m,\Sigma)'}{\partial\,\mathrm{vec}\,\Sigma} Y + \frac{\partial\,(\mathrm{vec}\,D(\Sigma))'}{\partial\,\mathrm{vec}\,\Sigma}\,\mathrm{vec}\,(YY') \right) (\mathrm{vec}\,(YY'))'\, l(Y,m,\Sigma)\upsilon(dY) = I_{G^2}$$

which is equivalent to

$$\frac{\partial A(m,\Sigma)}{\partial\,\mathrm{vec}\,\Sigma} (\mathrm{vec}(\Sigma + mm'))' + \frac{\partial C(m,\Sigma)'}{\partial\,\mathrm{vec}\,\Sigma} E_l\left[Y\,(\mathrm{vec}\,(YY'))'\right] + \frac{\partial\,(\mathrm{vec}\,D(\Sigma))'}{\partial\,\mathrm{vec}\,\Sigma} E_l\left[\mathrm{vec}\,(YY')\,(\mathrm{vec}\,(YY'))'\right] = I_{G^2}$$

Since $E_l\left[Y\,(\mathrm{vec}\,(YY'))'\right] = Cov_l\left[(Y, \mathrm{vec}\,(YY'))\right] + m\,(\mathrm{vec}\,(\Sigma + mm'))'$ and $E_l\left[\mathrm{vec}\,(YY')\,(\mathrm{vec}\,(YY'))'\right] = V_l\left[\mathrm{vec}\,(YY')\right] + \mathrm{vec}(\Sigma + mm')\,(\mathrm{vec}(\Sigma + mm'))'$, using Property 8(a) which may be written as

$$\frac{\partial A(m,\Sigma)}{\partial\,\mathrm{vec}\,\Sigma} (\mathrm{vec}(\Sigma + mm'))' + \frac{\partial C(m,\Sigma)'}{\partial\,\mathrm{vec}\,\Sigma} m\,(\mathrm{vec}(\Sigma + mm'))' + \frac{\partial\,(\mathrm{vec}\,D(\Sigma))'}{\partial\,\mathrm{vec}\,\Sigma}\,\mathrm{vec}(\Sigma + mm')\,(\mathrm{vec}(\Sigma + mm'))' = 0$$

by subtraction, we get

$$\frac{\partial C(m,\Sigma)'}{\partial\,\mathrm{vec}\,\Sigma} Cov_l\left[(Y, \mathrm{vec}\,(YY'))\right] + \frac{\partial\,(\mathrm{vec}\,D(\Sigma))'}{\partial\,\mathrm{vec}\,\Sigma} V_l\left[\mathrm{vec}\,(YY')\right] = I_{G^2}$$

or

$$\frac{\partial C(m,\Sigma)'}{\partial\,\mathrm{vec}\,\Sigma} Cov_l\left[(Y, \mathrm{vec}\,(YY'))\right] = I_{G^2} - \frac{\partial\,(\mathrm{vec}\,D(\Sigma))'}{\partial\,\mathrm{vec}\,\Sigma} V_l\left[\mathrm{vec}\,(YY')\right] \qquad \text{(B-4)}$$

(d) Substituting (B-3) into (B-4), we obtain

$$\begin{aligned} & -\frac{\partial\,(\mathrm{vec}\,D(\Sigma))'}{\partial\,\mathrm{vec}\,\Sigma} Cov_l\left[(\mathrm{vec}\,(YY')\,, Y)\right] \Sigma^{-1} Cov_l\left[(Y, \mathrm{vec}\,(YY'))\right] \\ = \; & I_{G^2} - \frac{\partial\,(\mathrm{vec}\,D(\Sigma))'}{\partial\,\mathrm{vec}\,\Sigma} V_l\left[\mathrm{vec}\,(YY')\right] \end{aligned}$$

or

$$I_{G^2} = \frac{\partial\,(\mathrm{vec}\,D(\Sigma))'}{\partial\,\mathrm{vec}\,\Sigma} \left(V_l\left[\mathrm{vec}\,(YY')\right] - Cov_l\left[(\mathrm{vec}\,(YY')\,, Y)\right] \Sigma^{-1} Cov_l\left[(Y, \mathrm{vec}\,(YY'))\right] \right)$$

Now, since $\mathrm{vec}\,(YY') = D_{uG}\,\mathrm{vech}\,(YY')$, we have that

$$\begin{aligned} & V_l\left[\mathrm{vec}\,(YY')\right] - Cov_l\left[(\mathrm{vec}\,(YY'),Y)\right]\Sigma^{-1}Cov_l\left[(Y,\mathrm{vec}\,(YY'))\right] \\ = \; & D_{uG}\left(V_l\left[\mathrm{vech}\,(YY')\right] - Cov_l\left[(\mathrm{vech}\,(YY'),Y)\right]\Sigma^{-1}Cov_l\left[(Y,\mathrm{vech}\,(YY'))\right]\right)D'_{uG}\end{aligned}$$

and thus

$$I_{G^2} = \frac{\partial\,(\mathrm{vec}\,D(\Sigma))'}{\partial\,\mathrm{vec}\,\Sigma}D_{uG}\begin{pmatrix} V_l\left[\mathrm{vech}\,(YY')\right] - Cov_l\left[(\mathrm{vech}\,(YY'),Y)\right] \\ \Sigma^{-1}Cov_l\left[(Y,\mathrm{vech}\,(YY'))\right]\end{pmatrix}D'_{uG}$$

or, post-multiplying both sides by $D_{uG}\left(D'_{uG}D_{uG}\right)^{-1}$,

$$D_{uG}\left(D'_{uG}D_{uG}\right)^{-1} = \frac{\partial\,(\mathrm{vec}\,D(\Sigma))'}{\partial\,\mathrm{vec}\,\Sigma}D_{uG}\begin{pmatrix} V_l\left[\mathrm{vech}\,(YY')\right] - Cov_l\left[(\mathrm{vech}\,(YY'),Y)\right] \\ \Sigma^{-1}Cov_l\left[(Y,\mathrm{vech}\,(YY'))\right]\end{pmatrix}$$

$(V_l\left[\mathrm{vech}\,(YY')\right] - Cov_l\left[(\mathrm{vech}\,(YY'),Y)\right]\Sigma^{-1}Cov_l\left[(Y,\mathrm{vech}\,(YY'))\right])$ is just the lower diagonal block of the inverse of the covariance matrix $V_l\left[\left((Y',\mathrm{vech}\,(YY'))'\right)'\right]$, it is thus positive definite if the latter matrix is positive definite. We may thus write

$$\frac{\partial\,(\mathrm{vec}\,D(\Sigma))'}{\partial\,\mathrm{vec}\,\Sigma}D_{uG} = D_{uG}\left(D'_{uG}D_{uG}\right)^{-1}\begin{pmatrix} V_l\left[\mathrm{vech}\,(YY')\right] - Cov_l\left[(\mathrm{vech}\,(YY'),Y)\right] \\ \Sigma^{-1}Cov_l\left[(Y,\mathrm{vech}\,(YY'))\right]\end{pmatrix}^{-1}$$

or, post-multiplying both sides by $\left(D'_{uG}D_{uG}\right)^{-1}D'_{uG} = D^{+}_{uG}$, since $D_{uG}D^{+}_{uG} = N_G$,

$$\frac{\partial\,(\mathrm{vec}\,D(\Sigma))'}{\partial\,\mathrm{vec}\,\Sigma}N_G = \left(D^{+}_{uG}\right)'\begin{pmatrix} V_l\left[\mathrm{vech}\,(YY')\right] - Cov_l\left[(\mathrm{vech}\,(YY'),Y)\right] \\ \Sigma^{-1}Cov_l\left[(Y,\mathrm{vech}\,(YY'))\right]\end{pmatrix}^{-1}D^{+}_{uG}$$

■

Appendix C

This appendix contains the proofs of Proposition 2, 6 and 10-13.

Proof of Proposition 2 ($G = 1$) It suffices to show that the condition is already necessary for a given choice $\widetilde{P}_o$ and $\widetilde{\mathcal{S}}$ of respectively P_o and $\mathcal{S}$ which satisfies the assumptions of the Proposition. Suppose that $\widetilde{P}_o$ is such that each Y_t is distributed with $E\left(Y_t|X_t\right) = m_o$ and $V\left(Y_t|X_t\right) = \sigma^2_o$, and that $\widetilde{\mathcal{S}}$ is specified as $\{m_t(X_t,\theta) = m\}$ and $\{\Omega_t(X_t,\theta) = \sigma^2\}$. Given regularity conditions R1-R5, R6′, from Theorem 3.5 of White (1994), we have $\hat{\theta}_n - \theta^*_n \to 0$ as $n \to \infty$ $a.s. - P_o$, where $\theta^*_n = \mathrm{Argmax}_{\theta\in\Theta}E\left(L_n(Y^n,X^n,\theta)\right)$, and $\theta^*_n \in int\,\Theta$. Further, given R7-R9, $\theta^*_n = \mathrm{Argmax}_{\theta\in\Theta}E\left(L_n(Y^n,X^n,\theta)\right)$ implies that $E\left(\frac{\partial}{\partial\theta}L_n(Y^n,X^n,\theta^*_n)\right) = 0$. Given the assumed structure for $\widetilde{P}_o$ and $\widetilde{\mathcal{S}}$, for all $n = 1,2,...$, we thus have $\theta^*_n = (m_o,\sigma^2_o)$ and

$$\frac{1}{n}\sum_{t=1}^{n}E\left[\frac{\partial}{\partial m}\ln f_t\left(Y_t,m_o,\sigma^2_o\right)\right] = 0 \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n}E\left[\frac{\partial}{\partial \sigma^2}\ln f_t\left(Y_t,m_o,\sigma^2_o\right)\right] = 0$$

These relations must hold for all $n = 1,2,...$ and any "regular", i.e., which satisfies the assumptions of the Proposition, choice of $\widetilde{P}_o, m_o, \sigma_o^2$, and f_t. It follows that for any such choice and for all $t = 1, 2, ...$, we must have

$$E\left[\frac{\partial}{\partial m} \ln f_t\left(Y_t, m_o, \sigma_o^2\right)\right] = 0 \quad \text{and} \quad E\left[\frac{\partial}{\partial \sigma^2} \ln f_t\left(Y_t, m_o, \sigma_o^2\right)\right] = 0$$

From this point, the proof of theorem 7 in Gourieroux-Monfort-Trognon (1984a) applies for each $t = 1, 2, ...$ such that, for all $t = 1, 2, ...$, we must have

$$\ln f_t\left(Y, m, \sigma^2\right) = A(m, \sigma^2) + B(Y) + C(m, \sigma^2)Y + D(m, \sigma^2)Y^2$$

■

Proof of Proposition 6 We proceed in two steps. First, we show that f_t has to belong to the restricted generalized linear exponential family. The proof of this first step is adapted from a similar proof given by Gourieroux-Monfort-Trognon (1984a) and White (1994) for PML1. Then, using this first result, we show that mean and variance parameters have to vary independently.

First step: It suffices to show that the condition is already necessary for a given choice $\widetilde{P}_o$ and $\widetilde{\mathcal{S}}$ of respectively P_o and $\mathcal{S}$ which satisfies the assumptions of the Proposition. For now, set $G = 1$ and suppose that $\widetilde{P}_o$ is such that each Y_t is distributed with $E\left(Y_t|X_t\right) = m_o$ and $V\left(Y_t|X_t\right) = \sigma_{ot}^2$, and that $\widetilde{\mathcal{S}}$ is specified as $\{m_t(X_t, \theta_1) = m\}$ and $\{\Omega_t(X_t, \theta_1, \theta_2) = \sigma^2\}$. Given regularity conditions R1-R5, R6′, from Theorem 3.5 of White (1994), we have $\hat{\theta}_n - \theta_n^* \to 0$ as $n \to \infty$ $a.s. - P_o$, where $\theta_n^* = \text{Argmax}_{\theta \in \Theta} E\left(L_n(Y^n, X^n, \theta)\right)$ and $\theta_n^* \in int\, \Theta$. Further, given R7-R9, $\theta_n^* = \text{Argmax}_{\theta \in \Theta} E\left(L_n(Y^n, X^n, \theta)\right)$ implies that $E\left(\frac{\partial}{\partial \theta} L_n(Y^n, X^n, \theta_n^*)\right) = 0$. Given the assumed structure for $\widetilde{P}_o$ and $\widetilde{\mathcal{S}}$, for all $n = 1, 2, ...$, we thus have $\theta_n^* = (m_o, \sigma_n^{2*})$ and

$$\frac{1}{n} \sum_{t=1}^{n} E\left[\frac{\partial}{\partial m} \ln f_t\left(Y_t, m_o, \sigma_n^{2*}\right)\right] = 0$$

This relation must hold for all $n = 1, 2, ...$ and any "regular", i.e., which satisfies the assumptions of the Proposition, choice of $\widetilde{P}_o, m_o, \sigma_{ot}^2$, and f_t. It follows that for any such choice and for all $t = 1, 2, ...$, we must have

$$E\left[\frac{\partial}{\partial m} \ln f_t\left(Y_t, m_o, \sigma_n^{2*}\right)\right] = 0$$

In particular, this has to be true if the support of each Y_t consists in two points y_t^1 and y_t^2 such that $-\infty < y_t^1 < m_o < y_t^2 < \infty$. Then, for any "regular" choice of y_t^1, y_t^2, m_o and f_t — once y_t^1, y_t^2 , m_o and f_t are chosen, σ_{ot}^2 and σ_n^{2*} are enforced —, and for all $t = 1, 2, ...$, we must have

$$m_o = p_t^1 y_t^1 + \left(1 - p_t^1\right) y_t^2 \tag{C-1}$$

where $\widetilde{P}_o[Y_t = y_t^1] = p_t^1$ and $\widetilde{P}_o[Y_t = y_t^2] = 1 - p_t^1$, and

$$E\left[\frac{\partial}{\partial m} \ln f_t\left(Y_t, m_o, \sigma_n^{2*}\right)\right]$$

$$= \; p_t^1 \frac{\partial \ln f_t\left(y_t^1, m_o, \sigma_n^{2*}\right)}{\partial m} + \left(1 - p_t^1\right) \frac{\partial \ln f_t\left(y_t^2, m_o, \sigma_n^{2*}\right)}{\partial m} = 0 \qquad \text{(C-2)}$$

Since from (C-1), we have that $p_t^1 = (y_t^2 - m_o)/(y_t^2 - y_t^1)$ and $1 - p_t^1 = (m_o - y_t^1)/(y_t^2 - y_t^1)$, (C-2) may be written as

$$(y_t^2 - m_o)\frac{\partial \ln f_t\left(y_t^1, m_o, \sigma_n^{2*}\right)}{\partial m} + (m_o - y_t^1)\frac{\partial \ln f_t\left(y_t^2, m_o, \sigma_n^{2*}\right)}{\partial m} = 0 \qquad \text{(C-3)}$$

So, we have that

$$\frac{\partial \ln f_t\left(y_t^2, m_o, \sigma_n^{2*}\right)}{\partial m} = (y_t^2 - m_o)\left(\frac{\partial \ln f_t\left(y_t^1, m_o, \sigma_n^{2*}\right)}{\partial m} / (y_t^1 - m_o)\right)$$

Now, fix y_t^1 and consider y_t^2, m_o and σ_n^{2*} as variables. Since the above relation must hold for any y_t^1, it follows that for any "regular" choice of y_t^1, y_t^2, m_o and f_t, and for all $t = 1, 2, ...$, we must have

$$\frac{\partial \ln f_t\left(y_t^2, m_o, \sigma_n^{2*}\right)}{\partial m} = (y_t^2 - m_o)\,\phi_t^2(m_o, \sigma_n^{2*}) \qquad \text{(C-4)}$$

where $\phi_t^2(m_o, \sigma_n^{2*}) = \frac{\partial \ln f_t\left(y_t^1, m_o, \sigma_n^{2*}\right)}{\partial m} / (y_t^1 - m_o)$ depends only on m_o and σ_n^{2*}.

The same reasoning establishes that, for any "regular" choice of y_t^1, y_t^2, m_o and f_t, and for all $t = 1, 2, ...$, we must have

$$\frac{\partial \ln f_t\left(y_t^1, m_o, \sigma_n^{2*}\right)}{\partial m} = (y_t^1 - m_o)\,\phi_t^1(m_o, \sigma_n^{2*}) \qquad \text{(C-5)}$$

where $\phi_t^1(m_o, \sigma_n^{2*}) = \frac{\partial \ln f_t\left(y_t^2, m_o, \sigma_n^{2*}\right)}{\partial m} / (y_t^2 - m_o)$ depends only on m_o and σ_n^{2*}.

Substituting (C-4) and (C-5) into (C-3), we get

$$(y_t^2 - m_o)(y_t^1 - m_o)\,\phi_t^1(m_o, \sigma_n^{2*}) - (y_t^1 - m_o)(y_t^2 - m_o)\,\phi_t^2(m_o, \sigma_n^{2*}) = 0$$

or

$$\phi_t^1(m_o, \sigma_n^{2*}) = \; \phi_t^2(m_o, \sigma_n^{2*}) = \phi_t(m_o, \sigma_n^{2*})$$

Hence, for any "regular" choice of y_t^i $(i = 1, 2)$, m_o and f_t, and for all $t = 1, 2, ...$, we must have

$$\frac{\partial \ln f_t\left(y_t^i, m_o, \sigma_n^{2*}\right)}{\partial m} = (y_t^i - m_o)\,\phi_t(m_o, \sigma_n^{2*}) \qquad \text{(C-6)}$$

The result that, for all $t = 1, 2, ...$, f_t has to belong to the restricted generalized linear exponential family

$$\ln f_t\left(Y, m, \sigma^2\right) = A_t(m, \sigma^2) + B_t(\sigma^2, Y) + C_t(m, \sigma^2)'Y$$

follows by recalling that, by assumption, $f_t\left(Y, m, \sigma^2\right)$ is a p.d.f. with $E(Y) = m$ and $V(Y) = \sigma^2$ and integrating both sides of (C-6) with respect to m on $int\,\mathcal{M}_t$ $(= int\,\Theta_1)$, a connected set, and extending to $\mathcal{M}_t$ $(= \Theta_1)$ using the continuity of $\ln f_t\left(y, m, \sigma^2\right)$ on $\mathcal{M}_t$ $(= \Theta_1)$. For the case $G > 1$, a similar proof is available by considering distributions whose supports consist in $G + 1$ independent points $y_t^1, ...y_t^{G+1}$ so that $\sum_{i=1}^{G=1} p_t^i y_t^i = m_o$.

Second step: Using the result of step one, it again suffices to show that the condition is already necessary for a given choice $\widetilde{P}_o$ and $\widetilde{\mathcal{S}}$ of respectively P_o and $\mathcal{S}$ which satisfies the assumptions of the Proposition. Suppose that $\widetilde{P}_o$ is such that each Y_t is distributed with $E\left(Y_t|X_t\right) = \theta_1^o$, $V\left(Y_t|X_t\right) = \Sigma_t^o$, and that $\widetilde{\mathcal{S}}$ is specified as $\{m_t(X_t, \theta_1) = \theta_1\}$ and $\{\Omega_t(X_t, \theta_1, \theta_2)\}$. As already outlined, given regularity conditions R1-R5, R6′, from Theorem 3.5 of White (1994), we have $\hat{\theta}_n - \theta_n^* \to 0$ as $n \to \infty$ $a.s. - P_o$, where $\theta_n^* = \text{Argmax}_{\theta\in\Theta} E\left(L_n(Y^n, X^n, \theta)\right)$ and $\theta_n^* \in int\,\Theta$. Further, given R7-R9, $\theta_n^* = \text{Argmax}_{\theta\in\Theta} E\left(L_n(Y^n, X^n, \theta)\right)$ implies that $E\left(\frac{\partial}{\partial\theta} L_n(Y^n, X^n, \theta_n^*)\right) = 0$. Given the assumed structure for $\widetilde{P}_o$ and $\widetilde{\mathcal{S}}$, for all $n = 1, 2, ...$, we thus have $\theta_n^* = \left(\theta_1^{o\prime}, \theta_{2_n}^{*\prime}\right)'$ and

$$\frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial}{\partial\theta_1} \ln f_t\left(Y_t, \theta_1^o, \Omega_t(X_t, \theta_1^o, \theta_{2_n}^*)\right)\right] = 0 \tag{C-7}$$

Since for all $t = 1, 2, ...$, from step one, f_t has to belong to the restricted generalized linear exponential family, letting Ω_t^{o*} stand for $\Omega_t(X_t, \theta_1^o, \theta_{2_n}^*)$ and recalling that, for L scalar, m and θ column vectors and Ω a matrix, by chain rule,

$$\begin{aligned} \frac{\partial}{\partial\theta} L\left(m\left(\theta\right), \Omega\left(\theta\right)\right) &= \left(\frac{\partial L}{\partial m'}\frac{\partial m}{\partial\theta'} + \frac{\partial L}{\partial\left(\operatorname{vec}\Omega\right)'}\frac{\partial \operatorname{vec}\Omega}{\partial\theta'}\right)' \\ &= \frac{\partial m'}{\partial\theta}\frac{\partial L}{\partial m} + \frac{\partial\left(\operatorname{vec}\Omega\right)'}{\partial\theta}\frac{\partial L}{\partial\operatorname{vec}\Omega} \end{aligned} \tag{C-8}$$

we have

$$\begin{aligned} & \frac{\partial}{\partial\theta_1} \ln f_t\left(Y_t, \theta_1^o, \Omega_t^{o*}\right) \\ = \; & \frac{\partial}{\partial\theta_1}\left[A_t(\theta_1^o, \Omega_t^{o*}) + B_t(\Omega_t^{o*}, Y_t) + C_t(\theta_1^o, \Omega_t^{o*})' Y_t\right] \\ = \; & \frac{\partial A_t(\theta_1^o, \Omega_t^{o*})}{\partial\theta_1} + \frac{\partial C_t(\theta_1^o, \Omega_t^{o*})'}{\partial\theta_1} Y_t + \frac{\partial\left(\operatorname{vec}\Omega_t^{o*}\right)'}{\partial\theta_1}\left(\begin{array}{c} \dfrac{\partial A_t(\theta_1^o, \Omega_t^{o*})}{\partial\operatorname{vec}\Omega_t} + \dfrac{\partial B_t(\Omega_t^{o*}, Y_t)}{\partial\operatorname{vec}\Omega_t} \\ + \dfrac{\partial C_t(\theta_1^o, \Omega_t^{o*})'}{\partial\operatorname{vec}\Omega_t} Y_t \end{array}\right) \end{aligned}$$

Given first order correct specification, $E\left(Y_t|X_t\right)$ by definition exists and the law of iterated expectations applies such that (C-7) may be written

$$\begin{aligned} 0 = \frac{1}{n}\sum_{t=1}^{n} E\Bigg[&\left(\frac{\partial A_t(\theta_1^o, \Omega_t^{o*})}{\partial\theta_1} + \frac{\partial C_t(\theta_1^o, \Omega_t^{o*})'}{\partial\theta_1} E(Y_t|X_t)\right) \\ &+ \frac{\partial\left(\operatorname{vec}\Omega_t^{o*}\right)'}{\partial\theta_1}\left(\frac{\partial A_t(\theta_1^o, \Omega_t^{o*})}{\partial\operatorname{vec}\Omega_t} + \frac{\partial B_t(\Omega_t^{o*}, Y_t)}{\partial\operatorname{vec}\Omega_t} + \frac{\partial C_t(\theta_1^o, \Omega_t^{o*})'}{\partial\operatorname{vec}\Omega_t} E(Y_t|X_t)\right)\Bigg] \end{aligned}$$

or

$$\begin{aligned} 0 = \frac{1}{n}\sum_{t=1}^{n} E\Bigg[&\left(\frac{\partial A_t(\theta_1^o, \Omega_t^{o*})}{\partial\theta_1} + \frac{\partial C_t(\theta_1^o, \Omega_t^{o*})'}{\partial\theta_1}\theta_1^o\right) \\ &+ \frac{\partial\left(\operatorname{vec}\Omega_t^{o*}\right)'}{\partial\theta_1}\left(\frac{\partial A_t(\theta_1^o, \Omega_t^{o*})}{\partial\operatorname{vec}\Omega_t} + \frac{\partial B_t(\Omega_t^{o*}, Y_t)}{\partial\operatorname{vec}\Omega_t} + \frac{\partial C_t(\theta_1^o, \Omega_t^{o*})'}{\partial\operatorname{vec}\Omega_t}\theta_1^o\right)\Bigg] \end{aligned}$$

Since this relation must hold for all $n = 1, 2, ...$ and any "regular" choice of $\widetilde{P}_o$, θ_1^o, Σ_t^o, f_t, $\Omega_t(.)$ and X_t (simply assume that $\{X_t\}$ is a sequence of degenerated random variables, i.e., a sequence of a.s. constants), it follows that for any such choice and for all $t = 1, 2, ...$, we must have

$$\begin{aligned} 0 &= \frac{\partial A_t(\theta_1^o, \Omega_t^{o*})}{\partial \theta_1} + \frac{\partial C_t(\theta_1^o, \Omega_t^{o*})'}{\partial \theta_1}\theta_1^o \\ &+ \frac{\partial\,(\operatorname{vec}\Omega_t^{o*})'}{\partial \theta_1}\left(\frac{\partial A_t(\theta_1^o, \Omega_t^{o*})}{\partial \operatorname{vec}\Omega_t} + E\left[\left.\frac{\partial B_t(\Omega_t^{o*}, Y_t)}{\partial \operatorname{vec}\Omega_t}\right| X_t\right] + \frac{\partial C_t(\theta_1^o, \Omega_t^{o*})'}{\partial \operatorname{vec}\Omega_t}\theta_1^o\right) \end{aligned} \tag{C-9}$$

Now, from Property 5(a), we have

$$\begin{aligned} 0 &= \frac{\partial A_t(\theta_1^o, \Omega_t^{o*})}{\partial \theta_1} + \frac{\partial C_t(\theta_1^o, \Omega_t^{o*})'}{\partial \theta_1}\theta_1^o \\ &+ \frac{\partial\,(\operatorname{vec}\Omega_t^{o*})'}{\partial \theta_1}\left(\frac{\partial A_t(\theta_1^o, \Omega_t^{o*})}{\partial \operatorname{vec}\Omega_t} + E_{\lambda_t^*}\left[\left.\frac{\partial B_t(\Omega_t^{o*}, Y_t)}{\partial \operatorname{vec}\Omega_t}\right| X_t\right] + \frac{\partial C_t(\theta_1^o, \Omega_t^{o*})'}{\partial \operatorname{vec}\Omega_t}\theta_1^o\right) \end{aligned} \tag{C-10}$$

where $E_{\lambda_t^*}[.|X_t]$ denotes expectation taken with respect to $\lambda_t(Y_t, X_t, \theta_1^o, \theta_{2_n}^*)$. Then, subtracting (C-10) from (C-9), we get

$$\frac{\partial\,(\operatorname{vec}\Omega_t^{o*})'}{\partial \theta_1}\left(E\left[\left.\frac{\partial B_t(\Omega_t^{o*}, Y_t)}{\partial \operatorname{vec}\Omega_t}\right| X_t\right] - E_{\lambda_t^*}\left[\left.\frac{\partial B_t(\Omega_t^{o*}, Y_t)}{\partial \operatorname{vec}\Omega_t}\right| X_t\right]\right) = 0 \tag{C-11}$$

The relation (C-11) must hold for any "regular" choice of $\widetilde{P}_o$, θ_1^o, Σ_t^o, f_t, $\Omega_t(.)$ and X_t. Since, according to Property 6, we cannot have $\partial B_t(\Omega_t^{o*}, Y_t)/\partial \operatorname{vec}\Omega_t = 0$, it follows that for any such choice and for all $t = 1, 2, ...$, we must have

$$\frac{\partial\,(\operatorname{vec}\Omega_t^{o*})'}{\partial \theta_1} = \frac{\partial\left(\operatorname{vec}\Omega_t(X_t, \theta_1^o, \theta_{2_n}^*)\right)'}{\partial \theta_1} = 0$$

which means that $\Omega_t(X_t, \theta_1, \theta_{2_n}^*) = \Omega_t(X_t, \theta_{2_n}^*)$, $\forall\ \theta_1 \in int\,\Theta_1$, $\forall\ \theta_{2_n}^* \in int\,\Theta_2$ and $\forall$ $X_t \in \mathcal{X}_t$, or in other words that, for all $t = 1, 2, ...$, $\Omega_t(X_t, \theta_1, \theta_2)$ does not depend on θ_1. ■

Proof of Proposition 10 From Proposition 9, we have that $\hat{\theta}_n - \theta_n^* \to 0$ as $n \to \infty$ $a.s. - P_o$ and $B_n^{*-1/2} A_n^* \sqrt{n}(\hat{\theta}_n - \theta_n^*) \xrightarrow{d} N(0, I_{k_\theta})$ where $\theta_n^* = \operatorname{Argmax}_{\theta\in\Theta} E(L_n(Y^n, X^n, \theta))$, $A_n^* = E(\frac{\partial^2}{\partial\theta\partial\theta'} L_n(Y^n, X^n, \theta_n^*))$, $B_n^* = V(n^{1/2}\frac{\partial}{\partial\theta} L_n(Y^n, X^n, \theta_n^*))$, so that $avar\ \hat{\theta}_n = C_n^* = A_n^{*-1} B_n^* A_n^{*-1}$. Now, since, for all $t = 1, 2, ...$, f_t belongs to the restricted quadratic exponential family, and thus to the restricted generalized linear exponential family, and $\mathcal{S}$ is as described in Assumption 1 and is first order correctly specified, according to the proof of Proposition 5, for all $n = 1, 2, ...$, we have $\theta_n^* = \left(\theta_1^{o\prime}, \theta_{2_n}^{*\prime}\right)'$ and

$$L_n(Y^n, X^n, \theta) = \frac{1}{n}\sum_{t=1}^{n} \ln f_t\left(Y_t, m_t(X_t, \theta_1), \Omega_t(X_t, \theta_2)\right) = \frac{1}{n}\sum_{t=1}^{n} \ln \lambda_t\left(Y_t, X_t, \theta_1, \theta_2\right)$$

where, letting m_t stand for $m_t(X_t, \theta_1)$ and Ω_t stand for $\Omega_t(X_t, \theta_2)$, for all $t = 1, 2, ...$,

$$\ln \lambda_t\left(Y_t, X_t, \theta_1, \theta_2\right) = A_t(m_t, \Omega_t) + B_t(Y_t) + C_t(m_t, \Omega_t)'Y_t + Y_t' D_t(\Omega_t) Y_t$$

Then, recalling the chain rule (C-8), we get

$$s_t^1 = \frac{\partial \ln \lambda_t (Y_t, X_t, \theta_1, \theta_2)}{\partial \theta_1} = \frac{\partial m_t'}{\partial \theta_1} \left(\frac{\partial A_t}{\partial m_t} + \frac{\partial C_t'}{\partial m_t} Y_t \right)$$

or, given Property 8(a) and 8(b),

$$s_t^1 = \frac{\partial m_t'}{\partial \theta_1} \Omega_t^{-1} (Y_t - m_t) \tag{C-12}$$

Similarly, noting that $Y_t' D_t(\Omega_t) Y_t = (\operatorname{vec} D_t(\Omega_t))' \operatorname{vec}(Y_t Y_t')$, we get

$$s_t^2 = \frac{\partial \ln \lambda_t (Y_t, X_t, \theta_1, \theta_2)}{\partial \theta_2} = \frac{\partial (\operatorname{vec} \Omega_t)'}{\partial \theta_2} \left(\frac{\partial A_t}{\partial \operatorname{vec} \Omega_t} + \frac{\partial C_t'}{\partial \operatorname{vec} \Omega_t} Y_t + \frac{\partial (\operatorname{vec} D_t)'}{\partial \operatorname{vec} \Omega_t} \operatorname{vec}(Y_t Y_t') \right)$$

or, given Property 8(a),

$$s_t^2 = \frac{\partial (\operatorname{vec} \Omega_t)'}{\partial \theta_2} \left(\frac{\partial C_t'}{\partial \operatorname{vec} \Omega_t} (Y_t - m_t) + \frac{\partial (\operatorname{vec} D_t)'}{\partial \operatorname{vec} \Omega_t} \operatorname{vec}(Y_t Y_t' - \Omega_t - m_t m_t') \right) \tag{C-13}$$

Further, from Property 8(d), we have

$$\begin{aligned} \frac{\partial (\operatorname{vec} D_t)'}{\partial \operatorname{vec} \Omega_t} N_G &= \left(D_{uG}^+\right)' \begin{pmatrix} V_{\lambda_t} [\operatorname{vech}(Y_t Y_t') | X_t] - Cov_{\lambda_t} [(\operatorname{vech}(Y_t Y_t'), Y_t) | X_t] \\ \Omega_t^{-1} Cov_{\lambda_t} [(Y_t, \operatorname{vech}(Y_t Y_t')) | X_t] \end{pmatrix}^{-1} D_{uG}^+ \\ &= \left(D_{uG}^+\right)' \left(M_t^4 - M_t^3 \Omega_t^{-1} M_t^{3\prime} \right)^{-1} D_{uG}^+ \end{aligned} \tag{C-14}$$

where $M_t^3 = Cov_{\lambda_t} [(\operatorname{vech}(Y_t Y_t'), Y_t) | X_t]$, $M_t^4 = V_{\lambda_t} [\operatorname{vech}(Y_t Y_t') | X_t]$, and $Cov_{\lambda_t} [.|X_t]$ and $V_{\lambda_t} [.|X_t]$ denotes respectively covariance and variance taken with respect to $\lambda_t (Y_t, X_t, \theta_1, \theta_2)$. Likewise, from Property 8(c), we have

$$\frac{\partial C_t'}{\partial \operatorname{vec} \Omega_t} = -\frac{\partial (\operatorname{vec} D_t)'}{\partial \operatorname{vec} \Omega_t} Cov_{\lambda_t} [(\operatorname{vec}(Y_t Y_t'), Y_t) | X_t] \Omega_t^{-1}$$

or, since $\operatorname{vec}(YY') = D_{uG} \operatorname{vech}(YY')$ and $N_G D_{uG} = D_{uG}$,

$$\begin{aligned} \frac{\partial C_t'}{\partial \operatorname{vec} \Omega_t} &= -\frac{\partial (\operatorname{vec} D_t)'}{\partial \operatorname{vec} \Omega_t} D_{uG} Cov_{\lambda_t} [(\operatorname{vech}(Y_t Y_t'), Y_t) | X_t] \Omega_t^{-1} \\ &= -\frac{\partial (\operatorname{vec} D_t)'}{\partial \operatorname{vec} \Omega_t} D_{uG} M_t^3 \Omega_t^{-1} = -\frac{\partial (\operatorname{vec} D_t)'}{\partial \operatorname{vec} \Omega_t} N_G D_{uG} M_t^3 \Omega_t^{-1} \end{aligned}$$

which, using (C-14) and the fact that $D_{uG}^+ D_{uG} = I_{\frac{1}{2}G(G+1)}$, is equivalent to

$$\begin{aligned} \frac{\partial C_t'}{\partial \operatorname{vec} \Omega_t} &= -\left(D_{uG}^+\right)' \left(M_t^4 - M_t^3 \Omega_t^{-1} M_t^{3\prime} \right)^{-1} D_{uG}^+ D_{uG} M_t^3 \Omega_t^{-1} \\ &= -\left(D_{uG}^+\right)' \left(M_t^4 - M_t^3 \Omega_t^{-1} M_t^{3\prime} \right)^{-1} M_t^3 \Omega_t^{-1} \end{aligned} \tag{C-15}$$

Since for A symmetric, $\operatorname{vec} A = N_G \operatorname{vec} A$, (C-13) may be written as

$$s_t^2 = \frac{\partial\left(\operatorname{vec}\Omega_t\right)'}{\partial\theta_2}\left(\frac{\partial C_t'}{\partial\operatorname{vec}\Omega_t}\left(Y_t - m_t\right) + \frac{\partial\left(\operatorname{vec}D_t\right)'}{\partial\operatorname{vec}\Omega_t}N_G\operatorname{vec}\left(Y_tY_t' - \Omega_t - m_tm_t'\right)\right)$$

or, substituting (C-14) and (C-15),

$$\begin{aligned} s_t^2 &= \frac{\partial\left(\operatorname{vec}\Omega_t\right)'}{\partial\theta_2}\left(D_{uG}^+\right)'\left(M_t^4 - M_t^3\Omega_t^{-1}M_t^{3\prime}\right)^{-1} \\ &\quad \left(D_{uG}^+\operatorname{vec}\left(Y_tY_t' - \Omega_t - m_tm_t'\right) - M_t^3\Omega_t^{-1}\left(Y_t - m_t\right)\right) \end{aligned}$$

Noting that, for A symmetric, $D_{uG}^+\operatorname{vec}A = \operatorname{vech}A$ and that, since Ω_t is symmetric, $D_{uG}^+\frac{\partial\operatorname{vec}\Omega_t}{\partial\theta_2'} = \frac{\partial\operatorname{vech}\Omega_t}{\partial\theta_2'}$, we finally get

$$\begin{aligned} s_t^2 &= \frac{\partial\left(\operatorname{vech}\Omega_t\right)'}{\partial\theta_2}\left(M_t^4 - M_t^3\Omega_t^{-1}M_t^{3\prime}\right)^{-1} \\ &\quad \left(\operatorname{vech}\left(Y_tY_t' - \Omega_t - m_tm_t'\right) - M_t^3\Omega_t^{-1}\left(Y_t - m_t\right)\right) \end{aligned} \tag{C-16}$$

s_t^{1*} and s_t^{2*} follow by evaluating (C-12), (C-13) and (C-16) at $\theta_n^* = \left(\theta_1^{o\prime}, \theta_{2_n}^{*\prime}\right)'$.
Recalling that $\operatorname{vec}ABC = (C' \otimes A)\operatorname{vec}B$, (C-12) may be written

$$s_t^1 = \frac{\partial m_t'}{\partial\theta_1}\Omega_t^{-1}\left(Y_t - m_t\right) = \operatorname{vec}s_t^1 = \left(\left(Y_t - m_t\right)' \otimes \frac{\partial m_t'}{\partial\theta_1}\right)\operatorname{vec}\Omega_t^{-1}$$

Then, we have that

$$\begin{aligned} \frac{\partial^2\ln\lambda_t\left(Y_t, X_t, \theta_1, \theta_2\right)}{\partial\theta_1\partial\theta_2'} &= \frac{\partial s_t^1}{\partial\theta_2'} = \left(\frac{\partial^2\ln\lambda_t\left(Y_t, X_t, \theta_1, \theta_2\right)}{\partial\theta_2\partial\theta_1'}\right)' \\ &= \left(\left(Y_t - m_t\right)' \otimes \frac{\partial m_t'}{\partial\theta_1}\right)\frac{\operatorname{vec}\Omega_t^{-1}}{\partial\theta_2'} \end{aligned}$$

Given first order correct specification, for all $t = 1, 2, ...$, by definition, $E\left(Y_t|X_t\right) = m_t\left(X_t, \theta_1^o\right) = m_t^o$, and the law of iterated expectations applies such that

$$E\left[\frac{\partial^2\ln\lambda_t\left(Y_t, X_t, \theta_1^o, \theta_{2_n}^*\right)}{\partial\theta_1\partial\theta_2'}\right] = E\left[\left(\left(E\left(Y_t|X_t\right) - m_t^o\right)' \otimes \frac{\partial m_t^{o\prime}}{\partial\theta_1}\right)\frac{\operatorname{vec}\Omega_t^{*^{-1}}}{\partial\theta_2'}\right] = 0$$

$A_{n_{12}}^* = A_{n_{21}}^{*\prime} = E\left[\frac{\partial^2}{\partial\theta_1\partial\theta_2'}L_n(Y^n, X^n, \theta_n^*)\right] = 0$ directly follows.

Recalling that, for m and θ column vectors, M and Ω matrices,

$$\frac{\partial}{\partial\theta'}\left[M(\theta)\Omega m(\theta)\right] = M(\theta)\Omega\frac{\partial m(\theta)}{\partial\theta'} + \left(m(\theta)'\Omega' \otimes I\right)\frac{\partial\operatorname{vec}M(\theta)}{\partial\theta'} \tag{C-17}$$

we have that

$$\frac{\partial^2 \ln \lambda_t (Y_t, X_t, \theta_1, \theta_2)}{\partial\theta_1 \partial\theta_1'} = \frac{\partial s_t^1}{\partial\theta_1'}$$

$$= -\frac{\partial m_t'}{\partial\theta_1}\Omega_t^{-1}\frac{\partial m_t}{\partial\theta_1'} + \left((Y_t - m_t)'\,\Omega_t^{-1} \otimes I_{k_{\theta_1}}\right)\frac{\partial}{\partial\theta_1'}\left[\operatorname{vec}\left(\frac{\partial m_t'}{\partial\theta_1}\right)\right]$$

Again, for all $t = 1, 2, ...$, given first order correct specification and according to the law of iterated expectations, we have

$$E\left[\left((Y_t - m_t^o)'\,\Omega_t^{*^{-1}} \otimes I_{k_{\theta_1}}\right)\frac{\partial}{\partial\theta_1'}\left[\operatorname{vec}\left(\frac{\partial m_t^{o\prime}}{\partial\theta_1}\right)\right]\right]$$

$$= E\left[\left((E(Y_t|X_t) - m_t^o)'\,\Omega_t^{*^{-1}} \otimes I_{k_{\theta_1}}\right)\frac{\partial}{\partial\theta_1'}\left[\operatorname{vec}\left(\frac{\partial m_t^{o\prime}}{\partial\theta_1}\right)\right]\right] = 0$$

such that

$$E\left[\frac{\partial^2 \ln \lambda_t \left(Y_t, X_t, \theta_1^o, \theta_{2_n}^*\right)}{\partial\theta_1 \partial\theta_1'}\right] = -E\left[\frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{*^{-1}}\frac{\partial m_t^o}{\partial\theta_1'}\right]$$

$A_{n_{11}}^*$ directly follows.

Similarly, using (C-17), we have that

$$\frac{\partial^2 \ln \lambda_t (Y_t, X_t, \theta_1, \theta_2)}{\partial\theta_2 \partial\theta_2'} = \frac{\partial s_t^2}{\partial\theta_2'}$$

$$= \left((Y_t - m_t)' \otimes I_{k_{\theta_2}}\right)\frac{\partial}{\partial\theta_2'}\left[\operatorname{vec}\left(\frac{\partial(\operatorname{vec}\Omega_t)'}{\partial\theta_2}\frac{\partial C_t'}{\partial\operatorname{vec}\Omega_t}\right)\right]$$

$$+\left((\operatorname{vec}(Y_tY_t' - \Omega_t - m_tm_t'))' \otimes I_{k_{\theta_2}}\right)\frac{\partial}{\partial\theta_2'}\left[\operatorname{vec}\left(\frac{\partial(\operatorname{vec}\Omega_t)'}{\partial\theta_2}\frac{\partial(\operatorname{vec}D_t)'}{\partial\operatorname{vec}\Omega_t}\right)\right]$$

$$-\frac{\partial(\operatorname{vec}\Omega_t)'}{\partial\theta_2}\frac{\partial(\operatorname{vec}D_t)'}{\partial\operatorname{vec}\Omega_t}\frac{\partial\operatorname{vec}\Omega_t}{\partial\theta_2'}$$

Once again, for all $t = 1, 2, ...$, given first order correct specification and according to the law of iterated expectations, we have

$$E\left[\left((Y_t - m_t^o)' \otimes I_{k_{\theta_2}}\right)\frac{\partial}{\partial\theta_2'}\left[\operatorname{vec}\left(\frac{\partial(\operatorname{vec}\Omega_t^*)'}{\partial\theta_2}\frac{\partial C_t^{*\prime}}{\partial\operatorname{vec}\Omega_t}\right)\right]\right]$$

$$= E\left[\left((E(Y_t|X_t) - m_t^o)' \otimes I_{k_{\theta_2}}\right)\frac{\partial}{\partial\theta_2'}\left[\operatorname{vec}\left(\frac{\partial(\operatorname{vec}\Omega_t^*)'}{\partial\theta_2}\frac{\partial C_t^{*\prime}}{\partial\operatorname{vec}\Omega_t}\right)\right]\right] = 0$$

such that

$$E\left[\frac{\partial^2 \ln \lambda_t \left(Y_t, X_t, \theta_1^o, \theta_{2_n}^*\right)}{\partial\theta_2 \partial\theta_2'}\right] = -E\left[\frac{\partial(\operatorname{vec}\Omega_t^*)'}{\partial\theta_2}\frac{\partial(\operatorname{vec}D_t^*)'}{\partial\operatorname{vec}\Omega_t}\frac{\partial\operatorname{vec}\Omega_t^*}{\partial\theta_2'} - \Delta_t^*\right]$$

where

$$\Delta_t^* = \left((\operatorname{vec}(Y_t Y_t' - \Omega_t^* - m_t^o m_t^{o\prime}))' \otimes I_{k_{\theta_2}}\right) \frac{\partial}{\partial \theta_2'} \left[\operatorname{vec}\left(\frac{\partial (\operatorname{vec}\Omega_t^*)'}{\partial \theta_2} \frac{\partial (\operatorname{vec} D_t^*)'}{\partial \operatorname{vec}\Omega_t}\right)\right]$$

The first expression of A_{n22}^* directly follows. Further, since Ω_t^* is symmetric, $N_G \frac{\partial \operatorname{vec}\Omega_t^*}{\partial \theta_2'} = \frac{\partial \operatorname{vec}\Omega_t^*}{\partial \theta_2'}$ and $D_{uG}^+ \frac{\partial \operatorname{vec}\Omega_t^*}{\partial \theta_2'} = \frac{\partial \operatorname{vech}\Omega_t^*}{\partial \theta_2'}$. Using (C-14) evaluated at $\theta_n^* = \left(\theta_1^{o\prime}, \theta_{2_n}^{*\prime}\right)'$, we thus have that

$$\begin{aligned}
& \frac{\partial (\operatorname{vec}\Omega_t^*)'}{\partial \theta_2} \frac{\partial (\operatorname{vec} D_t^*)'}{\partial \operatorname{vec}\Omega_t} \frac{\partial \operatorname{vec}\Omega_t^*}{\partial \theta_2'} \\
= & \frac{\partial (\operatorname{vec}\Omega_t^*)'}{\partial \theta_2} \frac{\partial (\operatorname{vec} D_t^*)'}{\partial \operatorname{vec}\Omega_t} N_G \frac{\partial \operatorname{vec}\Omega_t^*}{\partial \theta_2'} \\
= & \frac{\partial (\operatorname{vec}\Omega_t^*)'}{\partial \theta_2} \left(D_{uG}^+\right)' \left(M_t^{4*} - M_t^{3*} \Omega_t^{*^{-1}} M_t^{3*\prime}\right)^{-1} D_{uG}^+ \frac{\partial \operatorname{vec}\Omega_t^*}{\partial \theta_2'} \\
= & \frac{\partial (\operatorname{vech}\Omega_t^*)'}{\partial \theta_2} \left(M_t^{4*} - M_t^{3*} \Omega_t^{*^{-1}} M_t^{3*\prime}\right)^{-1} \frac{\partial \operatorname{vech}\Omega_t^*}{\partial \theta_2'}
\end{aligned}$$

The second expression of A_{n22}^* directly follows.

Finally, the structure of C_n^* follows from the block-diagonality of A_n^* while the expression of $B_{n_{ij}}^*$ directly follows by suitably partitioning and breaking down

$$\begin{aligned}
B_n^* &= V\left[n^{1/2} \frac{\partial}{\partial \theta} L_n(Y^n, X^n, \theta_n^*)\right] \qquad , \theta = (\theta_1', \theta_2')' \\
&= V\left[n^{-1/2} \sum_{t=1}^n \frac{\partial \ln \lambda_t\left(Y_t, X_t, \theta_1^o, \theta_{2_n}^*\right)}{\partial \theta}\right] \\
&= E\left[\left(n^{-1/2} \sum_{t=1}^n \frac{\partial \ln \lambda_t\left(Y_t, X_t, \theta_1^o, \theta_{2_n}^*\right)}{\partial \theta}\right)\left(n^{-1/2} \sum_{t=1}^n \frac{\partial \ln \lambda_t\left(Y_t, X_t, \theta_1^o, \theta_{2_n}^*\right)}{\partial \theta}\right)'\right]
\end{aligned}$$

where the last equality holds because, under the assumed regularity conditions, $\theta_n^* = \operatorname{Argmax}_{\theta \in \Theta} E\left(L_n(Y^n, X^n, \theta)\right)$ implies that

$$E\left[\frac{\partial}{\partial \theta} L_n(Y^n, X^n, \theta_n^*)\right] = E\left[\sum_{t=1}^n \frac{\partial \ln \lambda_t\left(Y_t, X_t, \theta_1^o, \theta_{2_n}^*\right)}{\partial \theta}\right] = 0$$

■

Proof of Proposition 11 Given all the assumptions of Proposition 9, since, for all $t = 1, 2, ...,$ f_t belongs to the restricted quadratic exponential family, and thus to the quadratic exponential family, and $\mathcal{S}$ is as described in Assumption 1 and is second order correctly specified, according to the proof of Proposition 1, for all $n = 1, 2, ...,$ we have $\theta_n^* = \theta^o = (\theta_1^{o\prime}, \theta_2^{o\prime})'$. Further, for all $t = 1, 2, ...,$ by definition, $E(Y_t|X_t) = m_t^o$ and $E(Y_t Y_t'|X_t) = V(Y_t|X_t) + E(Y_t|X_t) E(Y_t|X_t)' = (\Omega_t^o + m_t^o m_t^{o\prime})$, where $m_t^o = m_t(X_t, \theta_1^o)$ and $\Omega_t^o = \Omega_t(X_t, \theta_2^o)$, and the law of iterated expectations applies such that

$$
\begin{aligned}
&E\left(\Delta_t^o\right)\\
&= E\left[\left(\left(\operatorname{vec}\left(Y_t Y_t' - \Omega_t^o - m_t^o m_t^{o\prime}\right)\right)' \otimes I_{k_{\theta_2}}\right) \frac{\partial}{\partial \theta_2'}\left[\operatorname{vec}\left(\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial \theta_2}\frac{\partial\left(\operatorname{vec}D_t^o\right)'}{\partial \operatorname{vec}\Omega_t}\right)\right]\right]\\
&= E\left[\left(\left(\operatorname{vec}\left(E\left(Y_t Y_t'|X_t\right) - \Omega_t^o - m_t^o m_t^{o\prime}\right)\right)' \otimes I_{k_{\theta_2}}\right) \frac{\partial}{\partial \theta_2'}\left[\operatorname{vec}\left(\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial \theta_2}\frac{\partial\left(\operatorname{vec}D_t^o\right)'}{\partial \operatorname{vec}\Omega_t}\right)\right]\right]\\
&= 0
\end{aligned}
$$

$A_{n_{22}}^o$ directly follows from Proposition 10 by evaluating the remaining terms of $A_{n_{22}}^*$ at $\theta_n^* = (\theta_1^{o\prime}, \theta_2^{o\prime})'$. ■

Proof of Proposition 12 Given all the assumptions of Proposition 9, since, for all $t = 1, 2, ...,$ f_t belongs to the restricted quadratic exponential family and $\mathcal{S}$ is as described in Assumption 1 and is first order correctly specified, from Proposition 10, for all $n = 1, 2, ...,$ we have $\theta_n^* = \left(\theta_1^{o\prime}, \theta_{2_n}^{*\prime}\right)'$ and

$$
B_{n_{ij}}^* = \frac{1}{n}\sum_{t=1}^{n} E\left[s_t^{i*} s_t^{j*\prime}\right] + \frac{1}{n}\sum_{\tau=1}^{n-1}\sum_{t=\tau+1}^{n}\left(E\left[s_t^{i*} s_{t-\tau}^{j*\prime}\right] + E\left[s_{t-\tau}^{i*} s_t^{j*\prime}\right]\right), \quad i = 1, 2\,;\; j = 1, 2
$$

where

$$
\begin{aligned}
s_t^{1*} &= \frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{*^{-1}}\left(Y_t - m_t^o\right)\\
s_t^{2*} &= \frac{\partial\left(\operatorname{vec}\Omega_t^*\right)'}{\partial \theta_2}\left(\frac{\partial C_t^{*\prime}}{\partial \operatorname{vec}\Omega_t}\left(Y_t - m_t^o\right) + \frac{\partial\left(\operatorname{vec}D_t^*\right)'}{\partial \operatorname{vec}\Omega_t}\operatorname{vec}\left(Y_t Y_t' - \Omega_t^* - m_t^o m_t^{o\prime}\right)\right)
\end{aligned}
$$

Given first order correct specification and first order correct dynamic specification, for all $t = 1, 2, ...,$ by definition, $E\left(Y_t|X_t\right) = m_t\left(X_t, \theta_1^o\right) = m_t^o = E(Y_t|X_t, \Psi_{t-1})$, such that we have

$$
\begin{aligned}
E\left(s_t^{1*}|X_t\right) &= \frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{*^{-1}}\left(E\left(Y_t|X_t\right) - m_t^o\right)\\
&= E\left(s_t^{1*}|X_t, \Psi_{t-1}\right) = 0
\end{aligned}
$$

where $\Psi_{t-1} \equiv (Y_{t-1}, X_{t-1}, ..., Y_1, X_1)$. By the law of iterated expectations,

$$
E\left(s_t^{1*}|X_t, \Psi_{t-1}\right) = E\left(s_t^{1*}|\Psi_{t-1}\right) = E\left(s_t^{1*}\right) = 0 \qquad \text{(C-18)}
$$

Since s_t^{1*} is by definition measurable with respect to Ψ_t, $\{s_t^{1*}\}$ is a martingale difference sequence with respect to $\{\Psi_t\}$, so that s_t^{1*} is uncorrelated with its past values. We then have

$$
\begin{aligned}
B_{n_{11}}^* &= \overline{B}_{n_{11}}^* = \frac{1}{n}\sum_{t=1}^{n} E\left[s_t^{1*} s_t^{1*\prime}\right]\\
&= \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{*^{-1}}\left(Y_t - m_t^o\right)\left(Y_t - m_t^o\right)'\Omega_t^{*^{-1}}\frac{\partial m_t^o}{\partial \theta_1'}\right]
\end{aligned}
$$

or, using again the law of iterated expectations,

$$\begin{aligned}\overline{B}^{*}_{n_{11}} &= \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{*^{-1}} E\left[\left(Y_t - m_t^o\right)\left(Y_t - m_t^o\right)'\middle| X_t\right]\Omega_t^{*^{-1}}\frac{\partial m_t^o}{\partial \theta_1'}\right] \\ &= \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{*^{-1}}\Sigma_t^o\Omega_t^{*^{-1}}\frac{\partial m_t^o}{\partial \theta_1'}\right]\end{aligned}$$

$\overline{C}^{*}_{n_{11}} = A^{*^{-1}}_{n_{11}}\overline{B}^{*}_{n_{11}}A^{*^{-1}}_{n_{11}}$ directly follows from Proposition 10.

If, in addition, the semi-parametric model $\mathcal{S}$ is also second order correctly specified, from Proposition 11, for all $n = 1, 2, ...$, we have $\theta_n^* = (\theta_1^{o\prime}, \theta_2^{o\prime})'$ and, for all $t = 1, 2, ...$, by definition, $\Sigma_t^o = V(Y_t|X_t) = \Omega_t^o = \Omega_t(X_t, \theta_2^o)$, such that

$$\begin{aligned}B^{o}_{n_{11}} &= \overline{B}^{o}_{n_{11}} = \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{o^{-1}}\Sigma_t^o\Omega_t^{o^{-1}}\frac{\partial m_t^o}{\partial \theta_1'}\right] \\ &= \overline{\overline{B}}^{o}_{n_{11}} = \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{o^{-1}}\frac{\partial m_t^o}{\partial \theta_1'}\right] = -A^{o}_{n_{11}}\end{aligned}$$

where the last equality, as well as $\overline{C}^{o}_{n_{11}} = \overline{\overline{C}}^{o}_{n_{11}} = -A^{o^{-1}}_{n_{11}} = \overline{\overline{B}}^{o^{-1}}_{n_{11}} = \overline{B}^{o^{-1}}_{n_{11}}$ directly follows from Proposition 10 by evaluating all relevant terms at $\theta_n^* = (\theta_1^{o\prime}, \theta_2^{o\prime})'$.

Given the assumed regularity assumptions, according to Theorem 2.6 of Bates and White (1993), for proving that $\overline{C}^{*}_{n_{11}} - \overline{\overline{C}}^{o}_{n_{11}} \gg 0$, i.e., that $\overline{\overline{C}}^{o}_{n_{11}}$ is the minimum asymptotic covariance matrix of a RPML2 mean parameters estimator of a semi-parametric model $\mathcal{S}$ first order correctly specified and first order dynamically complete, it suffices to show that

$$-A^{*}_{n_{11}} = \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{*^{-1}}\frac{\partial m_t^o}{\partial \theta_1'}\right] = Cov\left[n^{-1/2}\sum_{t=1}^{n} s_t^{1*}, n^{-1/2}\sum_{t=1}^{n} s_t^{1o}\right]$$

where s_t^{1*} and $A^{*}_{n_{11}}$ are respectively the individual score and the expected hessian associated with an arbitrary RPML2 mean parameters estimator of a semi-parametric model $\mathcal{S}$ first order correctly specified and first order dynamically complete, and s_t^{1o} is the individual score associated with an arbitrary RPML2 mean parameters estimator of the same semi-parametric model $\mathcal{S}$ for the conditional mean and such that $\mathcal{S}$ is in addition also second order correctly specified. Now, given the martingale difference property of s_t^{1*}, and thus also of s_t^{1o}, and using the law of iterated expectations, we get

$$Cov\left[n^{-1/2}\sum_{t=1}^{n} s_t^{1*}, n^{-1/2}\sum_{t=1}^{n} s_t^{1o}\right] = \frac{1}{n}\sum_{t=1}^{n} E\left[s_t^{1*}s_t^{1o\prime}\right]$$

$$= \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{*^{-1}}\left(Y_t - m_t^o\right)\left(Y_t - m_t^o\right)'\Omega_t^{o^{-1}}\frac{\partial m_t^o}{\partial \theta_1'}\right]$$

$$= \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{*^{-1}} E\left[(Y_t - m_t^o)(Y_t - m_t^o)'\middle| X_t\right]\Omega_t^{o^{-1}}\frac{\partial m_t^o}{\partial\theta_1'}\right]$$

$$= \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{*^{-1}}\Omega_t^o\Omega_t^{o^{-1}}\frac{\partial m_t^o}{\partial\theta_1'}\right] = -A_{n_{11}}^*$$

and the proof is complete.

If, in addition, the semi-parametric model $\mathcal{S}$ is also second order dynamically complete, for all $t = 1, 2, ...$, by definition, we have that $V(Y_t|X_t) = \Omega_t^o = V(Y_t|X_t, \Psi_{t-1})$, or, since $V(Y_t|.) = E(Y_tY_t'|.) - E(Y_t|.)\,E(Y_t|.)'$ and $E(Y_t|X_t) = m_t^o = E(Y_t|X_t, \Psi_{t-1})$, $E(Y_tY_t'|X_t) = \Omega_t^o + m_t^o m_t^{o\prime} = E(Y_tY_t'|X_t, \Psi_{t-1})$, such that

$$E\left(s_t^{2o}|X_t\right)$$

$$= \frac{\partial(\operatorname{vec}\Omega_t^o)'}{\partial\theta_2}\left(\frac{\partial C_t^{o\prime}}{\partial\operatorname{vec}\Omega_t}\left(E\left(Y_t|X_t\right) - m_t^o\right) + \frac{\partial(\operatorname{vec}D_t^o)'}{\partial\operatorname{vec}\Omega_t}\operatorname{vec}\left(E(Y_tY_t'|X_t) - \Omega_t^o - m_t^o m_t^{o\prime}\right)\right)$$

$$= E\left(s_t^{2o}|X_t, \Psi_{t-1}\right) = 0$$

By the law of iterated expectations,

$$E\left(s_t^{2o}|X_t, \Psi_{t-1}\right) = E\left(s_t^{2o}|\Psi_{t-1}\right) = E\left(s_t^{2o}\right) = 0 \tag{C-19}$$

Letting s_t^o stand for $(s_t^{1o\prime}, s_t^{2o\prime})'$ and collecting (C-18) and (C-19), we have

$$E\left(s_t^o|\Psi_{t-1}\right) = E\left(s_t^o\right) = 0$$

As above, since s_t^o is by definition measurable with respect to Ψ_t, $\{s_t^o\}$ is a martingale difference sequence with respect to $\{\Psi_t\}$, so that s_t^o is uncorrelated with its past values. We then have

$$B_n^o = \overline{B}_n^o = \frac{1}{n}\sum_{t=1}^{n} E\left[s_t^o s_t^{o\prime}\right] \tag{C-20}$$

and thus in particular

$$B_{n_{12}}^o = \overline{B}_{n_{12}}^o = \frac{1}{n}\sum_{t=1}^{n} E\left[s_t^{1o} s_t^{2o\prime}\right] \quad \text{and} \quad B_{n_{22}}^o = \overline{B}_{n_{22}}^o = \frac{1}{n}\sum_{t=1}^{n} E\left[s_t^{2o} s_t^{2o\prime}\right]$$

$\overline{C}_{n_{12}}^o = A_{n_{11}}^{o^{-1}}\overline{B}_{n_{12}}^o A_{n_{22}}^{o^{-1}}$ and $\overline{C}_{n_{12}}^o = A_{n_{22}}^{o^{-1}}\overline{B}_{n_{22}}^o A_{n_{22}}^{o^{-1}}$ directly follow from Proposition 10 and 11 by evaluating all relevant terms at $\theta_n^* = (\theta_1^{o\prime}, \theta_2^{o\prime})'$.

If, in addition, the implicit parametric model $\mathcal{P}$ arising from the semi-parametric model $\mathcal{S}$ and the sequence $\{f_t\}$ is also third order correctly specified, for all $t = 1, 2, ...$, by definition, we have that $Cov\left[(\operatorname{vec}(Y_tY_t'), Y_t)| X_t\right] = Cov_{\lambda_t^o}\left[(\operatorname{vec}(Y_tY_t'), Y_t)| X_t\right]$. Then, since

$$B_{n_{12}}^o = \overline{B}_{n_{12}}^o = \frac{1}{n}\sum_{t=1}^{n} E\left[s_t^{1o} s_t^{2o\prime}\right]$$

$$= \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{o^{-1}}(Y_t - m_t^o)(Y_t - m_t^o)'\frac{\partial C_t^o}{\partial(\operatorname{vec}\Omega_t)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\right.$$

$$+\frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{o^{-1}}\left(Y_t-m_t^o\right)\left(\mathrm{vec}\left(Y_tY_t^\prime-\Omega_t^o-m_t^om_t^{o\prime}\right)\right)^\prime\frac{\partial\,\mathrm{vec}\,D_t^o}{\partial\left(\mathrm{vec}\,\Omega_t\right)^\prime}\frac{\partial\,\mathrm{vec}\,\Omega_t^o}{\partial\theta_2^\prime}\Bigg]$$

using the law of iterated expectations, we get

$$\begin{aligned}\overline{B}_{n_{12}}^o &= \frac{1}{n}\sum_{t=1}^n E\left[\frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{o^{-1}}\Omega_t^o\frac{\partial C_t^o}{\partial\left(\mathrm{vec}\,\Omega_t\right)^\prime}\frac{\partial\,\mathrm{vec}\,\Omega_t^o}{\partial\theta_2^\prime}\right.\\ &\quad\left.+\frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{o^{-1}}Cov_{\lambda_t^o}\left[\left(Y_t,\mathrm{vec}\left(Y_tY_t^\prime\right)\right)\mid X_t\right]\frac{\partial\,\mathrm{vec}\,D_t^o}{\partial\left(\mathrm{vec}\,\Omega_t\right)^\prime}\frac{\partial\,\mathrm{vec}\,\Omega_t^o}{\partial\theta_2^\prime}\right]\end{aligned}$$

Since from Property 8(c), we have

$$\frac{\partial C_t^o}{\partial\left(\mathrm{vec}\,\Omega_t\right)^\prime}=-\Omega_t^{o-1}Cov_{\lambda_t^o}\left[\left(Y_t,\mathrm{vec}\left(Y_tY_t^\prime\right)\right)\mid X_t\right]\frac{\partial\,\mathrm{vec}\,D_t^o}{\partial\left(\mathrm{vec}\,\Omega_t\right)^\prime}\qquad\text{(C-21)}$$

it is easily seen that

$$\overline{B}_{n_{12}}^o=\overline{\overline{B}}_{n_{12}}^o=0=A_{n_{12}}^o$$

where the last equality, as well as $\overline{\overline{C}}_{n_{12}}^o=0$ directly follow from Proposition 10 by evaluating relevant terms at $\theta_n^*=(\theta_1^{o\prime},\theta_2^{o\prime})^\prime$.

Finally, if, in addition, the implicit parametric model $\mathcal{P}$ arising from the semi-parametric model $\mathcal{S}$ and the sequence $\{f_t\}$ is also fourth order correctly specified, for all $t=1,2,...,$ by definition, we also have that $V\left[\left(\mathrm{vec}\left(Y_tY_t^\prime\right)\right)\mid X_t\right]=V_{\lambda_t^o}\left[\left(\mathrm{vec}\left(Y_tY_t^\prime\right)\right)\mid X_t\right]$. Then, since

$$\begin{aligned}B_{n_{22}}^o &= \overline{B}_{n_{22}}^o=\frac{1}{n}\sum_{t=1}^n E\left[s_t^{2o}s_t^{2o\prime}\right]\\ &= \frac{1}{n}\sum_{t=1}^n E\left[\frac{\partial\left(\mathrm{vec}\,\Omega_t^o\right)^\prime}{\partial\theta_2}\frac{\partial C_t^{o\prime}}{\partial\,\mathrm{vec}\,\Omega_t}\left(Y_t-m_t^o\right)\left(Y_t-m_t^o\right)^\prime\frac{\partial C_t^o}{\partial\left(\mathrm{vec}\,\Omega_t\right)^\prime}\frac{\partial\,\mathrm{vec}\,\Omega_t^o}{\partial\theta_2^\prime}\right.\\ &\quad+\frac{\partial\left(\mathrm{vec}\,\Omega_t^o\right)^\prime}{\partial\theta_2}\frac{\partial C_t^{o\prime}}{\partial\,\mathrm{vec}\,\Omega_t}\left(Y_t-m_t^o\right)\left(\mathrm{vec}\left(Y_tY_t^\prime-\Omega_t^o-m_t^om_t^{o\prime}\right)\right)^\prime\frac{\partial\,\mathrm{vec}\,D_t^o}{\partial\left(\mathrm{vec}\,\Omega_t\right)^\prime}\frac{\partial\,\mathrm{vec}\,\Omega_t^o}{\partial\theta_2^\prime}\\ &\quad+\frac{\partial\left(\mathrm{vec}\,\Omega_t^o\right)^\prime}{\partial\theta_2}\frac{\partial\left(\mathrm{vec}\,D_t^o\right)^\prime}{\partial\,\mathrm{vec}\,\Omega_t}\mathrm{vec}\left(Y_tY_t^\prime-\Omega_t^o-m_t^om_t^{o\prime}\right)\left(Y_t-m_t^o\right)^\prime\frac{\partial C_t^o}{\partial\left(\mathrm{vec}\,\Omega_t\right)^\prime}\frac{\partial\,\mathrm{vec}\,\Omega_t^o}{\partial\theta_2^\prime}\\ &\quad+\frac{\partial\left(\mathrm{vec}\,\Omega_t^o\right)^\prime}{\partial\theta_2}\frac{\partial\left(\mathrm{vec}\,D_t^o\right)^\prime}{\partial\,\mathrm{vec}\,\Omega_t}\mathrm{vec}\left(Y_tY_t^\prime-\Omega_t^o-m_t^om_t^{o\prime}\right)\\ &\quad\left.\left(\mathrm{vec}\left(Y_tY_t^\prime-\Omega_t^o-m_t^om_t^{o\prime}\right)\right)^\prime\frac{\partial\,\mathrm{vec}\,D_t^o}{\partial\left(\mathrm{vec}\,\Omega_t\right)^\prime}\frac{\partial\,\mathrm{vec}\,\Omega_t^o}{\partial\theta_2^\prime}\right]\end{aligned}$$

using the law of iterated expectations, we get

$$\begin{aligned}\overline{B}_{n_{22}}^o &= \frac{1}{n}\sum_{t=1}^n E\left[\frac{\partial\left(\mathrm{vec}\,\Omega_t^o\right)^\prime}{\partial\theta_2}\frac{\partial C_t^{o\prime}}{\partial\,\mathrm{vec}\,\Omega_t}\Omega_t^o\frac{\partial C_t^o}{\partial\left(\mathrm{vec}\,\Omega_t\right)^\prime}\frac{\partial\,\mathrm{vec}\,\Omega_t^o}{\partial\theta_2^\prime}\right.\\ &\quad+\frac{\partial\left(\mathrm{vec}\,\Omega_t^o\right)^\prime}{\partial\theta_2}\frac{\partial C_t^{o\prime}}{\partial\,\mathrm{vec}\,\Omega_t}Cov_{\lambda_t^o}\left[\left(Y_t,\mathrm{vec}\left(Y_tY_t^\prime\right)\right)\mid X_t\right]\frac{\partial\,\mathrm{vec}\,D_t^o}{\partial\left(\mathrm{vec}\,\Omega_t\right)^\prime}\frac{\partial\,\mathrm{vec}\,\Omega_t^o}{\partial\theta_2^\prime}\end{aligned}$$

$$+\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}D_t^o\right)'}{\partial\operatorname{vec}\Omega_t}Cov_{\lambda_t^o}\left[\left(\operatorname{vec}\left(Y_tY_t'\right),Y_t\right)\right|X_t\right]\frac{\partial C_t^o}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}$$
$$\left.+\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}D_t^o\right)'}{\partial\operatorname{vec}\Omega_t}V_{\lambda_t^o}\left[\left(\operatorname{vec}\left(Y_tY_t'\right)\right)\right|X_t\right]\frac{\partial\operatorname{vec}D_t^o}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\right]$$

Since from Property 8(c), we also have

$$Cov_{\lambda_t^o}\left[\left(\operatorname{vec}\left(Y_tY_t'\right),Y_t\right)\right|X_t\right]\frac{\partial C_t^o}{\partial\left(\operatorname{vec}\Omega_t\right)'}=I_{G^2}-V_{\lambda_t^o}\left[\left(\operatorname{vec}\left(Y_tY_t'\right)\right)\right|X_t\right]\frac{\partial\operatorname{vec}D_t^o}{\partial\left(\operatorname{vec}\Omega_t\right)'}\quad\text{(C-22)}$$

using (C-21) and (C-22), it is easily seen that $\overline{B}_{n22}^o$ may be simplified such that we finally get

$$\begin{aligned}\overline{B}_{n22}^o &= \overline{\overline{B}}_{n22}^o=\frac{1}{n}\sum_{t=1}^n E\left[\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}D_t^o\right)'}{\partial\operatorname{vec}\Omega_t}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\right]=-A_{n22}^o\\ &= \frac{1}{n}\sum_{t=1}^n E\left[\frac{\partial\left(\operatorname{vech}\Omega_t^o\right)'}{\partial\theta_2}\left(\overline{\overline{M}}_t^{4o}-\overline{\overline{M}}_t^{3o}\Omega_t^{o-1}\overline{\overline{M}}_t^{3o\prime}\right)^{-1}\frac{\partial\operatorname{vech}\Omega_t^o}{\partial\theta_2'}\right]\end{aligned}$$

where the equality $\overline{\overline{B}}_{n22}^o=-A_{n22}^o$, as well as $\overline{\overline{C}}_{n22}^o=-A_{n22}^{o-1}=\overline{\overline{B}}_{n22}^{o-1}$, directly follow from Proposition 10 and 11 by evaluating all relevant terms at $\theta_n^*=(\theta_1^{o\prime},\theta_2^{o\prime})'$ and the fact that $M_t^{3o}=\overline{\overline{M}}_t^{3o}$ and $M_t^{4o}=\overline{\overline{M}}_t^{4o}$ are now respectively equal to the actual third and fourth order conditional moments of Y_t given X_t, $Cov\left[\left(\operatorname{vech}\left(Y_tY_t'\right),Y_t\right)\right|X_t\right]$ and $V\left[\operatorname{vech}\left(Y_tY_t'\right)\right|X_t\right]$ obviously follow from (third and) fourth order correct specification.

As above, given the assumed regularity conditions, according to Theorem 2.6 of Bates and White (1993), for proving that $\overline{C}_n^o-\overline{\overline{C}}_n^o\gg 0$, i.e., that $\overline{\overline{C}}_n^o$ is the minimum asymptotic covariance matrix of a RPML2 estimator of a semi-parametric model $\mathcal{S}$ second order correctly specified and second order dynamically complete, it suffices to show that

$$-A_n^o=Cov\left[n^{-1/2}\sum_{t=1}^n s_t^o,n^{-1/2}\sum_{t=1}^n s_t^{o_{f_t^o}}\right]\quad\text{(C-23)}$$

where s_t^o and A_n^o are respectively the individual score and the expected hessian associated with an arbitrary RPML2 estimator of a semi-parametric model $\mathcal{S}$ second order correctly specified and second order dynamically complete, and $s_t^{o_{f_t^o}}$ is the individual score associated with an arbitrary RPML2 estimator of the same semi-parametric model $\mathcal{S}$ for the conditional mean and the conditional variance and such that the implicit parametric model $\mathcal{P}$ arising from $\mathcal{S}$ and the sequence $\{f_t^o\}$ is in addition also fourth order correctly specified. Now, given the martingale difference property of s_t^o, and thus also of $s_t^{o_{f_t^o}}$, we have

$$Cov\left[n^{-1/2}\sum_{t=1}^n s_t^o,n^{-1/2}\sum_{t=1}^n s_t^{o_{f_t^o}}\right]=\frac{1}{n}\sum_{t=1}^n E\left[s_t^o s_t^{o_{f_t^o}\prime}\right]$$

We proceed block-by-block. Since s_t^{1o}, and thus also $s_t^{1o_{f_t^o}}$, does not depend on the sequence $\{f_t\}$, we have

$$\frac{1}{n}\sum_{t=1}^{n}E\left[s_t^{1o}s_t^{1o_{f_t^o}\prime}\right]=\frac{1}{n}\sum_{t=1}^{n}E\left[s_t^{1o}s_t^{1o\prime}\right]=\overline{B}_{n_{11}}^{o}$$

and thus, since $\overline{B}_{n_{11}}^{o}=-A_{n_{11}}^{o}$, the upper diagonal block of (C-23) holds. Further, we have

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}E\left[s_t^{1o}s_t^{2o_{f_t^o}\prime}\right]\\ =\ &\frac{1}{n}\sum_{t=1}^{n}E\left[\frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{o^{-1}}\left(Y_t-m_t^o\right)\left(Y_t-m_t^o\right)'\frac{\partial C_t^{o_{f_t^o}}}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\right.\\ &\left.+\frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{o^{-1}}\left(Y_t-m_t^o\right)\left(\operatorname{vec}\left(Y_tY_t'-\Omega_t^o-m_t^om_t^{o\prime}\right)\right)'\frac{\partial\operatorname{vec}D_t^{o_{f_t^o}}}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\right]\end{aligned}$$

Using the law of iterated expectations, since $Cov\left[\left(\operatorname{vec}\left(Y_tY_t'\right),Y_t\right)\mid X_t\right]=Cov_{\lambda_t^{o_{f_t^o}}}\left[\left(\operatorname{vec}\left(Y_tY_t'\right),Y_t\right)\mid X_t\right]$, we get

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}E\left[s_t^{1o}s_t^{2o_{f_t^o}\prime}\right]\\ =\ &\frac{1}{n}\sum_{t=1}^{n}E\left[\frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{o^{-1}}\Omega_t^{o}\frac{\partial C_t^{o_{f_t^o}}}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\right.\\ &\left.+\frac{\partial m_t^{o\prime}}{\partial\theta_1}\Omega_t^{o^{-1}}Cov_{\lambda_t^{o_{f_t^o}}}\left[\left(Y_t,\operatorname{vec}\left(Y_tY_t'\right)\right)\mid X_t\right]\frac{\partial\operatorname{vec}D_t^{o_{f_t^o}}}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\right]\end{aligned}$$

Given (C-21), it is easily seen that

$$\frac{1}{n}\sum_{t=1}^{n}E\left[s_t^{1o}s_t^{2o_{f_t^o}\prime}\right]=0=A_{n_{12}}^{o}=\left(A_{n_{21}}^{o}\right)'=\left(\frac{1}{n}\sum_{t=1}^{n}E\left[s_t^{2o}s_t^{1o_{f_t^o}\prime}\right]\right)'$$

where the last equalities follow from the symmetry of the problem at hand. Thus (C-23) also holds for both off-diagonal blocks. Finally, we have

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}E\left[s_t^{2o}s_t^{2o_{f_t^o}\prime}\right]\\ =\ &\frac{1}{n}\sum_{t=1}^{n}E\left[\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial\theta_2}\frac{\partial C_t^{o\prime}}{\partial\operatorname{vec}\Omega_t}\left(Y_t-m_t^o\right)\left(Y_t-m_t^o\right)'\frac{\partial C_t^{o_{f_t^o}}}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\right.\\ &+\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial\theta_2}\frac{\partial C_t^{o\prime}}{\partial\operatorname{vec}\Omega_t}\left(Y_t-m_t^o\right)\left(\operatorname{vec}\left(Y_tY_t'-\Omega_t^o-m_t^om_t^{o\prime}\right)\right)'\frac{\partial\operatorname{vec}D_t^{o_{f_t^o}}}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\\ &+\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}D_t^o\right)'}{\partial\operatorname{vec}\Omega_t}\operatorname{vec}\left(Y_tY_t'-\Omega_t^o-m_t^om_t^{o\prime}\right)\left(Y_t-m_t^o\right)'\frac{\partial C_t^{o_{f_t^o}}}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\end{aligned}$$

$$+\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}D_t^o\right)'}{\partial\operatorname{vec}\Omega_t}\operatorname{vec}\left(Y_tY_t'-\Omega_t^o-m_t^om_t^{o\prime}\right)$$

$$\left.\left(\operatorname{vec}\left(Y_tY_t'-\Omega_t^o-m_t^om_t^{o\prime}\right)\right)'\frac{\partial\operatorname{vec}D_t^{o_{f_t^o}}}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\right]$$

Using the law of iterated expectations, since we also have $V\left[\left(\operatorname{vec}\left(Y_tY_t'\right)\right)|\,X_t\right]=V_{\lambda_t^{o_{f_t^o}}}\left[\left(\operatorname{vec}\left(Y_tY_t'\right)\right)|\,X_t\right]$, we get

$$\frac{1}{n}\sum_{t=1}^{n}E\left[s_t^{2o}s_t^{2o_{f_t^o}\prime}\right]=\frac{1}{n}\sum_{t=1}^{n}E\left[\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial\theta_2}\frac{\partial C_t^{o\prime}}{\partial\operatorname{vec}\Omega_t}\Omega_t^o\frac{\partial C_t^{o_{f_t^o}}}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\right.$$

$$+\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial\theta_2}\frac{\partial C_t^{o\prime}}{\partial\operatorname{vec}\Omega_t}Cov_{\lambda_t^{o_{f_t^o}}}\left[\left(Y_t,\operatorname{vec}\left(Y_tY_t'\right)\right)|\,X_t\right]\frac{\partial\operatorname{vec}D_t^{o_{f_t^o}}}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}$$

$$+\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}D_t^o\right)'}{\partial\operatorname{vec}\Omega_t}Cov_{\lambda_t^{o_{f_t^o}}}\left[\left(\operatorname{vec}\left(Y_tY_t'\right),Y_t\right)|\,X_t\right]\frac{\partial C_t^{o_{f_t^o}}}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}$$

$$\left.+\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}D_t^o\right)'}{\partial\operatorname{vec}\Omega_t}V_{\lambda_t^{o_{f_t^o}}}\left[\left(\operatorname{vec}\left(Y_tY_t'\right)\right)|\,X_t\right]\frac{\partial\operatorname{vec}D_t^{o_{f_t^o}}}{\partial\left(\operatorname{vec}\Omega_t\right)'}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\right]$$

Given (C-21) and (C-22), it is easily seen that this expression may be simplified such that we finally get

$$\frac{1}{n}\sum_{t=1}^{n}E\left[s_t^{2o}s_t^{2o_{f_t^o}\prime}\right]=\frac{1}{n}\sum_{t=1}^{n}E\left[\frac{\partial\left(\operatorname{vec}\Omega_t^o\right)'}{\partial\theta_2}\frac{\partial\left(\operatorname{vec}D_t^o\right)'}{\partial\operatorname{vec}\Omega_t}\frac{\partial\operatorname{vec}\Omega_t^o}{\partial\theta_2'}\right]=-A_{n22}^o$$

Thus (C-23) also holds for the lower diagonal block and the proof is complete.

■

Proof of Proposition 13 Given all the assumptions of Proposition 9, since, for all $t=1,2,...$, f_t belongs to the restricted quadratic exponential family and $\mathcal{S}$ is as described in Assumption 1 and is first order correctly specified, from Proposition 9 and 10, for all $n=1,2,...$, we have $\theta_n^*=\left(\theta_1^{o\prime},\theta_{2_n}^{*\prime}\right)'$ and

$$B_n^*=V\left(n^{1/2}\frac{\partial}{\partial\theta}L_n(Y^n,X^n,\theta_n^*)\right)=V\left(n^{-1/2}\sum_{t=1}^{n}s_t^*\right)$$

where $s_t^*=\left(s_t^{1*\prime},s_t^{2*\prime}\right)'$, or, because the observations are independent across t,

$$B_n^*=\frac{1}{n}\sum_{t=1}^{n}V\left(s_t^*\right)=\frac{1}{n}\sum_{t=1}^{n}E\left[s_t^*s_t^{*\prime}\right]-\frac{1}{n}\sum_{t=1}^{n}E\left(s_t^*\right)E\left(s_t^*\right)'$$

and thus

$$B_{n_{ij}}^*=\frac{1}{n}\sum_{t=1}^{n}E\left[s_t^{i*}s_t^{j*\prime}\right]-\frac{1}{n}\sum_{t=1}^{n}E\left(s_t^{i*}\right)E\left(s_t^{j*}\right)',\quad i=1,2\,;\ j=1,2$$

According to the proof of Proposition 12, given first order correct specification (and tautologically correct dynamic specification), $E\left(s_t^{1*}\right) = 0$ while, unless second order correct specification, $E\left(s_t^{2*}\right) \neq 0$. Consequently, we have

$$\begin{aligned}
B_{n_{12}}^* &= \overline{B}_{n_{12}}^* = \frac{1}{n}\sum_{t=1}^n E\left[s_t^{1*} s_t^{2*\prime}\right] \\
B_{n_{22}}^* &= \ddot{B}_{n_{22}}^* = \frac{1}{n}\sum_{t=1}^n E\left[s_t^{2*} s_t^{2*\prime}\right] - \frac{1}{n}\sum_{t=1}^n E\left(s_t^{2*}\right) E\left(s_t^{2*}\right)'
\end{aligned}$$

$\overline{C}_{n_{12}}^* = A_{n_{11}}^{*-1}\overline{B}_{n_{12}}^* A_{n_{22}}^{*-1}$, and $C_{n_{22}}^* = A_{n_{22}}^{*-1}\ddot{B}_{n_{22}}^* A_{n_{22}}^{*-1} \ll \overline{Q}_{n_{22}}^* = A_{n_{22}}^{*-1}\overline{B}_{n_{22}}^* A_{n_{22}}^{*-1}$ directly follow from Proposition 10 and the fact that $\frac{1}{n}\sum_{t=1}^n E\left(s_t^{2*}\right) E\left(s_t^{2*}\right)'$ is positive semi-definite, and thus that $\overline{B}_{n_{22}}^* \gg \ddot{B}_{n_{22}}^*$. ■

Appendix D

In this appendix, we first show that the asymptotic covariance matrix $\overline{C}_n^o$ of an RPML2 estimator of a semi-parametric model $\mathcal{S}$ second order correctly specified and second order dynamically complete is always greater or equal than the asymptotic covariance matrix $\overline{\overline{C}}_n^{oGMM}$ of the optimal GMM estimator associated with the first and second order conditional moments restrictions $E[(Y_t - m_t(X_t, \theta_1)) \,|X_t] = 0$ and $E[\text{vech}\left(Y_t Y_t' - \Omega_t\left(X_t, \theta_2\right) - m_t(X_t, \theta_1) m_t(X_t, \theta_1)'\right) |X_t] = 0$, i.e., that $\overline{C}_n^o - \overline{\overline{C}}_n^{oGMM} \gg 0$. Next we show that, if there exists a sequence of pseudo-densities $\{f_t^o\}$ such that the implicit parametric model $\mathcal{P}$ arising from the semi-parametric model $\mathcal{S}$ and the sequence $\{f_t^o\}$ is in addition also fourth order correctly specified, then the minimum asymptotic covariance matrix $\overline{\overline{C}}_n^o$ associated with an RMPL2 estimator based on it is equal to the semi-parametric efficiency bound $\overline{\overline{C}}_n^{oGMM}$, i.e., that $\overline{\overline{C}}_n^o = \overline{\overline{C}}_n^{oGMM}$. Before proceeding, note that the maintained assumption that the semi-parametric model $\mathcal{S}$ is jointly correctly specified and dynamically complete for the conditional mean and the conditional variance ensures that the conditional moments restrictions on which optimal GMM is assumed to be based are both dynamically complete and satisfied at $\theta = \left(\theta_1', \theta_2'\right)' = \left(\theta_1^{o\prime}, \theta_2^{o\prime}\right)'$.

Let $r_t = r_t(Y_t, X_t, \theta_1, \theta_2) = \left(r_t^{1\prime}, r_t^{2\prime}\right)'$, where $r_t^1 = r_t^1(Y_t, X_t, \theta_1) = Y_t - m_t(X_t, \theta_1)$ and $r_t^2 = r_t^2(Y_t, X_t, \theta_1, \theta_2) = \text{vech}(Y_t Y_t' - \Omega_t\left(X_t, \theta_2\right) - m_t(X_t, \theta_1) m_t(X_t, \theta_1)')$. According to Wooldrigdge (1994) (see also Newey (1993)), under usual regularity conditions, the optimal GMM estimator is given by

$$\hat{\theta}_n^{GMM} = \text{Argmin}_{\theta\in\Theta}\left(\frac{1}{n}\sum_{t=1}^n F_t^o\left(X_t\right)' r_t(Y_t, X_t, \theta_1, \theta_2)\right)'\left(\frac{1}{n}\sum_{t=1}^n F_t^o\left(X_t\right)' r_t(Y_t, X_t, \theta_1, \theta_2)\right)$$

where $F_t^o\left(X_t\right)' = P R_t^{o\prime} \overline{\overline{\Xi}}_t^{o^{-1}}$ are optimal instruments, P is any non singular $k_\theta \times k_\theta$ matrix and

$$R_t^o = E\left[\left.\frac{\partial r_t(Y_t, X_t, \theta_1^o, \theta_2^o)}{\partial \theta'}\right| X_t\right]$$

$$\overline{\overline{\Xi}}_t^o = E\left[r_t(Y_t, X_t, \theta_1^o, \theta_2^o) r_t(Y_t, X_t, \theta_1^o, \theta_2^o)' \,|\, X_t\right]$$

The asymptotic covariance matrix of this estimator is given by the semi-parametric efficiency bound

$$\overline{C}_n^{o_{GMM}} = \left(\frac{1}{n}\sum_{t=1}^{n} E\left[R_t^{o\prime}\overline{\overline{\Xi}}_t^{o^{-1}} R_t^o\right]\right)^{-1}$$

According to the notation of Proposition 12, we have

$$\overline{\overline{\Xi}}_t^{o^{-1}} = \begin{bmatrix} \Omega_t^o & \overline{\overline{M}}_t^{3o\prime} \\ \overline{\overline{M}}_t^{3o} & \overline{\overline{M}}_t^{4o} \end{bmatrix}^{-1} = \begin{bmatrix} \overline{\overline{\Xi}}_t^{o^{11}} & \overline{\overline{\Xi}}_t^{o^{21}\prime} \\ \overline{\overline{\Xi}}_t^{o^{21}} & \overline{\overline{\Xi}}_t^{o^{22}} \end{bmatrix}$$

where, by partitioned inverse,

$$\begin{aligned}
\overline{\overline{\Xi}}_t^{o^{11}} &= \left(\Omega_t^o - \overline{\overline{M}}_t^{3o\prime}\overline{\overline{M}}_t^{4o^{-1}}\overline{\overline{M}}_t^{3o}\right)^{-1} \\
\overline{\overline{\Xi}}_t^{o^{21}} &= -\left(\overline{\overline{M}}_t^{4o} - \overline{\overline{M}}_t^{3o}\Omega_t^{o^{-1}}\overline{\overline{M}}_t^{3o\prime}\right)^{-1}\overline{\overline{M}}_t^{3o}\Omega_t^{o^{-1}} = -\overline{\overline{\Xi}}_t^{o^{22}}\overline{\overline{M}}_t^{3o}\Omega_t^{o^{-1}} \\
\overline{\overline{\Xi}}_t^{o^{22}} &= \left(\overline{\overline{M}}_t^{4o} - \overline{\overline{M}}_t^{3o}\Omega_t^{o^{-1}}\overline{\overline{M}}_t^{3o\prime}\right)^{-1}
\end{aligned}$$

On the other hand, we have

$$R_t^o = \begin{bmatrix} R_t^{11o} & R_t^{12o} \\ R_t^{21o} & R_t^{22o} \end{bmatrix}$$

with

$$\begin{aligned}
R_t^{11o} &= E\left[\left.\frac{\partial r_t^1(Y_t, X_t, \theta_1^o)}{\partial \theta_1'}\right| X_t\right] = -\frac{\partial m_t^o}{\partial \theta_1'} \\
R_t^{12o} &= E\left[\left.\frac{\partial r_t^1(Y_t, X_t, \theta_1^o)}{\partial \theta_2'}\right| X_t\right] = 0 \\
R_t^{21o} &= E\left[\left.\frac{\partial r_t^2(Y_t, X_t, \theta_1^o, \theta_2^o)}{\partial \theta_1'}\right| X_t\right] = -2D_{uG}^{+}\left(I_G \otimes m_t^o\right)\frac{\partial m_t^o}{\partial \theta_1'} \\
R_t^{22o} &= E\left[\left.\frac{\partial r_t^2(Y_t, X_t, \theta_1^o, \theta_2^o)}{\partial \theta_2'}\right| X_t\right] = -\frac{\partial \operatorname{vech}\Omega_t^o}{\partial \theta_2'}
\end{aligned}$$

where R_t^{21o} follows from the fact that, because of the symmetry of the matrix $m_t m_t'$,

$$\frac{\partial \operatorname{vech}(m_t m_t')}{\partial \theta_1'} = D_{uG}^{+}\frac{\partial \operatorname{vec}(m_t m_t')}{\partial \theta_1'} = D_{uG}^{+}\left((I_G \otimes m_t) + (m_t \otimes I_G)\right)\frac{\partial m_t}{\partial \theta_1'}$$

or, since for any $G \times G$ matrix A and $G \times 1$ vector b, $\frac{1}{2}\left((A \otimes b) + (b \otimes A)\right) = N_G\left(A \otimes b\right) = N_G\left(b \otimes A\right)$, and $D_{uG}^{+}N_G = D_{uG}^{+}$,

$$\frac{\partial \operatorname{vech}(m_t m_t')}{\partial \theta_1'} = 2D_{uG}^{+}N_G\left(I_G \otimes m_t\right)\frac{\partial m_t}{\partial \theta_1'} = 2D_{uG}^{+}\left(I_G \otimes m_t\right)\frac{\partial m_t}{\partial \theta_1'}$$

Setting $P=-I_{k_\theta}$, a set of optimal instruments $F_t^o(X_t)'$ is given by

$$F_t^o(X_t)' = \begin{bmatrix} \dfrac{\partial m_t^{o\prime}}{\partial \theta_1} & 2\dfrac{\partial m_t^{o\prime}}{\partial \theta_1}\left(I_G \otimes m_t^{o\prime}\right)\left(D_{uG}^{+}\right)' \\ 0 & \dfrac{\partial\left(\text{vech}\,\Omega_t^o\right)'}{\partial \theta_2} \end{bmatrix} \begin{bmatrix} \Xi_t^{o^{11}} & \Xi_t^{o^{21}\prime} \\ \Xi_t^{o^{21}} & \Xi_t^{o^{22}} \end{bmatrix}$$

According to (C-12) and (C-16) given in the proof of Proposition 10, the score of an RPML2 estimator $\hat{\theta}_n$ may be written as

$$\begin{aligned} s_t &= \begin{bmatrix} s_t^1 \\ s_t^2 \end{bmatrix} = \begin{bmatrix} \dfrac{\partial m_t'}{\partial \theta_1}\Omega_t^{-1} r_t^1 \\ \dfrac{\partial\left(\text{vech}\,\Omega_t\right)'}{\partial \theta_2}\left(M_t^4 - M_t^3\Omega_t^{-1}M_t^{3\prime}\right)^{-1}\left(r_t^2 - M_t^3\Omega_t^{-1}r_t^1\right) \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{\partial m_t'}{\partial \theta_1} & 0 \\ 0 & \dfrac{\partial\left(\text{vech}\,\Omega_t\right)'}{\partial \theta_2} \end{bmatrix} \begin{bmatrix} \Omega_t^{-1} & 0 \\ \Xi_t^{21} & \Xi_t^{22} \end{bmatrix} \begin{bmatrix} r_t^1 \\ r_t^2 \end{bmatrix} \\ &= H_t(X_t,\theta_1,\theta_2)' r_t(Y_t,X_t,\theta_1,\theta_2) \end{aligned}$$

where $\Xi_t^{22} = \left(M_t^4 - M_t^3\Omega_t^{-1}M_t^{3\prime}\right)^{-1}$ and $\Xi_t^{21} = -\Xi_t^{22}M_t^3\Omega_t^{-1}$.

Now, consider the following just-identified GMM estimator based on an arbitrary chosen sequence of pseudo-densities $\{f_t\}$ belonging to restricted quadratic exponential families

$$\hat{\theta}_n^{RPML2} = \text{Argmin}_{\theta\in\Theta}\left(\frac{1}{n}\sum_{t=1}^{n} H_t^o(X_t)' r_t(Y_t,X_t,\theta_1,\theta_2)\right)'\left(\frac{1}{n}\sum_{t=1}^{n} H_t^o(X_t)' r_t(Y_t,X_t,\theta_1,\theta_2)\right)$$

where

$$H_t^o(X_t)' = \begin{bmatrix} \dfrac{\partial m_t^{o\prime}}{\partial \theta_1} & 0 \\ 0 & \dfrac{\partial\left(\text{vech}\,\Omega_t^o\right)'}{\partial \theta_2} \end{bmatrix} \begin{bmatrix} \Omega_t^{o^{-1}} & 0 \\ \Xi_t^{o^{21}} & \Xi_t^{o^{22}} \end{bmatrix}$$

$\hat{\theta}_n^{RPML2}$ is just a GMM (or IV) analogue of the RPML2 estimator $\hat{\theta}_n$ based on the chosen sequence of pseudo-densities $\{f_t\}$. Again according to Wooldrigdge (1994) and under usual regularity conditions, because $\hat{\theta}_n^{RPML2}$ is a just-identified estimator, its asymptotic covariance matrix is given by

$$C_n^{o^{RPML2}} = \left(\frac{1}{n}\sum_{t=1}^{n} E\left[H_t^o(X_t)' R_t^o\right]\right)^{-1}\frac{1}{n}\sum_{t=1}^{n} V\left[H_t^o(X_t)' r_t^o\right]\left(\frac{1}{n}\sum_{t=1}^{n} E\left[H_t^o(X_t)' R_t^o\right]'\right)^{-1}$$

Since $H_t^o(X_t)' r_t^o = s_t^o$ is equal to the score, evaluated at $\theta = (\theta_1^{o\prime}, \theta_2^{o\prime})'$, of the RPML2 estimator $\hat{\theta}_n$ based on the chosen sequence of pseudo-densities $\{f_t\}$, and since the semi-parametric model $\mathcal{S}$ is assumed second order correctly specified and second order dynamically complete, from Proposition 12, we have

$$\frac{1}{n}\sum_{t=1}^{n} V\left[H_t^o(X_t)' r_t^o\right] = \frac{1}{n}\sum_{t=1}^{n} E\left[s_t^o s_t^{o\prime}\right] = \overline{B}_n^o$$

Likewise, since $s_t = H_t(X_t, \theta_1, \theta_2)' r_t(Y_t, X_t, \theta_1, \theta_2)$, we have

$$\begin{aligned} A_n^o &= A_n^{o\prime} = \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial s_t^o}{\partial \theta'}\right] = \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial\left(H_t^{o\prime} r_t^o\right)}{\partial \theta'}\right] \\ &= \frac{1}{n}\sum_{t=1}^{n} E\left[H_t^{o\prime}\frac{\partial r_t^o}{\partial \theta'} + \left(r_t^{o\prime} \otimes I_{k_\theta}\right)\frac{\partial \operatorname{vec}\left(H_t^{o\prime}\right)}{\partial \theta'}\right] \qquad \text{(D-1)} \\ &= \frac{1}{n}\sum_{t=1}^{n} E\left[H_t^{o\prime}\frac{\partial r_t^o}{\partial \theta'}\right] = \frac{1}{n}\sum_{t=1}^{n} E\left[H_t^o(X_t)' R_t^o\right] \end{aligned}$$

where the last equality holds because, using the law of iterated expectations,

$$E\left[\left(r_t^{o\prime} \otimes I_{k_\theta}\right)\frac{\partial \operatorname{vec}\left(H_t^{o\prime}\right)}{\partial \theta'}\right] = E\left[\left(E\left(r_t^{o\prime}|X_t\right) \otimes I_{k_\theta}\right)\frac{\partial \operatorname{vec}\left(H_t^{o\prime}\right)}{\partial \theta'}\right] = 0$$

Thus, $C_n^{oRPML2} = A_n^{o-1}\overline{B}_n^o A_n^{o-1} = \overline{C}_n^o$, i.e., $\hat{\theta}_n^{RPML2}$ is asymptotically equivalent to the RPML2 estimator $\hat{\theta}_n$ based on the chosen sequence of pseudo-densities $\{f_t\}$. Since the only difference between $\hat{\theta}_n^{RPML2}$ and $\hat{\theta}_n^{GMM}$ stems from the choice of the instruments $(H_t^o(X_t) \neq F_t^o(X_t))$ and that the latter uses the optimal instruments, it follows that $C_n^{oRPML2} - \overline{\overline{C}}_n^{oGMM} \gg 0$, or because $C_n^{oRPML2} = \overline{C}_n^o$, $\overline{C}_n^o - \overline{\overline{C}}_n^{oGMM} \gg 0$.

Next, suppose that there exists a sequence of pseudo-densities $\{f_t^o\}$ such that the implicit parametric model $\mathcal{P}$ arising from the semi-parametric model $\mathcal{S}$ and the sequence $\{f_t^o\}$ is in addition also fourth order correctly specified. Because the pseudo conditional densities $\{\lambda_t^{f_t^o}(., X_t, \theta^o) = f_t^o(., m_t(X_t, \theta_1^o), \Omega_t(X_t, \theta_2^o))\}$ are jointly correct for the first four conditional moments of Y_t given X_t, we have

$$\begin{aligned} E\left[F_t^o(X_t)' r_t^o|X_t\right] &= E_{\lambda_t^{f_t^o}}\left[F_t^o(X_t)' r_t^o|X_t\right] \\ &= \int \left(F_t^o(X_t)' r_t(Y_t, X_t, \theta_1^o, \theta_2^o)\right)\lambda_t^{f_t^o}(Y_t, X_t, \theta^o)\, \upsilon_t(dY_t) = 0 \end{aligned}$$

Differentiating this equality with respect to θ', we get

$$\begin{aligned} 0 &= \int\left(F_t^o(X_t)'\frac{\partial r_t^o}{\partial \theta'} + F_t^o(X_t)' r_t^o \frac{\partial \ln \lambda_t^{f_t^o}(Y_t, X_t, \theta^o)}{\partial \theta'}\right)\lambda_t^{f_t^o}(Y_t, X_t, \theta^o)\, \upsilon_t(dY_t) \\ &= \int\left(F_t^o(X_t)'\frac{\partial r_t^o}{\partial \theta'} + F_t^o(X_t)' r_t^o s_t^{of_t^o\prime}\right)\lambda_t^{f_t^o}(Y_t, X_t, \theta^o)\, \upsilon_t(dY_t) \end{aligned}$$

or

$$E_{\lambda_t^{f_t^o}}\left[F_t^o(X_t)'\frac{\partial r_t^o}{\partial \theta'}\bigg| X_t\right] = -E_{\lambda_t^{f_t^o}}\left[F_t^o(X_t)' r_t^o s_t^{of_t^o\prime}\bigg| X_t\right] \qquad \text{(D-2)}$$

Setting for convenience $P = I_{k_\theta}$, $F_t^o(X_t)' = R_t^{o\prime}\overline{\overline{\Xi}}_t^{o^{-1}}$ and we have

$$E_{\lambda_t^{f_t^o}}\left[F_t^o(X_t)'\frac{\partial r_t^o}{\partial\theta'}\bigg|X_t\right] = E_{\lambda_t^{f_t^o}}\left[R_t^{o\prime}\overline{\overline{\Xi}}_t^{o^{-1}}\frac{\partial r_t^o}{\partial\theta'}\bigg|X_t\right] = R_t^{o\prime}\overline{\overline{\Xi}}_t^{o^{-1}}R_t^o$$

Further, since $s_t^{o_{f_t^o}\prime} = r_t^{o\prime}H_t^{o_{f_t^o}}(X_t)$, we have

$$\begin{aligned}
&-E_{\lambda_t^{f_t^o}}\left[F_t^o(X_t)' r_t^o s_t^{o_{f_t^o}\prime}\bigg|X_t\right]\\
=\;&-E_{\lambda_t^{f_t^o}}\left[R_t^{o\prime}\overline{\overline{\Xi}}_t^{o^{-1}} r_t^o r_t^{o\prime}H_t^{o_{f_t^o}}(X_t)\bigg|X_t\right] = -R_t^{o\prime}\overline{\overline{\Xi}}_t^{o^{-1}}E_{\lambda_t^{f_t^o}}\left[r_t^o r_t^{o\prime}|X_t\right]H_t^{o_{f_t^o}}(X_t)\\
=\;&-R_t^{o\prime}\overline{\overline{\Xi}}_t^{o^{-1}}E\left[r_t^o r_t^{o\prime}|X_t\right]H_t^{o_{f_t^o}}(X_t) = -R_t^{o\prime}\overline{\overline{\Xi}}_t^{o^{-1}}\overline{\overline{\Xi}}_t^o H_t^{o_{f_t^o}}(X_t) = -R_t^{o\prime}H_t^{o_{f_t^o}}(X_t)
\end{aligned}$$

Thus, according to (D-2), we get

$$\overline{\overline{C}}_n^{oGMM} = \left(\frac{1}{n}\sum_{t=1}^n E\left[R_t^{o\prime}\overline{\overline{\Xi}}_t^{o^{-1}}R_t^o\right]\right)^{-1} = \left(\frac{1}{n}\sum_{t=1}^n E\left[-R_t^{o\prime}H_t^{o_{f_t^o}}(X_t)\right]\right)^{-1}$$

Finally, since, according to (D-1), $-A_n^{o^{-1}} = \left(\frac{1}{n}\sum_{t=1}^n E\left[-R_t^{o\prime}H_t^{o_{f_t^o}}(X_t)\right]\right)^{-1}$, and from Proposition 12, $\overline{\overline{C}}_n^o = -A_n^{o^{-1}}$, we thus have $\overline{\overline{C}}_n^{oGMM} = \overline{\overline{C}}_n^o$.

Appendix E

This appendix contains the simplified forms[7], for the normal density, of the general expressions appearing in Proposition 10-13 as well as a convenient writing of sufficient conditions for the information equality $B_n^o = -A_n^o$ to hold when using this density as pseudo-densities $\{f_t\}$.

If, for all $t = 1, 2, ...,$ f_t is the multivariate normal density, we have

$$\begin{aligned}
L_n(Y^n, X^n, \theta_1, \theta_2) &= \frac{1}{n}\sum_{t=1}^n \ln f_t\left(Y_t, m_t(X_t,\theta_1), \Omega_t(X_t,\theta_2)\right) = \frac{1}{n}\sum_{t=1}^n \ln\lambda_t\left(Y_t, X_t, \theta_1, \theta_2\right)\\
&= \frac{1}{n}\sum_{t=1}^n\left(-\frac{G}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega_t| - \frac{1}{2}(Y_t - m_t)'\,\Omega_t^{-1}(Y_t - m_t)\right)
\end{aligned}$$

where $m_t = m_t(X_t, \theta_1)$ and $\Omega_t = \Omega_t(X_t, \theta_2)$.

Then, following Magnus-Neudecker (1988), letting u_t stand for $(Y_t - m_t)$, for the first order derivatives, we have

$$s_t^1 = \frac{\partial\ln\lambda_t(Y_t, X_t, \theta_1, \theta_2)}{\partial\theta_1} = \frac{\partial m_t'}{\partial\theta_1}\Omega_t^{-1}u_t \tag{E-1}$$

[7] See Chapter 3 for further simplifications of the derivatives with respect to the vector of variance parameters.

$$\begin{aligned}
s_t^2 &= \frac{\partial \ln \lambda_t\left(Y_t, X_t, \theta_1, \theta_2\right)}{\partial \theta_1} = -\frac{1}{2}\frac{\partial \left(\operatorname{vec} \Omega_t^{-1}\right)'}{\partial \theta_2} \operatorname{vec}(u_t u_t' - \Omega_t) \\
&= \frac{1}{2}\frac{\partial \left(\operatorname{vec} \Omega_t\right)'}{\partial \theta_2}\left(\Omega_t^{-1} \otimes \Omega_t^{-1}\right) \operatorname{vec}(u_t u_t' - \Omega_t) \qquad \text{(E-2)} \\
&= \frac{1}{2}\frac{\partial \left(\operatorname{vec} \Omega_t\right)'}{\partial \theta_2} \operatorname{vec}\left(\Omega_t^{-1}(u_t u_t' - \Omega_t)\Omega_t^{-1}\right)
\end{aligned}$$

Useful identities for decoding (E-2) as well as subsequent expressions, and linking them to the general expressions appearing in Proposition 10 are that, for the normal density, we have

$$\frac{\partial C\left(m, \Sigma\right)'}{\partial \operatorname{vec} \Sigma} = -\left(\Sigma^{-1} m \otimes \Sigma^{-1}\right) \quad \text{and} \quad \frac{\partial \left(\operatorname{vec} D\left(\Sigma\right)\right)'}{\partial \operatorname{vec} \Sigma} = \frac{1}{2}\left(\Sigma^{-1} \otimes \Sigma^{-1}\right)$$

while, on the other hand, we have

$$\frac{\partial \operatorname{vec} \Omega_t^{-1}}{\partial \theta_2'} = -\left(\Omega_t^{-1} \otimes \Omega_t^{-1}\right)\frac{\partial \operatorname{vec} \Omega_t}{\partial \theta_2'}$$

$$\operatorname{vec}\left(ABC\right) = \left(C' \otimes A\right) \operatorname{vec} B \quad \text{and} \quad \left(A \otimes B\right)\left(C \otimes D\right) = \left(AC\right) \otimes \left(BD\right)$$

Further, for the second order derivatives, we have

$$\begin{aligned}
h_t^{11} &= \frac{\partial^2 \ln \lambda_t\left(Y_t, X_t, \theta_1, \theta_2\right)}{\partial \theta_1 \partial \theta_1'} = \frac{\partial s_t^1}{\partial \theta_1'} \\
&= -\frac{\partial m_t'}{\partial \theta_1}\Omega_t^{-1}\frac{\partial m_t}{\partial \theta_1'} + \left(u_t'\Omega_t^{-1} \otimes I_{k_{\theta_1}}\right)\frac{\partial}{\partial \theta_1'}\left[\operatorname{vec}\left(\frac{\partial m_t'}{\partial \theta_1}\right)\right] \qquad \text{(E-3)} \\
h_t^{12} &= \frac{\partial^2 \ln \lambda_t\left(Y_t, X_t, \theta_1, \theta_2\right)}{\partial \theta_1 \partial \theta_2'} = \frac{\partial s_t^1}{\partial \theta_2'} = \left(\frac{\partial s_t^2}{\partial \theta_1'}\right)' = \left(\frac{\partial^2 \ln \lambda_t\left(Y_t, X_t, \theta_1, \theta_2\right)}{\partial \theta_2 \partial \theta_1'}\right)' = \left(h_t^{21}\right)' \\
&= \left(u_t' \otimes \frac{\partial m_t'}{\partial \theta_1}\right)\frac{\partial \operatorname{vec} \Omega_t^{-1}}{\partial \theta_2'} = -\left(u_t' \otimes \frac{\partial m_t'}{\partial \theta_1}\right)\left(\Omega_t^{-1} \otimes \Omega_t^{-1}\right)\frac{\partial \operatorname{vec} \Omega_t}{\partial \theta_2'} \\
&= -\left(u_t'\Omega_t^{-1} \otimes \frac{\partial m_t'}{\partial \theta_1}\Omega_t^{-1}\right)\frac{\partial \operatorname{vec} \Omega_t}{\partial \theta_2'} \\
h_t^{22} &= \frac{\partial^2 \ln \lambda_t\left(Y_t, X_t, \theta_1, \theta_2\right)}{\partial \theta_2 \partial \theta_2'} = \frac{\partial s_t^2}{\partial \theta_2'} \\
&= \frac{1}{2}\frac{\partial \left(\operatorname{vec} \Omega_t^{-1}\right)'}{\partial \theta_2}\frac{\partial \operatorname{vec} \Omega_t}{\partial \theta_2'} - \frac{1}{2}\left(\left(\operatorname{vec}(u_t u_t' - \Omega_t)\right)' \otimes I_{k_{\theta_1}}\right)\Upsilon_t \qquad \text{(E-4)} \\
&= -\frac{1}{2}\frac{\partial \left(\operatorname{vec} \Omega_t\right)'}{\partial \theta_2}\left(\Omega_t^{-1} \otimes \Omega_t^{-1}\right)\frac{\partial \operatorname{vec} \Omega_t}{\partial \theta_2'} - \frac{1}{2}\left(\left(\operatorname{vec}(u_t u_t' - \Omega_t)\right)' \otimes I_{k_{\theta_1}}\right)\Upsilon_t
\end{aligned}$$

where

$$\Upsilon_t = \frac{\partial}{\partial \theta_2'}\left[\operatorname{vec}\left(\frac{\partial \left(\operatorname{vec} \Omega_t^{-1}\right)'}{\partial \theta_2}\right)\right]$$

i.e., a $G^2k_{\theta_2} \times k_{\theta_2}$ matrix.

Simplified expressions of $s_t^* = (s_t^{1*\prime}, s_t^{2*\prime})'$ and $s_t^o = (s_t^{1o\prime}, s_t^{2o\prime})'$ follow by evaluating (E-1) and (E-2) at respectively $\theta_n^* = \left(\theta_1^{o\prime}, \theta_{2n}^{*\prime}\right)'$ and $\theta_n^o = \left(\theta_1^{o\prime}, \theta_2^{o\prime}\right)$. Obviously, $A_{n_{11}}^*$ and $A_{n_{11}}^o$ are unchanged while $A_{n_{12}}^* = A_{n_{12}}^o = 0$. Finally, from (E-4), we have

$$A_{n_{22}}^* = -\frac{1}{2n}\sum_{t=1}^{n} E\left[\begin{array}{c} \dfrac{\partial\,(\mathrm{vec}\,\Omega_t^*)'}{\partial\theta_2}\left(\Omega_t^{*^{-1}} \otimes \Omega_t^{*^{-1}}\right)\dfrac{\partial\,\mathrm{vec}\,\Omega_t^*}{\partial\theta_2'} \\ +\left((\mathrm{vec}(u_t^o u_t^{o\prime} - \Omega_t^*))' \otimes I_{k_{\theta_1}}\right)\Upsilon_t^* \end{array}\right]$$

and

$$A_{n_{22}}^o = -\frac{1}{2n}\sum_{t=1}^{n} E\left[\frac{\partial\,(\mathrm{vec}\,\Omega_t^o)'}{\partial\theta_2}\left(\Omega_t^{o^{-1}} \otimes \Omega_t^{o^{-1}}\right)\frac{\partial\,\mathrm{vec}\,\Omega_t^o}{\partial\theta_2'}\right]$$

From Proposition 12, for the information equality $B_n^o = -A_n^o$ to hold, it is sufficient that, in addition to second order correct specification and second order correct dynamic specification, the implicit parametric $\mathcal{P}$ be also jointly correctly specified for the third and the fourth order conditional moments. According to Wooldridge (1994), when using the normal density as pseudo-densities $\{f_t\}$, these latter conditions are respectively satisfied if, for all $t = 1, 2, ...$, we have

(a) $E\left[\mathrm{vec}(u_t^o u_t^{o\prime})u_t^{o\prime} \middle| X_t\right] = 0$

(b) $E\left[\mathrm{vec}(u_t^o u_t^{o\prime} - \Omega_t^o)\,(\mathrm{vec}(u_t^o u_t^{o\prime} - \Omega_t^o))' \middle| X_t\right] = 2N_G\,(\Omega_t^o \otimes \Omega_t^o)$

In the univariate case, (a) is the symmetry condition $E\left[u_t^{o3}|X_t\right] = 0$ and (b) is the familiar fourth order moment condition $E\left[\left(u_t^{o2} - \sigma_t^{o2}\right)^2|X_t\right] = 2\sigma_t^{o4}$. The above sufficient conditions are just the multivariate version of these assumptions, and they could hold for distributions other than multivariate normal.

Chapter 2

Robust pseudo-maximum likelihood of order two estimation and specification testing

2.1. Introduction

In Chapter 1, we studied the behavior of second order pseudo-maximum likelihood estimators of second order semi-parametric models under possible conditional variance misspecification. We showed that sufficient and essentially necessary conditions for such an estimator to be robust to conditional variance misspecification are (1) that the mean and variance parameters vary independently and (2) that the pseudo-likelihood used as criterion function belongs to a family of distributions that we called restricted quadratic exponential families. We entitled a second order pseudo-maximum likelihood estimator which satisfies these conditions RPML2. Further, we provided the limiting distribution properties of this class of estimators under different assumptions regarding the degree of misspecification present in the model.

As concluding comments we argued that implemented using the normal density as pseudo-densities — which is probably the only manageable way to implement it —, because of its relative simplicity and its potential efficiency, this estimator should be useful in a variety of situations. In particular, we suggested that it constitutes an attractive tool for implementing the natural sequential "bottom-up" model construction/specification testing strategy advocated by Wooldridge (1991a). The purpose of the present chapter is to prop up this assertion.

In the framework of second order semi-parametric models, a sequential "bottom-up" model construction/specification testing strategy basically means first concentrating on the conditional mean specification, and, once this first step completed, to further explore, for efficiency reasons and/or because it is of interest of its own, the conditional variance specification. This obviously requires both estimation and specification testing procedures which allow to concentrate on some aspects of the conditional distribution of interest without having to worry about possible misspecification of the aspects which are actually not under scrutiny.

For such a job, in particular as an alternative to quasi-generalized pseudo-maximum likelihood of order one (QGPML1), gaussian RPML2 appears as a very convenient go-between estimator. Because of its robustness, it indeed allows to get efficiency gains from approximately taking into account the scedastic structure of the data when in a first step concentrating on the conditional mean specification, while in the same time it provides a basis to further explore the conditional variance specification.

This chapter describes how, from this nice go-between estimator, to take advantage of the very powerful m-testing / Wooldridge's modified m-testing framework for testing, either with or without clear alternatives in mind, the specification of second order semi-parametric models. We sequentially consider nested, non-nested, Hausman-type and information matrix-type testing of the prominent hypotheses of first order correct specification and second order correct specification. We also cover the testing of first order and second order dynamic completeness. In all cases, maintained hypotheses of the tests are precisely stated and reduced to their minimum so that the validity of the tests usually requires no more than just the hypothesis of interest under the null. This is an essential point for the outcomes of the testing procedures to be as unambiguous as possible, and thus for their practical usefulness.

Much of the material of this chapter is built from White (1981,1982,1987,1990, 1994) and Wooldridge (1990,1991a,1991b). Some of the proposed test statistics seem to be new.

The analysis is organized in the following manner. Section 2.2 defines gaussian robust pseudo-maximum likelihood of order two estimation and outlines some of its properties as they follow from Chapter 1. Section 2.3 sets up the principle of specification testing via m-tests. Section 2.4 portrays the m-testing / Wooldridge's modified m-testing framework. Section 2.5 is concerned with specification testing of the conditional mean and Section 2.6 with specification testing of the conditional variance. Finally, Section 2.7 proposes some concluding comments.

2.2. Gaussian robust pseudo-maximum likelihood of order two estimation (GRPML2)

We adopt the same general multivariate nonlinear dynamic framework and notational conventions than in Chapter 1.

For the record, the observations are denoted by $\{(Y_t', Z_t')' : \ t = 1, 2, ...\}$, where Y_t is a $G \times 1$ vector and Z_t is a $(\nu - G) \times 1$ vector, and are assumed to be a realization of an unknown stochastic process to which it is referred as the "true data generating process" (true DGP) P_o. X_t stands for some subset of the information set $(Z_t, \tilde{W}_{t-1})$, where $\tilde{W}_{t-1} \equiv (Y_{t-1}, Z_{t-1}, ..., Y_1, Z_1)$ is the information available on Y and Z at time $t-1$, and $Y^n \equiv (Y_1, Y_2, ..., Y_n)$ and $X^n \equiv (X_1, X_2, ..., X_n)$ denote finite random samples of size n.

We suppose that interest lies in modelling the conditional expectation of Y_t given X_t and, either for efficiency reasons or because it is of interest of its own, the conditional variance of Y_t given X_t. Accordingly, we assume that some structural second order semi-parametric model $\widetilde{\mathcal{S}}$ is available for $E(Y_t|X_t)$ and $V(Y_t|X_t)$, $t =$

$1, 2, ...,$ X_t being defined as comprising all the variables which appear either in the conditional mean or in the conditional variance.

Robust pseudo-maximum likelihood of order two estimation of $\widetilde{\mathcal{S}}$ basically means discarding eventual structural cross-constraints between mean and variance parameters and specifying the pseudo-densities underlying the second order pseudo-maximum likelihood estimator as members of restricted quadratic exponential families. The first requirement imposes that we treat the structural model $\widetilde{\mathcal{S}}$ as if it were given by

$$\mathcal{S} \equiv \begin{cases} \{m_t(X_t, \theta_1) : \ \theta_1 \in \Theta_1\} \\ \{\Omega_t(X_t, \theta_2) : \ \theta_2 \in \Theta_2\} \end{cases}, \ t = 1, 2, ...$$

where the functions m_t are known $G \times 1$ vector functions which may depend on t, and the functions Ω_t are known $G \times G$ matrix functions which may also depend on t and are symmetric positive definite, and the $k_{\theta_1} \times 1$ vector of parameters θ_1 and the $k_{\theta_2} \times 1$ vector of parameters θ_2 vary independently on respectively Θ_1 and Θ_2. Obviously, if mean and variance parameters already varied independently in the structural model $\widetilde{\mathcal{S}}$, $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are identical.

As mentioned, the easiest way to fulfill the second requirement is to resort to the gaussian density as pseudo-densities. A gaussian robust pseudo-maximum likelihood of order two estimator $\hat{\theta}_n = \left(\hat{\theta}'_{1_n}, \hat{\theta}'_{2_n}\right)'$ (GRPML2) is then defined as a solution of

$$\text{Max}_{\theta \in \Theta} \ L_n(Y^n, X^n, \theta) \equiv \frac{1}{n} \sum_{t=1}^{n} \ln f\left(Y_t, m_t\left(X_t, \theta_1\right), \Omega_t\left(X_t, \theta_2\right)\right)$$

where

$$f\left(Y_t, m_t, \Omega_t\right) = -\frac{G}{2} \ln 2\pi - \frac{1}{2} \ln |\Omega_t| - \frac{1}{2} u'_t \Omega_t^{-1} u_t$$

with $u_t = Y_t - m_t$, $m_t = m_t\left(X_t, \theta_1\right)$, $\Omega_t = \Omega_t\left(X_t, \theta_2\right)$ and $\Theta = \Theta_1 \times \Theta_2$.

According to the results[1] of Chapter 1, under usual regularity conditions, if $\mathcal{S}$ is correctly specified for the conditional mean (first order correct specification), i.e., if for some $\theta_1^o \in \Theta_1$, $E(Y_t|X_t) = m_t(X_t, \theta_1^o)$, $t = 1, 2, ...,$ from Proposition 7, we have that

$$\hat{\theta}_{1_n} \xrightarrow{a.s.} \theta_1^o \text{ and } \hat{\theta}_{2_n} - \theta_{2_n}^* \xrightarrow{a.s.} 0, \text{ as } n \to \infty$$

where $\theta_{2_n}^*$ is a pseudo-true value. Further, regarding the mean parameters estimator $\hat{\theta}_{1_n}$, from Proposition 9 and 10, we also have that

$$\sqrt{n}(\hat{\theta}_{1_n} - \theta_1^o) = -A_{n_{11}}^{*^{-1}} n^{-1/2} \sum_{t=1}^{n} s_t^{1*} + o_{P_o}(1) \tag{2.1}$$

where

$$A_{n_{11}}^* = -\frac{1}{n} \sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1} \Omega_t^{*^{-1}} \frac{\partial m_t^o}{\partial \theta'_1}\right]$$

[1] We only outline here the properties of GRPML2 which will be used in the sequence of this Chapter. All its limiting distribution properties may be retrieved by using Proposition 9-13 and the expressions given in Appendix E of Chapter 1. We will detail them for the case at hand in Chapter 3.

and

$$s_t^{1*} = \frac{\partial m_t^{o\prime}}{\partial \theta_1} \Omega_t^{*^{-1}} (Y_t - m_t^o), \quad m_t^o = m_t(X_t, \theta_1^o), \quad \Omega_t^* = \Omega_t(X_t, \theta_{2_n}^*)$$

On the other hand, if $\mathcal{S}$ is in addition also correctly specified for the conditional variance (second order correct specification), i.e., if, in addition, for some $\theta_2^o \in \Theta_2$, $V(Y_t|X_t) = \Omega_t(X_t, \theta_2^o)$, $t = 1, 2, ...$, again from Proposition 7, we have that

$$\hat{\theta}_n \xrightarrow{a.s.} \theta^o, \text{ as } n \to \infty$$

where $\theta^o = (\theta_1^{o\prime}, \theta_2^{o\prime})'$. Now, regarding the variance parameters estimator $\hat{\theta}_{2_n}$, from Proposition 9, 10 and 11, we similarly have that

$$\sqrt{n}(\hat{\theta}_{2_n} - \theta_2^o) = -A_{n22}^{o^{-1}} n^{-1/2} \sum_{t=1}^{n} s_t^{2o} + o_{P_o}(1) \tag{2.2}$$

where, for the normal density,

$$A_{n22}^o = -\frac{1}{2n} \sum_{t=1}^{n} E\left[\frac{\partial\,(\mathrm{vec}\,\Omega_t^o)'}{\partial \theta_2} \left(\Omega_t^{o^{-1}} \otimes \Omega_t^{o^{-1}}\right) \frac{\partial\,\mathrm{vec}\,\Omega_t^o}{\partial \theta_2'}\right]$$

and

$$s_t^{2o} = \frac{1}{2}\frac{\partial\,(\mathrm{vec}\,\Omega_t^o)'}{\partial \theta_2} \left(\Omega_t^{o^{-1}} \otimes \Omega_t^{o^{-1}}\right) \mathrm{vec}(u_t^o u_t^{o\prime} - \Omega_t^o), \quad u_t^o = Y_t - m_t^o, \quad \Omega_t^o = \Omega_t(X_t, \theta_2^o)$$

To conclude this section, note that the first order conditions defining of the GRPML2 estimator are given by

$$\frac{\partial L_n(Y^n, X^n, \hat{\theta}_n)}{\partial \theta_1} = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \hat{m}_t'}{\partial \theta_1} \hat{\Omega}_t^{-1} \hat{u}_t = 0 \tag{2.3}$$

$$\frac{\partial L_n(Y^n, X^n, \hat{\theta}_n)}{\partial \theta_2} = \frac{1}{2n} \sum_{t=1}^{n} \frac{\partial \left(\mathrm{vec}\,\hat{\Omega}_t\right)'}{\partial \theta_2} \left(\hat{\Omega}_t^{-1} \otimes \hat{\Omega}_t^{-1}\right) \mathrm{vec}(\hat{u}_t \hat{u}_t' - \hat{\Omega}_t) = 0 \tag{2.4}$$

where $\hat{u}_t = Y_t - \hat{m}_t$, $\hat{m}_t = m_t(X_t, \hat{\theta}_{1_n})$ and $\hat{\Omega}_t = \Omega_t(X_t, \hat{\theta}_{2_n})$.

2.3. Specification testing via m-tests

Let us summarize the situation. We have data which are supposed to arise from an unknown DGP P_o. Because we are interested in explaining Y_t in terms of X_t, we make assumptions about the two first conditional moments of Y_t given X_t. On the other hand, we have an estimation procedure which is known to deliver an n-root consistent estimator of either the mean or the mean and variance parameters depending on the extent of the correct specification of our tentative model. The question is now: how to check the extent to which our tentative model is actually correctly specified?

The null hypothesis that the conditional mean is correctly specified is equivalent to the null

$$\mathrm{H}_0^m : E(u_t^o|X_t) = 0 \text{ for some } \theta_1^o \in \Theta_1, \quad t = 1, 2, ...$$

If H_0^m is true, for any "regular" $G \times q$ matrix functions $F_t^m(X_t)$ depending on X_t — and possibly on some vector of nuisance parameters, see below —, by the law of iterated expectations, we must have that

$$E\left[F_t^m(X_t)' u_t^o\right] = 0, \quad t = 1, 2, ...$$

while we may generally expect that

$$E\left[F_t^m(X_t)' u_t^o\right] \neq 0, \quad t = 1, 2, ...$$

whenever H_0^m is false.

Both θ_1^o and the expectations — taken with respect to the DGP P_o — above are unknown. However, $\frac{1}{n}\sum_{t=1}^n E[F_t^m(X_t)' u_t^o]$ can usually be consistently estimated. This suggests that a test of first order correct specification may be undertaken by looking at the empirical covariances

$$\hat{\Phi}_n = \frac{1}{n}\sum_{t=1}^n F_t^m(X_t)' \hat{u}_t$$

$\hat{\theta}_{1_n}$ being a consistent estimator of θ_1^o under H_0^m. When H_0^m is true, $\hat{\Phi}_n$ should be close to zero while we may generally expect it to be far from zero otherwise.

This way of looking at specification testing is just an example of what it is sometimes referred to as 'specification testing via m-tests'. The m-testing framework provides a very powerful unified framework for specification testing. Virtually all specification tests — actually most usual tests— may indeed be viewed as m-tests. The m-testing framework was first suggested by Newey (1985) and Tauchen (1985) for the detection of misspecification in the context of maximum likelihood models. It has been further developed by White (1987, 1990, 1994), Wooldridge (1990, 1991a, 1991b) and Bollerslev-Wooldridge (1992). Related works may be found in Bierens (1994), among others.

Testing that the model is in addition also correctly specified for the conditional variance may be performed along the same lines. The null hypothesis is now equivalent to the null

$$\mathrm{H}_0^v : E(u_t^o|X_t) = 0 \text{ and } E\left[\mathrm{vec}(u_t^o u_t^{o\prime} - \Omega_t^o)|X_t\right] = 0 \text{ for some } \theta^o \in \Theta, \quad t = 1, 2, ...$$

Given the nested nature of H_0^m and H_0^v, according to the sequential "bottom-up" model construction/specification testing strategy evoked in the introduction, it seems sensible to emphasize the construction of statistics for testing departures from $E\left[\mathrm{vec}(u_t^o u_t^{o\prime} - \Omega_t^o)|X_t\right] = 0$, i.e., the conditional variance, letting first order correct specification test statistics to take care of detecting departures from H_0^m.

So, as above, if H_0^v is true, for any "regular" $G^2 \times q$ matrix functions $F_t^v(X_t)$ depending on X_t — and again possibly on some vector of nuisance parameters —, we must have that

$$E\left[F_t^v\left(X_t\right)' \operatorname{vec}(u_t^o u_t^{o\prime} - \Omega_t^o)\right] = 0, \quad t = 1, 2, ...$$

and this similarly suggests looking at the empirical covariances

$$\hat{\Phi}_n = \frac{1}{n}\sum_{t=1}^{n} F_t^v\left(X_t\right)' \operatorname{vec}(\hat{u}_t\hat{u}_t' - \hat{\Omega}_t)$$

$\hat{\theta}_n$ being a consistent estimator of θ^o under H_0^v.

The implementation of the above specification testing scheme requires three ingredients. First consistent estimators under the null hypotheses. Although they are by no means the only possible ones, GRPML2 provides such estimators.

Secondly relevant choices for the matrix functions F_t^m and F_t^v. The choice of the misspecification indicators $\hat{\Phi}_n$ is of course crucial: it determines the directions of the departures from the null in which the tests will have power. Appropriate choices of misspecification indicators, in particular as they follow from a variety of popular specification tests, are provided in Section 2.5 and 2.6, for respectively conditional mean and conditional variance testing.

The last needed ingredient is a statistical decision rule for deciding how far from zero a value of the misspecification indicator $\hat{\Phi}_n$ constitutes an evidence of misspecification. This question of how far from zero is too far may be answered asymptotically by finding the limiting distribution of $\hat{\Phi}_n$. That is the purpose of next section.

2.4. The m-testing / Wooldridge's modified m-testing framework

In this section, we outline abstract results following from the m-testing framework and adapted to our testing problem. We will make use of them in Section 2.5 and 2.6.

According to the previous section, as a general setting, suppose that interest lies in testing the null hypothesis

$$\mathrm{H}_0 : E\left[r_t\left(Y_t, X_t, \varphi^o\right)|X_t\right] = 0 \text{ for some } \varphi^o \in \Theta_\varphi, \quad t = 1, 2, ...$$

where the possibly time-varying functions r_t are known $l \times 1$ vector functions of the data and some $k_\varphi \times 1$ vector of parameters.

As it will be shown to be relevant below, following Wooldridge (1990), consider testing H_0 through $q \times 1$ empirical moment restrictions of the form

$$\hat{\Phi}_n = \frac{1}{n}\sum_{t=1}^{n} \phi_t\left(Y_t, X_t, \hat{\varphi}_n, \hat{\pi}_n\right)$$

$$= \frac{1}{n}\sum_{t=1}^{n} W_t\left(X_t, \hat{\varphi}_n, \hat{\pi}_n\right)' \Lambda_t(X_t, \hat{\varphi}_n, \hat{\pi}_n)^{-1} r_t\left(Y_t, X_t, \hat{\varphi}_n\right) \tag{2.5}$$

where the functions W_t are known $l \times q$ matrix functions which may depend on t, the functions Λ_t are known $l \times l$ matrix functions which may also depend on t and are symmetric positive definite, $\hat{\varphi}_n$ is a n-root consistent estimator of φ^o under H_0 and $\hat{\pi}_n$ is assumed to be, under H_0, a n-root consistent estimator of some non-stochastic sequence of $k_\pi \times 1$ vectors of pseudo-true values $\{\pi_n^* : n = 1, 2, ...\}$. Note that the π_n^* need not have any meaningful interpretation under H_0. Hereafter, we will refer to the W_t as the indicator matrices, to the Λ_t — which could simply be identity matrices — as the weighting matrices, and to π as the nuisance parameters.

Following the White's (1987, 1990, 1994) general treatment of m-testing, to use $\hat{\Phi}_n$ as a basis for testing the null H_0 entails finding the limiting distribution of $\sqrt{n}\hat{\Phi}_n$ under H_0. Under usual regularity conditions, a standard mean value expansion of $\sqrt{n}\hat{\Phi}_n$ around $(\varphi^{o\prime}, \pi_n^{*\prime})'$ yields

$$\sqrt{n}\hat{\Phi}_n = \sqrt{n}\Phi_n^{o*} + G_{n_\varphi}^{o*}\sqrt{n}\left(\hat{\varphi}_n - \varphi^o\right) + G_{n_\pi}^{o*}\sqrt{n}\left(\hat{\pi}_n - \pi_n^*\right) + o_{P_o}(1) \tag{2.6}$$

where

$$\Phi_n^{o*} = \frac{1}{n}\sum_{t=1}^{n}\phi_t^{o*}\ , \qquad \phi_t^{o*} = \phi_t\left(Y_t, X_t, \varphi^o, \pi_n^*\right)$$

$$G_{n_\varphi}^{o*} = \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial \phi_t^{o*}}{\partial \varphi'}\right] \quad \text{and} \quad G_{n_\pi}^{o*} = \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial \phi_t^{o*}}{\partial \pi'}\right]$$

Letting W_t^{o*}, Λ_t^{o*}, and r_t^o respectively denote $W_t\left(X_t, \varphi^o, \pi_n^*\right)$, $\Lambda_t(X_t, \varphi^o, \pi_n^*)$ and $r_t\left(Y_t, X_t, \varphi^o\right)$, and recalling that $\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec} B$, we have

$$\frac{\partial \phi_t^{o*}}{\partial \varphi'} = W_t^{o*\prime}\Lambda_t^{o*^{-1}}\frac{\partial r_t^o}{\partial \varphi'} + \left(r_t^{o\prime} \otimes I_l\right)\frac{\partial}{\partial \varphi'}\left[\operatorname{vec}\left(W_t^{o*\prime}\Lambda_t^{o*^{-1}}\right)\right] \tag{2.7}$$

$$\frac{\partial \phi_t^{o*}}{\partial \pi'} = \left(r_t^{o\prime} \otimes I_l\right)\frac{\partial}{\partial \pi'}\left[\operatorname{vec}\left(W_t^{o*\prime}\Lambda_t^{o*^{-1}}\right)\right] \tag{2.8}$$

Further, assuming that the $l \times k_\varphi$ matrix functions R_t of conditional expectations

$$R_t\left(X_t, \varphi^o\right) = E\left[\left.\frac{\partial r_t\left(Y_t, X_t, \varphi^o\right)}{\partial \varphi'}\right| X_t\right], \quad t = 1, 2, ... \tag{2.9}$$

are of known form under H_0, since $E(r_t^o|X_t) = 0$, from (2.7) and (2.8), applying the law of iterated expectations, we get

$$G_{n_\varphi}^{o*} = \frac{1}{n}\sum_{t=1}^{n} E\left[W_t^{o*\prime}\Lambda_t^{o*^{-1}}R_t^o\right] \quad \text{and} \quad G_{n_\pi}^{o*} = 0 \tag{2.10}$$

where $R_t^o = R_t(X_t, \varphi^o)$. (2.6) then collapses to

$$\sqrt{n}\hat{\Phi}_n = \sqrt{n}\Phi_n^{o*} + G_{n_\varphi}^{o*}\sqrt{n}\left(\hat{\varphi}_n - \varphi^o\right) + o_{P_o}(1) \tag{2.11}$$

According to (2.11), the limiting distribution of $\sqrt{n}\hat{\Phi}_n$ under H_0 thus depends on the limiting distribution of the estimator $\hat{\varphi}_n$ but not on the one of the nuisance parameters estimator $\hat{\pi}_n$. It however depends on the pseudo-true values π_n^*.

As it will again be shown to be relevant below — GRPML2 estimators indeed satisfy such a condition — let us further assume that the n-root consistent estimator $\hat{\varphi}_n$ of φ^o satisfies the first order expansion

$$\sqrt{n}\left(\hat{\varphi}_n - \varphi^o\right) = -\left(\frac{1}{n}\sum_{t=1}^{n} E\left[R_t^{o\prime}\Lambda_t^{o*^{-1}}R_t^o\right]\right)^{-1} n^{-1/2}\sum_{t=1}^{n}\hat{R}_t^{\prime}\hat{\Lambda}_t^{-1}\hat{r}_t + o_{P_o}(1) \quad (2.12)$$

where, $\hat{R}_t = R_t\left(X_t, \hat{\varphi}_n\right)$, $\hat{\Lambda}_t = \Lambda_t(X_t, \hat{\varphi}_n, \hat{\pi}_n)$ and $\hat{r}_t = r_t\left(Y_t, X_t, \hat{\varphi}_n\right)$. (2.12) is typically satisfied when

$$\sum_{t=1}^{n} R_t(X_t,\varphi)^{\prime}\Lambda_t(X_t,\varphi,\hat{\pi}_n)^{-1} r_t\left(Y_t, X_t, \varphi\right) = 0 \quad (2.13)$$

are the first order conditions that defines $\hat{\varphi}_n$.

Substituting (2.10) and (2.12) into (2.11), we get

$$\begin{aligned}\sqrt{n}\hat{\Phi}_n &= n^{-1/2}\sum_{t=1}^{n}\left(W_t^{o*} - R_t^o P_n^{o*}\right)^{\prime}\Lambda_t^{o*^{-1}} r_t^o + o_{P_o}(1) \\ &= n^{-1/2}\sum_{t=1}^{n}\xi_t^{o*} + o_{P_o}(1) \quad (2.14)\end{aligned}$$

where

$$\begin{aligned}\xi_t^{o*} &= \left(W_t^{o*} - R_t^o P_n^{o*}\right)^{\prime}\Lambda_t^{o*^{-1}} r_t^o \\ P_n^{o*} &= \left(\sum_{t=1}^{n} E\left[R_t^{o\prime}\Lambda_t^{o*^{-1}}R_t^o\right]\right)^{-1}\sum_{t=1}^{n} E\left[R_t^{o\prime}\Lambda_t^{o*^{-1}}W_t^{o*}\right]\end{aligned}$$

Usually, a central limit theorem will ensure that the sum $n^{-1/2}\sum_{t=1}^{n}\xi_t^{o*}$ is asymptotically normal with zero mean and covariance matrix K_n^{o*}. Then, under standard regularity conditions, a test of H_0 may be based on the asymptotic chi-square statistic

$$\mathcal{M}_n = n\hat{\Phi}_n^{\prime}\hat{K}_n^{-1}\hat{\Phi}_n \xrightarrow{d} \chi^2(q)$$

where $\hat{K}_n$ is a consistent estimator of

$$K_n^{o*} = \frac{1}{n}\sum_{t=1}^{n} E\left[\xi_t^{o*}\xi_t^{o*\prime}\right] + \frac{1}{n}\sum_{\tau=1}^{n-1}\sum_{t=\tau+1}^{n}\left(E\left[\xi_t^{o*}\xi_{t-\tau}^{o*\prime}\right] + E\left[\xi_{t-\tau}^{o*}\xi_t^{o*\prime}\right]\right)$$

The validity of the test statistic $\mathcal{M}_n$ crucially relies on the assumed limiting distribution property (2.12) for the estimator $\hat{\varphi}_n$. Wooldrige (1990) proposed a clever modification of the above m-testing scheme which allows to perform the testing of H_0 without having to worry about the limiting distribution properties of the estimator used to estimate φ^o under the null. Its basic idea is to modify the original misspecification indicator $\hat{\Phi}_n$ in order to get rid of its influence.

Wooldridge (1990) considers the modified misspecification indicator

$$\begin{aligned}\hat{\Phi}_n^w &= \frac{1}{n}\sum_{t=1}^{n}\phi_t^w\left(Y_t,X_t,\hat{\varphi}_n,\hat{\pi}_n\right)\\ &= \frac{1}{n}\sum_{t=1}^{n}\Big(W_t\left(X_t,\hat{\varphi}_n,\hat{\pi}_n\right)-R_t\left(X_t,\hat{\varphi}_n\right)\hat{P}_n\Big)'\Lambda_t(X_t,\hat{\varphi}_n,\hat{\pi}_n)^{-1}r_t\left(Y_t,X_t,\hat{\varphi}_n\right)\end{aligned}$$

where

$$\hat{P}_n=\left(\sum_{t=1}^{n}\hat{R}_t'\hat{\Lambda}_t^{-1}\hat{R}_t\right)^{-1}\sum_{t=1}^{n}\hat{R}_t'\hat{\Lambda}_t^{-1}\hat{W}_t$$

with $\hat{W}_t=W_t\left(X_t,\hat{\varphi}_n,\hat{\pi}_n\right)$, the other quantities having already been defined.

The fundamental result regarding this modified misspecification indicator is that if we apply essentially the same reasoning than above (with $\hat{P}_n$ viewed as an additional nuisance parameter), according to Theorem 2.1 of Wooldridge (1990) (see also White (1994)), we now get that

$$\sqrt{n}\hat{\Phi}_n^w=\sqrt{n}\hat{\Phi}_n^{wo*}+o_{P_o}(1) \tag{2.15}$$

where

$$\hat{\Phi}_n^{wo*}=\frac{1}{n}\sum_{t=1}^{n}\phi_t^{wo*}\ ,\qquad \phi_t^{wo*}=\phi_t^w\left(Y_t,X_t,\varphi^o,\pi_n^*\right)=\left(W_t^{o*}-R_t^oP_n^{o*}\right)'\Lambda_t^{o*^{-1}}r_t^o$$

i.e.,

$$\sqrt{n}\hat{\Phi}_n^w=n^{-1/2}\sum_{t=1}^{n}\xi_t^{o*}+o_{P_o}(1) \tag{2.16}$$

In other words, according to (2.15), and unlike $\sqrt{n}\hat{\Phi}_n$ — see equation (2.11) —, the limiting distribution of $\sqrt{n}\hat{\Phi}_n^w$ does no longer depends on the limiting distribution of the estimator used to estimate φ^o under the null. Further, comparing (2.14) and (2.16), it appears that $\sqrt{n}\hat{\Phi}_n^w$ and $\sqrt{n}\hat{\Phi}_n$ are asymptotically equivalent — obviously under the null, but also under local alternatives, see Wooldrige (1990) — while they (trivially) are numerically equal if both $\hat{\Phi}_n^w$ and $\hat{\Phi}_n$ are computed at an estimator $\hat{\varphi}_n$ which satisfies the first order conditions (2.13).

So, from (2.16) and provided that standard regularity conditions hold, a test of H_0 may alternatively be based on the asymptotic chi-square statistic

$$\mathcal{M}_n^w=n\hat{\Phi}_n^{w\prime}\hat{K}_n^{-1}\hat{\Phi}_n^w\xrightarrow{d}\chi^2(q)$$

Now, unlike $\mathcal{M}_n$, $\mathcal{M}_n^w$ is valid whatever the n-root consistent estimator used to estimate φ^o under the null.

To sum up, any test of H_0 originally intended to be performed based on a misspecification indicator of the kind of $\hat{\Phi}_n$ and using an estimator $\hat{\varphi}_n$ of φ^o satisfying the first order expansion (2.12), so that $\mathcal{M}_n$ is the relevant "standard" m-test statistic, may equivalently — from a local alternatives point of view — be undertaken through the Wooldridge's modified $\mathcal{M}_n^w$ statistic using any n-root consistent estimator of φ^o under the null. Further, if the estimator $\hat{\varphi}_n$ also satisfies the first order conditions (2.13) and the same estimator $\hat{K}_n$ of K_n^{o*} is used, then $\mathcal{M}_n^w$ evaluated at $\hat{\varphi}_n$ will yield a statistic numerically equal — and thus trivially also equivalent under global alternatives — to $\mathcal{M}_n$. In this case, $\mathcal{M}_n^w$ then appears as a particular way to compute $\mathcal{M}_n$ which has the additional property to remain valid — and locally equivalent to $\mathcal{M}_n$ — when evaluated at any alternative n-root consistent estimator of φ^o. $\mathcal{M}_n^w$ is thus in this case particularly appealing. This are exactly the situations that we will encounter hereafter.

2.5. Testing the conditional mean

In this section, we are concerned with testing the null that the conditional mean is correctly specified, i.e., testing the null

$$\mathrm{H}_0^m : E\left(Y_t|X_t\right) = m_t(X_t, \theta_1^o) \text{ for some } \theta_1^o \in \Theta_1, \quad t = 1, 2, ...$$

against the alternative

$$\mathrm{H}_1^m : \mathrm{H}_0^m \text{ is false}$$

Hereafter, we outline misspecification indicators suitable for testing H_0^m against auxiliary nested alternatives, auxiliary non-nested alternatives, as well as for testing H_0^m without resorting to explicit alternatives. As an extension, misspecification indicators for testing the dynamic completeness of the conditional mean specification are also discussed.

All considered misspecification indicators are special cases — for various choices of $\underline{\hat{S}}_n$ and W_t^m — of

$$\underline{\hat{\Phi}}_n^m = \underline{\hat{S}}_n \hat{\Phi}_n^m \tag{2.17}$$

where

$$\hat{\Phi}_n^m = \frac{1}{n} \sum_{t=1}^{n} W_t^m(X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n}, \hat{\gamma}_n)' \Omega_t(X_t, \hat{\theta}_{2_n})^{-1} u_t(Y_t, X_t, \hat{\theta}_{1_n}) \tag{2.18}$$

$u_t(Y_t, X_t, \hat{\theta}_{1_n}) = Y_t - m_t(X_t, \hat{\theta}_{1_n})$, the W_t^m are $G \times q$ indicator matrix functions, $\hat{\theta}_{1_n}$ and $\hat{\theta}_{2_n}$ are the GRPML2 estimators of model $\mathcal{S}$, $\hat{\gamma}_n$ is, under H_0^m, a n-root consistent estimator of some non-stochastic sequence of $k_\gamma \times 1$ vectors of pseudo-true values $\{\gamma_n^* : n = 1, 2, ...\}$ and $\underline{\hat{S}}_n$ is a stochastic $p \times q$ $(p \leq q)$ selection matrix converging in probability under H_0^m to some non-stochastic sequence $\{\underline{S}_n^* : n = 1, 2, ...\}$.

$\hat{\Phi}_n^m$ is obviously a special case of (2.5) with $\hat{r}_t = \hat{u}_t = u_t(Y_t, X_t, \hat{\theta}_{1_n})$, $\hat{\Lambda}_t = \hat{\Omega}_t = \Omega_t(X_t, \hat{\theta}_{2_n})$, $\hat{W}_t = \hat{W}_t^m = W_t^m(X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n}, \hat{\gamma}_n)$, $\hat{\varphi}_n = \hat{\theta}_{1_n}$, $\hat{\pi}_n = \left(\hat{\theta}_{2_n}', \hat{\gamma}_n'\right)'$ — the vector of variance parameters θ_2 thus appears as a nuisance parameter — and

$l = G$. The matrices R_t of conditional expectations (2.9) are here equal to

$$R_t\left(X_t, \theta_1^o\right) = E\left[\left.\frac{\partial u_t\left(Y_t, X_t, \theta_1^o\right)}{\partial \theta_1'}\right| X_t\right] = -\frac{\partial m_t\left(X_t, \theta_1^o\right)}{\partial \theta_1'}$$

In view of (2.1) and (2.3), it is easily seen that the GRPML2 mean parameters estimator $\hat{\theta}_{1_n}$ satisfies both the first order expansion (2.12) and the first order conditions (2.13). An $\mathcal{M}_n^w$-like test statistic is thus the most appealing to look at. Using the general results of Section 2.4, and noting that, because $\underline{\hat{S}}_n - \underline{S}_n^* = o_{P_o}(1)$, we have

$$\underline{\hat{S}}_n \sqrt{n} \hat{\Phi}_n^{m^w} = \underline{S}_n^* \sqrt{n} \hat{\Phi}_n^{m^w} + o_{P_o}(1)$$

it may be readily checked that the relevant $\mathcal{M}_n^w$-like statistic for testing H_0^m based on $\underline{\hat{\Phi}}_n^m$ is given by

$$\mathcal{M}_n^{m^w} = n \hat{\Phi}_n^{m^w\prime} \underline{\hat{S}}_n' \left[\underline{\hat{S}}_n \hat{K}_n^m \underline{\hat{S}}_n'\right]^{-1} \underline{\hat{S}}_n \hat{\Phi}_n^{m^w} \xrightarrow{d} \chi^2(p) \tag{2.19}$$

where

$$\hat{\Phi}_n^{m^w} = \frac{1}{n} \sum_{t=1}^n \hat{\xi}_t^m = \frac{1}{n} \sum_{t=1}^n \left(\hat{W}_t^m - \hat{R}_t^m \hat{P}_n^m\right)' \hat{\Omega}_t^{-1} \hat{u}_t$$

with[2]

$$\hat{R}_t^m = \frac{\partial m_t(X_t, \hat{\theta}_{1_n})}{\partial \theta_1'}, \quad \hat{P}_n^m = \left(\sum_{t=1}^n \hat{R}_t^{m\prime} \hat{\Omega}_t^{-1} \hat{R}_t^m\right)^{-1} \sum_{t=1}^n \hat{R}_t^{m\prime} \hat{\Omega}_t^{-1} \hat{W}_t^m$$

and $\hat{K}_n^m$ is a consistent estimator of

$$K_n^{m^{o*}} = \frac{1}{n} \sum_{t=1}^n E\left[\xi_t^{m^{o*}} \xi_t^{m^{o*}\prime}\right] + \frac{1}{n} \sum_{\tau=1}^{n-1} \sum_{t=\tau+1}^n \left(E\left[\xi_t^{m^{o*}} \xi_{t-\tau}^{m^{o*}\prime}\right] + E\left[\xi_{t-\tau}^{m^{o*}} \xi_t^{m^{o*}\prime}\right]\right)$$

where

$$\xi_t^{m^{o*}} = \left(W_t^{m^{o*}} - R_t^{m^o} P_n^{m^{o*}}\right)' \Omega_t^{*^{-1}} u_t^o$$

$$\begin{aligned} W_t^{m^{o*}} &= W_t^m\left(X_t, \theta_1^o, \theta_{2_n}^*, \hat{\gamma}_n^*\right), \quad R_t^{m^o} = \frac{\partial m_t(X_t, \theta_1^o)}{\partial \theta_1'} \\ P_n^{m^{o*}} &= \left(\sum_{t=1}^n E\left[R_t^{m^o\prime} \Omega_t^{*^{-1}} R_t^{m^o}\right]\right)^{-1} \sum_{t=1}^n E\left[R_t^{m^o\prime} \Omega_t^{*^{-1}} W_t^{m^{o*}}\right] \end{aligned}$$

Evaluated at $\hat{\theta}_{1_n}$ (and any given $\hat{\pi}_n = (\hat{\theta}_{2_n}', \hat{\gamma}_n')'$), $\mathcal{M}_n^{m^w}$ is identical to its standard $\mathcal{M}_n$-like counterpart — obtained by replacing $\hat{\Phi}_n^{m^w}$ by $\hat{\Phi}_n^m$ in (2.19) — and remains valid and asymptotically locally equivalent to it whatever the n-root consistent estimator used to estimate θ_1^o (and $\hat{\pi}_n^* = \left(\hat{\theta}_{2_n}^{*\prime}, \hat{\gamma}_n^{*\prime}\right)'$). Because the asymptotic behavior of $\mathcal{M}_n^{m^w}$ depends on the nuisance parameters pseudo-true values $\hat{\pi}_n^* = \left(\hat{\theta}_{2_n}^{*\prime}, \hat{\gamma}_n^{*\prime}\right)'$,

[2] Note that changing the sign of $\hat{R}_t^m$ does not affect the product $\hat{R}_t^m \hat{P}_n^m$.

remark that not using the GRPML2 estimator $\hat{\theta}_n = \left(\hat{\theta}_{1_n}', \hat{\theta}_{2_n}'\right)'$ will however usually yield a different — although valid — test statistic since, if the conditional variance $\{\Omega_t(X_t, \theta_2)\}$ is misspecified, $\hat{\theta}_{2_n}^*$ itself depends on the chosen alternative estimator used for its estimation. The same is obviously true for the other nuisance parameters vector γ.

In the general case, obtaining a consistent estimator of $K_n^{m^{o*}}$ entails the same difficulties than those evoked in Chapter 1 regarding consistent estimation of the variance B_n^* of the score of RPML2 estimators. We shall not discuss them here. As in Chapter 1, we refer the reader to White (1994), Wooldridge (1994) or Pötscher-Prucha (1997) for both a general discussion and references.

Getting a consistent estimator of $K_n^{m^{o*}}$ is considerably simplified if model $\mathcal{S}$ is dynamically complete for conditional mean (first order dynamically complete), i.e., if we have that

$$E(Y_t|X_t) = E(Y_t|X_t, \Psi_{t-1}), \quad t = 1, 2, ... \tag{2.20}$$

where $\Psi_{t-1} \equiv (Y_{t-1}, X_{t-1}, ..., Y_1, X_1)$ is the information available at time $t-1$. When (2.20) holds, under H_0^m, u_t^o is a martingale difference sequence with respect to $\{\Psi_t\}$, and thus so does $\xi_t^{m^{o*}}$, so that $\xi_t^{m^{o*}}$ is uncorrelated with its past values. $K_n^{m^{o*}}$ then collapses to

$$\overline{K}_n^{m^{o*}} = \frac{1}{n}\sum_{t=1}^{n} E\left[\xi_t^{m^{o*}} \xi_t^{m^{o*}\prime}\right]$$

and, under usual regularity conditions, a consistent estimator of it is simply given by

$$\frac{1}{n}\sum_{t=1}^{n} \hat{\xi}_t^m \hat{\xi}_t^{m\prime} \tag{2.21}$$

The auxiliary assumption (2.20) trivially holds if the observations are independent across t as in cross-section or panel data. Note that whenever (2.20) holds, $\mathcal{M}_n^{m^w}$ may be computed as n minus the residual sum of squares ($= nR_u^2$, R_u^2 being the uncentered R-squared) of the OLS regression

$$1 = \left[\hat{u}_t'\hat{\Omega}_t^{-1/2}\hat{\Omega}_t^{-1/2}\left(\hat{W}_t^m - \hat{R}_t^m\hat{P}_n^m\right)\underline{\hat{S}}_n'\right] b + \text{residuals}, \quad t = 1, 2, ...$$

where the $\hat{\Omega}_t^{-1/2}\left(\hat{W}_t^m - \hat{R}_t^m\hat{P}_n^m\right)\underline{\hat{S}}_n'$ may themselves be computed as the $G \times p$ matrix residuals of the OLS multivariate regression

$$\hat{\Omega}_t^{-1/2}\hat{W}_t^m\underline{\hat{S}}_n' = \left[\hat{\Omega}_t^{-1/2}\hat{R}_t^m\right] P + \text{residuals}, \quad t = 1, 2, ...$$

This way of computing $\mathcal{M}_n^{m^w}$ is particularly convenient in the univariate case. Its second step is however is less appealing in the multivariate case, unless $\hat{\Omega}_t^{-1/2}$ has a simple form.

2.5.1. Testing against nested alternatives

The most popular way to perform specification testing is to embed the model of interest in a more general auxiliary model in such a way that the former appears as a special case of latter when some parameter restrictions hold. The adequacy of the null model may then be assessed by checking through a Lagrange Multiplier (LM) or score type test if these restrictions are congruent with the data. The classical LM approach to misspecification testing is extensively treated in Godfrey (1988). The general model is labelled 'auxiliary' in the sense that it is usually instrumental: it is selected in the hope of obtaining reasonable power against departures from the null which are in its 'direction' or 'neighborhood'. It is only used as an auxiliary nested alternative.

A convenient and useful form for such an auxiliary nested alternative to the null conditional mean specification is

$$\mathrm{H}_1^{m\prime} : E\left(Y_t|X_t\right) = m_t^a(X_t, \theta_1^o, \alpha^o) \text{ for some } \left(\theta_1^{o\prime}, \alpha^{o\prime}\right)' \in \Theta_1 \times \Theta_\alpha, \quad t = 1, 2, ...$$

where α is a $k_\alpha \times 1$ vector of auxiliary mean parameters, and for some constant vector $c \in \Theta_\alpha$, we have

$$m_t^a(X_t, \theta_1, c) = m_t(X_t, \theta_1), \quad t = 1, 2, ...$$

Testing the null H_0^m against H_1^m using the auxiliary alternative $\mathrm{H}_1^{m\prime}$ now means testing the null that $\alpha^o = c$. Based on the GRPML2 estimator — actually on any RPML2 estimator —, a LM-type test of H_0^m against the auxiliary alternative $\mathrm{H}_1^{m\prime}$ yields the misspecification indicator

$$\begin{aligned}
\hat{\Phi}_n &= \frac{\partial L_n^a(Y^n, X^n, \hat{\theta}_{1n}, c, \hat{\theta}_{2_n})}{\partial \alpha} \\
&= \frac{1}{n} \sum_{t=1}^{n} \frac{\partial m_t^a(X_t, \hat{\theta}_{1_n}, c)'}{\partial \alpha} \Omega_t(X_t, \hat{\theta}_{2_n})^{-1} \left(Y_t - m_t^a(X_t, \hat{\theta}_{1_n}, c)\right) \\
&= \frac{1}{n} \sum_{t=1}^{n} \frac{\partial m_t^a(X_t, \hat{\theta}_{1_n}, c)'}{\partial \alpha} \Omega_t(X_t, \hat{\theta}_{2_n})^{-1} \left(Y_t - m_t(X_t, \hat{\theta}_{1_n})\right) \\
&= \frac{1}{n} \sum_{t=1}^{n} \frac{\partial m_t^a(X_t, \hat{\theta}_{1_n}, c)'}{\partial \alpha} \Omega_t(X_t, \hat{\theta}_{2_n})^{-1} u_t(Y_t, X_t, \hat{\theta}_{1_n})
\end{aligned}$$

where $L_n^a(Y^n, X^n, \theta_1, \alpha, \theta_2) = \frac{1}{n} \sum_{t=1}^{n} \ln f\left(Y_t, m_t^a(X_t, \theta_1, \alpha), \Omega_t\left(X_t, \theta_2\right)\right).$

$\hat{\Phi}_n$ may then be checked through the $\mathcal{M}_n^{m^w}$ statistic (2.19) by setting $\underline{\hat{S}}_n = I_q$, $\hat{W}_t^m = \frac{\partial m_t^a(X_t, \hat{\theta}_{1_n}, c)}{\partial \alpha'}$ and $p = q = k_\alpha$. The obtained test statistic is valid regardless of distributional and/or conditional variance misspecification and may in addition be implemented using any n-root consistent estimator of the parameters θ_1 and θ_2. Thus, if we except standard regularity conditions, the validity of the test requires no more than just the null hypothesis of interest H_0^m. Using a plausible specification for the second order conditional moments is just a way (or an attempt) to boost the power of the test. Under correct specification of the conditional variance, the test is asymptotically equivalent to the (non-robust to second order misspecification) Engle's (1982,1984) classical LM testing procedures. Note finally that $\alpha^o = c$ is allowed to be on the boundary of its parameter space Θ_α.

2.5.2. Testing against non-nested alternatives

Rather than testing the null H_0^m against an auxiliary nested alternative, we may wish to test it against a non-nested auxiliary model. Let such an auxiliary non-nested alternative to the null conditional mean specification be

$$\text{H}_1^{m\prime} : E\left(Y_t|X_t\right) = \mu_t(X_t, \beta^o) \text{ for some } \beta^o \in \Theta_\beta, \quad t = 1, 2, ...$$

where β is a $k_\beta \times 1$ vector of parameters, and suppose that some n-root consistent estimator $\hat{\beta}_n$ of β^o under $\text{H}_1^{m\prime}$ is available. It may be, but need not to be, a GRPML2 estimator. Provided that usual regularity conditions hold, under H_0^m, $\hat{\beta}_n$ will converge to some non-stochastic sequence of $k_\beta \times 1$ vectors of pseudo-true values $\{\beta_n^* : n = 1, 2, ...\}$.

Non-nested hypotheses testing may be performed along different lines (see for example Gourieroux-Monfort (1989)). This simplest and most popular one is due to Davidson-Mackinnon (1981). Its basic idea is to transform the non-nested hypotheses testing problem into a nested one by resorting to an artificial compound model. Consider the auxiliary artificially nested alternative

$$\text{H}_1^{m\prime\prime} : E\left(Y_t|X_t\right) = (1 - \lambda^o) m_t(X_t, \theta_1^o) + \lambda^o \mu_t(X_t, \beta^o) \text{ for some } \left(\theta_1^{o\prime}, \lambda^o, \beta^{o\prime}\right)' \in \Theta_a,$$

$t = 1, 2, ...$, where λ is a scalar parameter and $\Theta_a = \Theta_1 \times \Theta_\lambda \times \Theta_\beta$.

Now, similarly to the previous section, testing the null H_0^m against H_1^m using the artificial auxiliary alternative $\text{H}_1^{m\prime\prime}$ means testing the null that $\lambda^o = 0$. We however face a new problem: under H_0^m, β^o is not identified. Following Davidson-Mackinnon (1981), this may be overcome by using its consistent estimator $\hat{\beta}_n$ under $\text{H}_1^{m\prime}$. Putting this trick into service, based on the GRPML2 estimator, a LM-like test of H_0^m against the artificial auxiliary alternative $\text{H}_1^{m\prime\prime}$ yields the scalar misspecification indicator

$$\begin{aligned}
\hat{\Phi}_n &= \frac{\partial L_n^a(Y^n, X^n, \hat{\theta}_{1n}, 0, \hat{\beta}_n, \hat{\theta}_{2_n})}{\partial \lambda} \\
&= \frac{1}{n} \sum_{t=1}^{n} \frac{\partial m_t^a(X_t, \hat{\theta}_{1n}, 0, \hat{\beta}_n)'}{\partial \lambda} \Omega_t(X_t, \hat{\theta}_{2_n})^{-1} \left(Y_t - m_t^a(X_t, \hat{\theta}_{1n}, 0, \hat{\beta}_n)\right) \\
&= \frac{1}{n} \sum_{t=1}^{n} \left(\mu_t(X_t, \hat{\beta}_n) - m_t(X_t, \hat{\theta}_{1_n})\right)' \Omega_t(X_t, \hat{\theta}_{2_n})^{-1} \left(Y_t - m_t(X_t, \hat{\theta}_{1_n})\right) \\
&= \frac{1}{n} \sum_{t=1}^{n} \left(\mu_t(X_t, \hat{\beta}_n) - m_t(X_t, \hat{\theta}_{1_n})\right)' \Omega_t(X_t, \hat{\theta}_{2_n})^{-1} u_t(Y_t, X_t, \hat{\theta}_{1_n})
\end{aligned}$$

where $L_n^a(Y^n, X^n, \theta_1, \lambda, \beta, \theta_2) = \frac{1}{n} \sum_{t=1}^{n} \ln f\left(Y_t, m_t^a(X_t, \theta_1, \lambda, \beta), \Omega_t\left(X_t, \theta_2\right)\right)$ with $m_t^a(X_t, \theta_1, \lambda, \beta) = (1 - \lambda) m_t(X_t, \theta_1) + \lambda \mu_t(X_t, \beta)$.

As above, $\hat{\Phi}_n$ may be checked through the $\mathcal{M}_n^{m^w}$ statistic (2.19) by setting $\underline{\hat{S}}_n = 1$, $\hat{W}_t^m = \mu_t(X_t, \hat{\beta}_n) - m_t(X_t, \hat{\theta}_{1_n})$ and $p = q = 1$. Likewise, the test statistic is robust to distributional and/or conditional variance misspecification and may be implemented using any n-root consistent estimator of θ_1, θ_2 and β. Note that $\hat{\Phi}_n$ is the misspecification indicator which naturally arises from the Cox (1961,1962) procedure for testing gaussian models with non-nested mean and identical variance (see White (1994)).

2.5.3. Testing without explicit alternatives: Hausman and information matrix type tests

Relevant auxiliary alternatives to H_0^m are not always obvious. On the other hand, we might wish to test H_0^m without resorting to such explicit alternatives, in hope of getting power against a broader, less targeted, spectrum of departures from the null. The standard way to do that is based on the Hausman (1978) approach to specification testing. It may also be done by resorting to White (1982) information matrix type tests.

2.5.3.1. Hausman type tests

If $\mathcal{S}$ is correctly specified for the conditional mean, two different consistent estimators of θ_1^o, say $\hat{\theta}_{1_n}$ and an other n-root consistent estimator $\underline{\hat{\theta}}_{1_n}$, should give about the same results. If they do not, then misspecification is evident. This suggests that a test of H_0^m may be based on a misspecification indicator of the form

$$\hat{\Phi}_n = S(\hat{\theta}_{1_n} - \underline{\hat{\theta}}_{1_n}) \tag{2.22}$$

where the $p \times k_{\theta_1}$ ($p \leq k_{\theta_1}$) non-stochastic selection matrix S allows to focus on particular elements or linear combinations of elements of $(\hat{\theta}_{1_n} - \underline{\hat{\theta}}_{1_n})$. A test based on $\hat{\Phi}_n$ will have power against any alternative H_1^m for which $\hat{\theta}_{1_n}$ and $\underline{\hat{\theta}}_{1_n}$ converge to different pseudo-true values.

A natural candidate for $\underline{\hat{\theta}}_{1_n}$ is the GRPML2 mean estimator — or equivalently a QGPML1 estimator — of the model

$$\underline{\mathcal{S}} \equiv \begin{cases} \{m_t(X_t, \theta_1) : \ \theta_1 \in \Theta_1\} \\ \{\Sigma_t(X_t, \delta) : \ \delta \in \Theta_\delta\} \end{cases}, \ t = 1, 2, ...$$

where the $G \times G$ (symmetric positive definite) matrix functions Σ_t are alternative specifications for the conditional variances $V(Y_t|X_t)$ and δ is a $k_\delta \times 1$ vector of parameters which varies independently of θ_1. Let $\{\delta_n^* : n = 1, 2, ...\}$ denote the non-stochastic sequence of vectors of pseudo-true values to which the GRPML2 — or the auxiliary variance parameters estimator associated to QGPML1 — n-root consistent estimator $\hat{\delta}_n$ converges under H_0^m.

The misspecification indicator (2.22) is not precisely of our standard form (2.17). However, following White (1982,1994) (see also Ruud (1984)), a misspecification indicator of the form (2.17) yielding a test asymptotically equivalent to the one which could directly be obtained from (2.22) may be derived.

According to (2.3), the score associated to $\underline{\hat{\theta}}_{1_n}$ is given by

$$\frac{1}{n}\sum_{t=1}^{n} \underline{s}_t^1(\underline{\hat{\theta}}_{1_n}, \hat{\delta}_n) = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial m_t(X_t, \underline{\hat{\theta}}_{1_n})'}{\partial \theta_1} \Sigma_t(X_t, \hat{\delta}_n)^{-1} u_t(Y_t, X_t, \underline{\hat{\theta}}_{1_n}) = 0 \qquad (2.23)$$

Now, consider evaluating the score (2.23) associated to $\underline{\hat{\theta}}_{1_n}$ at $\hat{\theta}_{1_n}$. Under H_0^m and usual regularity conditions, using standard arguments, a mean value expansion of $n^{-1/2}\sum_{t=1}^{n} \underline{s}_t^1(\hat{\theta}_{1_n}, \hat{\delta}_n)$ at $\left(\underline{\hat{\theta}}_{1_n}', \hat{\delta}_n'\right)'$ gives

$$n^{-1/2}\sum_{t=1}^{n} \underline{s}_t^1(\hat{\theta}_{1_n}, \hat{\delta}_n) = n^{-1/2}\sum_{t=1}^{n} \underline{s}_t^1(\underline{\hat{\theta}}_{1_n}, \hat{\delta}_n) + \underline{A}_{n_{11}}^* \sqrt{n}(\hat{\theta}_{1_n} - \underline{\hat{\theta}}_{1_n}) + o_{P_o}(1)$$

where

$$\underline{A}_{n_{11}}^* = \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial \underline{s}_t^1(\theta_1^o, \delta_n^*)}{\partial \theta_1'}\right] = -\frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1} \Sigma_t^{*-1} \frac{\partial m_t^o}{\partial \theta_1'}\right], \quad \Sigma_t^* = \Sigma_t(X_t, \delta_n^*)$$

or, given (2.23),

$$\sqrt{n}(\hat{\theta}_{1_n} - \underline{\hat{\theta}}_{1_n}) = \underline{A}_{n_{11}}^{*^{-1}} n^{-1/2}\sum_{t=1}^{n} \underline{s}_t^1(\hat{\theta}_{1_n}, \hat{\delta}_n) + o_{P_o}(1)$$

and further,

$$\sqrt{n}(\hat{\theta}_{1_n} - \underline{\hat{\theta}}_{1_n}) = \underline{\hat{A}}_{n_{11}}^{-1} n^{-1/2}\sum_{t=1}^{n} \underline{s}_t^1(\hat{\theta}_{1_n}, \hat{\delta}_n) + o_{P_o}(1) \qquad (2.24)$$

where

$$\underline{\hat{A}}_{n_{11}} = -\frac{1}{n}\sum_{t=1}^{n} \frac{\partial m_t(X_t, \hat{\theta}_{1_n})'}{\partial \theta_1} \Sigma_t(X_t, \hat{\delta}_n)^{-1} \frac{\partial m_t(X_t, \hat{\theta}_{1_n})}{\partial \theta_1'}$$

According to (2.22) and (2.24), we thus have

$$\sqrt{n}\hat{\Phi}_n = S\sqrt{n}(\hat{\theta}_{1_n} - \underline{\hat{\theta}}_{1_n}) = S\underline{\hat{A}}_{n_{11}}^{-1} n^{-1/2}\sum_{t=1}^{n} \underline{s}_t^1(\hat{\theta}_{1_n}, \hat{\delta}_n) + o_{P_o}(1)$$

In other words, a test based on $\hat{\Phi}_n$ may, from an asymptotic point of view, equivalently be based on the misspecification indicator

$$\begin{aligned}\hat{\Phi}_n^h &= S\underline{\hat{A}}_{n_{11}}^{-1}\frac{1}{n}\sum_{t=1}^{n} \frac{\partial m_t(X_t, \hat{\theta}_{1_n})'}{\partial \theta_1} \Sigma_t(X_t, \hat{\delta}_n)^{-1} u_t(Y_t, X_t, \hat{\theta}_{1_n}) \qquad (2.25)\\ &= S\underline{\hat{A}}_{n_{11}}^{-1}\frac{1}{n}\sum_{t=1}^{n} \frac{\partial m_t(X_t, \hat{\theta}_{1_n})'}{\partial \theta_1} \Sigma_t(X_t, \hat{\delta}_n)^{-1}\Omega_t(X_t, \hat{\theta}_{2_n})\Omega_t(X_t, \hat{\theta}_{2_n})^{-1} u_t(Y_t, X_t, \hat{\theta}_{1_n})\end{aligned}$$

$\hat{\Phi}_n^h$ may here be checked through the $\mathcal{M}_n^{m^w}$ statistic (2.19) by setting $\underline{\hat{S}}_n = S\left(\frac{1}{n}\sum_{t=1}^n \frac{\partial m_t(X_t,\hat{\theta}_{1n})'}{\partial\theta_1}\Sigma_t(X_t,\hat{\delta}_n)^{-1}\frac{\partial m_t(X_t,\hat{\theta}_{1n})}{\partial\theta_1'}\right)^{-1}$ and $\hat{W}_t^m = \Omega_t(X_t,\hat{\theta}_{2_n})\Sigma_t(X_t,\hat{\delta}_n)^{-1}$ $\frac{\partial m_t(X_t,\hat{\theta}_{1n})}{\partial\theta_1'}$. As all $\mathcal{M}_n^{m^w}$ statistic, the validity of this Hausman type test requires no more than just the null hypothesis of interest H_0^m and, although quite paradoxical for such a type of test, it may be implemented using any n-root consistent estimator of the parameters θ_1, θ_2 and δ. The test statistic is asymptotically equivalent to comparing (some linear combination S of) two multivariate weighted nonlinear least squares (MWNLS) estimators of θ_1^o, one with weights $\{\Omega_t(X_t,\theta_{2_n}^*)^{-1}\}$ and the other with weights $\{\Sigma_t(X_t,\delta_n^*)^{-1}\}$. If the $\Sigma_t(X_t,\delta)$ are set equal to I_G, $\underline{\hat{\theta}}_{1_n}$ is just the standard (unweighted) multivariate nonlinear least squares estimator of θ_1^o. In this case, the implementation of the test requires no additional estimators than the ones needed for estimating the null model. Note finally that if the selection matrix S is set equal to $I_{k_{\theta_1}}$(or any other non-singular square matrix), then the entire term $\underline{\hat{S}}_n$ may be dropped from the statistic (where now $p = q = k_{\theta_1}$) without affecting it.

2.5.3.2. Information matrix type tests

In Chapter 1 (Proposition 10), we saw that correct specification of the conditional mean implies the block-diagonality between mean and variance parameters of the expected hessian of RPML2 estimators, i.e., that

$$E\left[\frac{\partial^2 L_n(Y^n,X^n,\theta_n^*)}{\partial\theta_1\partial\theta_2'}\right] = 0, \quad \theta_n^* = \left(\theta_1^{o\prime},\theta_{2_n}^{*\prime}\right)' \tag{2.26}$$

a feature which ensures that the asymptotic distribution of $\hat{\theta}_{1_n}$ does not depend on the fact that $\theta_{2_n}^*$ is estimated, and conversely.

This suggests that a test of H_0^m might be based on checking that the empirical counterpart of (2.26) is indeed close to zero. Let θ_2^r denote the r-th component of θ_2. Each of the k_{θ_2} columns ($r = 1,...,k_{\theta_2}$) of $\frac{\partial^2}{\partial\theta_1\partial\theta_2'}L_n(Y^n,X^n,\theta)$ is given by

$$\begin{aligned}
&\frac{\partial^2 L_n(Y^n,X^n,\theta)}{\partial\theta_1\partial\theta_2^r}\\
=\;& \frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial\theta_2^r}\left[\frac{\partial m_t(X_t,\theta_1)'}{\partial\theta_1}\Omega_t(X_t,\theta_2)^{-1}u_t(Y_t,X_t,\theta_1)\right]\\
=\;& \frac{1}{n}\sum_{t=1}^n \frac{\partial m_t(X_t,\theta_1)'}{\partial\theta_1}\frac{\partial\Omega_t(X_t,\theta_2)^{-1}}{\partial\theta_2^r}u_t(Y_t,X_t,\theta_1) \qquad (2.27)\\
=\;& -\frac{1}{n}\sum_{t=1}^n \frac{\partial m_t(X_t,\theta_1)'}{\partial\theta_1}\Omega_t(X_t,\theta_2)^{-1}\frac{\partial\Omega_t(X_t,\theta_2)}{\partial\theta_2^r}\Omega_t(X_t,\theta_2)^{-1}u_t(Y_t,X_t,\theta_1)
\end{aligned}$$

Letting the $G \times k_{\theta_1} k_{\theta_2}$ matrix functions $F_t(X_t, \theta_1, \theta_2)$ be defined as

$$F_t(X_t, \theta_1, \theta_2)' = \begin{bmatrix} \frac{\partial m_t(X_t,\theta_1)'}{\partial \theta_1} \Omega_t(X_t, \theta_2)^{-1} \frac{\partial \Omega_t(X_t,\theta_2)}{\partial \theta_2^1} \\ \frac{\partial m_t(X_t,\theta_1)'}{\partial \theta_1} \Omega_t(X_t, \theta_2)^{-1} \frac{\partial \Omega_t(X_t,\theta_2)}{\partial \theta_2^2} \\ \vdots \\ \frac{\partial m_t(X_t,\theta_1)'}{\partial \theta_1} \Omega_t(X_t, \theta_2)^{-1} \frac{\partial \Omega_t(X_t,\theta_2)}{\partial \theta_2^{k_{\theta_2}}} \end{bmatrix}, \quad t = 1, 2, ...$$

a relevant misspecification indicator for checking (2.26) may be written

$$\hat{\Phi}_n = S \frac{1}{n} \sum_{t=1}^{n} F_t(X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n})' \Omega_t(X_t, \hat{\theta}_{2_n})^{-1} u_t(Y_t, X_t, \hat{\theta}_{1_n})$$

where the $p \times k_{\theta_1} k_{\theta_2}$ $(p \leq k_{\theta_1} k_{\theta_2})$ non-stochastic selection matrix S allows to focus on particular elements or linear combinations of elements of the second term which is equal to $\text{vec}(-\frac{\partial^2}{\partial \theta_1 \partial \theta_2'} L_n(Y^n, X^n, \hat{\theta}_n))$.

$\hat{\Phi}_n$ may then be checked through the $\mathcal{M}_n^{m^w}$ statistic (2.19) by setting $\underline{\hat{S}}_n = S$ and $\hat{W}_t^m = F_t(X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n})$. As usual, only H_0^m and n-root consistent estimators of the parameters θ_1 and θ_2 are required for the test to be valid. The test statistic admits a simple interpretation. Comparing (2.27) and (2.25), it is readily seen that for $S = I_{(k_{\theta_1} \times k_{\theta_2})}$ it simply amounts to jointly performing k_{θ_2} Hausman type tests, each of which being asymptotically equivalent to comparing (with $S = I_{k_{\theta_1}}$) the MWNLS of θ_1^o with weights $\left\{\Omega_t(X_t, \theta_{2_n}^*)^{-1}\right\}$ and the MWNLS of θ_1^o with "endogeneously determined" weights $\left\{\frac{\partial}{\partial \theta_2^\tau}\left[\Omega_t(X_t, \theta_{2_n}^*)^{-1}\right]\right\}$. The test statistic will thus have power against any alternative H_1^m for which at least one of these $k_{\theta_2}+1$ estimators converge to a pseudo-true value different from the one of the others. Note that such a test can not always be undertaken. This is for example the case if the conditional variance is specified as $\{\Omega_t(X_t, \theta_2) = \sigma^2 I_G\}$.

2.5.4. Testing dynamic completeness

We saw in Chapter 1 (Proposition 12) that making inference about the mean parameters — which implies getting a consistent estimator of their asymptotic covariance matrix — of a first order correctly specified semi-parametric model $\mathcal{S}$ estimated by RPML2 is considerably simplified if the conditional mean is also dynamically complete. Accordingly, we here concentrate on testing the null

$$\mathrm{H}_0^{md} : \mathrm{H}_0^m \text{ holds and } E\left(Y_t | X_t\right) = E\left(Y_t | X_t, \Psi_{t-1}\right), \quad t = 1, 2, ...$$

where, for the record, $\Psi_{t-1} \equiv (Y_{t-1}, X_{t-1}, ..., Y_1, X_1)$ is the information available at time $t-1$.

As briefly discussed in Section 2.3 for the conditional variance, given the nested nature of H_0^m and H_0^{md}, according to Wooldridge's (1991a) sequential "bottom-up" model construction/specification testing strategy, it seems sensible to emphasize the construction of statistics that have power against the alternative

$$\mathrm{H}_1^{md} : \mathrm{H}_0^m \text{ holds but } \mathrm{H}_0^{md} \text{ is false}$$

H_0^{md} is equivalent to

$$\mathrm{H}_0^{md} : E(Y_t|X_t, \Psi_{t-1}) = m_t(X_t, \theta_1^o) \text{ for some } \theta_1^o \in \Theta_1, \quad t = 1, 2, ...$$

From a testing point of view, H_0^{md} is not conceptually different from the H_0^m : it simply enlarges the information set with respect to which the conditional mean $\{m_t(X_t, \theta_1)\}$ is assumed to be correctly specified. We may then proceed exactly in the same way than above for testing H_0^m. The only difference is that we may now unambiguously — i.e., without relying on more assumptions than just the null hypothesis of interest — take advantage of the fact that under H_0^{md} the simple estimator (2.21) is consistent for the asymptotic covariance matrix $K_n^{m^{o*}}$.

The most general ways to check dynamic completeness of the conditional mean are either to look at autocorrelation in the errors u_t or to resort to a White (1987,1994) dynamic information matrix type test.

Looking at a multivariate AR(κ) process for u_t means using as an auxiliary nested alternative to the null conditional mean specification

$$\mathrm{H}_1^{md\prime} : E(Y_t|X_t, \Psi_{t-1}) = m_t(X_t, \theta_1^o) + D_1^o u_{t-1}^o + ... + D_\kappa^o u_{t-\kappa}^o \text{ for some } a^o \in \Theta_a,$$

$t = 1, 2, ...$, where $\kappa \geq 1$ is a integer that determines the maximum autocorrelation of u_t to be examined, $a^o = (\theta_1^{o\prime}, (\mathrm{vec}\, D_1^o)', ..., (\mathrm{vec}\, D_\kappa^o)')'$, $\Theta_a = \Theta_1 \times \Theta_{D_1} \times ... \times \Theta_{D_\kappa}$ and the D_i, $i = 1, ..., \kappa$, are $G \times G$ matrices of auxiliary mean parameters. Let for now $t = 1$ denote the $(\kappa + 1)$-th observation and define $n_\kappa = n - \kappa$.

As in Section 2.5.1, testing the null H_0^{md} against H_1^{md} using the auxiliary alternative $\mathrm{H}_1^{md\prime}$ means testing the null that $D_1^o = ... = D_\kappa^o = 0$. Based on the GRPML2 estimator, a LM-type test yields the misspecification indicator

$$\begin{aligned}
\hat{\Phi}_{n_\kappa}^{AR} &= \frac{\partial L_{n_\kappa}^a(Y^n, X^n, \hat{\theta}_{1n}, 0, ..., 0, \hat{\theta}_{2_n})}{\partial\left((\mathrm{vec}\, D_1)', ..., (\mathrm{vec}\, D_\kappa)'\right)'} \\
&= \frac{1}{n_\kappa}\sum_{t=1}^{n_\kappa} \frac{\partial m_t^a(X_t, \Psi_{t-1}, \theta_1, 0, ..., 0)'}{\partial\left((\mathrm{vec}\, D_1)', ..., (\mathrm{vec}\, D_\kappa)'\right)'} \Omega_t(X_t, \hat{\theta}_{2_n})^{-1}\left(Y_t - m_t^a(X_t, \Psi_{t-1}, \theta_1, 0, ..., 0)\right) \\
&= \frac{1}{n_\kappa}\sum_{t=1}^{n_\kappa} F_t^{AR}(X_t, \Psi_{t-1}, \hat{\theta}_{1_n})'\Omega_t(X_t, \hat{\theta}_{2_n})^{-1}\left(Y_t - m_t(X_t, \hat{\theta}_{1_n})\right) \\
&= \frac{1}{n_\kappa}\sum_{t=1}^{n_\kappa} F_t^{AR}(X_t, \Psi_{t-1}, \hat{\theta}_{1_n})'\Omega_t(X_t, \hat{\theta}_{2_n})^{-1} u_t(Y_t, X_t, \hat{\theta}_{1_n})
\end{aligned}$$

where $L_{n_\kappa}^a(Y^n, X^n, \theta_1, D_1, ..., D_\kappa, \theta_2) = \frac{1}{n_\kappa}\sum_{t=1}^{n_\kappa} \ln f(Y_t, m_t^a(X_t, \Psi_{t-1}, \theta_1, D_1, ..., D_\kappa),$ $\Omega_t(X_t, \theta_2))$ with $m_t^a(X_t, \Psi_{t-1}, \theta_1, D_1, ..., D_\kappa) = m_t(X_t, \theta_1) + D_1 u_{t-1} + ... + D_\kappa u_{t-\kappa}$ and the $G \times \kappa G^2$ matrix functions $F_t^{AR}(X_t, \Psi_{t-1}, \theta_1)$ are defined as

$$F_t^{AR}(X_t, \Psi_{t-1}, \theta_1) = \left(u_{t-1}', ..., u_{t-\kappa}'\right) \otimes I_G, \quad t = 1, 2, ...$$

with $u_t = u_t(Y_t, X_t, \theta_1)$.

As already outlined, under H_0^{md}, u_t^o is a martingale difference sequence with respect to $\{\Psi_t\}$, and thus so does the score s_t^{1*}, so that s_t^{1*} is uncorrelated with its past values. Accordingly, for all $\kappa \geq 1$, we must have

$$E\left[s_t^{1*}\left(s_{t-1}^{1*\prime}, ..., s_{t-\kappa}^{1*\prime}\right)\right] = 0, \quad t = 1, 2, ... \tag{2.28}$$

Then, choosing some integer κ and vectorizing (2.28), a test of H_0^{md} may alternatively be based on the misspecification indicator

$$\hat{\Phi}_{n_\kappa}^{IM} = \frac{1}{n_\kappa}\sum_{t=1}^{n_\kappa} F_t^{IM}(X_t, \Psi_{t-1}, \hat{\theta}_{1_n}, \hat{\theta}_{2_n})'\Omega_t(X_t, \hat{\theta}_{2_n})^{-1}u_t(Y_t, X_t, \hat{\theta}_{1_n})$$

where the $G \times \kappa k_{\theta_1}^2$ matrix functions $F_t^{IM}(X_t, \Psi_{t-1}, \theta_1, \theta_2)$ are defined as

$$F_t^{IM}(X_t, \Psi_{t-1}, \theta_1, \theta_2) = \frac{\partial m_t(X_t, \theta_1)}{\partial \theta_1'}\left(\left(s_{t-1}^{1\prime}, ..., s_{t-\kappa}^{1\prime}\right) \otimes I_{k_{\theta_1}}\right), \quad t = 1, 2, ...$$

with $s_t^1 = \frac{\partial m_t(X_t,\theta_1)'}{\partial \theta_1}\Omega_t(X_t, \theta_2)^{-1}u_t(Y_t, X_t, \theta_1)$.

Both $\hat{\Phi}_{n_\kappa}^{AR}$ and $\hat{\Phi}_{n_\kappa}^{IM}$ may be checked through the $\mathcal{M}_n^{m^w}$ statistic (2.19) by setting for $\hat{\Phi}_{n_\kappa}^{AR}$, $\hat{W}_t^m = F_t^{AR}(X_t, \Psi_{t-1}, \hat{\theta}_{1_n})$ and $p = q = \kappa G^2$, and for $\hat{\Phi}_{n_\kappa}^{IM}$, $\hat{W}_t^m = F_t^{IM}(X_t, \Psi_{t-1}, \hat{\theta}_{1_n}, \hat{\theta}_{2_n})$ and $p = q = \kappa k_{\theta_1}^2$. In both cases, $\underline{\hat{S}}_n = I_p$ — if wished, a selection matrix may straightforwardly be introduced —, $n = n_\kappa$, $t = 1$ denotes the $(\kappa+1)$-th observation and $\hat{K}_n^m$ is the simple estimator (2.21). As usual, the test statistics are robust to distributional and/or conditional variance misspecification and may be implemented using any n-root consistent estimator of θ_1, θ_2. The choice between using $\hat{\Phi}_{n_\kappa}^{AR}$ or $\hat{\Phi}_{n_\kappa}^{IM}$ may be done on the grounds of computational convenience but should also take into account their relative degree of freedom κG^2 and $\kappa k_{\theta_1}^2$. When both are very large, it may be wise to resort to a selection matrix.

2.6. Testing the conditional variance

We now turn our attention to testing the correct specification of the conditional variance. According to the Wooldridge's (1991a) sequential "bottom-up" model construction/specification testing strategy, this entails testing the null

$$\mathrm{H}_0^v : \mathrm{H}_0^m \text{ holds and } V\left(Y_t|X_t\right) = \Omega_t(X_t, \theta_2^o) \text{ for some } \theta_2^o \in \Theta_2, \quad t = 1, 2, ...$$

against the alternative

$$\mathrm{H}_1^v : \mathrm{H}_0^m \text{ holds but } \mathrm{H}_0^v \text{ is false}$$

As for conditional mean testing, hereafter, we outline misspecification indicators suitable for testing H_0^v against auxiliary nested alternatives, auxiliary non-nested alternatives, as well as for testing H_0^v without resorting to explicit alternatives. Misspecification indicators for testing the dynamic completeness of the conditional variance specification are also discussed.

All considered misspecification indicators are special cases — for various choices of $\underline{\hat{S}}_n$ and W_t^v — of

$$\underline{\hat{\Phi}}_n^v = \underline{\hat{S}}_n \hat{\Phi}_n^v \tag{2.29}$$

where

$$\hat{\Phi}_n^v = \frac{1}{n}\sum_{t=1}^{n} W_t^v(X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n}, \hat{\gamma}_n)'\Gamma_t(X_t, \hat{\theta}_{2_n})^{-1} v_t(Y_t, X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n}) \tag{2.30}$$

$v_t(Y_t, X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n}) = \text{vec}\left(u_t(Y_t, X_t, \hat{\theta}_{1_n}) u_t(Y_t, X_t, \hat{\theta}_{1_n})' - \Omega_t(X_t, \hat{\theta}_{2_n})\right)$, $\Gamma_t(X_t, \hat{\theta}_{2_n}) = \Omega_t(X_t, \hat{\theta}_{2_n}) \otimes \Omega_t(X_t, \hat{\theta}_{2_n})$ — note that $\Gamma_t^{-1} = \Omega_t^{-1} \otimes \Omega_t^{-1}$ —, the W_t^v are $G^2 \times q$ indicator matrix functions, $\hat{\theta}_{1_n}$ and $\hat{\theta}_{2_n}$ are the GRPML2 estimators of the null model $\mathcal{S}$, $\hat{\gamma}_n$ is, under H_0^v, a n-root consistent estimator of some non-stochastic sequence of $k_\gamma \times 1$ vectors of pseudo-true values $\{\gamma_n^* : n = 1, 2, ...\}$ and $\underline{\hat{S}}_n$ is a stochastic $p \times q$ $(p \leq q)$ selection matrix converging in probability under H_0^v to some non-stochastic sequence $\{\underline{S}_n^* : n = 1, 2, ...\}$.

$\hat{\Phi}_n^v$ is obviously again a special case of (2.5) with $\hat{r}_t = \hat{v}_t = v_t(Y_t, X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n})$, $\hat{\Lambda}_t = \hat{\Gamma}_t = \Gamma_t(X_t, \hat{\theta}_{2_n})$, $\hat{W}_t = \hat{W}_t^v = W_t^v(X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n}, \hat{\gamma}_n)$, $\hat{\varphi}_n = \left(\hat{\theta}_{1_n}', \hat{\theta}_{2_n}'\right)'$, $\hat{\pi}_n = \hat{\gamma}_n$ and $l = G^2$. The matrices R_t of conditional expectations (2.9) are here equal to

$$R_t(X_t, \theta_1^o, \theta_2^o) = \left[R_t^1(X_t, \theta_1^o, \theta_2^o) \,\vdots\, R_t^2(X_t, \theta_1^o, \theta_2^o)\right]$$

where

$$R_t^1(X_t, \theta_1^o, \theta_2^o) = E\left[\left.\frac{\partial v_t\left(Y_t, X_t, \theta_1^o, \theta_2^o\right)}{\partial \theta_1'}\right| X_t\right] = 0 \tag{2.31}$$

$$R_t^2(X_t, \theta_1^o, \theta_2^o) = E\left[\left.\frac{\partial v_t\left(Y_t, X_t, \theta_1^o, \theta_2^o\right)}{\partial \theta_2'}\right| X_t\right] = -\frac{\partial\, \text{vec}\, \Omega_t(X_t, \theta_2^o)}{\partial \theta_2'} \tag{2.32}$$

(2.31) follows from

$$\frac{\partial v_t\left(Y_t, X_t, \theta_1, \theta_2\right)}{\partial \theta_1'} = \left(u_t(Y_t, X_t, \theta_1) \otimes I_G\right) + \left(I_G \otimes u_t(Y_t, X_t, \theta_1)\right) \frac{\partial m_t(X_t, \theta_1)}{\partial \theta_1'}$$

and the fact that, under H_0^v, $E(u_t^o | X_t) = 0$.

According to the reasoning underlying (2.6)-(2.10)-(2.11), (2.31) means that although θ_1 does appear in $\hat{v}_t$, the limiting distribution of $\sqrt{n}\hat{\Phi}_n^v$ under H_0^v does actually not depend on the one of $\hat{\theta}_{1_n}$. In other words, in terms of (2.5), we may proceed as if $\hat{\theta}_{1_n}$ were actually known and fixed at θ_1^o, and consequently redefine $\hat{\varphi}_n = \hat{\theta}_{2_n}$, $\hat{\pi}_n = \left(\hat{\theta}_{1_n}', \hat{\gamma}_n'\right)'$, i.e., consider $\hat{\theta}_{1_n}$ as a nuisance parameters vector.

Now, keeping this in mind — which implies that $R_t\left(X_t, \theta_1^o, \theta_2^o\right)$ is redefined as equal to $R_t^2\left(X_t, \theta_1^o, \theta_2^o\right)$ alone —, in view of (2.2) and (2.4), it is easily seen that the GRPML2 mean parameters estimator $\hat{\theta}_{2_n}$ satisfies both the first order expansion (2.12) and the first order conditions (2.13). As for conditional mean testing, an $\mathcal{M}_n^w$-like test statistic is thus the most appealing to look at. Collecting all this, applying the same trick than in Section 2.5 and using the general results of Section 2.4, it may be readily checked that the relevant $\mathcal{M}_n^w$-like statistic for testing H_0^v based on $\underline{\hat{\Phi}}_n^v$ is given by

$$\mathcal{M}_n^{v^w} = n\hat{\Phi}_n^{v^w\prime}\underline{\hat{S}}_n' \left[\underline{\hat{S}}_n\hat{K}_n^v\underline{\hat{S}}_n'\right]^{-1}\underline{\hat{S}}_n\hat{\Phi}_n^{v^w} \xrightarrow{d} \chi^2(p) \tag{2.33}$$

where

$$\hat{\Phi}_n^{v^w} = \frac{1}{n}\sum_{t=1}^n \hat{\xi}_t^v = \frac{1}{n}\sum_{t=1}^n \left(\hat{W}_t^v - \hat{R}_t^v\hat{P}_n^v\right)'\hat{\Gamma}_t^{-1}\hat{v}_t$$

with[3]

$$\hat{R}_t^v = \frac{\partial \operatorname{vec}\Omega_t(X_t,\hat{\theta}_{2_n})}{\partial\theta_2'}, \quad \hat{P}_n^v = \left(\sum_{t=1}^n \hat{R}_t^{v\prime}\hat{\Gamma}_t^{-1}\hat{R}_t^v\right)^{-1}\sum_{t=1}^n \hat{R}_t^{v\prime}\hat{\Gamma}_t^{-1}\hat{W}_t^v$$

and $\hat{K}_n^v$ is a consistent estimator of

$$K_n^{v^{o*}} = \frac{1}{n}\sum_{t=1}^n E\left[\xi_t^{v^{o*}}\xi_t^{v^{o*}\prime}\right] + \frac{1}{n}\sum_{\tau=1}^{n-1}\sum_{t=\tau+1}^n \left(E\left[\xi_t^{v^{o*}}\xi_{t-\tau}^{v^{o*}\prime}\right] + E\left[\xi_{t-\tau}^{v^{o*}}\xi_t^{v^{o*}\prime}\right]\right)$$

where

$$\xi_t^{v^{o*}} = \left(W_t^{v^{o*}} - R_t^{v^o}P_n^{v^{o*}}\right)'\Gamma_t^{o^{-1}}v_t^o$$

$$\begin{aligned}
\Gamma_t^o &= \Gamma_t(X_t,\theta_2^o), \qquad v_t^o = v_t(Y_t,X_t,\theta_1^o,\theta_2^o) \\
W_t^{v^{o*}} &= W_t^v\left(X_t,\theta_1^o,\theta_2^o,\hat{\gamma}_n^*\right), \quad R_t^{v^o} = \frac{\partial \operatorname{vec}\Omega_t(X_t,\theta_2^o)}{\partial\theta_2'} \\
P_n^{v^{o*}} &= \left(\sum_{t=1}^n E\left[R_t^{v^o\prime}\Gamma_t^{o^{-1}}R_t^{v^o}\right]\right)^{-1}\sum_{t=1}^n E\left[R_t^{v^o\prime}\Gamma_t^{o^{-1}}W_t^{v^{o*}}\right]
\end{aligned}$$

Evaluated at $\hat{\theta}_{2_n}$ (and any given $\hat{\pi}_n = (\hat{\theta}_{1_n}', \hat{\gamma}_n')'$), $\mathcal{M}_n^{v^w}$ is identical to its standard $\mathcal{M}_n$-like counterpart — obtained by replacing $\hat{\Phi}_n^{v^w}$ by $\hat{\Phi}_n^v$ in (2.33) — and remains valid and asymptotically locally equivalent to it whatever the n-root consistent estimator used to estimate θ_2^o (and $\hat{\pi}_n^{o*} = (\theta_1^o, \hat{\gamma}_n^{*\prime})'$).

Here, a sufficient auxiliary assumption for $K_n^{v^{o*}}$ to collapse to

$$\overline{K}_n^{v^{o*}} = \frac{1}{n}\sum_{t=1}^n E\left[\xi_t^{v^{o*}}\xi_t^{v^{o*}\prime}\right]$$

such that, under usual regularity conditions, the simple estimator

$$\frac{1}{n}\sum_{t=1}^n \hat{\xi}_t^v\hat{\xi}_t^{v\prime} \tag{2.34}$$

is a consistent estimator of it, is that model $\mathcal{S}$ be second order dynamically complete,

[3] As in Section 2.5, changing the sign of $\hat{R}_t^v$ does not affect the product $\hat{R}_t^v\hat{P}_n^v$.

i.e., that we have

$$E(Y_t|X_t) = E(Y_t|X_t, \Psi_{t-1}) \text{ and } V(Y_t|X_t) = V(Y_t|X_t, \Psi_{t-1}), \quad t = 1, 2, ... \quad (2.35)$$

Indeed, when (2.35) holds — it trivially does if the observations are independent across t as in cross-section or panel data —, under H_0^v, v_t^o is a martingale difference sequence with respect to $\{\Psi_t\}$, and thus so does $\xi_t^{v^{o*}}$, so that $\xi_t^{v^{o*}}$ is uncorrelated with its past values. In this case, as above, $\mathcal{M}_n^{v^w}$ may be computed as n minus the residual sum of squares ($= nR_u^2$, R_u^2 being the uncentered R-squared) of the OLS regression

$$1 = \left[\hat{v}_t'\hat{\Gamma}_t^{-1/2}\hat{\Gamma}_t^{-1/2}\left(\hat{W}_t^v - \hat{R}_t^v\hat{P}_n^v\right)\underline{\hat{S}}_n'\right] b + \text{residuals}, \quad t = 1, 2, ...$$

where the $\hat{\Gamma}_t^{-1/2}\left(\hat{W}_t^v - \hat{R}_t^v\hat{P}_n^v\right)\underline{\hat{S}}_n' = \left(\hat{\Omega}_t^{-1/2} \otimes \hat{\Omega}_t^{-1/2}\right)\left(\hat{W}_t^v - \hat{R}_t^v\hat{P}_n^v\right)\underline{\hat{S}}_n'$ may themselves be computed as the $G^2 \times p$ matrix residuals of the OLS multivariate regression

$$\left(\hat{\Omega}_t^{-1/2} \otimes \hat{\Omega}_t^{-1/2}\right)\hat{W}_t^v\underline{\hat{S}}_n' = \left[\left(\hat{\Omega}_t^{-1/2} \otimes \hat{\Omega}_t^{-1/2}\right)\hat{R}_t^v\right] P + \text{residuals}, \quad t = 1, 2, ...$$

This way of computing $\mathcal{M}_n^{v^w}$ is again particularly convenient in the univariate case. Its second step is however quite useless in the multivariate case.

2.6.1. Testing against nested alternatives

As for conditional mean testing, a convenient and useful auxiliary nested alternative to the null conditional variance specification is

$$\mathrm{H}_1^{v\prime} : V(Y_t|X_t) = \Omega_t^a(X_t, \theta_2^o, \alpha^o) \text{ for some } (\theta_2^{o\prime}, \alpha^{o\prime})' \in \Theta_2 \times \Theta_\alpha, \quad t = 1, 2, ...$$

where α is a $k_\alpha \times 1$ vector of auxiliary variance parameters, and for some constant vector $c \in \Theta_\alpha$, we have

$$\Omega_t^a(X_t, \theta_2, c) = \Omega_t(X_t, \theta_2), \quad t = 1, 2, ...$$

such that testing the null H_0^v against H_1^v using the auxiliary alternative $\mathrm{H}_1^{v\prime}$ means testing the null that $\alpha^o = c$. Based on the GRPML2 estimator — here, using an other RPML2 estimator would yield an other misspecification indicator —, a LM-type test of H_0^v against the auxiliary alternative $\mathrm{H}_1^{v\prime}$ yields the misspecification indicator

$$\begin{aligned}
\hat{\Phi}_n &= \frac{\partial L_n^a(Y^n, X^n, \hat{\theta}_{1n}, \hat{\theta}_{2_n}, c)}{\partial \alpha} \\
&= \frac{1}{n}\sum_{t=1}^n \frac{\partial\left(\operatorname{vec}\Omega_t^a(X_t, \hat{\theta}_{2_n}, c)\right)'}{\partial \alpha}\Gamma_t^a(X_t, \hat{\theta}_{2_n}, c)^{-1}\operatorname{vec}(\hat{u}_t\hat{u}_t' - \Omega_t^a(X_t, \hat{\theta}_{2_n}, c)) \\
&= \frac{1}{n}\sum_{t=1}^n \frac{\partial\left(\operatorname{vec}\Omega_t^a(X_t, \hat{\theta}_{2_n}, c)\right)'}{\partial \alpha}\Gamma_t(X_t, \hat{\theta}_{2_n})^{-1}\operatorname{vec}(\hat{u}_t\hat{u}_t' - \Omega_t(X_t, \hat{\theta}_{2_n}))
\end{aligned}$$

$$= \frac{1}{n}\sum_{t=1}^{n}\frac{\partial\left(\operatorname{vec}\Omega_t^a(X_t,\hat{\theta}_{2_n},c)\right)'}{\partial\alpha}\Gamma_t(X_t,\hat{\theta}_{2_n})^{-1}v_t(Y_t,X_t,\hat{\theta}_{1_n},\hat{\theta}_{2_n})$$

where $L_n^a(Y^n, X^n, \theta_1, \theta_2, \alpha) = \frac{1}{n}\sum_{t=1}^n \ln f\left(Y_t, m_t(X_t,\theta_1), \Omega_t^a\left(X_t,\theta_2,\alpha\right)\right)$ and $\Gamma_t^a(X_t,\theta_2,\alpha) = \Omega_t^a\left(X_t,\theta_2,\alpha\right)\otimes\Omega_t^a\left(X_t,\theta_2,\alpha\right)$.

$\hat{\Phi}_n$ may then be checked through the $\mathcal{M}_n^{v^w}$ statistic (2.33) by setting $\underline{\hat{S}}_n = I_q$, $\hat{W}_t^v = \frac{\partial \operatorname{vec}\Omega_t^a(X_t,\hat{\theta}_{2_n},c)}{\partial\alpha'}$ and $p = q = k_\alpha$. The obtained test statistic is valid regardless of distributional misspecification and may in addition be implemented using any n-root consistent estimator of the parameters θ_1 and θ_2. Thus, if we except standard regularity conditions, the validity of the test again requires no more than just the null hypothesis of interest H_0^v. Under normality, the test is asymptotically equivalent to the (non-robust to non-normality, or more precisely non-robust to departures from the third and fourth order moments of the normal distribution) Breush-Pagan's (1980) classical second order gaussian LM testing procedures. Note finally that, as above, $\alpha^o = c$ is allowed to be on the boundary of its parameter space Θ_α. This is especially useful in the present case since conditional variances necessarily imply non-negativity restrictions, so that testing a null which lies on the boundary of Θ_α is not seldom. A classical example is testing for one-way error components in panel data.

2.6.2. Testing against non-nested alternatives

Consider the following auxiliary non-nested alternative to the null conditional variance specification

$$\mathrm{H}_1^{v\prime} : V\left(Y_t|X_t\right) = \Sigma_t(X_t,\delta^o) \text{ for some } \delta^o \in \Theta_\delta, \quad t = 1,2,...$$

where δ is a $k_\delta \times 1$ vector of parameters, and suppose that some n-root consistent estimator $\hat{\delta}_n$ of δ^o under $\mathrm{H}_1^{v\prime}$ is available. It may be, but need not to be, a GRPML2 estimator. Provided that standard regularity conditions hold, under H_0^v, $\hat{\delta}_n$ will converge to some non-stochastic sequence of $k_\delta \times 1$ vectors of pseudo-true values $\{\delta_n^* : n = 1,2,...\}$.

As in Section 2.5.2, this non-nested hypotheses testing problem may be transformed into a nested one by resorting to the auxiliary artificially nested alternative

$$\mathrm{H}_1^{v\prime\prime} : V\left(Y_t|X_t\right) = (1-\lambda^o)\Omega_t(X_t,\theta_2^o) + \lambda^o\Sigma_t(X_t,\delta^o) \text{ for some } \left(\theta_2^{o\prime},\lambda^o,\delta^{o\prime}\right)' \in \Theta_a,$$

$t = 1,2,...$, where λ is a scalar parameter and $\Theta_a = \Theta_2 \times \Theta_\lambda \times \Theta_\delta$, such that testing the null H_0^v against H_1^v using the artificial auxiliary alternative $\mathrm{H}_1^{v\prime\prime}$ now means testing the null that $\lambda^o = 0$. Putting the Davidson-Mackinnon (1981) trick intended to overcome the non-identifiability of δ^o under H_0^v into service, based on the GRPML2 estimator, a LM-like test of H_0^v against the artificial auxiliary alternative $\mathrm{H}_1^{m\prime\prime}$ yields the scalar misspecification indicator

$$\hat{\Phi}_n = \frac{\partial L_n^a(Y^n, X^n, \hat{\theta}_{1n}, \hat{\theta}_{2_n}0, \hat{\delta}_n)}{\partial\lambda}$$

$$= \frac{1}{n}\sum_{t=1}^{n}\frac{\partial\left(\operatorname{vec}\Omega_t^a(X_t,\hat\theta_{2_n},0,\hat\delta_n)\right)'}{\partial\lambda}\Gamma_t^a(X_t,\hat\theta_{2_n},0,\hat\delta_n)^{-1}\operatorname{vec}(\hat u_t\hat u_t'-\Omega_t^a(X_t,\hat\theta_{2_n},0,\hat\delta_n))$$

$$= \frac{1}{n}\sum_{t=1}^{n}\left(\operatorname{vec}(\Sigma_t(X_t,\hat\delta_n)-\Omega_t(X_t,\hat\theta_{2_n}))\right)'\Gamma_t(X_t,\hat\theta_{2_n})^{-1}\operatorname{vec}(\hat u_t\hat u_t'-\Omega_t(X_t,\hat\theta_{2_n}))$$

$$= \frac{1}{n}\sum_{t=1}^{n}\left(\operatorname{vec}(\Sigma_t(X_t,\hat\delta_n)-\Omega_t(X_t,\hat\theta_{2_n}))\right)'\Gamma_t(X_t,\hat\theta_{2_n})^{-1}v_t(Y_t,X_t,\hat\theta_{1_n},\hat\theta_{2_n})$$

where $L_n^a(Y^n,X^n,\theta_1,\theta_2,\lambda,\delta) = \frac{1}{n}\sum_{t=1}^{n}\ln f\left(Y_t,m_t(X_t,\theta_1),\Omega_t^a\left(X_t,\theta_2,\lambda,\delta\right)\right)$ with $\Omega_t^a\left(X_t,\theta_2,\lambda,\delta\right) = (1-\lambda)\Omega_t(X_t,\theta_2)+\lambda\Sigma_t(X_t,\delta))$ and $\Gamma_t^a(X_t,\theta_2,\lambda,\delta) = \Omega_t^a\left(X_t,\theta_2,\lambda,\delta\right) \otimes \Omega_t^a\left(X_t,\theta_2,\lambda,\delta\right)$.

Contrary to the non-nested mean testing case, the misspecification indicator $\hat\Phi_n$ is not the one which arises from the Cox (1961,1962) procedure for testing gaussian models with identical mean and non-nested variance. The scalar misspecification indicator arising from the Cox (1961,1962) approach is (see White (1994) and recall that $\operatorname{vec}(ABC) = (C'\otimes A)\operatorname{vec}B$, so that for any symmetric non-singular matrix A, $\operatorname{vec}A^{-1} = (A^{-1}\otimes A^{-1})\operatorname{vec}A$)

$$\begin{aligned}\hat\Phi_n^c &= \frac{1}{n}\sum_{t=1}^{n}\left(\operatorname{vec}(\Sigma_t(X_t,\hat\delta_n)^{-1}-\Omega_t(X_t,\hat\theta_{2_n})^{-1})\right)'v_t(Y_t,X_t,\hat\theta_{1_n},\hat\theta_{2_n})\\
&= \frac{1}{n}\sum_{t=1}^{n}\left(\Gamma_t(X_t,\hat\theta_{2_n})\operatorname{vec}\Sigma_t(X_t,\hat\delta_n)^{-1}-\operatorname{vec}\Omega_t(X_t,\hat\theta_{2_n})\right)'\\
&\qquad\Gamma_t(X_t,\hat\theta_{2_n})^{-1}v_t(Y_t,X_t,\hat\theta_{1_n},\hat\theta_{2_n})\\
&= \frac{1}{n}\sum_{t=1}^{n}\left(\operatorname{vec}(\Omega_t(X_t,\hat\theta_{2_n})\Sigma_t(X_t,\hat\delta_n)^{-1}\Omega_t(X_t,\hat\theta_{2_n})-\Omega_t(X_t,\hat\theta_{2_n}))\right)'\\
&\qquad\Gamma_t(X_t,\hat\theta_{2_n})^{-1}v_t(Y_t,X_t,\hat\theta_{1_n},\hat\theta_{2_n})\end{aligned}$$

Tests of H_0^v against $\mathrm{H}_1^{v\prime}$ may be based on either $\hat\Phi_n$ or $\hat\Phi_n^c$ and performed through $\mathcal{M}_n^{vw}$ statistic (2.33) by setting for $\hat\Phi_n$, $\hat W_t^v = \operatorname{vec}(\Sigma_t(X_t,\hat\delta_n)-\Omega_t(X_t,\hat\theta_{2_n}))$, and for $\hat\Phi_n^c$, $\hat W_t^v = \operatorname{vec}(\Omega_t(X_t,\hat\theta_{2_n})\Sigma_t(X_t,\hat\delta_n)^{-1}\Omega_t(X_t,\hat\theta_{2_n})-\Omega_t(X_t,\hat\theta_{2_n}))$. In both cases, $\underline{\hat S}_n = 1$ and $p = q = 1$. Both test statistics are robust to distributional misspecification and may be implemented using any n-root consistent estimator of θ_1, θ_2 and δ. We may quite naturally expect tests based $\hat\Phi_n^c$ to be generally more powerful than tests based on $\hat\Phi_n$.

2.6.3. Testing without alternatives : Hausman and information matrix type tests

As for the conditional mean, conditional variance testing without resorting to explicit alternatives may be performed through Hausman and information matrix type tests.

2.6.3.1. Hausman type tests

Following the arguments of Section 2.5.3.1, a Hausman type test of H_0^v here means checking a misspecification indicator of the form

$$\hat{\Phi}_n = S(\hat{\theta}_{2_n} - \hat{\underline{\theta}}_{2_n}) \tag{2.36}$$

where $\hat{\underline{\theta}}_{2_n}$ is some n-root consistent estimator of θ_2^o alternative to $\hat{\theta}_{2_n}$ and S is a $p \times k_{\theta_2}$ ($p \leq k_{\theta_2}$) non-stochastic selection matrix. Such a test will have power against any alternative H_1^v for which $\hat{\theta}_{2_n}$ and $\hat{\underline{\theta}}_{2_n}$ converge to different pseudo-true values.

As suggested by the form of first order conditions (2.4), the GRPML2 variance parameters estimator $\hat{\theta}_{2_n}$ may be shown to be asymptotically equivalent to the MWNLS estimator — or any other QGPML1 estimator — with weights $\left\{\Gamma_t(X_t, \check{\theta}_{2_n})^{-1}\right\}$ of the G^2-variate nonlinear regression

$$\mathrm{vec}(u_t(Y_t, X_t, \check{\theta}_{1_n})u_t(Y_t, X_t, \check{\theta}_{1_n})') = \mathrm{vec}\,\Omega_t(X_t, \theta_2) + \text{residuals}, \quad t = 1, 2, ... \tag{2.37}$$

where $\check{\theta}_{1_n}$ and $\check{\theta}_{2_n}$ are arbitrary preliminary n-root consistent estimators of θ_1^o and θ_2^o.

This suggest as a natural generic choice for $\hat{\underline{\theta}}_{2_n}$ MWNLS estimators of (2.37) using alternative weights of the form $\left\{\Gamma_t^{\Sigma}(X_t, \hat{\delta}_n)^{-1}\right\}$, where $\Gamma_t^{\Sigma}(X_t, \delta) = \Sigma_t\,(X_t, \delta) \otimes \Sigma_t\,(X_t, \delta)$, the $\Sigma_t\,(X_t, \delta)$ are alternative (necessarily misspecified) specifications for $V(Y_t|X_t)$ and $\hat{\delta}_n$ is some n-root consistent estimator which converges, under H_0^v, to some non-stochastic sequence of $k_\delta \times 1$ vectors of pseudo-true values $\{\delta_n^* : n = 1, 2, ...\}$. The easiest alternative estimator $\hat{\underline{\theta}}_{2_n}$ is simply obtained by setting $\Sigma_t\,(X_t, \delta) = I_G$.

Now, deriving a misspecification indicator of the form (2.30) yielding a test asymptotically equivalent to the one which could directly be obtained from (2.36) may be done along the same lines than in Section 2.5.3.1. The score associated with $\hat{\underline{\theta}}_{2_n}$ as defined above is given by (using $\hat{\theta}_{1_n}$ as a preliminary estimator of θ_1^o)

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} \underline{s}_t^2(\hat{\underline{\theta}}_{2_n}, \hat{\theta}_{1_n}, \hat{\delta}_n) \\
= \;& \frac{1}{n}\sum_{t=1}^{n} \frac{\partial\left(\mathrm{vec}\,\Omega_t(X_t, \hat{\underline{\theta}}_{2_n})\right)'}{\partial\theta_2}\Gamma_t^{\Sigma}(X_t, \hat{\delta}_n)^{-1} v_t(Y_t, X_t, \hat{\theta}_{1_n}, \hat{\underline{\theta}}_{2_n}) = 0
\end{aligned} \tag{2.38}$$

Consider similarly evaluating (2.38) at $\hat{\theta}_{2_n}$. Under H_0^v and usual regularity conditions, using standard arguments, a mean value expansion of $n^{-1/2}\sum_{t=1}^{n} \underline{s}_t^2(\hat{\theta}_{2_n}, \hat{\theta}_{1_n}, \hat{\delta}_n)$ at $\left(\hat{\underline{\theta}}_{2_n}', \hat{\theta}_{1_n}', \hat{\delta}_n'\right)'$ gives

$$n^{-1/2}\sum_{t=1}^{n} \underline{s}_t^2(\hat{\theta}_{2_n}, \hat{\theta}_{1_n}, \hat{\delta}_n) = n^{-1/2}\sum_{t=1}^{n} \underline{s}_t^2(\hat{\underline{\theta}}_{2_n}, \hat{\theta}_{1_n}, \hat{\delta}_n) + \underline{A}_{n22}^*\sqrt{n}(\hat{\theta}_{2_n} - \hat{\underline{\theta}}_{2_n}) + o_{P_o}(1)$$

where

$$\underline{A}^*_{n22} = \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial \underline{s}^2_t(\theta^o_2, \theta^o_1, \delta^*_n)}{\partial\theta'_2}\right] = -\frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial\,(\mathrm{vec}\,\Omega^o_t)'}{\partial\theta_2}\Gamma^\Sigma_t(X_t, \delta^*_n)^{-1}\frac{\partial\,\mathrm{vec}\,\Omega^o_t}{\partial\theta'_2}\right]$$

or, given (2.38),

$$\sqrt{n}(\hat{\theta}_{2_n} - \underline{\hat{\theta}}_{2_n}) = \underline{A}^{*^{-1}}_{n22} n^{-1/2}\sum_{t=1}^{n} \underline{s}^2_t(\hat{\theta}_{2_n}, \hat{\theta}_{1_n}, \hat{\delta}_n) + o_{P_o}(1)$$

and further,

$$\sqrt{n}(\hat{\theta}_{2_n} - \underline{\hat{\theta}}_{2_n}) = \underline{\hat{A}}^{-1}_{n22} n^{-1/2}\sum_{t=1}^{n} \underline{s}^2_t(\hat{\theta}_{2_n}, \hat{\theta}_{1_n}, \hat{\delta}_n) + o_{P_o}(1)$$

where

$$\underline{\hat{A}}_{n22} = -\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\left(\mathrm{vec}\,\Omega_t(X_t, \hat{\theta}_{2_n})\right)'}{\partial\theta_2}\Gamma^\Sigma_t(X_t, \hat{\delta}_n)^{-1}\frac{\partial\,\mathrm{vec}\,\Omega_t(X_t, \hat{\theta}_{2_n})}{\partial\theta'_2}$$

such that we have

$$\sqrt{n}\hat{\Phi}_n = S\sqrt{n}(\hat{\theta}_{2_n} - \underline{\hat{\theta}}_{2_n}) = S\underline{\hat{A}}^{-1}_{n22} n^{-1/2}\sum_{t=1}^{n} \underline{s}^2_t(\hat{\theta}_{2_n}, \hat{\theta}_{1_n}, \hat{\delta}_n) + o_{P_o}(1)$$

Thus, a test based on $\hat{\Phi}_n$ may, from an asymptotic point of view, equivalently be based on the misspecification indicator

$$\begin{aligned}
\hat{\Phi}^h_n &= S\underline{\hat{A}}^{-1}_{n22}\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\left(\mathrm{vec}\,\Omega_t(X_t, \hat{\theta}_{2_n})\right)'}{\partial\theta_2}\Gamma^\Sigma_t(X_t, \hat{\delta}_n)^{-1} v_t(Y_t, X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n}) \\
&= S\underline{\hat{A}}^{-1}_{n22}\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\left(\mathrm{vec}\,\Omega_t(X_t, \hat{\theta}_{2_n})\right)'}{\partial\theta_2}\Gamma^\Sigma_t(X_t, \hat{\delta}_n)^{-1}\Gamma_t(X_t, \hat{\theta}_{2_n}) \\
&\qquad \Gamma_t(X_t, \hat{\theta}_{2_n})^{-1} v_t(Y_t, X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n})
\end{aligned}$$

$\hat{\Phi}^h_n$ may here be checked through the $\mathcal{M}^{v^w}_n$ statistic (2.33) by setting $\underline{\hat{S}}_n = S\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\left(\mathrm{vec}\,\Omega_t(X_t,\hat{\theta}_{2_n})\right)'}{\partial\theta_2}\Gamma^\Sigma_t(X_t, \hat{\delta}_n)^{-1}\frac{\partial\,\mathrm{vec}\,\Omega_t(X_t,\hat{\theta}_{2_n})}{\partial\theta'_2}\right)^{-1}$ and $\hat{W}^v_t = \Gamma_t(X_t, \hat{\theta}_{2_n})\Gamma^\Sigma_t(X_t, \hat{\delta}_n)^{-1}\frac{\partial\,\mathrm{vec}\,\Omega_t(X_t,\hat{\theta}_{2_n})}{\partial\theta'_2}$. As usual, the validity of this Hausman type test requires no more than just the null hypothesis of interest H^v_0 and may be implemented using any n-root consistent estimator of θ_1, θ_2 and δ. Regardless of the used estimators, the test statistic is asymptotically equivalent to comparing (some linear combination S of) two MWNLS estimators of the regression model $\mathrm{vec}(u^o_t u^{o\prime}_t) = \mathrm{vec}\,\Omega_t(X_t, \theta_2) +$ residuals, the first with weights $\{\Gamma_t(X_t, \theta^o_2)^{-1}\}$ and the second with weights $\left\{\Gamma^\Sigma_t(X_t, \delta^*_n)^{-1}\right\}$. When the $\Sigma_t(X_t, \delta)$ are set equal to I_G, the implementation of the test requires no additional estimators than the ones needed for estimating the null model. Note finally that if the selection matrix S is set equal to $I_{k_{\theta_2}}$ (or any other non-singular square matrix), then the entire term $\underline{\hat{S}}_n$ may as above be dropped from the statistic (where now $p = q = k_{\theta_2}$) without affecting it.

2.6.3.2. Information matrix type tests

Using the law of iterated expectations, it is readily seen that, under H_0^v, we must have

$$\frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{o^{-1}} u_t^o u_t^{o\prime} \Omega_t^{o^{-1}} \frac{\partial m_t^o}{\partial \theta_1'}\right] = \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{o^{-1}} \frac{\partial m_t^o}{\partial \theta_1'}\right] \tag{2.39}$$

According to the results of Chapter 1 (Proposition 10 and 12), (2.39) is one of the two sufficient conditions for the mean parameters information matrix equality $B_{n_{11}}^o = -A_{n_{11}}^o$ associated with RPML2 estimators to hold, the other one being that

$$E\left[\left(n^{-1/2}\sum_{t=1}^{n} s_t^{1o}\right)\left(n^{-1/2}\sum_{t=1}^{n} s_t^{1o}\right)'\right] = \frac{1}{n}\sum_{t=1}^{n} E\left[s_t^{1o} s_t^{1o\prime}\right]$$

a property which holds if $\mathcal{S}$ is further first order dynamically complete.

Rearranging (2.39), we get

$$\frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t^{o\prime}}{\partial \theta_1}\Omega_t^{o^{-1}}\left(u_t^o u_t^{o\prime} - \Omega_t^o\right)\Omega_t^{o^{-1}}\frac{\partial m_t^o}{\partial \theta_1'}\right] = 0 \tag{2.40}$$

or, by vectorizing,

$$\frac{1}{n}\sum_{t=1}^{n} E\left[\left(\frac{\partial m_t^{o\prime}}{\partial \theta_1} \otimes \frac{\partial m_t^{o\prime}}{\partial \theta_1}\right)\left(\Omega_t^{o^{-1}} \otimes \Omega_t^{o^{-1}}\right)\mathrm{vec}(u_t^o u_t^{o\prime} - \Omega_t^o)\right] = 0$$

A test of conditional variance correct specification may then be based on a misspecification indicator of the form

$$\hat{\Phi}_n = S\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial m_t(X_t,\hat{\theta}_{1_n})}{\partial \theta_1'} \otimes \frac{\partial m_t(X_t,\hat{\theta}_{1_n})}{\partial \theta_1'}\right)' \Gamma_t(X_t,\hat{\theta}_{2_n})^{-1} v_t(Y_t,X_t,\hat{\theta}_{1_n},\hat{\theta}_{2_n})$$

where the $p \times k_{\theta_1}^2$ $(p \leq (k_{\theta_1}^2 + k_{\theta_1})/2)$ non-stochastic selection matrix S allows to focus on some of the non-redundant elements or linear combinations of the non-redundant elements of the second term.

$\hat{\Phi}_n$ may then again be checked through the $\mathcal{M}_n^{vw}$ statistic (2.33) by setting $\underline{\hat{S}}_n = S$ and $\hat{W}_t^v = \frac{\partial m_t(X_t,\hat{\theta}_{1_n})}{\partial \theta_1'} \otimes \frac{\partial m_t(X_t,\hat{\theta}_{1_n})}{\partial \theta_1'}$. As usual, only H_0^v and n-root consistent estimators of the parameters θ_1 and θ_2 are required for the test to be valid. Because of the presence by construction of redundant elements in the second term of $\hat{\Phi}_n$, it is worth emphasizing that S can never be set equal to an identity matrix. Note that, as the information matrix type test of the conditional mean, the present test statistic admits a Hausman type test interpretation. It indeed essentially amounts to comparing two consistent estimators, under H_0^v, of $-A_{n_{11}}^o$, the two estimators simply being the empirical counterparts of respectively the left-hand side and the right-hand side of (2.39). The test statistic will thus have power against any alternative H_1^v for which these two estimators converge to different pseudo-true values.

2.6.4. Testing dynamic completeness

It follows from Chapter 1 (Proposition 12) that when in addition to be second order correctly specified and first order dynamically complete, $\mathcal{S}$ is also dynamically complete for the conditional variance, then the entire asymptotic covariance matrix of RPML2 estimators may always be readily estimated. So, following again the Wooldridge's (1991a) sequential "bottom-up" model construction/specification testing strategy, we finally concentrate on testing the null

$$\mathrm{H}_0^{vd} : \mathrm{H}_0^{md} \text{ and } \mathrm{H}_0^v \text{ hold, and } V\left(Y_t|X_t\right) = V\left(Y_t|X_t, \Psi_{t-1}\right), \quad t = 1, 2, ...$$

against the alternative

$$\mathrm{H}_1^{vd} : \mathrm{H}_0^{md} \text{ and } \mathrm{H}_0^v \text{ hold but } \mathrm{H}_0^{vd} \text{ is false}$$

H_0^{vd} is equivalent to

$$\mathrm{H}_0^{vd} : \mathrm{H}_0^{md} \text{ holds and } V(Y_t|X_t, \Psi_{t-1}) = \Omega_t(X_t, \theta_2^o) \text{ for some } \theta_2^o \in \Theta_2, \quad t = 1, 2, ...$$

As in Section 2.6.1, for testing H_0^{vd}, we may proceed in the same way than above for testing H_0^v and likewise now unambiguously take advantage of the fact that under H_0^{vd} the simple estimator (2.34) is consistent for the asymptotic covariance matrix $K_n^{v^{o*}}$.

As for the conditional mean, the most general ways to check dynamic completeness of the conditional variance are either to look at autocorrelation in the non-redundant terms of the errors v_t or to resort to a White (1987,1994) dynamic information matrix type test.

Looking at a multivariate AR (κ) process for the non-redundant terms of v_t — i.e., at an multivariate ARCH (κ) type process — means using as an auxiliary nested alternative to the null conditional variance specification

$$\mathrm{H}_1^{vd\prime} : V(Y_t|X_t, \Psi_{t-1}) = \Omega_t^a(X_t, \Psi_{t-1}, \theta_2^o, D_1, ..., D_\kappa, \theta_1^o) \text{ for some } a^o \in \Theta_a, \quad t = 1, 2, ...$$

with

$$\operatorname{vech} \Omega_t^a(X_t, \Psi_{t-1}, \theta_2^o, D_1^o, ..., D_\kappa^o, \theta_1^o) = \operatorname{vech} \Omega_t(X_t, \theta_2^o) + D_1^o \underline{v}_{t-1}^o + ... + D_\kappa^o \underline{v}_{t-\kappa}^o$$
$$\underline{v}_t^o = \operatorname{vech}(u_t(Y_t, X_t, \theta_1^o) u_t(Y_t, X_t, \theta_1^o)' - \Omega_t(X_t, \theta_2^o))$$

where $\kappa \geq 1$ is a integer that determines the maximum autocorrelation of $\underline{v}_t$ to be examined, $a^o = (\theta_2^{o\prime}, (\operatorname{vec} D_1^o)', ..., (\operatorname{vec} D_\kappa^o)', \theta_1^{o\prime})'$, $\Theta_a = \Theta_2 \times \Theta_{D_1} \times ... \times \Theta_{D_\kappa} \times \Theta_1$, θ_1^o is the true value ensuring first order correct specification and the D_i, $i = 1, ..., \kappa$, are $(G^2 + G)/2 \times (G^2 + G)/2$ matrices of auxiliary variance parameters. Note that this alternative specification contains links between mean and variance parameters.

Let for now $t=1$ denote the $(\kappa+1)$-th observation and define $n_\kappa = n-\kappa$.

Now, testing the null H_0^{vd} against H_1^{vd} using the auxiliary alternative $\mathrm{H}_1^{vd\prime}$ means testing the null that $D_1^o = ... = D_\kappa^o = 0$. Based on the GRPML2 estimator, a LM-type test yields the misspecification indicator

$$\begin{aligned}
\hat{\Phi}_{n_\kappa}^{AR} &= \frac{\partial L_{n_\kappa}^a(Y^n, X^n, \hat{\theta}_{1n}, \hat{\theta}_{2_n}, 0, ..., 0)}{\partial\left((\operatorname{vec} D_1)', ..., (\operatorname{vec} D_\kappa)'\right)'} \\
&= \frac{1}{n_\kappa}\sum_{t=1}^{n_\kappa} \frac{\partial\left(\operatorname{vec}\Omega_t^a(X_t, \Psi_{t-1}, \hat{\theta}_{2_n}, 0, ..., 0, \hat{\theta}_{1n})\right)'}{\partial\left((\operatorname{vec} D_1)', ..., (\operatorname{vec} D_\kappa)'\right)'}\Gamma_t^a(X_t, \Psi_{t-1}, \hat{\theta}_{2_n}, 0, ..., 0, \hat{\theta}_{1n})^{-1} \\
&\qquad \operatorname{vec}(\hat{u}_t\hat{u}_t' - \Omega_t^a(X_t, \Psi_{t-1}, \hat{\theta}_{2_n}, 0, ..., 0, \hat{\theta}_{1n})) \\
&= \frac{1}{n_\kappa}\sum_{t=1}^{n_\kappa} \frac{\partial\left(\operatorname{vec}\Omega_t^a(X_t, \Psi_{t-1}, \hat{\theta}_{2_n}, 0, ..., 0, \hat{\theta}_{1n})\right)'}{\partial\left((\operatorname{vec} D_1)', ..., (\operatorname{vec} D_\kappa)'\right)'}\Gamma_t(X_t, \hat{\theta}_{2_n})^{-1} \\
&\qquad \operatorname{vec}(\hat{u}_t\hat{u}_t' - \Omega_t(X_t, \hat{\theta}_{2_n})) \\
&= \frac{1}{n_\kappa}\sum_{t=1}^{n_\kappa} F_t^{AR}(X_t, \Psi_{t-1}, \hat{\theta}_{2_n}, \hat{\theta}_{1_n})'\Gamma_t(X_t, \hat{\theta}_{2_n})^{-1} v_t(Y_t, X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n})
\end{aligned}$$

where $L_{n_\kappa}^a(Y^n, X^n, \theta_1, \theta_2, D_1, ..., D_\kappa) = \frac{1}{n_\kappa}\sum_{t=1}^{n_\kappa} \ln f(Y_t, m_t(X_t, \theta_1), \Omega_t^a(X_t, \Psi_{t-1}, \theta_2, D_1, ..., D_\kappa, \theta_1))$ with $\operatorname{vech}\Omega_t^a(X_t, \Psi_{t-1}, \theta_2, D_1, ..., D_\kappa, \theta_1) = \operatorname{vech}\Omega_t(X_t, \theta_2) + D_1\underline{v}_{t-1} + ... + D_\kappa\underline{v}_{t-\kappa}$ and $\underline{v}_t = \operatorname{vech}(u_t(Y_t, X_t, \theta_1)u_t(Y_t, X_t, \theta_1)' - \Omega_t(X_t, \theta_2))$, $\Gamma_t^a(X_t, \Psi_{t-1}, \theta_2, D_1, ..., D_\kappa, \theta_1) = \Omega_t^a(X_t, \Psi_{t-1}, \theta_2, D_1, ..., D_\kappa, \theta_1) \otimes \Omega_t^a(X_t, \Psi_{t-1}, \theta_2, D_1, ..., D_\kappa, \theta_1)$ and the $G^2 \times \kappa(G^2+G)^2/4$ matrix functions $F_t^{AR}(X_t, \Psi_{t-1}, \theta_2, \theta_1)$ are defined as

$$F_t^{AR}(X_t, \Psi_{t-1}, \theta_2, \theta_1) = D_{uG}\left(\left(\underline{v}_{t-1}', ..., \underline{v}_{t-\kappa}'\right) \otimes I_{(G^2+G)/2}\right), \quad t = 1, 2, ...$$

where D_{uG} denotes the $G^2 \times (G^2+G)/2$ duplication matrix, i.e., for the record, a matrix such that, for any symmetric $G \times G$ matrix A, $D_{uG}\operatorname{vech} A = \operatorname{vec} A$.

As outlined above, under H_0^{vd}, v_t^o is a martingale difference sequence with respect to $\{\Psi_t\}$, and thus so does the score s_t^{2o}, so that s_t^{2o} is uncorrelated with its past values. Accordingly, for all $\kappa \geq 1$, we must have

$$E\left[s_t^{2o}\left(s_{t-1}^{2o\prime}, ..., s_{t-\kappa}^{2o\prime}\right)\right] = 0, \quad t = 1, 2, ... \tag{2.41}$$

Then, choosing some integer κ and vectorizing (2.41), a test of H_0^{vd} may alternatively be based on the misspecification indicator

$$\hat{\Phi}_{n_\kappa}^{IM} = \frac{1}{n_\kappa}\sum_{t=1}^{n_\kappa} F_t^{IM}(X_t, \Psi_{t-1}, \hat{\theta}_{1_n}, \hat{\theta}_{2_n})'\Gamma_t(X_t, \hat{\theta}_{2_n})^{-1} v_t(Y_t, X_t, \hat{\theta}_{1_n}, \hat{\theta}_{2_n})$$

where the $G^2 \times \kappa k_{\theta_2}^2$ matrix functions $F_t^{IM}(X_t, \Psi_{t-1}, \theta_1, \theta_2)$ are defined as

$$F_t^{IM}(X_t, \Psi_{t-1}, \theta_1, \theta_2) = \frac{\partial \operatorname{vec}\Omega_t(X_t, \theta_2)}{\partial\theta_2'}\left(\left(s_{t-1}^{2\prime}, ..., s_{t-\kappa}^{2\prime}\right) \otimes I_{k_{\theta_2}}\right), \quad t = 1, 2, ...$$

with $s_t^2 = \frac{\partial(\mathrm{vec}\,\Omega_t(X_t,\theta_2))'}{\partial\theta_2}\Gamma_t(X_t,\theta_2)^{-1}v_t(Y_t,X_t,\theta_1,\theta_2)$.

Both $\hat{\Phi}^{AR}_{n_\kappa}$ and $\hat{\Phi}^{IM}_{n_\kappa}$ may again be checked through the $\mathcal{M}^{v^w}_n$ statistic (2.33) by setting for $\hat{\Phi}^{AR}_{n_\kappa}$, $\hat{W}^v_t = F^{AR}_t(X_t,\Psi_{t-1},\hat{\theta}_{2_n},\hat{\theta}_{1_n})$ and $p = q = \kappa(G^2+G)^2/4$, and for $\hat{\Phi}^{IM}_{n_\kappa}$, $\hat{W}^v_t = F^{IM}_t(X_t,\Psi_{t-1},\hat{\theta}_{1_n},\hat{\theta}_{2_n})$ and $p = q = \kappa k^2_{\theta_2}$. In both cases, $\underline{\hat{S}}_n = I_p$ — if wished, a selection matrix may again straightforwardly be introduced —, $n = n_\kappa$, $t = 1$ denotes the $(\kappa+1)$-th observation and $\hat{K}^v_n$ is the simple estimator (2.34). As usual, the test statistics are robust to distributional misspecification and may be implemented using any n-root consistent estimator of θ_1, θ_2. As for the conditional mean, the choice between using $\hat{\Phi}^{AR}_{n_\kappa}$ or $\hat{\Phi}^{IM}_{n_\kappa}$ may be done on the grounds of computational convenience but should also take into account their relative degree of freedom $\kappa(G^2+G)^2/4$ and $\kappa k^2_{\theta_2}$. When both are very large, it may be wise to resort to a selection matrix.

2.7. Concluding comments

This chapter concentrated on the question of how to check, after having estimated it by some method known to be robust to conditional misspecification, the extent to which a tentative second order semi-parametric model $\mathcal{S}$ is actually correctly specified. We surveyed a large spectrum of m-type diagnostic tests, primarily built on the GRPML2 estimator but yielding valid and asymptotically locally equivalent tests if implemented using any alternative n-root consistent estimator. Because of the nested nature the null hypotheses and the fact that the validity of all test statistics requires no more than the nulls of interest, they provide ways to quite comprehensively check — and hopefully unambiguously identify eventual departures from — the prominent aspects of the model specification.

The choice of which aspects — i.e., the conditional mean, the conditional variance, as well as their dynamic completeness — of the model specification to look at is up to the researcher and depends on the problem at hand, so does the choice of the misspecification indicators for checking the retained aspects. In most cases, all aspects are likely to be of interest and, at least for conditional mean and conditional variance testing (there are less possibilities when testing dynamic completeness), it seems sensible to resort to more than one misspecification indicator. Typically, an extensive investigation of the conditional mean and the conditional variance should be based on both Hausman or information matrix type misspecification indicator(s) and misspecification indicator(s) designed to check the null against plausible auxiliary (nested or non-nested) alternatives.

Once the misspecification indicators are chosen, we may proceed by performing individual and/or joint tests. For both conditional mean and conditional variance testing, joint tests may readily be constructed by appropriately stacking the 'individual' misspecification indicators. In both cases, it simply means forming 'joint' W_t indicator matrices by horizontally concatenating the indicator matrices associated to the 'individual' misspecification indicators. This for example allows to jointly perform a test against several nested and/or non-nested auxiliary alternatives along with Hausman and/or information matrix type diagnostic(s). Note that this similarly allows to perform joint tests of first (resp. second) order correct specification and first (resp. second) order dynamic completeness. In this latter case however we will no longer be able to distinguish the source — misspecification of the conditional moment or dynamic incompleteness — of departure from the joint null. The same problem obviously arises if, following the lines of Bollerslev-Wooldridge (1992), joint tests of the conditional mean and the conditional variance are undertaken. Accordingly, and this is the essence of the Wooldridge's (1991a) "bottom-up" model construction/specification testing strategy, we suspect that most empirical researchers will prefer to test the prominent aspects of the model specification separately. Likewise, because it may provide useful, although possibly misleading, information about the source(s) of departure from the null, we believe that they will also prefer checking individually the chosen misspecification indicators associated to the different aspects of the model specification. From a formal point of view, a joint induced test with bounded asymptotic size of the overall null of interest may then be carried out by using a Bonferroni approach: the joint induced test consists in accepting the overall null underlying the, say q, separate tests if and only if all the separate tests are accepted, and in rejecting it if one or more of the q separate tests is rejected. If each separate test has asymptotic size α_r, the Bonferroni inequality ensures that the joint induced test will have true asymptotic size at most equal to $\alpha = \sum_{r=1}^{q} \alpha_r$ (see Savin (1980,1984)), so that choosing for example $\alpha_r = \alpha/q$ will yield a joint induced test with true asymptotic size at most equal to α. If such a approach is followed, from a empirical point of view, according to Wooldridge (1991a), we believe that a good practice is to report the computed individual — or "partially joint" — test statistics along with their usual p-value and let the readers draw their own conclusions. Note that the above Bonferroni approach may also be used in the reverse manner for gaining insights about the direction(s) in which misspecification detected by a genuine joint test may lie.

The above diagnostic tests essentially deals with checking if the model is not in some way "underparametrized". An other question of interest is to see if it is not "overparametrized", if it may not be simplified. Provided that a consistent estimator of the asymptotic covariance matrix of the GRPML2 estimator may be obtained, this may of course readily be checked through classical Wald tests, and so do the eventual cross-constraints between mean and variance parameters, which have been discarded for robustness, contained in the structural model $\widetilde{\mathcal{S}}$.

To conclude, some important considerations closely related to a remark already made in Chapter 1. We outlined in Chapter 1, and recalled in Section 2.2 of this chapter, that the set of conditioning variables X_t underlying model $\mathcal{S}$ must be defined as comprising all the variables which appear either in the conditional mean or in the conditional variance. So, for example, if the conditional mean specification depends on say variables X_t^1 and the conditional variance specification depends on say variables X_t^2, then X_t must be defined as $X_t \equiv (X_t^1, X_t^2)$. This means that testing for example the conditional mean through a Hausman[4] or information matrix type test actually entails testing the null hypothesis H_0^m that there exists some θ_1^o such that $E(Y_t|X_t^1, X_t^2) = m_t(X_t^1, \theta_1^o)$, and not simply that for some θ_1^o we have $E(Y_t|X_t^1) = m_t(X_t^1, \theta_1^o)$. It is worth further emphasizing that whenever an auxiliary

[4] Where the alternative conditional variance specification $\{\Sigma_t\}$ underlying the estimator $\hat{\underline{\theta}}_{1_n}$ only depends on variables $X_t \equiv (X_t^1, X_t^2)$. If it is not the case — but there is a priori no good reason to do that — , then the actual null H_0^m will be different.

(nested or non-nested) alternative model is involved, the set of conditioning variables X_t must be defined as comprising not only all the variables which appear either in the conditional mean or in the conditional variance, but also the variables which appear in the auxiliary alternative model. Continuing the above example, when checking the conditional mean against some auxiliary (nested or non-nested) alternative specified as function of say the extended set of variables (X_t^1, X_t^3), X_t must be defined as $X_t \equiv (X_t^1, X_t^2, X_t^3)$. In such a test, the actual null hypothesis H_0^m is thus no longer that there exists some θ_1^o such that $E(Y_t|X_t^1, X_t^2) = m_t(X_t^1, \theta_1^o)$, but that for some θ_1^o we have $E(Y_t|X_t^1, X_t^2, X_t^3) = m_t(X_t^1, \theta_1^o)$. In other words, in conditional mean testing, specifying the conditional variance, or, when resorting to an explicit alternative, the auxiliary alternative model, as function of variables which do not appear in the conditional mean modifies the content of the null hypothesis of first order correct specification. The same reasoning obviously applies when considering conditional variance testing. To be aware of that is crucial for suitably designing and correctly interpreting the test statistics. So, in the examples just given, rejecting the null hypotheses H_0^m does not signify that the conditional mean is misspecified with respect to its own set of conditioning variables X_t^1: we might well have that $E(Y_t|X_t^1) = m_t(X_t, \theta_1^o)$ for some θ_1^o holds while neither $E(Y_t|X_t^1, X_t^2) = m_t(X_t, \theta_1^o)$ nor $E(Y_t|X_t^1, X_t^2, X_t^3) = m_t(X_t, \theta_1^o)$ do. Likewise, to give another example of possible misinterpretation, when testing two non-nested models (in mean or variance) specified as function of different variables, rejecting each model against the other does not mean that the models are misspecified with respect to their own set of conditioning variables. When considered with respect to their own set of conditioning variables, both might actually be correctly specified.

Chapter 3

A full heteroscedastic one-way error components model allowing for incomplete panels: second order pseudo-maximum likelihood estimation and specification testing

3.1. Introduction

As pointed out in our general introduction, heteroscedasticity seems to be endemic in work with microeconomic cross-section data. Basically, heteroscedasticity may be viewed as a symptom arising from the fact that the degree to which an economic relationship may explain actual individual observations is likely to depend on their specific characteristics. Put in other words, it may be viewed as a symptom of variable heterogeneity across individuals. A primary and well known source of heteroscedasticity stems from differences in the "size characteristic" (the level of the variables in the relationship) of the observations. This kind of heteroscedasticity is purely mechanical. It is simply a consequence of the assumed additive disturbance structure of the classical regression model. It is generally tackled by performing a logarithmic transformation of the dependent variable. However, even after accounting in such a way for differences in size, numerous cases remain where we can not expect the error variance to be constant. First, there is no a priori reason to believe that the logarithmic specification postulating similar percentage variation across observations is relevant. In the production field for example, observations for lower outputs firms seem likely to evoke larger variances (see Batalgi-Griffin (1988)). On the other hand, the error variance may also systematically vary across observations of similar size. For example, the variance of firms profits might depend upon product diversification or research and development expenditures. Likewise, the variance of firms outputs might depend upon their capitalistic intensity and so on. Note that in practice, these different sources of heteroscedasticity may be simultaneously present.

Obviously, there is no reason to expect the heteroscedasticity problems associated with microeconomic panel data to be markedly different from those encountered in work with cross-section data. Although innocuous in terms of consistency, when not taken into account, the violation of the standard second order assumptions of the classical one-way error components regression model implies inefficient estimation and, undoubtedly more alarming, makes the usual textbook inferential procedures, including popular specification tests such as the Hausman test, invalid. Further, it at least casts doubt on the heuristic interpretation of the model. Nonetheless, except for a few papers mentioned below, the issue of heteroscedasticity seems somewhat to have been ignored in the literature related to panel data error components models.

Seemingly, the first authors who dealt with the problem were Mazodier-Trognon (1978). Subsequent contributions[1] in the area include Verbon (1980), Rao-Kaplan-Cochran (1981), Magnus (1982), Arellano (1987), Baltagi-Griffin (1988), Baltagi (1988) - Wansbeek (1989), Randolf (1988a), Li-Stengos (1994) and Muus-Wansbeek (1994). Into the framework of the classical one-way error components regression model[2], the issues considered by these papers can be summarized as follows[3]. Both Mazodier-Trognon (1978) and Baltagi-Griffin (1988) are concerned with estimation of a model allowing for changing variances of the individual-specific error term across individuals, i.e., assume that, if we write the error components $\varepsilon_{it} = \mu_i + \nu_{it}$, $\nu_{it} \sim \text{IID}(0, \sigma^2_\nu)$ while $\mu_i \sim \text{ID}(0, \sigma^2_{\mu_i})$. Rao-Kaplan-Cochran (1981), Magnus (1982) and Baltagi (1988) - Wansbeek (1989) adopt a symmetrical opposite specification, allowing for changing variances of the general error term across individuals, i.e., assume that $\nu_{it} \sim \text{ID}(0, \sigma^2_{\nu_i})$ while $\mu_i \sim \text{IID}(0, \sigma^2_\mu)$. This specification is a particular case of the Swamy's (1970) random coefficient model where only the intercept parameter is assumed to be random. Verbon (1980) is interested in Lagrange Multiplier testing of the standard normally distributed homoscedastic model against the heteroscedastic alternative $\nu_{it} \sim \text{NID}(0, \sigma^2_{\nu_i})$ and $\mu_i \sim \text{NID}(0, \sigma^2_{\mu_i})$, where $\sigma^2_{\nu_i}$ and $\sigma^2_{\mu_i}$ are, up to a multiplicative constant, identical parametric functions of a (row) vector of time-invariant explanatory variables Z_i, i.e., $\sigma^2_{\nu_i} = \sigma^2_\nu \phi(Z_i\gamma)$ and $\sigma^2_{\mu_i} = \sigma^2_\mu \phi(Z_i\gamma)$. Randolf (1988a) concentrates on supplying an observation-by-observation data transformation for a full heteroscedastic error components model assuming that $\nu_{it} \sim \text{ID}(0, \sigma^2_{\nu_{it}})$ and $\mu_i \sim \text{ID}(0, \sigma^2_{\mu_i})$. Provided that the variances $\sigma^2_{\nu_{it}}$ and $\sigma^2_{\mu_i}$ are known, this transformation allows generalized least squares estimates to be obtained from ordinary least squares. Li-Stengos (1994) deals with adaptive estimation of an error components model supposing heteroscedasticity of unknown form for the general error term, i.e., assumes that $\mu_i \sim \text{IID}(0, \sigma^2_\mu)$ while $\nu_{it} \sim \text{ID}(0, \sigma^2_{\nu_{it}})$, where $\sigma^2_{\nu_{it}}$ is a nonparametric function $\phi(Z_{it})$ of a vector of explanatory variables Z_{it}[4]. Finally, in the context of the fixed effects model (within estimator), Arellano (1987) and Muus-Wansbeek (1994) outline heteroscedasticity-consistent (allowing for a rich variety of heteroscedasticity and serial correlation patterns) covariance

[1] We do not include Graag (1993) in this list since he works with a quite unusual bilinear model.

[2] Mazodier-Trognon (1978), Verbon (1980) and Magnus (1982) deal with the problem of heteroscedasticity in a more general framework than the simple one-way error components model: the former treats the problem in the context of the two-way error components model while the latters consider it (in the case of Magnus, anecdotally) in the context of a multivariate (SURE) error components model.

[3] Below, ID means "independently distributed", IID "identically independently distributed", and NID "normally independently distributed".

[4] Close to this specification is the statistical model underlying the Randolf's (1988b) empirical study of housing depreciation. In this paper, it is assumed that $\mu_i \sim \text{IID}(0, \sigma^2_\mu)$ while $\nu_{it} \sim \text{ID}(0, \sigma^2_{\nu_{it}})$, where $\sigma^2_{\nu_{it}}$ is a linear function of a vector of explanatory variables, this linear variance function arising from a special random coefficient assumption a la Hildreth-Houck (1968). The model is estimated (and tested) by standard gaussian ML methods.

matrix estimators. Except Randolf (1988a) and Rao-Kaplan-Cochran (1981), all the mentioned papers assume data arising from balanced (complete) panels.

In this chapter, we are concerned with estimation and specification testing of a full heteroscedastic one-way error components linear regression model specified in the spirit of Randolf (1988a). In short, we assume that the (conditional) variances $\sigma^2_{\nu_{it}}$ and $\sigma^2_{\mu_i}$ are distinct parametric functions of, respectively, (row) vectors of explanatory variables Z^1_{it} and Z^2_i, i.e., $\sigma^2_{\nu_{it}} = \phi_\nu(Z^1_{it}\gamma_1)$ and $\sigma^2_{\mu_i} = \phi_\mu(Z^2_i\gamma_2)$. Further, we treat the model in the context of incomplete (or unbalanced) panels. This specification differs from the previously proposed formulations of heteroscedastic error components models as it simultaneously embodies three basic characteristics. First, heteroscedasticity distinctly applies to both individual-specific and general error components. Second, (nonlinear) variance functions are parametrically specified. Finally, the model allows for incomplete panels .

Explicitly allowing for incomplete panels is an obviously desirable feature. Indeed, at least for micro-data, incompleteness is rather the rule than the exception. Further, as noted by Wansbeek-Kapteyn (1989), an unbalanced panel dataset makes most of the results obtained in the error components literature inapplicable. A common procedure to overcome this problem is to drop from the original panel the individuals for which the observations are not complete and carry out the estimation on a complete sub-panel. However, as discussed in Mátyás-Lovrics (1991) and Baltagi-Chang (1994), when the sample size is moderate, this procedure may incur considerable loss of efficiency.

Specifying parametric variance functions also presents some attractive features. First, this strategy avoids incidental parameter (and thus consistency) problems arising from any attempt to model changing variances by grouped heteroscedasticity when the number of individual units is large but the number of observations per individual is small, i.e., in typical microeconomic panel datasets. This is particularly obvious if we want heteroscedasticity to apply to both the individual-specific and general error components of the model. Of course, following this strategy requires that we are able (or willing) to pick up the variables which enter into the variance functions as well as the variance functions themselves. Second, provided that the functional forms of the variance functions are judiciously chosen, it prevents problems due to estimated variances being negative or zero. As a matter of fact, Baltagi-Griffin (1988) reports negative variance estimates and numerical problems are mentioned in Randolf's (1988b) empirical study. Finally, since the conditional variance estimates may have intrinsic values of their own as indicators of the between and within individual heterogeneity, parametric forms are convenient for ease of interpretation.

The heuristic background for allowing heteroscedasticity to distinctly apply to both individual-specific and general error components is the following. Just as the composite error term in panel data, in cross-section data the error term reflects both variations between individuals and variations between repeated observations of an individual (within variations). The only difference is that in the latter case, there is no way to disentangle the two effects. Thus, all we said about the possible sources of heteroscedasticity in cross-section may be roughly applied to the panel data composite error term. Then, the remaining question is to determine the most plausible within and between scedastic patterns underlying a given overall (cross-section like) heteroscedastic structure, i.e. to assess the origin — within and/or between — of the composite disturbance variance heterogeneity. Clearly, the answer depends upon the situation at hand. Consider heteroscedasticity arising from differences in size. In this case, both error terms may be expected heteroscedastic, presumably (but not necessarily) according to parallel patterns. Indeed, assuming homoscedasticity for one of the two error terms would amount to considering that the unobservable effects associated with this term are all of the same (absolute) magnitude whatever the size of the individual units. This is very unrealistic[5]. As a matter of fact, this argument is implicitly acknowledged whenever a transformation of the dependent variable is used for solving heteroscedasticity problems (the transformation alters the distribution of both error terms). Likewise, if size-related heteroscedasticity still prevails after having transformed the dependent variable, the same reasoning should apply, although in this situation the two scedastic patterns might be substantially divergent. When heteroscedasticity may not be directly associated with size, it seems much more difficult to say anything general. Hanging on the nature of the relationship under investigation, either only one of the two or both error terms might be expected heteroscedastic. Note that in the latter case, their variances might further depend upon different sets of variables. Collecting all these considerations, as a general setting it thus appears sensible to allow heteroscedasticity to distinctly apply to both individual-specific and general error components. Doing this means adopting an a priori quite flexible parametrization allowing for variable heterogeneity both in the between and within dimensions.

Such a full heteroscedastic one-way error components model is nothing more than a static multivariate second order semi-parametric model. The general results obtained in Chapter 1 and 2 may thus be exploited for its estimation and specification testing. On the grounds of its ability to straightforwardly handle incomplete (unbalanced) panels, its robustness to distributional misspecification and possible misspecification of the heterogeneity, its computational convenience and its potential efficiency, we argue for estimating this model by gaussian pseudo-maximum likelihood of order two. Consequently, we provide all the required ingredients needed for its practical implementation, and review its limiting properties and asymptotic covariance matrix estimation under the major assumptions of practical interest regarding the degree of misspecification present in the model.

Then, as an adapted handy synthesis of the general results obtained in Chapter 2, we review the different ways in which the correct specification of the prominent aspects of the assumed full heteroscedastic one-way error components model may be tested. So, are succinctly surveyed potentially useful nested, non-nested, Hausman and information matrix type diagnostic tests of both the mean and the variance specification.

Since the estimation of the model, although quite straightforward, is computationally expensive[6], it seems wise to check its potential relevance before undertaking

[5] Graag (1993) argues from the same reasoning to justify its bilinear heteroscedastic panel data model.

[6] Actually, it is mainly expensive in programming time. At this respect, easy-to-use procedures for Gauss including gaussian pseudo-maximum likelihood estimation and comprehensive specification testing of the model are available (free of charge) upon request from the author. To run properly, they require Gauss for Windows v.3.2 and the Gauss Optimization Application Module v.3.1.

the estimation procedure. In order to meet this prerequisite, using the general results obtained in Chapter 2, we finally derive a simple pseudo Lagrange Multiplier (LM) test statistic (based on OLS residuals) for jointly testing the null of no individual effects and homoscedasticity against the alternative of (possibly heteroscedastic) random individual effects and general form of heteroscedasticity (a set of locally equivalent alternatives) in the usual white noise error term. If independence of the errors is assumed under the null, the joint test statistic turns out to be simply the sum of two asymptotically independent pseudo LM statistics, allowing for easily gaining insights about the direction(s) in which misspecification detected by the joint statistic may lie.

The chapter proceeds as follows. Section 3.2 describes the model under consideration. Section 3.3 provides all the required ingredients for performing gaussian pseudo-maximum likelihood of order 2 estimation and discusses practical ways for obtaining the estimator as well as its limiting properties and asymptotic covariance matrix estimation. In this section, special attention is given to provide matrix expressions such that they only include matrices of moderate size and that they can be straightforwardly implemented with a matrix-oriented programming language. Section 3.4 deals with specification testing of the model. Preliminary pseudo Lagrange multiplier testing is developed in section 3.5. Finally, concluding comments are offered in Section 3.6. As in the previous chapters, matrix calculus notational conventions are in accordance with those of Magnus-Neudecker (1986,1988).

3.2. The model

We consider the following one-way error components linear regression model

$$Y_{it} = X_{it}\beta + \varepsilon_{it}, \quad \varepsilon_{it} = \mu_i + \nu_{it}, \qquad i = 1, 2, ..., n\ ; \quad t = 1, 2, ..., T_i \tag{3.1}$$

where Y_{it}, ε_{it}, μ_i and ν_{it} are scalars, X_{it} is a $1 \times k_\beta$ vector of explanatory variables (the first element being a constant) and $\beta \in \Theta_\beta$ is a $k_\beta \times 1$ vector of parameters. The index i refers to the n individuals and the index t to the (repeated) observations (over time) of each individual i. The total number of observations is $N = \sum_{i=1}^{n} T_i$. The observations are assumed to be independently distributed across individuals.

Stacking the T_i observations of each individual i, (3.1) yields the multivariate linear regression model

$$Y_i = X_i\beta + \varepsilon_i, \quad \varepsilon_i = e_{T_i}\mu_i + \nu_i, \qquad i = 1, 2, ..., n \tag{3.2}$$

where e_{T_i} is a $T_i \times 1$ vector of ones, Y_i, ν_i and ε_i are $T_i \times 1$ vectors, and X_i is a $T_i \times k_\beta$ matrix of explanatory variables.

Let Z_i^1 denote a $T_i \times k_{\gamma_1}$ matrix of explanatory variables (the first column being a constant), Z_{it}^1 stand for the t-th row of Z_i^1, and Z_i^2 be a $1 \times k_{\gamma_2}$ vector of explanatory variables (the first element being a constant). For all i, t and t', the error terms ν_{it} and μ_i are tentatively assumed to satisfy the assumptions

$$E(\nu_{it}|X_i, Z_i^1, Z_i^2) = 0, \qquad E(\mu_i|X_i, Z_i^1, Z_i^2) = 0 \tag{3.3}$$

$$E\left(\nu_{it}\nu_{it'}|X_i, Z_i^1, Z_i^2\right) = 0 \ (t' \neq t), \quad E\left(\mu_i\nu_{it}|X_i, Z_i^1, Z_i^2\right) = 0 \tag{3.4}$$

$$V\left(\nu_{it}|X_i, Z_i^1, Z_i^2\right) = \sigma^2_{\nu_{it}} = \phi_\nu(Z_{it}^1\gamma_1), \quad V\left(\mu_i|X_i, Z_i^1, Z_i^2\right) = \sigma^2_{\mu_i} = \phi_\mu(Z_i^2\gamma_2) \tag{3.5}$$

where $\phi_\mu(.)$ and $\phi_\nu(.)$ are arbitrary non-indexed (strictly) positive twice continuously differentiable functions while $\gamma_1 \in \Theta_{\gamma_1}$ and $\gamma_2 \in \Theta_{\gamma_2}$ are, respectively, $k_{\gamma_1} \times 1$ and $k_{\gamma_2} \times 1$ vectors of parameters which vary independently of each other and independently of β. Hereafter, we will denote by $\gamma = \left(\gamma_1', \gamma_2'\right)'$ the vector of variance-specific parameters and $\theta = (\beta', \gamma')'$ will stand for the entire set of parameters.

The regressors appearing in the conditional variances (3.5) may (and usually will) be related to the X_i. Different choices are possible for the variance functions $\phi_\nu(.)$ and $\phi_\mu(.)$, see for example Breush-Pagan (1979) and Harvey (1976). Among them, the multiplicative heteroscedasticity formulation investigated in Harvey (1976) appears particularly attractive. It simply means taking $\phi_\nu(.) = \phi_\mu(.) = \exp(.)$.

Under (3.3)-(3.5), ε_i is easily seen to satisfy

$$\begin{aligned} &E(\varepsilon_i|X_i, Z_i^1, Z_i^2) = 0, \qquad i = 1, 2, ..., n \\ &V(\varepsilon_i|X_i, Z_i^1, Z_i^2) = \Omega_i = \text{diag}\left(\phi_\nu(Z_i^1\gamma_1)\right) + J_{T_i}\, \phi_\mu(Z_i^2\gamma_2) \end{aligned} \tag{3.6}$$

where $J_{T_i} = e_{T_i}e_{T_i}'$ and, for a $T_i \times 1$ vector x , the non-indexed function $\phi(x)$ denotes a $T_i \times 1$ vector containing the element-by-element transformation $\phi(.)$ of the elements of x.

The model may thus be written as

$$\begin{aligned} &E(Y_i|X_i, Z_i^1, Z_i^2) = X_i\beta, \qquad i = 1, 2, ..., n \\ &V(Y_i|X_i, Z_i^1, Z_i^2) = \Omega_i = \text{diag}\left(\phi_\nu(Z_i^1\gamma_1)\right) + J_{T_i}\, \phi_\mu(Z_i^2\gamma_2) \end{aligned} \tag{3.7}$$

This (possibly unbalanced) second order semi-parametric model obviously contains the standard (homoscedastic) one-way error components linear regression model as a special case. Following the definitions of Chapter 1, it will be said correctly specified for the conditional mean if, for some true-value $\beta^o \in \Theta_\beta$, we have $E(Y_i|X_i, Z_i^1, Z_i^2) = X_i\beta^o$, $i = 1, 2, ..., n$. Likewise, it will be said correctly specified for the conditional variance if, for some true-value $\gamma^o = (\gamma_1^{o\prime}, \gamma_2^{o\prime})' \in \Theta_{\gamma_1} \times \Theta_{\gamma_2}$, we have $V(Y_i|X_i, Z_i^1, Z_i^2) = \Omega_i^o = \text{diag}\left(\phi_\nu(Z_i^1\gamma_1^o)\right) + J_{T_i}\, \phi_\mu(Z_i^2\gamma_2^o)$, $i = 1, 2, ..., n$.

Note that these definitions of correct specification implicitly embody the ignorability of the selection mechanism (or missing data generating mechanism) giving rise to the eventual incompleteness of the panel dataset : if, to use Verbeek-Nijmans' (1996) terminology, the selection mechanism is not ignorable, model (3.7) will usually be misspecified. For a formal account of the concept of ignorability, see Verbeek-Nijmans (1996). Note also that, both conditional mean and conditional variance correct specification are defined with respect to the entire set of conditioning variables $CV_i \equiv (X_i, Z_i^1, Z_i^2)$, and not only with respect to the variables which actually enter in the mean or the variance specification. In other words, specifying the variances as functions of variables which do not enter in the mean may dismantle an original conditional mean correct specification, and vice versa.

3.3. Pseudo-maximum likelihood of order two estimation

The most popular estimator of the standard one-way error components model is undoubtedly the feasible generalized least squares (FGLS) estimator. In the present context where the disturbances are fully heteroscedastic, this gaussian quasi-generalized pseudo-maximum likelihood of order one estimator is no longer so attractive. Indeed, such an estimator requires a preliminary consistent ($n \to \infty$, T_i bounded) estimator of the conditional variance parameters appearing in the Ω_i. Given the general functional forms adopted for the variance functions — the problem would be different if the variance functions were assumed linear —, no simple, i.e., avoiding nonlinear optimization, two-step procedure for obtaining such a consistent estimator seems conceivable. In contrast, the gaussian pseudo-maximum likelihood of order two (GPML2) estimator also requires nonlinear optimization but simultaneously provides mean and variance parameters estimates. On the other hand, according to Chapter 1, because the normal density belongs to restricted quadratic exponential families and the model contains no functional links between mean and variance parameters, as FGLS, GPML2 is not only robust to distributional misspecification but also to conditional variance misspecification. So, from a statistical point of view, GPML2 has essentially the same properties than FGLS regarding the mean parameters — in this linear case, the GPML2 mean parameters estimator is actually just an FGLS estimator where the variance parameters are "endogenously" determined — while it offers additional by-product properties regarding the variance parameters. Among them, under favorable circumstances, the variance estimator may be asymptotically efficient. This is appreciable as far as the conditional variance estimates have intrinsic values of their own as indicators of the heterogeneity. Finally adding to these characteristics that it readily allows to handle incomplete panels, GPML2 thus appears as a very attractive — both from a computational and statistical point of view — estimation procedure.

Gaussian maximum likelihood estimation of the standard (homoscedastic) complete and incomplete panel one-way error components models are, among others, respectively discussed in Breush (1987) and in Baltagi-Chang (1994). In the following sub-section, we provide the basic ingredients needed for performing GPML2 estimation of model (3.7), namely the pseudo log-likelihood function and its derivatives. Subsequently, we discuss practical ways for obtaining the GPML2 estimates and detail the limiting properties and asymptotic covariance matrix estimation of the estimator.

3.3.1. The pseudo log-likelihood function and its derivatives

The GPML2 estimator $\hat{\theta}_n = \left(\beta_n', \gamma_{1_n}', \gamma_{2_n}'\right)'$ of model (3.7) is defined is a solution of

$$\text{Max}_{\theta \in \Theta}\, L_n(\beta, \gamma_1, \gamma_2) \equiv \frac{1}{n} \sum_{i=1}^{n} L_i(Y_i | X_i, Z_i^1, Z_i^2; \beta, \gamma_1, \gamma_2) \qquad (3.8)$$

where $\Theta = \Theta_\beta \times \Theta_{\gamma_1} \times \Theta_{\gamma_2}$ and the (conditional) pseudo log-likelihood functions

$L_i(Y_i|X_i, Z_i^1, Z_i^2; \beta, \gamma_1, \gamma_2)$ are

$$L_i(Y_i|X_i, Z_i^1, Z_i^2; \beta, \gamma_1, \gamma_2) = -\frac{T_i}{2}\ln 2\pi - \frac{1}{2}\ln|\Omega_i| - \frac{1}{2}u_i'\Omega_i^{-1}u_i$$

with $u_i = Y_i - X_i\beta$.

Analytical expressions are available for $|\Omega_i|$ and Ω_i^{-1}. These are given by

$$\begin{aligned}
|\Omega_i| &= (b_i)^{T_i}|C_i|\left(1+\operatorname{tr}C_i^{-1}\right) = \left(\prod_{t=1}^{T_i} a_{it}\right)\left(1+e_{T_i}'\bar{c}_i\right) \\
\Omega_i^{-1} &= \frac{1}{b_i}\left(C_i^{-1} - \frac{1}{1+\operatorname{tr}C_i^{-1}}\left(C_i^{-1}J_{T_i}C_i^{-1}\right)\right) \\
&= \operatorname{diag}(\bar{a}_i) - \frac{1}{b_i\left(1+e_{T_i}'\bar{c}_i\right)}\bar{c}_i\bar{c}_i'
\end{aligned}$$

where

$$\begin{aligned}
b_i &= \phi_\mu(Z_i^2\gamma_2) & C_i &= \operatorname{diag}(c_i) \\
c_i &= \frac{1}{b_i}\phi_\nu(Z_i^1\gamma_1) & \bar{c}_i &= e_{T_i} \div c_i \\
a_i &= \phi_\nu(Z_i^1\gamma_1) & \bar{a}_i &= e_{T_i} \div a_i
\end{aligned}$$

a_{it} being the t-th element of a_i and $\div$ indicating an element-by-element division. Note that according to this notation, $\Omega_i = b_i\left(C_i + J_{T_i}\right)$.

Following Magnus (1978,1988), we obtain for the first derivatives of L_n

$$\frac{\partial L_n}{\partial\theta} = \frac{1}{n}\sum_{i=1}^{n} s_i^\theta, \qquad s_i^\theta = \begin{bmatrix} s_i^\beta \\ s_i^{\gamma_1} \\ s_i^{\gamma_2} \end{bmatrix}$$

with

$$s_i^\beta = X_i'\Omega_i^{-1}u_i \tag{3.9}$$

$$\begin{aligned}
s_i^{\gamma_p} &= -\frac{1}{2}\left(\frac{\partial\operatorname{vec}\Omega_i^{-1}}{\partial\gamma_p'}\right)'\operatorname{vec}(u_iu_i' - \Omega_i) \qquad (p = 1,2) \qquad (3.10) \\
&= \frac{1}{2}\left(\frac{\partial\operatorname{vec}\Omega_i}{\partial\gamma_p'}\right)'\left(\Omega_i^{-1}\otimes\Omega_i^{-1}\right)\operatorname{vec}(u_iu_i' - \Omega_i) \\
&= \frac{1}{2}\left(\frac{\partial\operatorname{vec}\Omega_i}{\partial\gamma_p'}\right)'\operatorname{vec}\left(\Omega_i^{-1}(u_iu_i' - \Omega_i)\Omega_i^{-1}\right) \\
&= -\frac{1}{2}\sum_{r=1}^{k_{\gamma_p}}\operatorname{tr}\left(\frac{\partial\Omega_i^{-1}}{\partial\gamma_p^r}(u_iu_i' - \Omega_i)\right)e_{k_{\gamma_p}}^r \\
&= \frac{1}{2}\sum_{r=1}^{k_{\gamma_p}}\operatorname{tr}\left(\Omega_i^{-1}\frac{\partial\Omega_i}{\partial\gamma_p^r}\Omega_i^{-1}(u_iu_i' - \Omega_i)\right)e_{k_{\gamma_p}}^r
\end{aligned}$$

where $e^r_{k_{\gamma_p}}$ is a $k_{\gamma_p} \times 1$ vector with a one in the r-th place and zeros elsewhere, i.e., the r-th column of a $k_{\gamma_p} \times k_{\gamma_p}$ identity matrix, and γ_p^r is the r-th component of γ_p.

Some useful identities for decoding (3.10) as well as subsequent expressions are given by

$$\begin{aligned} &\frac{\partial \operatorname{vec} \Omega_i^{-1}}{\partial \gamma_p'} = -\left(\Omega_i^{-1} \otimes \Omega_i^{-1}\right) \frac{\partial \operatorname{vec} \Omega_i}{\partial \gamma_p'} && \frac{\partial \Omega_i^{-1}}{\partial \gamma_p^r} = -\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \gamma_p^r} \Omega_i^{-1} \\ &\operatorname{vec} \Omega_i^{-1} = \left(\Omega_i^{-1} \otimes \Omega_i^{-1}\right) \operatorname{vec} \Omega_i && \operatorname{vec}(ABC) = (C' \otimes A) \operatorname{vec} B \\ &\frac{\partial \operatorname{vec} \Omega_i^{-1}}{\partial \gamma_p'} = \sum_{r=1}^{k_{\gamma_p}} \operatorname{vec}\left(\frac{\partial \Omega_i^{-1}}{\partial \gamma_p^r}\right) e^{r\prime}_{k_{\gamma_p}} && \operatorname{tr}(A'B) = (\operatorname{vec} A)' \operatorname{vec} B \end{aligned} \tag{3.11}$$

where A, B, C are conformable matrices.

The derivatives of $\operatorname{vec}\Omega_i$ with respect to γ_1' and γ_2' are

$$\frac{\partial \operatorname{vec} \Omega_i}{\partial \gamma_1'} = \operatorname{diag}\left(\operatorname{vec}\left(\operatorname{diag}\left(\phi_\nu'(Z_i^1 \gamma_1)\right)\right)\right) \left(Z_i^1 \otimes e_{T_i}\right) \tag{3.12}$$

$$\frac{\partial \operatorname{vec} \Omega_i}{\partial \gamma_2'} = \phi_\mu'(Z_i^2 \gamma_2) \operatorname{vec}(J_{n_i}) Z_i^2 = \phi_\mu'(Z_i^2 \gamma_2) \left(e_{T_i} \otimes e_{T_i}\right) Z_i^2 \tag{3.13}$$

and the derivatives of Ω_i with respect to γ_1^r and γ_2^r are

$$\frac{\partial \Omega_i}{\partial \gamma_1^r} = \operatorname{diag}\left(\phi_\nu'(Z_i^1 \gamma_1) \odot Z_i^{1^r}\right) \tag{3.14}$$

$$\frac{\partial \Omega_i}{\partial \gamma_2^r} = \phi_\mu'(Z_i^2 \gamma_2) Z_i^{2^r} J_{T_i} \tag{3.15}$$

where $\phi_\nu'(x)$ and $\phi_\mu'(x)$ denote the first derivatives of $\phi_\nu(x)$ and $\phi_\mu(x)$ with respect to x, $Z_i^{1^r}$ is the r-th column of the matrix of explanatory variables Z_i^1, $\odot$ stands for the Hadamard product, i.e., an element-by-element multiplication, and $Z_i^{2^r}$ is the r-th column of the row vector of explanatory variables Z_i^2.

Note that if the multiplicative heteroscedasticity formulation is adopted for both $\phi_\nu(.)$ and $\phi_\mu(.)$, then, in (3.12)-(3.15), $\phi_\nu'(.)$ and $\phi_\mu'(.)$ are simply equal to exp(.).

On the other hand, following again Magnus (1978,1988), we obtain for the hessian matrix of L_n

$$\frac{\partial^2 L_n}{\partial \theta \partial \theta'} = \frac{1}{n} \sum_{i=1}^{n} h_i^{\theta\theta}, \qquad h_i^{\theta\theta} = \begin{bmatrix} h_i^{\beta\beta} & h_i^{\beta\gamma_1} & h_i^{\beta\gamma_2} \\ h_i^{\gamma_1\beta} & h_i^{\gamma_1\gamma_1} & h_i^{\gamma_1\gamma_2} \\ h_i^{\gamma_2\beta} & h_i^{\gamma_2\gamma_1} & h_i^{\gamma_2\gamma_2} \end{bmatrix}$$

with

$$h_i^{\beta\beta} = \frac{\partial s_i^\beta}{\partial \beta'} = -X_i' \Omega_i^{-1} X_i \tag{3.16}$$

$$h_i^{\beta\gamma_p} = \frac{\partial s_i^\beta}{\partial \gamma_p'} = \left(\frac{\partial s_i^{\gamma_p}}{\partial \beta'}\right)' = h_i^{\gamma_p\beta\prime} \qquad (p = 1, 2) \tag{3.17}$$

$$
\begin{aligned}
&= (u_i' \otimes X_i') \frac{\partial \operatorname{vec} \Omega_i^{-1}}{\partial \gamma_p'} = -(u_i' \otimes X_i') \left(\Omega_i^{-1} \otimes \Omega_i^{-1} \right) \frac{\partial \operatorname{vec} \Omega_i}{\partial \gamma_p'} \\
&= - \left(u_i' \Omega_i^{-1} \otimes X_i' \Omega_i^{-1} \right) \frac{\partial \operatorname{vec} \Omega_i}{\partial \gamma_p'} \\
&= \sum_{r=1}^{k_{\gamma_p}} \left(X_i' \frac{\partial \Omega_i^{-1}}{\partial \gamma_p^r} u_i \right) e_{k_{\gamma_p}}^{r\prime} = - \sum_{r=1}^{k_{\gamma_p}} \left(X_i' \Omega_i^{-1} \frac{\partial \Omega_i}{\partial \gamma_p^r} \Omega_i^{-1} u_i \right) e_{k_{\gamma_p}}^{r\prime}
\end{aligned}
$$

$$
\begin{aligned}
h_i^{\gamma_p \gamma_q} &= \frac{\partial s_i^{\gamma_p}}{\partial \gamma_q'} = \left(\frac{\partial s_i^{\gamma_q}}{\partial \gamma_p'} \right)' = h_i^{\gamma_q \gamma_p \prime} \qquad (p = 1, 2\,; \quad q = 1, 2) \qquad (3.18) \\
&= \frac{1}{2} \left(\frac{\partial \operatorname{vec} \Omega_i^{-1}}{\partial \gamma_p'} \right)' \frac{\partial \operatorname{vec} \Omega_i}{\partial \gamma_q'} - \frac{1}{2} \left((\operatorname{vec}(u_i u_i' - \Omega_i))' \otimes I_{k_{\gamma_p}} \right) \Upsilon_i^{\gamma_p \gamma_q} \\
&= -\frac{1}{2} \left(\frac{\partial \operatorname{vec} \Omega_i}{\partial \gamma_p'} \right)' \left(\Omega_i^{-1} \otimes \Omega_i^{-1} \right) \frac{\partial \operatorname{vec} \Omega_i}{\partial \gamma_q'} - \frac{1}{2} \left((\operatorname{vec}(u_i u_i' - \Omega_i))' \otimes I_{k_{\gamma_p}} \right) \Upsilon_i^{\gamma_p \gamma_q} \\
&= \frac{1}{2} \sum_{r=1}^{k_{\gamma_p}} \sum_{s=1}^{k_{\gamma_q}} \left(\operatorname{tr} \left(\frac{\partial \Omega_i^{-1}}{\partial \gamma_p^r} \frac{\partial \Omega_i}{\partial \gamma_q^s} \right) - \operatorname{tr} \left((u_i u_i' - \Omega_i) \frac{\partial^2 \Omega_i^{-1}}{\partial \gamma_p^r \partial \gamma_q^s} \right) \right) e_{k_{\gamma_p}}^r e_{k_{\gamma_q}}^{s\prime} \\
&= -\frac{1}{2} \sum_{r=1}^{k_{\gamma_p}} \sum_{s=1}^{k_{\gamma_q}} \left(\operatorname{tr} \left(\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \gamma_p^r} \Omega_i^{-1} \frac{\partial \Omega_i}{\partial \gamma_q^s} \right) + \operatorname{tr} \left((u_i u_i' - \Omega_i) \frac{\partial^2 \Omega_i^{-1}}{\partial \gamma_p^r \partial \gamma_q^s} \right) \right) e_{k_{\gamma_p}}^r e_{k_{\gamma_q}}^{s\prime}
\end{aligned}
$$

where $I_{k_{\gamma_p}}$ is a $k_{\gamma_p} \times k_{\gamma_p}$ identity matrix,

$$
\Upsilon_i^{\gamma_p \gamma_q} = \frac{\partial \operatorname{vec} \left(\frac{\partial \operatorname{vec} \Omega_i^{-1}}{\partial \gamma_p'} \right)'}{\partial \gamma_q'} = \sum_{r=1}^{k_{\gamma_p}} \sum_{s=1}^{k_{\gamma_q}} \operatorname{vec} \left(\frac{\partial^2 \Omega_i^{-1}}{\partial \gamma_p^r \partial \gamma_q^s} \right) \otimes \left(e_{k_{\gamma_p}}^r e_{k_{\gamma_q}}^{s\prime} \right) \qquad (3.19)
$$

i.e., a $T_i^2 k_{\gamma_p} \times k_{\gamma_q}$ matrix,

$$
\frac{\partial^2 \Omega_i^{-1}}{\partial \gamma_p^r \partial \gamma_q^s} = \Omega_i^{-1} \left(2 \frac{\partial \Omega_i}{\partial \gamma_p^r} \Omega_i^{-1} \frac{\partial \Omega_i}{\partial \gamma_q^s} - \frac{\partial^2 \Omega_i}{\partial \gamma_p^r \partial \gamma_q^s} \right) \Omega_i^{-1} \qquad (3.20)
$$

and the needed derivatives not yet given are[7]

$$
\begin{aligned}
\frac{\partial^2 \Omega_i}{\partial \gamma_1^r \partial \gamma_1^s} &= \operatorname{diag} \left(\phi_\nu'' \left(Z_i^1 \gamma_1 \right) \odot Z_i^{1^r} \odot Z_i^{1^s} \right) \\
\frac{\partial^2 \Omega_i}{\partial \gamma_1^r \partial \gamma_2^s} &= 0 = \frac{\partial^2 \Omega_i}{\partial \gamma_2^r \partial \gamma_1^s} \\
\frac{\partial^2 \Omega_i}{\partial \gamma_2^r \partial \gamma_2^s} &= \phi_\mu'' \left(Z_i^2 \gamma_2 \right) Z_i^{2^r} Z_i^{2^s} J_{T_i}
\end{aligned}
$$

where $\phi_\nu''(x)$ and $\phi_\mu''(x)$ denote the second derivatives of $\phi_\nu(x)$ and $\phi_\mu(x)$ with respect

[7] Note that (3.20) relies on the symmetry of the matrices at hand. Notice also that an expression similar to the first one given in (3.11) may be derived for the second derivatives. However, this relation is only of theoretical interest since it implies very large and thus untractable go-between matrices.

to x. If the multiplicative heteroscedasticity formulation is adopted for both $\phi_\nu(.)$ and $\phi_\mu(.)$, $\phi''_\nu(.)$ and $\phi''_\mu(.)$ are again simply equal to exp(.).

In addition to the relations given by (3.11) and (3.19), for verifying the equality of the alternative forms of (3.17) and (3.18), it is worth noting that

$$(\text{vec}(u_i u_i' - \Omega_i))' \otimes I_{k_{\gamma_p}} = \sum_{r=1}^{k_{\gamma_p}} (\text{vec}(u_i u_i' - \Omega_i))' \otimes \left(e^r_{k_{\gamma_p}} e^{r\prime}_{k_{\gamma_p}}\right)$$

and

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$$

where A, B, C and D are conformable matrices.

3.3.2. Numerical optimization and starting values

For obtaining the GPML2 estimates, we need two more ingredients: a numerical algorithm for maximizing (3.8) and a set of starting values.

In the context of the standard complete panel one-way error components model, Breush (1987) discusses an iterated GLS procedure. Although applicable in very general situations (see Magnus (1978)), in the present case it is not very attractive since it implies at each step the (numerical) resolution of the highly nonlinear set of equations defined by the first-order conditions $\frac{\partial L_n}{\partial \gamma_p} = 0$ $(p = 1, 2)$.

As alternatives, we can use either a Newton or quasi-Newton (secant methods) algorithm. While the former requires the computation of the first and second derivatives, the latter (for example, the so-called Davidson-Fletcher-Powell and Broyden-Fletcher-Goldfard-Shanno methods) requires only the computation of the first derivatives (see Quandt (1983) or Harvey (1981)). In the present case, a variant of the Newton method appears particularly appealing, namely the scoring method. This variant simply means substituting the hessian used in the Newton algorithm by the empirical counterpart of its expectation $A^o_{n\theta\theta}$ under conditional mean and conditional variance correct specification.

Using the law of iterated expectation and the fact that under conditional mean and conditional variance correct specification we have $E(u^o_i | X_i, Z^1_i, Z^2_i) = 0$ and $E\left((u^o_i u^{o\prime}_i - \Omega^o_i) | X_i, Z^1_i, Z^2_i\right)$, it is easily checked that

$$A^o_{n\theta\theta} = \frac{1}{n} \sum_{i=1}^{n} E\left[h^{\theta\theta o}_i\right], \quad E\left[h^{\theta\theta o}_i\right] = E\left[\underline{h}^{\theta\theta o}_i\right] = E \begin{bmatrix} \underline{h}^{\beta\beta o}_i & 0 & 0 \\ 0 & \underline{h}^{\gamma_1\gamma_1 o}_i & \underline{h}^{\gamma_1\gamma_2 o}_i \\ 0 & \underline{h}^{\gamma_2\gamma_1 o}_i & \underline{h}^{\gamma_2\gamma_2 o}_i \end{bmatrix}$$

where

$$\begin{aligned} \underline{h}^{\beta\beta}_i &= h^{\beta\beta}_i = -X_i' \Omega_i^{-1} X_i \\ \underline{h}^{\gamma_p\gamma_q}_i &= \underline{h}^{\gamma_q\gamma_p\prime}_i \qquad (p = 1, 2\,; \quad q = 1, 2) \end{aligned} \tag{3.21}$$

$$= -\frac{1}{2}\left(\frac{\partial \operatorname{vec}\Omega_i}{\partial \gamma_p'}\right)'\left(\Omega_i^{-1}\otimes\Omega_i^{-1}\right)\frac{\partial \operatorname{vec}\Omega_i}{\partial \gamma_q'}$$

$$= -\frac{1}{2}\sum_{r=1}^{k_{\gamma_p}}\sum_{s=1}^{k_{\gamma_q}} \operatorname{tr}\left(\Omega_i^{-1}\frac{\partial\Omega_i}{\partial\gamma_p^r}\Omega_i^{-1}\frac{\partial\Omega_i}{\partial\gamma_q^s}\right) e_{k_{\gamma_p}}^r e_{k_{\gamma_q}}^{s\prime}$$

and the superscript '^{o}' denotes quantities evaluated at $\theta^o = (\beta^{o\prime}, \gamma^{o\prime})'$.

The scoring method thus means replacing in the Newton algorithm $\frac{\partial^2 L_n}{\partial\theta\partial\theta'} = \frac{1}{n}\sum_{i=1}^n h_i^{\theta\theta}$ by $\frac{1}{n}\sum_{i=1}^n \underline{h}_i^{\theta\theta}$. The latter is considerably simpler : it is block-diagonal and only involves first derivatives. It will be a good approximation of the hessian if the model is correctly specified and θ is not too far from θ^o. According to our experience, even under quite severe misspecification, provided that all quantities are analytically computed, the scoring method generally converges in less time (more computation time per iteration but fewer iterations) than the secant methods. Further, since the empirical expected hessian is always negative semidefinite, it is numerically stable.

Note that from a computational point of view the "vec" and the "trace" formulations of all the above expressions are not at all equivalent. Indeed, hanging on the number of observations per individual and on the number of variables entering into the variance functions, they may entail quite substantially different computation times to complete. For example, using Gauss, if $T_i = 2$ ($\forall\, i$) and $k_{\gamma_1} = k_{\gamma_2} = 10$, the computational times ratios between the "trace" and the "vec" formulations (i.e., "trace"over "vec") are about equal to 6.8 for the gradient and 30.7 for the empirical expected hessian. On the contrary, if $T_i = 10$ ($\forall\, i$) and $k_{\gamma_1} = k_{\gamma_2} = 2$, the same ratio are respectively about equal to 0.52 and 0.47. Not taking this fact into account when practically implementing the estimation procedure would be very inefficient.

A sensible set of starting values for the above algorithms may be computed by proceeding as follows.

(1) Obtain the $\hat{\underline{\beta}}$ and $\widehat{\alpha} = (\widehat{\alpha}_1, ..., \widehat{\alpha}_i, ..., \widehat{\alpha}_n)$ OLS estimates of the dummy variables model $Y_i = \alpha_i + \underline{X}_i\underline{\beta} + \upsilon_i$ $(i = 1, 2, ..., n)$, where $\underline{X}_i$ is the same as X_i except its dropped first column[8]. At this stage, $\hat{\underline{\beta}}$ and the mean of the $\widehat{\alpha}_i$, i.e., $\overline{\alpha} = \frac{1}{n}\sum_{i=1}^n \hat{\alpha}_i$, provide initial values for β.

(2) Run the OLS regression $\phi_\nu^{-1}(\hat{u}_{it}^2) = Z_{it}^1\gamma_1 + \upsilon_{it}$ $(i = 1, 2, ..., n;\ t = 1, 2, ..., T_i)$, where $\hat{u}_{it} = Y_{it} - \hat{\alpha}_i - \underline{X}_{it}\hat{\underline{\beta}}$ and $\phi_\nu^{-1}(.)$ is the (supposed well-defined) inverse function of $\phi_\nu(.)$. The non-intercept parameters of $\hat{\gamma}_1$ and the intercept parameter of $\hat{\gamma}_1$ minus γ_{1_c}, where γ_{1_c} is an intercept correction term[9], give initial values for γ_1. The "optimal" value of the intercept correction term γ_{1_c} depends of course upon the functional form $\phi_\nu^{-1}(.)$ and the actual distribution of the ν_{it}. In the case of the multiplicative heteroscedasticity formulation where $\phi_\nu^{-1}(.)$ is simply equal to ln(.), a sensible choice is[10] $\gamma_{1_c} = -1.2704$.

[8] As usual, $\hat{\underline{\beta}}$ may be computed as $\hat{\underline{\beta}} = (\sum_{i=1}^n \underline{X}_i' M_{T_i}\underline{X}_i)^{-1}\sum_{i=1}^n \underline{X}_i' M_{T_i} Y_i$ (WOLS estimator) and $\widehat{\alpha}_i = \frac{1}{T_i}e_{T_i}'(Y_i - \underline{X}_i\hat{\underline{\beta}})$, where $M_{T_i} = I_{T_i} - \frac{1}{T_i}J_{T_i}$, i.e., a within transformation matrix. See Balestra (1996) for details.

[9] The "desirability" of an intercept correction of $\hat{\gamma}_1$ arises from the fact that in the regression $\phi_\nu^{-1}(\hat{u}_{it}^2) = Z_{it}^1\gamma_1 + \upsilon_{it}$, even if we supposed that $\hat{u}_{it}$ is equal to the true disturbance ν_{it}, the (conditional) expectation of the error term υ_{it} is not necessarily zero, not even necessarily a constant.

(3) Finally, run the OLS regression $\phi_\mu^{-1}((\hat{\alpha}_i-\overline{\alpha})^2) = Z_i^2\gamma_2+\upsilon_i$ $(i=1,2,...,n)$, where $\phi_\mu^{-1}(.)$ is the (supposed well-defined) inverse function of $\phi_\mu(.)$. According to the same reasoning than above, the non-intercept parameters of $\hat{\gamma}_2$ and the intercept parameter of $\hat{\gamma}_2$ minus γ_{2_c}, where γ_{2_c} is an intercept correction term, give initial values for γ_2. In the case of the multiplicative heteroscedasticity formulation where ϕ_μ^{-1} is again equal to ln(.), γ_{2_c} should also be set to -1.2704.

Note that a simpler alternative to the step 2 and 3 is workable. It merely consists in computing the "mean variance components" $\widehat{\sigma}_\nu^2 = \frac{1}{N}\sum_{i=1}^n\sum_{t=1}^{T_i}\hat{u}_{it}^2$ and $\widehat{\sigma}_\mu^2 = \frac{1}{n}\sum_{i=1}^n(\hat{\alpha}_i-\overline{\alpha})^2$. The inverse function values $\phi_\nu^{-1}(\widehat{\sigma}_\nu^2)$ and $\phi_\mu^{-1}(\widehat{\sigma}_\mu^2)$ may then be used for the first elements (intercepts) of γ_1 and γ_2, their remaining elements being set to zero.

To conclude this section, two final remarks. First, as shown by Maddala (1971) and further discussed in Breush (1987) for the standard one-way error components model, (3.8) may allow multiple local maxima. Therefore, in practice, it is wise to check for this potential problem by starting the optimization from different sets of initial values. Second, we want to stress the fact that GPML2 estimation of the model is not as cumbersome as it may appear at first sight. Actually, although expensive in programming time, it does not at all entail impractical computational time.

3.3.3. Limiting properties of the estimator and asymptotic covariance matrix estimation

As already pointed out, since the model contains no functional links between mean and variance parameters and the normal density belongs to restricted quadratic exponential families, the GPML2 estimator defined by (3.8) is actually a robust pseudo-maximum likelihood of order two estimator (RPML2, or more precisely GRPML2). In the balanced case, the general results of Chapter 1 thus directly apply. Provided that suitable regularity conditions hold, they may likewise be shown to hold in the unbalanced case.

So, following Proposition 9 and 13, whatever the misspecification of the model, we have that

$$\hat{\theta}_n - \theta_n^* \xrightarrow{a.s.} 0, \text{ as } n \to \infty \ (T_i \text{ bounded})$$

and

$$\sqrt{n}(\hat{\theta}_n - \theta_n^*) \approx \mathrm{N}(0, C_n^*)$$

where $\theta_n^* = \left(\beta_n^{*\prime}, \gamma_{1_n}^{*\prime}, \gamma_{2_n}^{*\prime}\right)'$ and

$$C_n^* = A_n^{*^{-1}}\ddot{B}_n^* A_n^{*^{-1}}$$

[10] This follows from the fact that $E[\ln(\nu_{it}^2)-\ln(\sigma_{\nu_{it}}^2)] = E[\ln(\nu_{it}^2/\sigma_{\nu_{it}}^2)] = -1.2704$ if $\nu_{it} \sim \mathrm{N}(0,\sigma_{\nu_{it}}^2)$. See Harvey (1976).

with

$$A_n^* = \frac{1}{n}\sum_{i=1}^{n} E\left[h_i^{\theta\theta *}\right], \quad \ddot{B}_n^* = \frac{1}{n}\sum_{i=1}^{n} E\left[s_i^{\theta *} s_i^{\theta * \prime}\right] - U_n^*, \quad U_n^* = \frac{1}{n}\sum_{i=1}^{n} E\left(s_i^{\theta *}\right) E\left(s_i^{\theta *}\right)'$$

the superscript '*' denoting quantities evaluated at θ_n^*.

In other words, regardless of arbitrary misspecification, $\hat{\theta}_n$ is consistent for some pseudo-true value θ_n^* and asymptotically normally distributed with asymptotic covariance matrix C_n^*. Note that the relatively simple form of $\ddot{B}_n^*$ follows from the assumption that the observations are independent across individuals.

According to Proposition 7 and 12, depending on the extent of the correct specification of our tentative model, GPML2 will yield a consistent estimator of either the mean or the mean and variance parameters true values, and the form of its asymptotic covariance matrix — and consequently, the way it may be estimated — will accordingly change. Three cases are of practical interest. They are reviewed hereafter.

The easiest — and unfortunately less likely — situation is when not only both the conditional mean and the conditional variance are correctly specified, but in addition normality also holds. In this case, GPML2 is just a standard maximum likelihood estimator and we have the standard results that

$$\hat{\theta}_n \xrightarrow{a.s.} \theta^o, \text{ as } n \to \infty \ (T_i \text{ bounded})$$

and

$$\sqrt{n}(\hat{\theta}_n - \theta^o) \approx \mathrm{N}(0, \bar{C}_n^o)$$

with

$$\bar{C}_n^o = \begin{bmatrix} \bar{C}^o_{n_{\beta\beta}} & 0 \\ 0 & \bar{C}^o_{n_{\gamma\gamma}} \end{bmatrix} = \begin{bmatrix} -A^{o^{-1}}_{n_{\beta\beta}} & 0 \\ 0 & -A^{o^{-1}}_{n_{\gamma\gamma}} \end{bmatrix}$$

where

$$A^o_{n_{\beta\beta}} = \frac{1}{n}\sum_{i=1}^{n} E\left[\underline{h}_i^{\beta\beta o}\right], \quad A^o_{n_{\gamma\gamma}} = \frac{1}{n}\sum_{i=1}^{n} E\left[\underline{h}_i^{\gamma\gamma o}\right] \text{ and } \underline{h}_i^{\gamma\gamma} = \begin{bmatrix} \underline{h}_i^{\gamma_1\gamma_1} & \underline{h}_i^{\gamma_1\gamma_2} \\ \underline{h}_i^{\gamma_2\gamma_1} & \underline{h}_i^{\gamma_2\gamma_2} \end{bmatrix}$$

The asymptotic covariance matrix $\bar{C}_n^o$ may then simply be estimated by replacing $A^o_{n_{\beta\beta}}$ and $A^o_{n_{\gamma\gamma}}$ by their empirical counterpart $\hat{A}_{n_{\beta\beta}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\underline{h}}_i^{\beta\beta}$ and $\hat{A}_{n_{\gamma\gamma}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\underline{h}}_i^{\gamma\gamma}$, where the superscript '$\hat{\ }$' denotes quantities evaluated at $\hat{\theta}_n$. Note that normality is actually not required for the above result to hold : it is sufficient that the third and the fourth order conditional moments of the observations correspond to those of the gaussian distribution.

Distributional misspecification does not affect the consistency of the estimator. It however complicates the form of its asymptotic covariance matrix. So, if the model is only correctly specified for the conditional mean and the conditional variance, we still have that

$$\hat{\theta}_n \xrightarrow{a.s.} \theta^o, \text{ as } n \to \infty \ (T_i \text{ bounded})$$

but now

$$\sqrt{n}(\hat{\theta}_n - \theta^o) \approx \mathrm{N}(0, C_n^o)$$

with

$$C_n^o = \begin{bmatrix} C_{n_{\beta\beta}}^o & C_{n_{\beta\gamma}}^o \\ C_{n_{\beta\gamma}}^{o\prime} & C_{n_{\gamma\gamma}}^o \end{bmatrix} = \begin{bmatrix} -A_{n_{\beta\beta}}^{o^{-1}} & A_{n_{\beta\beta}}^{o^{-1}} B_{n_{\beta\gamma}}^o A_{n_{\gamma\gamma}}^{o^{-1}} \\ A_{n_{\gamma\gamma}}^{o^{-1}} B_{n_{\gamma\beta}}^o A_{n_{\beta\beta}}^{o^{-1}} & A_{n_{\gamma\gamma}}^{o^{-1}} B_{n_{\gamma\gamma}}^o A_{n_{\gamma\gamma}}^{o^{-1}} \end{bmatrix}$$

where

$$B_{n_{\beta\gamma}}^o = \frac{1}{n}\sum_{i=1}^n E\left[s_i^{\beta o} s_i^{\gamma o\prime}\right] = B_{n_{\gamma\beta}}^{o\prime}, \quad B_{n_{\gamma\gamma}}^o = \frac{1}{n}\sum_{i=1}^n E\left[s_i^{\gamma o} s_i^{\gamma o\prime}\right] \quad \text{and} \quad s_i^\gamma = \left(s_i^{\gamma_1\prime}, s_i^{\gamma_2\prime}\right)'$$

If nothing changes for the mean parameters, getting consistent estimators of $C_{n_{\beta\gamma}}^o$ and $C_{n_{\gamma\gamma}}^o$ now requires using heteroscedasticity-consistent like estimators which may be computed by replacing $B_{n_{\beta\gamma}}^o$ and $B_{n_{\gamma\gamma}}^o$ by their empirical counterpart $\hat{B}_{n_{\beta\gamma}} = \frac{1}{n}\sum_{i=1}^n \hat{s}_i^\beta \hat{s}_i^{\gamma\prime}$ and $\hat{B}_{n_{\gamma\gamma}} = \frac{1}{n}\sum_{i=1}^n \hat{s}_i^\gamma \hat{s}_i^{\gamma\prime}$, and $A_{n_{\beta\beta}}^o$ and $A_{n_{\gamma\gamma}}^o$ by the quantities outlined above.

According to the robustness property of RPML2, the consistency of the mean parameters estimator $\hat{\beta}_n$ is not dismantled by possible misspecification of the assumed scedastic pattern of the data. But the whole asymptotic covariance matrix is seriously affected. So, if the model is correctly specified for the conditional mean but misspecified for the conditional variance, we then have that

$$\hat{\beta}_n \xrightarrow{a.s.} \beta^o \text{ and } \hat{\gamma}_n - \gamma_n^* \xrightarrow{a.s.} 0, \text{ as } n \to \infty \ (T_i \text{ bounded})$$

where $\gamma_n^* = \left(\gamma_{1_n}^{*\prime}, \gamma_{2_n}^{*\prime}\right)'$, while

$$\sqrt{n}(\hat{\theta}_n - \theta_n^{o*}) \approx \mathrm{N}(0, C_n^{o*})$$

with $\theta_n^{o*} = \left(\beta^{o\prime}, \gamma_n^{*\prime}\right)'$ and

$$C_n^{o*} = \begin{bmatrix} C_{n_{\beta\beta}}^{o*} & C_{n_{\beta\gamma}}^{o*} \\ C_{n_{\beta\gamma}}^{o*\prime} & C_{n_{\gamma\gamma}}^{o*} \end{bmatrix} = \begin{bmatrix} A_{n_{\beta\beta}}^{o*^{-1}} B_{n_{\beta\beta}}^{o*} A_{n_{\beta\beta}}^{o*^{-1}} & A_{n_{\beta\beta}}^{o*^{-1}} B_{n_{\beta\gamma}}^{o*} A_{n_{\gamma\gamma}}^{o*^{-1}} \\ A_{n_{\gamma\gamma}}^{o*^{-1}} B_{n_{\gamma\beta}}^{o*} A_{n_{\beta\beta}}^{o*^{-1}} & A_{n_{\gamma\gamma}}^{o*^{-1}} \ddot{B}_{n_{\gamma\gamma}}^{o*} A_{n_{\gamma\gamma}}^{o*^{-1}} \end{bmatrix}$$

where

$$A_{n_{\beta\beta}}^{o*} = \frac{1}{n}\sum_{i=1}^n E\left[h_i^{\beta\beta o*}\right], \quad A_{n_{\gamma\gamma}}^{o*} = \frac{1}{n}\sum_{i=1}^n E\left[h_i^{\gamma\gamma o*}\right], \quad h_i^{\gamma\gamma} = \begin{bmatrix} h_i^{\gamma_1\gamma_1} & h_i^{\gamma_1\gamma_2} \\ h_i^{\gamma_2\gamma_1} & h_i^{\gamma_2\gamma_2} \end{bmatrix}$$

$$B_{n_{\beta\beta}}^{o*} = \frac{1}{n}\sum_{i=1}^n E\left[s_i^{\beta o*} s_i^{\beta o*\prime}\right], \quad B_{n_{\beta\gamma}}^{o*} = \frac{1}{n}\sum_{i=1}^n E\left[s_i^{\beta o*} s_i^{\gamma o*\prime}\right] = B_{n_{\gamma\beta}}^{o*\prime}$$

$$\ddot{B}_{n_{\gamma\gamma}}^{o*} = \frac{1}{n}\sum_{i=1}^n E\left[s_i^{\gamma o*} s_i^{\gamma o*\prime}\right] - U_{n_{\gamma\gamma}}^{o*}, \quad U_{n_{\gamma\gamma}}^{o*} = \frac{1}{n}\sum_{i=1}^n E\left(s_i^{\gamma o*}\right) E\left(s_i^{\gamma o*}\right)'$$

the superscript '$o*$' denoting quantities evaluated at θ_n^{o*}.

Now, consistent estimation of both $C^{o*}_{n_{\beta\beta}}$ and $C^{o*}_{n_{\beta\gamma}}$ requires heteroscedasticity-consistent like estimators. A consistent estimator of $C^{o*}_{n_{\beta\beta}}$ is obtained by replacing $A^{o*}_{n_{\beta\beta}}$ by $\hat{A}_{n_{\beta\beta}}$ given above (this is because $h^{\beta\beta}_i = \underline{h}^{\beta\beta}_i$) and $B^{o*}_{n_{\beta\beta}}$ by $\hat{B}_{n_{\beta\beta}} = \frac{1}{n}\sum_{i=1}^{n} \hat{s}^{\beta}_i \hat{s}^{\beta\prime}_i$. Likewise, a consistent estimator of $C^{o*}_{n_{\beta\gamma}}$ is obtained by replacing $A^{o*}_{n_{\beta\beta}}$ by $\hat{A}_{n_{\beta\beta}}$, $B^{o*}_{n_{\beta\gamma}}$ by $\hat{B}_{n_{\beta\gamma}}$ given above, and $A^{o*^{-1}}_{n_{\gamma\gamma}}$ by its empirical counterpart $\check{A}_{n_{\gamma\gamma}} = \frac{1}{n}\sum_{i=1}^{n} \hat{h}^{\gamma\gamma}_i$. Remark that $\check{A}_{n_{\gamma\gamma}}$ and $\hat{A}_{n_{\gamma\gamma}}$ are quite different: if the latter only involves first derivatives, the former contains both first and second derivatives, and is thus significantly more burdensome to compute. Unless the observations are IID and the panel dataset is balanced, a consistent estimator of $C^{o*}_{n_{\gamma\gamma}}$ may usually not be obtained. However, a consistent estimator of its upper bound $Q^{o*}_{n_{\gamma\gamma}} = A^{o*^{-1}}_{n_{\gamma\gamma}}(\ddot{B}^{o*}_{n_{\gamma\gamma}} + U^{o*}_{n_{\gamma\gamma}})A^{o*^{-1}}_{n_{\gamma\gamma}}$ ($Q^{o*}_{n_{\gamma\gamma}} \gg C^{o*}_{n_{\gamma\gamma}}$) may simply be computed as $\check{A}^{-1}_{n_{\gamma\gamma}}\hat{B}_{n_{\gamma\gamma}}\check{A}^{-1}_{n_{\gamma\gamma}}$, where $\hat{B}_{n_{\gamma\gamma}}$ is as given above. This for example allows to perform a valid under conditional variance misspecification conservative — i.e., with asymptotic true size necessarily inferior to its specified nominal size — (joint) Wald test of the nullity of the non-intercept parameter of γ_1 and γ_2, that is to say, to readily — but not unambiguously — check if, as tentatively assumed, the observations actually exhibit some heteroscedasticity-like pattern related to the Z^1_i and Z^2_i explanatory variables. Note that the possibility to easily compute a consistent estimator of an upper bound of C^*_n also holds under arbitrary misspecification.

Following the logic underlying robust to conditional variance misspecification estimation, in empirical practice, although somewhat computationally more burdensome, it seems wise, at least in first investigations, to routinely compute the asymptotic covariance matrix according to the latter outlined scheme, i.e., as

$$\check{C}_n = \begin{bmatrix} \hat{A}^{-1}_{n_{\beta\beta}}\hat{B}_{n_{\beta\beta}}\hat{A}^{-1}_{n_{\beta\beta}} & \hat{A}^{-1}_{n_{\beta\beta}}\hat{B}_{n_{\beta\gamma}}\check{A}^{-1}_{n_{\gamma\gamma}} \\ \check{A}^{-1}_{n_{\gamma\gamma}}\hat{B}_{n_{\gamma\beta}}\hat{A}^{-1}_{n_{\beta\beta}} & \check{A}^{-1}_{n_{\gamma\gamma}}\hat{B}_{n_{\gamma\gamma}}\check{A}^{-1}_{n_{\gamma\gamma}} \end{bmatrix} \tag{3.22}$$

$\check{C}_n$ obviously yielding a consistent estimator of the asymptotic covariance matrix of all parameters if the model is actually correctly specified to a larger extent than just the conditional mean. A more precise estimator may subsequently be used if, according to specification testing, the model actually proves to be correctly specified to a larger extent than just the conditional mean.

3.4. Specification testing

The GPML2 estimator of model (3.7) thus delivers a consistent estimator of either the mean or the mean and variance parameters depending on the extent of the correct specification of our tentative model. The question is then: to which extent is our tentative model actually correctly specified?

Quite obviously, as for the consistency and limiting distribution properties outlined above, the general results regarding specification testing derived in Chapter 2 directly apply in the balanced case. Provided that suitable regularity conditions hold, they may likewise be shown to hold in the unbalanced case ($n \to \infty$, T_i bounded).

Hereafter, as an adapted handy synthesis, we succinctly review how to perform potentially useful conditional mean and conditional variance diagnostic tests of the model.

3.4.1. Conditional mean diagnostic tests

Testing the null hypothesis that the conditional mean is correctly specified means testing

$$\mathrm{H}_0^m : E(Y_i|X_i, Z_i^1, Z_i^2) = X_i\beta^o, \text{ for some } \beta^o \in \Theta_\beta, \quad i = 1, 2, ..., n$$

Following the lines of Chapter 2, H_0^m may be tested using auxiliary nested alternatives, auxiliary non-nested alternatives, or without resorting to explicit alternatives, through Hausman and information matrix type tests.

In all cases, such conditional mean diagnostic tests basically amount to checking, for given choices of $T_i \times q$ indicator matrices $\hat{W}_i^m$, that misspecification indicators of the form

$$\hat{\Phi}_n^m = \frac{1}{n}\sum_{i=1}^{n} \hat{W}_i^{m\prime}\hat{\Omega}_i^{-1}\hat{u}_i$$

are not significantly different from zero.

In the present context, given the assumed independence of the observations across individuals and the linearity of the null model, an $\mathcal{M}_n^w$-type test statistic — i.e., Wooldridge's modified m-test — for checking $\hat{\Phi}_n^m$ is given by the asymptotic chi-squared statistic

$$\begin{aligned}\mathcal{M}_n^{m^w} = & \left(\sum_{i=1}^{n}\left(\hat{W}_i^m - X_i\hat{P}_n^m\right)'\hat{\Omega}_i^{-1}\hat{u}_i\right)' \\ & \left(\sum_{i=1}^{n}\left(\hat{W}_i^m - X_i\hat{P}_n^m\right)'\hat{\Omega}_i^{-1}\hat{u}_i\hat{u}_i'\hat{\Omega}_i^{-1}\left(\hat{W}_i^m - X_i\hat{P}_n^m\right)\right)^{-1} \\ & \left(\sum_{i=1}^{n}\left(\hat{W}_i^m - X_i\hat{P}_n^m\right)'\hat{\Omega}_i^{-1}\hat{u}_i\right) \xrightarrow{d} \chi^2(q)\end{aligned}$$

which is equal to n minus the residual sum of squares ($= nR_u^2$, R_u^2 being the uncentered R-squared) of the OLS regression

$$1 = \left[\hat{u}_i'\hat{\Omega}_i^{-1}\left(\hat{W}_i^m - X_i\hat{P}_n^m\right)\right] b + \text{residuals}, \quad i = 1, 2, ..., n$$

the number of degrees of freedom q being equal to the size of the indicator $\hat{\Phi}_n^m$ and

$$\hat{P}_n^m = \left(\sum_{i=1}^{n} X_i'\hat{\Omega}_i^{-1}X_i\right)^{-1}\sum_{i=1}^{n} X_i'\hat{\Omega}_i^{-1}\hat{W}_i^m$$

The prominent characteristics of conditional mean diagnostic tests implemented through $\mathcal{M}_n^{m^w}$ are twofold. First, they yield valid tests of H_0^m regardless distributional and conditional variance misspecification. In other words, since they do not rely on other assumptions than the null itself, a rejection may effectively be attributed to a failure of H_0^m to hold. Second, they may be implemented using any consistent estimator of β^o, γ_n^* and, when involved, additional nuisance parameters δ, under H_0^m. We know that the GPML2 estimator $\hat{\theta}_n = \left(\hat{\beta}_n', \gamma_n'\right)'$ satisfies these consistency requirements. But others estimators, e.g., FGLS, satisfying these consistency requirements may thus also alternatively be used.

Following Section 2.5.1, conditional mean diagnostic tests of H_0^m against auxiliary nested alternatives of the form

$$\mathrm{H}_1^m : E(Y_i|X_i, Z_i^1, Z_i^2) = m_i^a(X_i, Z_i^1, Z_i^2, \beta^o, \alpha^o), \text{ for some } (\beta^{o\prime}, \alpha^{o\prime})' \in \Theta_a,$$

$i = 1, 2, ..., n$, where $\Theta_a = \Theta_\beta \times \Theta_\alpha$, α is a $k_\alpha \times 1$ vector of additional parameters, and for some constant vector $c \in \Theta_\alpha$

$$m_i^a(X_i, Z_i^1, Z_i^2, \beta, c) = X_i\beta, \quad i = 1, 2, ..., n$$

i.e., pseudo Lagrange multiplier tests that, under H_1^m, $\alpha^o = c$, are obtained by setting $\hat{W}_t^m = \frac{\partial m_i^a(X_i, Z_i^1, Z_i^2, \hat{\beta}_n, c)}{\partial \alpha'}$. When, as quite natural here, the auxiliary nested alternative specification takes the linear form

$$m_i^a(X_i, Z_i^1, Z_i^2, \beta, \alpha) = X_i\beta + G_i\alpha, \quad i = 1, 2, ..., n$$

where the $T_i \times k_\alpha$ matrices G_i are functions of the set of conditioning variables $CV_i \equiv (X_i, Z_i^1, Z_i^2)$, $\frac{\partial m_i^a(X_i, Z_i^1, Z_i^2, \hat{\beta}_n, c)}{\partial \alpha'}$ is simply equal to G_i and the test is a usual variable addition test. A common relevant choice of G_i is then (some of) the squares and/or the cross-products of (some of) the CV_i variables. Note by the way that the above general formulation includes the well-known Ramsey's (1969) RESET test for nonlinearity (see also Mackinnon-Magee (1990)). In its most popular form, it simply means setting $\hat{W}_t^m = X_i\hat{\beta}_n \odot X_i\hat{\beta}_n$.

According to Section 2.5.2, Davidson-Mackinnon (1981) type tests of H_0^m against auxiliary non-nested alternatives like

$$\mathrm{H}_1^m : E(Y_i|X_i, Z_i^1, Z_i^2, X_i^a) = g_i(X_i, Z_i^1, Z_i^2, \delta^o), \text{ for some } \delta^o \in \Theta_\delta, \quad i = 1, 2, ..., n$$

where δ is a $k_\delta \times 1$ vector of parameters, are, on the other hand, obtained by setting $\hat{W}_t^m = g_i(X_i, Z_i^1, Z_i^2, \hat{\delta}_n) - X_i\hat{\beta}_n$, where $\hat{\delta}_n$ is any consistent estimator of δ^o under H_1^m, e.g., the nonlinear least squares (NLS) estimator or, in the linear case, the OLS estimator. Because obvious appealing choices of $g_i(.)$ are rarely available, this kind of test of H_0^m is unlikely to be performed routinely.

One of the equivalent forms of the popular Hausman specification test of the standard homoscedastic model is based on comparing the (non-intercept) FGLS and OLS estimators of β^o (see for example Baltagi (1995)). This strongly suggests considering a generalized — allowing for any choice of S and robust to conditional variance misspecification — Hausman type test of H_0^m based on checking, for some chosen selection matrix S, the misspecification indicator

$$\hat{\Phi}_n^m = S(\hat{\beta}_n - \hat{\beta}_n^{OLS})$$

Following Section 2.5.3.1, a test asymptotically equivalent to checking the above misspecification indicator is procured by setting $\hat{W}_t^m = \hat{\Omega}_i X_i \hat{Q}_n^{-1} S'$, where $\hat{Q}_n = \sum_{i=1}^n X_i' X_i$. For the suitable choice of S and under standard conditions — i.e., homoscedasticity and conditional variance correct specification —, this test, which will have power against any alternative H_1^m for which $\hat{\beta}_n$ and $\hat{\beta}_n^{OLS}$ converge to different pseudo-true values, is asymptotically equivalent to its standard textbook counterpart[11]. Note that, contrary to the standard case, heteroscedasticity (and incompleteness) usually allows to include all β parameters as part of the Hausman test without yielding a singular statistic.

Finally, according to Section 2.5.3.2, an information matrix type test of H_0^m based on checking, for some chosen selection matrix S, the misspecification indicator

$$\hat{\Phi}_n^m = S \frac{1}{n} \sum_{i=1}^n \operatorname{vec} \hat{h}_i^{\beta\gamma}, \qquad h_i^{\beta\gamma} = \left[\begin{array}{cc} h_i^{\beta\gamma_1} & h_i^{\beta\gamma_2} \end{array} \right]$$

i.e., essentially based on checking the block diagonality between mean and variance parameters of the hessian, is obtained by setting $\hat{W}_t^m = \hat{F}_i S'$, where

$$\hat{F}_i' = \left[\begin{array}{c} X_i' \hat{\Omega}_i^{-1} \frac{\partial \hat{\Omega}_i}{\partial \gamma_1^1} \\ \vdots \\ X_i' \hat{\Omega}_i^{-1} \frac{\partial \hat{\Omega}_i}{\partial \gamma_1^{k_{\gamma_1}}} \\ X_i' \hat{\Omega}_i^{-1} \frac{\partial \hat{\Omega}_i}{\partial \gamma_2^1} \\ \vdots \\ X_i' \hat{\Omega}_i^{-1} \frac{\partial \hat{\Omega}_i}{\partial \gamma_2^{k_{\gamma_2}}} \end{array} \right]$$

This kind of test, which may also be interpreted as simultaneously performing several Hausman type tests and which will have power against any alternative H_1^m for which the block diagonality of the hessian fails, is a quite natural complement to the above Hausman test for testing H_0^m without resorting to explicit alternatives. Note that if the multiplicative heteroscedasticity formulation is adopted for both $\phi_\nu(.)$ and $\phi_\mu(.)$, one of the two matrix elements $X_i' \hat{\Omega}_i^{-1} \frac{\partial \hat{\Omega}_i}{\partial \gamma_1^1}$ and $X_i' \hat{\Omega}_i^{-1} \frac{\partial \hat{\Omega}_i}{\partial \gamma_2^1}$ of $\hat{F}_i$ is redundant, yielding for S set to an identity matrix a singular statistic, and must thus be discarded.

To conclude this succinct review, a remark. In empirical practice, it is quite usual to test the null model by specifying an explicit auxiliary alternative which includes variables which are not functions of the original set of conditioning variables CV_i. This does not modify the way in which testing against explicit alternatives is implemented. It is however important to be aware that, in such a case, we are no longer only testing the null H_0^m but instead the null $\mathrm{H}_0^{m^a}$: H_0^m holds and

[11] For another robust to conditional variance misspecification version of the standard textbook Hausman test, see Arellano (1993).

$E(Y_i|X_i, Z_i^1, Z_i^2) = E(Y_i|X_i, Z_i^1, Z_i^2, G_i^a)$, $i = 1, 2, ..., n$, where G_i^a denotes the variables which are not functions of CV_i. In other words, we are jointly testing that H_0^m holds and that the additional G_i^a variables are irrelevant as conditioning variables for the expectation of Y_i. We thus must be careful in interpreting such a specification test: H_0^m might well hold while $\mathrm{H}_0^{m^a}$ does not.

3.4.2. Conditional variance diagnostic tests

Testing the null hypothesis that the conditional variance is correctly specified entails testing the null

$$\mathrm{H}_0^v : \begin{cases} \mathrm{H}_0^m \text{ holds and, for some } \gamma^o \in \Theta_{\gamma_1} \times \Theta_{\gamma_2}, \\ V(Y_i|X_i, Z_i^1, Z_i^2) = \mathrm{diag}\,(\phi_\nu(Z_i^1\gamma_1^o)) + J_{T_i}\,\phi_\mu(Z_i^2\gamma_2^o), \quad i = 1, 2, ..., n \end{cases}$$

Following again the lines of Chapter 2, as for the conditional mean, H_0^v may be tested using auxiliary nested alternatives, auxiliary non-nested alternatives, or without resorting to explicit alternatives, through Hausman and information matrix type tests, and in all cases, it basically amounts to checking, for given choices of $T_i^2 \times q$ indicator matrices $\hat{W}_i^v$, misspecification indicators which are similarly of the form

$$\hat{\Phi}_n^v = \frac{1}{n} \sum_{i=1}^{n} \hat{W}_i^{v\prime} \hat{\Gamma}_i^{-1} \hat{v}_i$$

where

$$\Gamma_i^{-1} = \left(\hat{\Omega}_i^{-1} \otimes \hat{\Omega}_i^{-1}\right) \quad \text{ and } \quad \hat{v}_i = \mathrm{vec}(\hat{u}_i\hat{u}_i' - \hat{\Omega}_i)$$

Given the assumed independence across i, an $\mathcal{M}_n^w$-type test statistic for checking $\hat{\Phi}_n^v$ is likewise given by the asymptotic chi-squared statistic

$$\begin{aligned} \mathcal{M}_n^{v^w} \;=\; & \left(\sum_{i=1}^{n} \left(\hat{W}_i^v - \tfrac{\partial \mathrm{vec}\,\hat{\Omega}_i}{\partial \gamma'} \hat{P}_n^v\right)' \hat{\Gamma}_i^{-1} \hat{v}_i\right)' \\ & \left(\sum_{i=1}^{n} \left(\hat{W}_i^v - \tfrac{\partial \mathrm{vec}\,\hat{\Omega}_i}{\partial \gamma'} \hat{P}_n^v\right)' \hat{\Gamma}_i^{-1} \hat{v}_i \hat{v}_i' \hat{\Gamma}_i^{-1} \left(\hat{W}_i^v - \tfrac{\partial \mathrm{vec}\,\hat{\Omega}_i}{\partial \gamma'} \hat{P}_n^v\right)\right)^{-1} \qquad (3.23) \\ & \left(\sum_{i=1}^{n} \left(\hat{W}_i^v - \tfrac{\partial \mathrm{vec}\,\hat{\Omega}_i}{\partial \gamma'} \hat{P}_n^v\right)' \hat{\Gamma}_i^{-1} \hat{v}_i\right) \xrightarrow{d} \chi^2(q) \end{aligned}$$

which is equal to n minus the residual sum of squares ($= nR_u^2$) of the OLS regression

$$1 = \left[\hat{v}_i' \hat{\Gamma}_i^{-1} \left(\hat{W}_i^v - \tfrac{\partial \mathrm{vec}\,\hat{\Omega}_i}{\partial \gamma'} \hat{P}_n^v\right)\right] b + \text{residuals}, \quad i = 1, 2, ..., n$$

the number of degrees of freedom q being again equal the size of the indicator $\hat{\Phi}_n^v$ and

$$\frac{\partial\,\mathrm{vec}\,\hat{\Omega}_i}{\partial \gamma'} = \left[\; \frac{\partial\,\mathrm{vec}\,\hat{\Omega}_i}{\partial \gamma_1'} \quad \frac{\partial\,\mathrm{vec}\,\hat{\Omega}_i}{\partial \gamma_2'} \;\right]$$

$$\hat{P}_n^v = \left(\sum_{i=1}^{n} \left(\frac{\partial \operatorname{vec} \hat{\Omega}_i}{\partial \gamma'} \right)' \hat{\Gamma}_i^{-1} \frac{\partial \operatorname{vec} \hat{\Omega}_i}{\partial \gamma'} \right)^{-1} \sum_{i=1}^{n} \left(\frac{\partial \operatorname{vec} \hat{\Omega}_i}{\partial \gamma'} \right)' \hat{\Gamma}_i^{-1} \hat{W}_i^v$$

In short, we have the same general structure than for conditional mean testing but, unfortunately, it involves more complicated expressions. In this respect, it is worth noting that cleverly using the simplifying tricks outlined in Section 3.3.1 — in particular the identities (3.11) — may significantly alleviate the computational burden of the test statistics reviewed below.

Similarly to conditional mean diagnostic tests, conditional variance diagnostic tests performed through $\mathcal{M}_n^{v^w}$ may be implemented using any consistent estimator of β^o, γ^o and, when involved, additional nuisance parameters δ, under H_0^v, i.e., the GPML2 estimator is not required. Likewise, they do not rely on other assumptions than H_0^v itself, i.e., they are robust to distributional misspecification. A rejection may thus effectively be attributed to a failure of H_0^v to hold. Given the nested nature of H_0^m and H_0^v, the robustness to conditional variance misspecification of the diagnostic tests of H_0^m and the fact that following diagnostic tests of H_0^v concentrates on detecting departures in the second order moments, if no misspecification has been detected by conditional mean diagnostic tests, a rejection of H_0^v may sensibly be attributed to conditional variance misspecification: situations where conditional variance diagnostic tests detect a misspecification in the mean which has not been detected by conditional mean diagnostic tests are likely to be rare in practice.

Now, according to Section 2.6.1, conditional variance diagnostic tests of H_0^v against auxiliary nested alternatives of the form

$$\mathrm{H}_1^v : \begin{cases} \mathrm{H}_0^m \text{ holds and, for some } (\gamma^{o\prime}, \alpha^{o\prime})' \in \Theta_{\gamma_1} \times \Theta_{\gamma_2} \times \Theta_{\alpha}, \\ V(Y_i|X_i, Z_i^1, Z_i^2) = \Omega_i^a(X_i, Z_i^1, Z_i^2, \gamma^o, \alpha^o), \quad i = 1, 2, ..., n \end{cases}$$

where α is a $k_\alpha \times 1$ vector of additional parameters, and for some constant vector $c \in \Theta_\alpha$

$$\Omega_i^a(X_i, Z_i^1, Z_i^2, \gamma, c) = \operatorname{diag}\left(\phi_\nu(Z_i^1\gamma_1)\right) + J_{T_i}\,\phi_\mu(Z_i^2\gamma_2), \quad i = 1, 2, ..., n$$

i.e., pseudo Lagrange multiplier tests that, under H_1^v, $\alpha^o = c$, are obtained by setting $\hat{W}_i^v = \frac{\partial \operatorname{vec} \Omega_i^a(X_i, Z_i^1, Z_i^2, \hat{\gamma}_n, c)}{\partial \alpha'}$. If the auxiliary nested alternative takes, as natural in the present context, the semi-linear form

$$\Omega_i^a(X_i, Z_i^1, Z_i^2, \gamma, \alpha) = \operatorname{diag}\left(\phi_\nu(Z_i^1\gamma_1 + G_i^1\alpha_1)\right) + J_{T_i}\,\phi_\mu(Z_i^2\gamma_2 + G_i^2\alpha_2)$$

where $\alpha = (\alpha_1', \alpha_2')'$ and the $T_i \times k_{\alpha_1}$ matrices G_i^1 and the $1 \times k_{\alpha_2}$ matrices G_i^2 are functions of the set of conditioning variables CV_i, the appropriate indicator matrices $\hat{W}_i^v$ are given by

$$\hat{W}_i^v = \left[\ \hat{W}_i^{v1} \quad \hat{W}_i^{v2}\ \right]$$

with

$$\hat{W}_i^{v1} = \operatorname{diag}\left(\operatorname{vec}\left(\operatorname{diag}\left(\phi_\nu'(Z_i^1\hat{\gamma}_{1_n})\right)\right)\right)\left(G_i^1 \otimes e_{T_i}\right)$$

$$
\begin{aligned}
&= \sum_{r=1}^{k_{\alpha_1}} \operatorname{vec}\left(\frac{\partial \hat{\Omega}_t^a}{\partial \alpha_1^r}\right) e_{k_{\alpha_1}}^{r\prime}, \quad \frac{\partial \hat{\Omega}_t^a}{\partial \alpha_1^r} = \operatorname{diag}\left(\phi_\nu'(Z_i^1 \hat{\gamma}_{1_n}) \odot G_i^{1^r}\right) \\
\hat{W}_i^{v2} &= \phi_\mu'(Z_i^2 \hat{\gamma}_{2_n}) \operatorname{vec}(J_{n_i})\, G_i^2 \\
&= \sum_{r=1}^{k_{\alpha_2}} \operatorname{vec}\left(\frac{\partial \hat{\Omega}_t^a}{\partial \alpha_2^r}\right) e_{k_{\alpha_2}}^{r\prime}, \quad \frac{\partial \hat{\Omega}_t^a}{\partial \alpha_2^r} = \phi_\mu'(Z_i^2 \hat{\gamma}_{2_n}) G_i^{2^r} J_{T_i}
\end{aligned}
$$

where $G_i^{1^r}$ and $G_i^{2^r}$ denote the r-th column of respectively G_i^1 and G_i^2. As for the conditional mean, common relevant choices of G_i^1 and G_i^2 are then (some of) the squares and/or the cross-products of (some of) the CV_i variables.

On the other hand, following Section 2.6.2, Davidson-Mackinnon (1981) and Cox (1961,1962) type tests of H_0^v against auxiliary non-nested alternatives like

$$
\mathrm{H}_1^v : \begin{cases} \mathrm{H}_0^m \text{ holds and, for some } \delta^o \in \Theta_\delta \\ V(Y_i|X_i, Z_i^1, Z_i^2) = \Sigma_i(X_i, Z_i^1, Z_i^2, \delta^o), \quad i = 1, 2, ..., n \end{cases}
$$

where δ is a $k_\delta \times 1$ vector of parameters, are obtained by respectively setting $\hat{W}_i^v = \operatorname{vec}(\hat{\Sigma}_i - \hat{\Omega}_i)$ and $\hat{W}_i^v = \operatorname{vec}(\hat{\Omega}_i \hat{\Sigma}_i^{-1} \hat{\Omega}_i - \hat{\Omega}_i)$, where $\hat{\Sigma}_i = \Sigma_i(X_i, Z_i^1, Z_i^2, \hat{\delta}_n)$ and $\hat{\delta}_n$ is any consistent estimator of δ^o under H_1^m, e.g., the multivariate NLS (MNLS) estimator of the T_i^2-variate nonlinear regression $\operatorname{vec}(\hat{u}_i \hat{u}_i') = \operatorname{vec} \Sigma_i(X_i, Z_i^1, Z_i^2, \delta) +$ residuals, $i = 1, 2, ..., n$. The Cox form of the test is probably generally more powerful. Be that as it may, such tests may for example be used for checking the chosen variance functions $\phi_\nu(.)$ and $\phi_\mu(.)$ against some alternative choices, or more radically the assumed form of the heterogeneity against some other non-nested specification allowing for variable heterogeneity.

As for conditional mean testing, in both the nested and non-nested case, the way to perform the tests is unchanged if the auxiliary alternative includes variables which are not functions of the original set of conditioning variables CV_i. But likewise the tested null hypothesis is modified. It here takes the form $\mathrm{H}_0^{v^a}$: H_0^v holds and, both $E(Y_i|X_i, Z_i^1, Z_i^2) = E(Y_i|X_i, Z_i^1, Z_i^2, G_i^a)$ and $V(Y_i|X_i, Z_i^1, Z_i^2) = V(Y_i|X_i, Z_i^1, Z_i^2, G_i^a)$, $i = 1, 2, ..., n$, where G_i^a denotes the variables which are not functions of CV_i. In other words, besides H_0^v, it further assumes that the additional variables G_i^a are irrelevant as conditioning variables for the variance but also the expectation of Y_i.

Testing H_0^v through a Hausman type test requires to choose a consistent estimator of γ^o alternative to $\hat{\gamma}_n$. The GPML2 estimator $\hat{\gamma}_n$ may be shown to be asymptotically equivalent to the multivariate weighted NLS (MWNLS) estimator with weigths $\left\{\ddot{\Gamma}_i^{-1}\right\}$ of the T_i^2-variate nonlinear regression $\operatorname{vec}(\ddot{u}_i \ddot{u}_i') = \operatorname{vec}(\operatorname{diag}(\phi_\nu(Z_i^1 \gamma_1)) + J_{T_i}\, \phi_\mu(Z_i^2 \gamma_2)) +$ residuals, $i = 1, 2, ..., n$, where the superscript '$\ddot{\ }$' denoting quantities evaluated at any preliminary consistent estimator of β^o and γ^o. A simple and natural alternative to it is thus the standard (unweighted) MNLS estimator, say $\hat{\underline{\gamma}}_n$, of the same regression. According to Section 2.6.3.1, a test asymptotically equivalent to checking, for some chosen selection matrix S, the misspecification indicator

$$
\hat{\Phi}_n^v = S(\hat{\gamma}_n - \hat{\underline{\gamma}}_n)
$$

is procured by setting $\hat{W}_i^v = \hat{\Gamma}_i \frac{\partial \operatorname{vec} \hat{\Omega}_i}{\partial \gamma'} \hat{Q}_n^{-1} S'$, where $\hat{Q}_n = \sum_{i=1}^n \left(\frac{\partial \operatorname{vec} \hat{\Omega}_i}{\partial \gamma'} \right)' \frac{\partial \operatorname{vec} \hat{\Omega}_i}{\partial \gamma'}$. As all Hausman type tests, this test will have power against any alternative H_1^v for which $\hat{\gamma}_n$ and $\hat{\underline{\gamma}}_n$ converge to different pseudo-true values.

Finally, following Section 2.5.3.2, an information matrix type test of H_0^v based on checking, for some chosen selection matrix S which at least removes its otherwise obvious redundant elements, the misspecification indicator

$$\hat{\Phi}_n^v = S \frac{1}{n} \sum_{i=1}^n \operatorname{vec} \left(X_i' \hat{\Omega}_i^{-1} \hat{u}_i \hat{u}_i' \hat{\Omega}_i^{-1} X_i - X_i' \hat{\Omega}_i^{-1} X_i \right)$$

i.e., essentially based on checking the information matrix equality $B_{n_{\beta\beta}}^o = -A_{n_{\beta\beta}}^o$ of the mean parameters, is obtained by setting $\hat{W}_t^v = (X_i \otimes X_i) S'$. This latter way of testing H_0^v without resorting to explicit alternatives, which seems generally more powerful than the above Hausman type test, will clearly have power against any alternative H_1^v for which the mean parameters information matrix equality fails. Contrary to the information matrix type test of the conditional mean for which it does not seem to be a fruitful strategy, in our experience, setting S such that $\hat{\Phi}_n^v$ is simply the sum of all non-redundant elements of the information matrix equality appears quite appealing.

3.5. A preliminary joint pseudo Lagrange multiplier test

In the framework of the one-way error components model, several test statistics are available for testing the presence of individual effects. A comprehensive survey of these tests may be found in Mouton-Randolf (1989) and Baltagi-Chang-Li (1992). Almost all these tests are derived under normality and one-directional in the sense that they are designed to test against only one departure from the null[12]. Notable exceptions are the joint gaussian LM test for serial correlation and random individual effects of Baltagi-Li (1991) and the distribution-free test statistic recently derived by Li-Stengos (1994) for testing the presence of random individual effects while allowing for arbitrary pattern of heteroscedasticity under the null.

As outlined in the introduction, we are here interested in deriving a simple OLS residuals based distribution-free test statistic for checking the potential relevance of our general model before undertaking the estimation procedure, i.e., in jointly testing the null hypothesis of no individual effects and homoscedasticity against the alternative of random individual effects and heteroscedasticity.

Such a test may be expressed as testing the null

$$\mathrm{H}_0 : \left\{ \begin{array}{l} E(Y_i | X_i, \underline{Z}_i) = X_i \beta^o \\ V(Y_i | X_i, \underline{Z}_i) = \sigma_\nu^{o2} I_{T_i} \end{array} \right. , \text{ for some } \left(\beta^{o\prime}, \sigma_\nu^{o2} \right)' \in \Theta_\beta \times \Theta_{\sigma_\nu^2}, \quad i = 1, 2, ..., n$$

[12] In the context of the gaussian two-way error components model, two-directional tests are available for jointly testing the presence of both individual and time effects. See Breush-Pagan (1980) and Baltagi-Li (1990). Test statistics are also available for testing the presence of individual effects while assuming time effects under the null, and vice versa. See Baltagi-Chang-Li (1992).

against the auxiliary nested alternative

$$\mathrm{H}_1 : \begin{cases} E(Y_i|X_i,\underline{Z}_i) = X_i\beta^o \\ V(Y_i|X_i,\underline{Z}_i) = \sigma_\nu^{o2}\,\mathrm{diag}\left(\phi(\underline{Z}_i\underline{\gamma}^o)\right) + \sigma_\mu^{o2} J_{T_i} \end{cases}, \text{ for some } \left(\beta^{o\prime},\sigma_\nu^{o2},\sigma_\mu^{o2},\underline{\gamma}^{o\prime}\right)' \in \Theta_a,$$

$i = 1, 2, ..., n$, where $\Theta_a = \Theta_\beta \times \Theta_{\sigma_\nu^2} \times \Theta_{\sigma_\mu^2} \times \Theta_{\underline{\gamma}}$, $\underline{Z}_i$ denotes a $T_i \times k_{\underline{\gamma}}$ matrix of explanatory variables (without intercept), $\underline{\gamma}$ is a $k_{\underline{\gamma}} \times 1$ vector of parameters, σ_ν^2 and σ_μ^2 are scalar variance parameters and $\phi(.)$ is an arbitrary non-indexed (strictly) positive twice continuously differentiable function satisfying $\phi(0) = 1$ and $\phi^{'}(0) \neq 0$, $\phi'(x)$ denoting the first derivative of $\phi(x)$ with respect to x. Notice that the multiplicative heteroscedasticity link function exp(.) satisfies these assumptions.

The formulation of the test deserves some comments. First note that it maintains the hypothesis that the conditional mean is correctly specified with respect to the set of conditioning variables $CV_i \equiv (X_i, \underline{Z}_i)$. Note further that the alternative H_1 is less general than model (3.7): contrary to it, it does not allow for heteroscedasticity in the individual-specific error term. This would be irrelevant for the present purpose of the test statistic[13]. We want however to stress that this does not mean that the joint test statistic will be insensitive to heteroscedasticity in the individual-specific dimension. Indeed, such a heteroscedasticy pattern implies non-constant diagonal elements in the conditional variance, a feature which may be captured through the general error term variance function. So, by suitably picking up $\underline{Z}_i$, the joint test statistic could actually be specifically designed for testing against a true data generating process exhibiting heteroscedasticy only in the individual-specific error term. Be that as it may, further observe that, regarding the general error term, the alternative allows for a quite broad class of heteroscedastic models. In particular, it includes some versions of the random coefficient model (see Breush-Pagan (1979)).

Letting Ω_i^a stand for the alternative specification $\sigma_\nu^2\,\mathrm{diag}\left(\phi(\underline{Z}_i\underline{\gamma})\right) + \sigma_\mu^2 J_{T_i}$, according to the previous section, a pseudo LM test of the joint null that[14] $\sigma_\mu^{o2} = 0$ and $\underline{\gamma}^o = 0$ is for the case at hand based on the $(1+k_{\underline{\gamma}}) \times 1$ misspecification indicator

$$\tilde{\Phi}_n = \frac{1}{n}\sum_{i=1}^{n} \tilde{W}_i'\tilde{\Gamma}_i^{-1}\tilde{v}_i$$

with

$$\tilde{W}_i = \frac{\partial\,\mathrm{vec}\,\tilde{\Omega}_i^a}{\partial\left(\sigma_\mu^2,\underline{\gamma}'\right)}\ ,\quad \tilde{\Gamma}_i = \tilde{\sigma}_\nu^2 I_{T_i} \otimes \tilde{\sigma}_\nu^2 I_{T_i}\ ,\quad \tilde{v}_i = \mathrm{vec}(\tilde{u}_i\tilde{u}_i' - \tilde{\sigma}_\nu^2 I_{T_i}) \tag{3.24}$$

---

[13] A joint test of $\sigma_\mu^{o2} = 0$ and $\underline{\gamma}^o = 0$ allowing for heteroscedasticity in the individual-specific error term under the alternative could be derived by using the Davidson-Mackinnon (1981) trick outlined in Chapter 2 to overcome the non-identifiability of the individual-specific variance parameters under the null. The test statistic would however then require a consistent estimator of these parameters under the alternative. In other words, checking the potential relevance of our general model before estimating it would actually entail first estimating it in some way.

[14] Note that, as outlined in Chapter 2, the fact that σ_μ^{o2} is on the border of its parameter space does not entail any problem.

where the superscript '$\tilde{\ }$' denotes quantities evaluated under H_0 and

$$\begin{aligned}\tilde{W}_i &= \left[\begin{array}{cc} \dfrac{\partial \operatorname{vec} \tilde{\Omega}_i^a}{\partial \sigma_\mu^2} & \dfrac{\partial \operatorname{vec} \tilde{\Omega}_i^a}{\partial \underline{\gamma}'} \end{array} \right] \\ &= \left[\begin{array}{cc} \operatorname{vec} J_{T_i} & \tilde{\sigma}_\nu^2 \phi'(0) \operatorname{diag}(\operatorname{vec} I_{T_i}) (\underline{Z}_i \otimes e_{T_i}) \end{array} \right]\end{aligned}$$

An m-test — either standard or Wooldridge's modified — that $\tilde{\Phi}_n$ is not too far from zero is equivalent to an m-test that $\tilde{S}_n \tilde{\Phi}_n$, where $\tilde{S}_n$ is a $(1 + k_{\underline{\gamma}}) \times (1 + k_{\underline{\gamma}})$ non-singular matrix, is not too far from zero. By adequately defining $\tilde{S}_n$, we may thus get rid of the multiplicative constant $\tilde{\sigma}_\nu^2 \phi'(0)$ appearing in $\frac{\partial \operatorname{vec} \tilde{\Omega}_i^a}{\partial \underline{\gamma}'}$. In other words, the precise form of $\phi(.)$ is irrelevant for deriving the present test statistic: provided of course that it satisfies the assumptions we made about it, any choice of $\phi(.)$ leads to the same statistic. They are locally equivalent alternatives.

So, the relevant misspecification indicator to consider is actually

$$\tilde{\Phi}_n^{I_r H} = \frac{1}{n} \sum_{i=1}^n W_i^{I_r H \prime} \tilde{\Gamma}_i^{-1} \tilde{v}_i$$

with

$$W_i^{I_r H} = \left[\begin{array}{cc} \operatorname{vec} J_{T_i} & \operatorname{diag}(\operatorname{vec} I_{T_i}) (\underline{Z}_i \otimes e_{T_i}) \end{array} \right] \tag{3.25}$$

From this point, we might proceed by roughly computing the test statistic (3.23) with the relevant quantities. This however does not yield a simple statistic, as we are looking for.

Considerable simplification may be obtained if, on one hand, we restrict our attention to an $\mathcal{M}_n$-type test , i.e., a standard m-test, of $\tilde{\Phi}_n^{I_r H}$, and, on the other hand and much more importantly, we are willing to further assume that, under H_0, the errors $u_{it}^o = \nu_{it}$ are conditionally independently — but not necessarily identically — distributed across t with fourth order conditional moments $E(\nu_{it}^4 | X_i, \underline{Z}_i) = \delta_{it}^o$, $i = 1, ..., n$.

According to Section 2.4 and 2.6, for the case at hand, the $\mathcal{M}_n$-type test of $\tilde{\Phi}_n^{I_r H}$ is given by the asymptotic chi-square statistic

$$\mathcal{PLM}_n^{I_r H} = n \tilde{\Phi}_n^{I_r H \prime} \tilde{K}_n^{-1} \tilde{\Phi}_n^{I_r H} \xrightarrow{d} \chi^2(1 + k_{\underline{\gamma}}) \tag{3.26}$$

where $\tilde{K}_n$ is a consistent estimator of

$$K_n^o = \frac{1}{n \sigma_\nu^{o8}} \sum_{i=1}^n E\left[\left(W_i^{I_r H} - R_i P_n \right)' v_i^o v_i^{o\prime} \left(W_i^{I_r H} - R_i P_n \right) \right]$$

with

$$\begin{aligned} R_i &= \operatorname{vec} I_{T_i} \qquad (3.27) \\ P_n &= \left(\sum_{i=1}^n E\left[R_i' R_i \right] \right)^{-1} \sum_{i=1}^n E\left[R_i' W_i^{I_r H} \right] \end{aligned}$$

Restricting our attention to this $\mathcal{M}_n$-type test means that we have to use the GPML2 estimator of the variance parameters[15] — and not another consistent estimator unless it has the same limiting distribution — of the null model for the test statistic to be valid. This is here perfectly innocuous since the GPML2 estimators of the null model are simply the OLS estimator $\tilde{\beta}_n$ and the standard ML variance estimator $\tilde{\sigma}_\nu^2 = \frac{1}{N}\sum_{i=1}^n \sum_{t=1}^{T_i} \tilde{u}_{it}^2$, i.e., the simplest estimators of the null model. Let for now the superscript '~' denotes quantities evaluated at these estimators.

Using (3.24) and (3.25), it may be verified that $\tilde{\Phi}_n^{I_r H}$ is equal to

$$\tilde{\Phi}_n^{I_r H} = \begin{bmatrix} \dfrac{1}{n\tilde{\sigma}_\nu^4} \displaystyle\sum_{i=1}^n \left((\tilde{u}_i' e_{T_i})^2 - \tilde{\sigma}_\nu^2 T_i \right) \\ \dfrac{1}{n\tilde{\sigma}_\nu^4} \displaystyle\sum_{i=1}^n \underline{Z}_i' \left(\tilde{u}_i^{\hat{}2} - \tilde{\sigma}_\nu^2 e_{T_i} \right) \end{bmatrix} \tag{3.28}$$

where $\tilde{u}_i^{\hat{}2} = \tilde{u}_i \odot \tilde{u}_i$.

Following Magnus-Neudecker (1986), the additional auxiliary assumption that, under H_0, the errors $u_{it}^o = \nu_{it}$ are conditionally independently distributed across t with fourth order conditional moments $E(\nu_{it}^4 | X_i, \underline{Z}_i) = \delta_{it}^o$, implies that we have

$$E\left[v_i^o v_i^{o\prime} | X_i, \underline{Z}_i\right] = \sigma_\nu^{o4} I_{T_i^2} + \sigma_\nu^{o4} K_{T_i T_i} + \operatorname{diag}\left(\operatorname{vec}\left(\operatorname{diag}(\delta_i^o)\right) - 3\sigma_\nu^{o4} \operatorname{vec} I_{T_i}\right) \tag{3.29}$$

where $\delta_i^o = \left(\delta_{i1}^o, ..., \delta_{iT_i}^o\right)'$ and $K_{T_i T_i}$ is the $T_i^2 \times T_i^2$ commutation matrix, i.e., a matrix such that, for any $T_i \times T_i$ matrix A, $K_{T_i T_i} \operatorname{vec} A = \operatorname{vec} A'$. Under conditional normality, $\delta_{it}^o = 3\sigma_\nu^{o4}$ and the last term of (3.29) is zero.

Noting that $K_{T_i T_i} W_i^{I_r H} = W_i^{I_r H}$ and $K_{T_i T_i} R_i = R_i$, under (3.29), K_n^o then collapses to

$$\begin{aligned} &\bar{K}_n^o \\ = &\frac{1}{n\sigma_\nu^{o8}} \sum_{i=1}^n E\left[\left(W_i^{I_r H} - R_i P_n\right)' E\left[v_i^o v_i^{o\prime} | X_i, \underline{Z}_i\right] \left(W_i^{I_r H} - R_i P_n\right)\right] \\ = &\frac{2}{n\sigma_\nu^{o4}} \sum_{i=1}^n E\left[\left(W_i^{I_r H} - R_i P_n\right)' \left(W_i^{I_r H} - R_i P_n\right)\right] \\ &+ \frac{1}{n\sigma_\nu^{o8}} \sum_{i=1}^n E\left[\left(W_i^{I_r H} - R_i P_n\right)' \operatorname{diag}\left(\operatorname{vec}\left(\operatorname{diag}(\delta_i^o)\right) - 3\sigma_\nu^{o4} \operatorname{vec} I_{T_i}\right) \left(W_i^{I_r H} - R_i P_n\right)\right] \end{aligned}$$

Now, consider the matrix $\widetilde{\bar{K}_n^o}$ defined as

$$\widetilde{\bar{K}_n^o} = \frac{1}{n\sigma_\nu^{o8}} \sum_{i=1}^n E\left[2\sigma_\nu^{o4} \tilde{M}_{i1} + \tilde{M}_{i2}^o - 3\sigma_\nu^{o4} \tilde{M}_{i3}\right]$$

[15] For the mean parameters, it does not matter. See Section 2.6.

where

$$\begin{aligned}
\tilde{M}_{i1} &= \left(W_i^{I_rH} - R_i\tilde{P}_n\right)' \left(W_i^{I_rH} - R_i\tilde{P}_n\right) \\
\tilde{M}_{i2}^o &= \left(W_i^{I_rH} - R_i\tilde{P}_n\right)' \operatorname{diag}\left(\operatorname{vec}\left(\operatorname{diag}(\delta_i^o)\right)\right) \left(W_i^{I_rH} - R_i\tilde{P}_n\right) \\
\tilde{M}_{i3} &= \left(W_i^{I_rH} - R_i\tilde{P}_n\right)' \operatorname{diag}\left(\operatorname{vec} I_{T_i}\right) \left(W_i^{I_rH} - R_i\tilde{P}_n\right)
\end{aligned}$$

and

$$\tilde{P}_n = \left(\sum_{i=1}^n R_i' R_i\right)^{-1} \sum_{i=1}^n R_i' W_i^{I_rH}, \tag{3.30}$$

$\widetilde{\bar{K}_n^o}$ is the same than $\bar{K}_n^o$ except that P_n has been replaced by its consistent estimator $\tilde{P}_n$, so that $\widetilde{\bar{K}_n^o}$ converges to $\bar{K}_n^o$. Using (3.25), (3.27) and (3.30), it may be seen that

$$\left(W_i^{I_rH} - R_i\tilde{P}_n\right) = \left[\begin{array}{cc} \operatorname{vec}\left(J_{T_i} - I_{T_i}\right) & \operatorname{diag}\left(\operatorname{vec} I_{T_i}\right)\left(\ddot{\underline{Z}}_i \otimes e_{T_i}\right) \end{array}\right] \tag{3.31}$$

where $\ddot{\underline{Z}}_i = \underline{Z}_i - \frac{1}{N}\sum_{i=1}^n e_{T_i}'\underline{Z}_i$, thus variables expressed in deviations from their (entire) sample mean.

Then, using (3.31), it may be checked that

$$\begin{aligned}
\tilde{M}_{i1} &= \begin{bmatrix} (T_i^2 - T_i) & 0 \\ 0 & \ddot{\underline{Z}}_i'\ddot{\underline{Z}}_i \end{bmatrix} \\
\tilde{M}_{i2}^o &= \begin{bmatrix} 0 & 0 \\ 0 & \ddot{\underline{Z}}_i' \operatorname{diag}(\delta_i^o)\ddot{\underline{Z}}_i \end{bmatrix} \\
\tilde{M}_{i3} &= \begin{bmatrix} 0 & 0 \\ 0 & \ddot{\underline{Z}}_i'\ddot{\underline{Z}}_i \end{bmatrix}
\end{aligned}$$

such that

$$\widetilde{\bar{K}_n^o} = \frac{1}{n\sigma_\nu^{o8}} \sum_{i=1}^n E \begin{bmatrix} 2\sigma_\nu^{o4}\left(T_i^2 - T_i\right) & 0 \\ 0 & \ddot{\underline{Z}}_i'\left(\operatorname{diag}(\delta_i^o) - \sigma_\nu^{o4} I_{T_i}\right)\ddot{\underline{Z}}_i \end{bmatrix}$$

In other words, under the null and the auxiliary assumption (3.29), the two components of $\tilde{\Phi}_n^{I_rH}$ are asymptotically independently distributed. A consistent estimator of $\widetilde{\bar{K}_n^o}$, and thus also of $\bar{K}_n^o$, may be computed as

$$\tilde{K}_n = \begin{bmatrix} \dfrac{2}{n\tilde{\sigma}_\nu^4}\displaystyle\sum_{i=1}^n (T_i^2 - T_i) & 0 \\ 0 & \dfrac{1}{n\tilde{\sigma}_\nu^8}\displaystyle\sum_{i=1}^n \ddot{\underline{Z}}_i'\left(\operatorname{diag}(\tilde{u}_i^{\hat{}4}) - \tilde{\sigma}_\nu^4 I_{T_i}\right)\ddot{\underline{Z}}_i \end{bmatrix} \tag{3.32}$$

where $\tilde{u}_i\hat{}^4 = \tilde{u}_i \odot \tilde{u}_i \odot \tilde{u}_i \odot \tilde{u}_i$.

Collecting (3.26), (3.28) and (3.32), the $\mathcal{PLM}_n^{I_rH}$ statistic thus turns out to simply be

$$\mathcal{PLM}_n^{I_rH} = \mathcal{PLM}_n^{I_r} + \mathcal{PLM}_n^{H} \xrightarrow{d} \chi^2(1 + k_{\underline{\gamma}}) \qquad (3.33)$$

where

$$\mathcal{PLM}_n^{I_r} = \frac{1}{2} \frac{\left(\left(\frac{1}{\tilde{\sigma}_\nu^2} \sum_{i=1}^n (\tilde{u}_i' e_{T_i})^2 \right) - N \right)^2}{\left(\sum_{i=1}^n T_i \right)^2 - N} \xrightarrow{d} \chi^2(1)$$

$$\mathcal{PLM}_n^{H} = \left(\sum_{i=1}^n \underline{Z}_i' \left(\tilde{u}_i\hat{}^2 - \tilde{\sigma}_\nu^2 e_{T_i} \right) \right)' \left(\sum_{i=1}^n \ddot{\underline{Z}}_i' \left(\text{diag}(\tilde{u}_i\hat{}^4) - \tilde{\sigma}_\nu^4 I_{T_i} \right) \ddot{\underline{Z}}_i \right)^{-1} \left(\sum_{i=1}^n \underline{Z}_i' \left(\tilde{u}_i\hat{}^2 - \tilde{\sigma}_\nu^2 e_{T_i} \right) \right) \xrightarrow{d} \chi^2(k_{\underline{\gamma}})$$

i.e., the sum of two asymptotically independent pseudo-LM statistics asymptotically distributed as, respectively, $\chi^2(1)$ and $\chi^2(k_{\underline{\gamma}})$.

$\mathcal{PLM}_n^{I_r}$ is nothing else than the incomplete panel version of the Breush-Pagan (1980) standard LM test for one-way error components derived in Baltagi-Li (1990). The balanced version of this standard LM test was shown to be robust to non-normality by Honda (1985). For deriving its result, Honda (1985) assumed IID errors under the null. The present result shows that robustness to non-normality also holds both in the unbalanced case and under the weaker assumption of independently but not necessarily identically distributed errors under H_0.

$\mathcal{PLM}_n^{H}$ contains as special cases well-known tests for heteroscedasticity. So, if the fourth order conditional moments of ν_{it} are further assumed constant under H_0, i.e., for all i and t, $E(\nu_{it}^4 | X_i, \underline{Z}_i) = \delta^o$, $\mathcal{PLM}_n^{H}$ collapses to the Koenker's (1981) statistic

$$\mathcal{PLM}_n^{H^K} = \frac{1}{\tilde{\delta} - \tilde{\sigma}_\nu^4} \left(\sum_{i=1}^n \underline{Z}_i' \left(\tilde{u}_i\hat{}^2 - \tilde{\sigma}_\nu^2 e_{T_i} \right) \right)' \left(\sum_{i=1}^n \ddot{\underline{Z}}_i' \ddot{\underline{Z}}_i \right)^{-1} \left(\sum_{i=1}^n \underline{Z}_i' \left(\tilde{u}_i\hat{}^2 - \tilde{\sigma}_\nu^2 e_{T_i} \right) \right)$$

where $\tilde{\delta}$ is a consistent estimator of δ^o. Likewise, if the ν_{it} are further assumed conditionally normal under H_0, i.e., for all i and t, $E(\nu_{it}^4 | X_i, \underline{Z}_i) = 3\sigma_\nu^{o4}$, we obtain the standard Breush-Pagan's (1979) statistic

$$\mathcal{PLM}_n^{H^{BP}} = \frac{1}{2\tilde{\sigma}_\nu^4} \left(\sum_{i=1}^n \underline{Z}_i' \left(\tilde{u}_i\hat{}^2 - \tilde{\sigma}_\nu^2 e_{T_i} \right) \right)' \left(\sum_{i=1}^n \ddot{\underline{Z}}_i' \ddot{\underline{Z}}_i \right)^{-1} \left(\sum_{i=1}^n \underline{Z}_i' \left(\tilde{u}_i\hat{}^2 - \tilde{\sigma}_\nu^2 e_{T_i} \right) \right)$$

For practical purpose, it is worth noting that a statistic asymptotically equivalent to $\mathcal{PLM}_n^{H}$ may be computed as N minus the residual sum of squares $(= NR_u^2)$

of the OLS regression

$$1 = \left[\left(\tilde{u}_{it}^2 - \tilde{\sigma}_\nu^2\right) \ddot{\underline{Z}}_{it}\right] b + \text{residuals}, \qquad i = 1, 2, ..., n\ ; \quad t = 1, 2, ..., T_i$$

where $\ddot{\underline{Z}}_{it}$ denotes t-th row of $\ddot{\underline{Z}}_i$. This statistic — which is a genuine Wooldridge's (1990) modified m-test — was outlined in Wooldridge (1990). Using this latter regression-based form of $\mathcal{PLM}_n^H$, provided that $\mathcal{PLM}_n^{I_r}$ is available from some standard software[16], $\mathcal{PLM}_n^{I_rH}$ is quite easy to implement.

A nice by-product of the additive structure of $\mathcal{PLM}_n^{I_rH}$ is that, although not unambiguously, it readily allows for gaining insights about the direction(s) in which misspecification detected by the joint statistic may lie by looking at the one-directional statistics $\mathcal{PLM}_n^{I_r}$ and $\mathcal{PLM}_n^H$. From a formal point of view, this may be done by using a Bonferroni approach in the reverse manner: the direction(s) in which misspecification detected by the joint test at asymptotic size α lies may tentatively be identified as given by the one-directional test statistic(s) rejected at asymptotic size $\alpha/2$ (on this approach, see Savin (1980,1984) and Bera-Jarque (1982))[17].

Being specifically designed for this purpose, we may expect $\mathcal{PLM}_n^{I_rH}$ to have good power for detecting heteroscedasticity and one-way error components like patterns in the second order moments of the data. On the contrary, we may not really expect it to exhibit good power against misspecification of the conditional mean, which is also part of the null hypothesis H_0. So, although also possibly due to size distortion arising from a lack of independence of the errors as assumed under the null, a rejection of the joint test along with some evidence that the rejection stems from its two components may be viewed as providing some support for tentatively looking at the general model (3.7). It is however worth stressing that it is very tentative: rejection of the joint statistic might well actually arise from misspecification of the mean and/or size distortion while, under the alternative, each one-directional statistic may be "contaminated" by a departure from the null in the other direction.

3.6. Concluding comments

This chapter proposed an extension of the standard one-way error components linear regression model allowing for heteroscedasticity in both the individual-specific and general error terms and, using the general results of the previous chapters, provided a comprehensive robust to distributional and conditional variance misspecification integrated inferential framework for its estimation and specification testing.

We believe that this model and its accompanying robust inferential methods should be useful for analyzing short, possibly unbalanced, microeconomic panel datasets. On the one hand, from an economic point of view, it offers an intuitively appealing way for modelling variable heterogeneity in both the between and within

[16] It is for example computed by LIMDEP.

[17] Note that asymptotic independence implies that a joint induced test based on the separate statistic $\mathcal{PLM}_n^{I_r}$ and $\mathcal{PLM}_n^H$ which consists in rejecting the joint null H_0 if one or more of the separate statistics $\mathcal{PLM}_n^{I_r}$ and $\mathcal{PLM}_n^H$ are rejected at asymptotic significance levels α_{I_r} and α_H, has an exact asymptotic size equal to $\alpha = \alpha_{I_r} + \alpha_H - \alpha_{I_r}\alpha_H$. At the usual significance levels (e.g., 0.05), taking the standard Bonferroni solution $\alpha_{I_r} = \alpha_H = \alpha/2$ is thus meaningful.

dimensions, so that if it actually proves to be correctly specified, it may provide very interesting information about the heterogeneity of the economic relationship under consideration. On the other hand, from a more statistical point of view, this specification embodies the scedastic characteristics which are the most likely to be observed when dealing with microeconomic panel data: autocorrelation in the time-series dimension and heteroscedasticity in the cross-section dimension. So, even if it actually proves to be second order misspecified, besides also providing some (possibly misleading) insights about the heterogeneity of the observations, it nevertheless allows to (eventually) get efficiency gains — both for estimation and testing of the conditional mean — from approximately taking into account the scedastic structure of the data. At this level, the robustness to conditional variance property of the outlined estimation and testing procedures is of course essential.

In this latter perspective, the fact that the proposed preliminary joint pseudo Lagrange multiplier test might not be very reliable in identifying heteroscedasticity is not crucial: its purpose is just to give some insights about the potential relevance of right away considering a heteroscedastic model. If it is not the case, we may more simply first look at the standard homoscedastic model. The estimation of the model may then be undertaken either by GPML2 or FGLS (for the unbalanced case, see for example Baltagi (1985,1994)). Whatever the choice — as outlined, to be valid, the diagnostic tests do not require the GMPL2 estimator but only consistent estimators under the null[18] —, the homoscedastic model may likewise be tested in mean and variance through the outlined diagnostic tests, and in particular tested against our general heteroscedastic model as an auxiliary nested alternative.

To conclude, let us point out that we considered a linear specification in the mean and a semi-linear specification in the variance mainly for simplicity and because it is the most likely to be used in practice. Using the results of Chapters 1 and 2, and following the lines of this chapter, it is straightforward to extend the present results to fully nonlinear specifications. Finally, regarding the implementation of the diagnostic tests, it worth recalling what we said at the end of Chapter 2. An extensive investigation of the conditional mean and conditional variance of the model should be based on both Hausman or information matrix type misspecification indicator(s) and misspecification indicator(s) designed to check the null against plausible auxiliary (nested or non-nested) alternatives. Further, because it may provide useful, although possibly misleading, information about the source(s) of departure from the null, it is probably a good strategy to check individually the chosen misspecification indicators associated to the different aspects of the model specification, keeping in mind that, from a formal point of view, joint induced tests with bounded asymptotic size of the null of interest may be carried out by using a Bonferroni approach.

[18] Although probably not commendable in small samples, note that mean and variance diagnostic tests of the standard model could actually be implemented by roughly resorting to the OLS estimator of β along with the simplest OLS residuals based estimators of the variance components.

Chapter 4

An empirical illustration: production functions estimation and testing

4.1. Introduction

When proposing an extension of a well-established model and statistical tools to deal with as in Chapter 3, some questions naturally come out: what is its empirical significance? how does its estimation and testing work in practice? The purpose of this chapter is to exemplify the potential usefulness of the proposed full heteroscedastic one-way error components model and its accompanying robust inferential methods through an empirical illustration consisting in production functions estimation and specification testing. This illustration is based on a strongly unbalanced panel dataset of 824 french firms observed over the period 1979 - 1988.

Since the seminal work of Cobb and Douglas (1928), considerable progress has been made by production theory and econometric methods, and data availability has rapidly grown, both in quantity and quality, shifting in particular from the macro level to the more relevant micro level. But the question raised by Douglas (1948) in its presidential address at the American Economic Association is still on the agenda: are there laws of production? Put in other words, is it possible to get a satisfactory, i.e., correctly specified, empirical counterpart to the indisputably fruitful theoretical concept of production function.

Our goal here is certainly not to discuss the numerous issues associated to this vast question. For an up-to-date discussion and references, see for example Mairesse (1988) and Griliches-Mairesse (1990,1995). More modestly, we concentrate on estimating and testing at an inter-sectorial level the correctness of the specification of two simple — commonly used in empirical practice — transcendental logarithmic (translog) production models, one taking into account differences in the "quality" of labor and the other not. This empirical illustration suggests (a) that heteroscedasticity-related problems are likely to be present when estimating this kind of production models using (cross-section or) panel data, (b) that the proposed full heteroscedastic one-way error components model and its accompanying robust inferential methods may offer a sensible, although imperfect, way to deal with it, and finally (c) that the set of proposed specification tests allows to get interesting insights about the empirical correctness of these simple models. In this latter respect, we will see that the more detailed model does not turn out to be the most appropriate.

This chapter is organized as follows. Section 4.2 describes the data and the tested production models. Section 4.3 reports and discusses the obtained empirical results. Finally, Section 4.4 proposes some concluding comments.

4.2. Data and tested models

The data originally come from a panel dataset constituted by the "Marchés et Stratégie d'Entreprises" division of INSEE[1]. The present dataset is actually a cleaned subset of this original dataset[2]. It contains 5 201 observations and, as already outlined, consists in a strongly unbalanced panel dataset of 824 french firms observed over the period 1979-1988. About one third only of the firms are observed over the entire period. The mean number of observations per firm is about 6.31.

Nine out of the fourteen sectors which compose the NAP 15 Classification are represented in the sample (sectorial code in parentheses): agricultural and food industries (02), energy production and distribution (03), intermediate goods industries (04), equipment goods industries (05), consumption goods industries (06), construction and civil engineering industries (07), trade (08), transport and telecommunications (09), market services (10). The sectors not represented in the sample are: agriculture, real estate renting and leasing, insurance, financial institutions, non-market services.

The number of firms and the number of observations per represented sector in the sample are reported in Table 1. As shown in this table, the bulk of the observations (about 76 %) actually belongs to the four sectors 04, 05, 06 and 08.

Table 1: Sectorial composition of the dataset (1979 - 1988)

Sector	Nb. of firms	Nb. of obs.	Nb. of obs. / firms.
02	52	377	7.25
03	14	84	6.00
04	144	843	5.85
05	184	1 240	6.74
06	162	990	6.11
07	48	329	6.85
08	157	885	5.64
09	20	137	6.85
10	67	316	4.72

The definition of the variables used hereafter in the production models are the following:

[1] For a precise description of this panel dataset, which is actually built upon different data sources, see Blanchard P. and al. (1996).

[2] I wish to thank Patrick Sevestre for kindly providing the dataset and Pierre Blanchard for furnishing it to me in a convenient format.

- va : value added deflated by an NAP 40 sector-specific price index (base : 1980).
- k : stock of capital.
- l : total number of workers ($l = l_s + l_{us}$).
- l_s : number of skilled workers.
- l_{us}: number of unskilled workers.

The stock of capital variable has been constructed by INSEE.

Table 2 reports some descriptive statistics for the entire sample. As it may be seen, the observations are extremely dispersed : the largest firm employs almost 32 000 workers while the smallest only 19, the capitalistic intensity varies from 10 to more than 3 200, while the proportion of skilled workers in the labor force ranges from 3 % to almost 85 %. Clearly, large firms are over-represented. As usual in this kind of dataset, the observation variability essentially lies in the between (across individuals) dimension. Note that the same extreme dispersion is observed at the sectorial level.

Table 2 : Descriptive statistics (5201 obs., 1979-1988)

Variable	Mean	Std. dev.	Min.	Max.
va	160 233	303 652	1 451	4 506 209
k	343 105	874 939	602	10 441 233
l	1 117	1 870	19	31 762
l_s	313	703	4	11 076
l_{us}	803	1 290	8	23 375
k/l	260.9	291.3	10.3	3217.6
l_s/l	0.283	0.157	0.030	0.844

We are thus interested in estimating and testing two simple translog production function models, one taking into account differences in the "quality" — skilled versus unskilled workers— of labor and the other not. These two models are defined hereafter. The notation is the same than in Section 3.2. The same statistical assumptions — equations (3.3)-(3.5) — are tentatively assumed to hold. Both models are thus second order semi-parametric models like (3.7).

- Model I :

$$V_{it} = \beta_{(sc\times t)} + \beta^k K_{it} + \beta^l L_{it} + \beta^{kk} K_{it}^2 + \beta^{ll} L_{it}^2 + \beta^{kl} K_{it} L_{it} + \mu_i + \nu_{it}$$

with

$$\begin{aligned}\sigma^2_{\nu_{it}} &= \exp\left(\gamma + \gamma^k K_{it} + \gamma^l L_{it}\right)\\ \sigma^2_{\mu_i} &= \exp\left(\alpha + \alpha^k \bar{K}_i + \alpha^l \bar{L}_i\right)\end{aligned}$$

where

$$V_{it} = \ln va_{it}, \quad K_{it} = (\ln k_{it} - \ln k^*), \quad L_{it} = (\ln l_{it} - \ln l^*),$$

$$\bar{K}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} K_{it}, \quad \bar{L}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} L_{it}$$

- Model II:

$$\begin{aligned} V_{it} &= \beta_{(sc\times t)} + \beta^k K_{it} + \beta^{l_s} L_{s_{it}} + \beta^{l_{us}} L_{us_{it}} + \beta^{kk} K_{it}^2 + \beta^{l_s l_s} L_{s_{it}}^2 + \beta^{l_{us} l_{us}} L_{us_{it}}^2 \\ &\quad + \beta^{kl_s} K_{it} L_{s_{it}} + \beta^{kl_{us}} K_{it} L_{us_{it}} + \beta^{l_s l_{us}} L_{s_{it}} L_{us_{it}} + \mu_i + \nu_{it} \end{aligned}$$

with

$$\begin{aligned} \sigma^2_{\nu_{it}} &= \exp\left(\gamma + \gamma^k K_{it} + \gamma^{l_s} L_{s_{it}} + \gamma^{l_{us}} L_{us_{it}}\right) \\ \sigma^2_{\mu_i} &= \exp\left(\alpha + \alpha^k \bar{K}_i + \alpha^{l_s} \bar{L}_{s_i} + \alpha^{l_{us}} \bar{L}_{us_i}\right) \end{aligned}$$

where, similarly,

$$L_{s_{it}} = (\ln l_{s_{it}} - \ln l_s^*), \quad L_{us_{it}} = (\ln l_{us_{it}} - \ln l_{us}^*)$$

$$\bar{L}_{s_i} = \frac{1}{T_i}\sum_{t=1}^{T_i} L_{s_{it}}, \quad \bar{L}_{us_i} = \frac{1}{T_i}\sum_{t=1}^{T_i} L_{us_{it}}$$

In both models the subscript '$(sc\times t)$' attached to the intercept parameter $\beta_{(sc\times t)}$ means that we actually let the intercept be sectorial and time-period specific. Each model thus contains 90 dummies (9 sectors × 10 periods). This allows for sector-specific productivity growth patterns. Our primary interest is to see whether a common parametrization holds for the non-intercept parameters.

The explanatory variables are centered so that the estimated values of β^k and β^l (resp. β^k, β^{l_s} and $\beta^{l_{us}}$) reported below may directly be interpreted as the elasticities of the value added with respect to capital and labor (resp. capital, skilled labor and unskilled labor) at $k = k^*$ and $l = l^*$ (resp. $k = k^*$, $l_s = l_s^*$ and $l_{us} = l_{us}^*$). We set k^* and l^* (resp. k^*, l_s^* and l_{us}^*) at their entire sample means as given in Table 2.

For both the individual-specific and general error variance functions, and for both models, we adopt the Harvey's (1976) multiplicative heteroscedasticity formulation. In the general error variance functions, the explanatory variables are simply taken as the (log of the) different inputs. Taking the individual mean values of the (log of the) different inputs as explanatory variables in the individual-specific variance functions is mainly a pragmatic choice. It appears sensible as far as the observation variability prominently lies in the between dimension. Be that as it may, these choices allow the variances to change according to both size and input ratios.

Model I and model II clearly consider different sets of conditioning variables as explanatory variables for the two first conditional moments of (the log of) the value added. The set of conditioning variables of model I is $CV_{\text{I}} \equiv (D_{sc}, D_t, k, l) \equiv (D_{sc}, D_t, k, (l_s + l_{us}))$ while the one of model II is $CV_{\text{II}} \equiv (D_{sc}, D_t, k, l_s, l_{us})$, where D_{sc} and D_t denote respectively sectorial ($sc = 1, ..., 9$) and time ($t = 1, ..., 10$) dummies. The latter is obviously larger than the former. Generally speaking, considering different sets of conditioning variables yields different models, so that it is always possible, at least in principle, that all considered models turn out to be simultaneously correctly specified. In the present case however, given the maintained functional forms, they are essentially incompatible. In other words, we may expect to find out misspecified at least one of the two models. Note finally that in both models the conditional variance is specified as functions of variables which do appear in the conditional mean, so that it does not alter the conditioning set with respect to which the conditional mean is tentatively assumed to be correctly specified. The converse is nevertheless not true.

4.3. Empirical results

We proceed in three steps. We first report preliminary estimation and testing of the models. We then outline the results obtained from GPML2 estimation. Finally, we deal with specification testing.

4.3.1. Preliminary estimation end testing

As yardstick, we first provide the obtained results from OLS and within OLS (WOLS) estimation of model I and II. They are given in Table 3 and 4. The reported standard errors are heteroscedasticity-robust[3].

Table 3 : OLS and within OLS estimation of model I

	OLS			WOLS		
Variable	Coefficient	Std. error	t-ratio	Coefficient	Std. error	t-ratio
K	0.2222	0.0155	14.34	0.2637	0.0460	5.73
L	0.7780	0.0219	35.62	0.7097	0.0377	18.84
K^2	0.0422	0.0075	5.64	0.0625	0.0119	5.26
L^2	0.0308	0.0141	2.18	0.0723	0.0180	4.01
KL	-0.0748	0.0190	-3.89	-0.1372	0.0261	-5.26

Heteroscedasticity-robust standard errors

Table 4 : OLS and within OLS estimation of model II

	OLS			WOLS		
Variable	Coefficient	Std. error	t-ratio	Coefficient	Std. error	t-ratio
K	0.2039	0.0138	14.73	0.2462	0.0457	5.38
L_s	0.4244	0.0142	29.98	0.2083	0.0159	13.07
L_{us}	0.3759	0.0176	21.35	0.5136	0.0271	18.97
K^2	0.0375	0.0068	5.49	0.0494	0.0118	4.19
L_s^2	0.1071	0.0079	13.64	0.0637	0.0068	9.33
L_{us}^2	0.0814	0.0119	6.87	0.1100	0.0102	10.77
KL_s	-0.0210	0.0119	-1.76	-0.0134	0.0109	-1.23
KL_{us}	-0.0452	0.0146	-3.09	-0.0827	0.0189	-4.38
L_sL_{us}	-0.1671	0.0138	-12.14	-0.1299	0.0132	-9.83

Heteroscedasticity-robust standard errors

[3] Actually, since we are in a multivariate framework, we should say heteroscedasticity and autocorrelation robust.

The R-squared the OLS regressions of model I and II are respectively equal to 0.952 and 0.961. Thus, considering the "quality" — skilled versus unskilled — of the labor force does not really add a lot in terms of explanatory power.

Under correct specification of the conditional mean, the coefficients obtained from OLS and WOLS should not be too different. At first sight, taking into account the reported standard errors, they do not seem too dramatically different for model I. This does not appear to be the case for model II, in particular regarding the coefficients of L_s and L_{us}. All this already suggests that model I, although coarser, might be more appropriate than model II. We shall return to this point below when considering the specification tests.

Is it worth considering full heteroscedastic models like model I and II? As suggested in Section 3.5, the preliminary pseudo-LM statistic $\mathcal{PLM}_n^{I_r H}$ along with its one-directional components $\mathcal{PLM}_n^{I_r}$ and $\mathcal{PLM}_n^{H}$ may provide some interesting insights about this question. Table 5 reports the results of their computation for model I and II. The retained variables for their $\mathcal{PLM}_n^{H}$ components are logically the variables appearing in their general error variance functions.

Table 5 : Preliminary pseudo-LM tests of model I and II

		$\mathcal{PLM}_n^{I_r H}$	$\mathcal{PLM}_n^{I_r}$	$\mathcal{PLM}_n^{H}$
	Stat.	11 016.6	10 981.0	35.6
Model I	D.f.	3	1	2
	p-value	0.0000	0.0000	0.0000
	Stat.	8 194.3	8 118.9	75.4
Model II	D.f.	4	1	3
	p-value	0.0000	0.0000	0.0000

The pseudo-LM statistic $\mathcal{PLM}_n^{I_r H}$ turns out to drastically reject the joint null of no individual effects and homoscedasticity in both models. Looking at their one-directional components reveals that, although the values of the joint statistics are mainly explained by their individual effects components, homoscedasticity is also strongly rejected in both models. This provides support for indeed looking at the full heteroscedastic models.

4.3.2. GPML2 estimation

The results of GPML2 estimation of model I and II are reported in Table 6 and 7. In both cases, the standard errors of the parameters are computed as it was suggested to routinely compute them in Section 3.3.3, i.e., as given in (3.22). For the record, this yields correct standard errors for the conditional mean parameters and an upper bound of the true standard errors for the conditional variance parameters if the models are correctly specified for the conditional mean but misspecified for the conditional variance. On the other hand, this yields correct standard errors of all parameters if the models are correctly specified for both the conditional mean and the conditional variance.

Based on these computed covariance matrix of the parameters — and thus on the hypothesis which underlies their validity —, Table 8 further reports some Wald tests: a test of the appropriateness of the nested Cobb-Douglas formulation for the conditional mean, a test of the joint significance of the dummies introduced in the models, and a test of heteroscedasticity (checking the significance of the non-intercept parameters of both individual-specific and general variance functions).

Table 6: GPML2 estimation of model I

Variable	Coefficient	Std. error	t-ratio	p-value
K	0.2487	0.0188	13.26	0.0000
L	0.7367	0.0244	30.21	0.0000
K^2	0.0547	0.0072	7.58	0.0000
L^2	0.0572	0.0132	4.35	0.0000
KL	-0.1137	0.0176	-6.48	0.0000
$\sigma^2_{\nu_{it}} = \exp(.)$				
const.	-4.1997	0.0541	-77.65	0.0000
K	0.1870	0.0582	3.21	0.0013
L	-0.2482	0.0849	-2.92	0.0035
$\sigma^2_{\mu_i} = \exp(.)$				
const.	-2.5213	0.0732	-34.43	0.0000
$\bar{K}$	0.1676	0.0610	2.74	0.0060
$\bar{L}$	-0.1709	0.0799	-2.14	0.0325

Standard errors computed according to (3.22)

Table 7: GPML2 estimation of model II

Variable	Coefficient	Std. error	t-ratio	p-value
K	0.2414	0.0182	13.27	0.0000
L_s	0.2351	0.0131	17.98	0.0000
L_{us}	0.5034	0.0176	28.62	0.0000
K^2	0.0488	0.0071	6.89	0.0000
L_s^2	0.0694	0.0063	10.99	0.0000
L_{us}^2	0.1057	0.0088	11.97	0.0000
KL_s	-0.0194	0.0089	-2.19	0.0284
KL_{us}	-0.0747	0.0134	-5.56	0.0000
L_sL_{us}	-0.1340	0.0102	-13.15	0.0000
$\sigma^2_{\nu_{it}} = \exp(.)$				
const.	-4.1923	0.0558	-75.14	0.0000
K	0.1868	0.0555	3.37	0.0008
L_s	0.0132	0.0641	0.21	0.8370
L_{us}	-0.2607	0.0736	-3.54	0.0004
$\sigma^2_{\mu_i} = \exp(.)$				
const.	-2.5171	0.0717	-35.12	0.0000
$\bar{K}$	0.1773	0.0618	2.87	0.0041
$\bar{L}_s$	0.3415	0.0694	4.92	0.0000
$\bar{L}_{us}$	-0.4823	0.0792	-6.09	0.0000

Standard errors computed according to (3.22)

Table 8 : Wald tests of model I and II (GPML2 estimation)

	Model I			Model II		
	Stat.	D.f.	p-value	stat.	D.f.	p-value
H_0: Cobb-Douglas	57.8	3	0.0000	384.3	6	0.0000
H_0: no sector × time effects	2 307.9	89	0.0000	1 796.7	89	0.0000
H_0: no heteroscedasticity	18.9	4	0.0008	65.8	6	0.0000

Regarding the mean parameters, first note that, for model I, the GMPL2 estimates are roughly at "equal distance" between the OLS and WOLS estimates — and thus not very different of them —, while in the case of model II, they are much more similar to WOLS than to OLS. This reinforces the idea that model I seems more appropriate than model II.

Further, again regarding the mean parameters, according to Table 8, both the simple Cobb-Douglas formulation and the hypothesis of no 'sector × time' effects are strongly rejected in the two models. With respect to the 'sector × time' effects, note that separate tests suggest that sector-specific, time-specific and sector-time interactions are all responsible of the rejection of the joint 'sector × time' effects tests, and that in both models.

Turning our attention to the variance parameters, according again to Table 8 — recall however that they are not valid (conservative) tests of the variance parameters if the mean is misspecified —, it appears that heteroscedasticity-like patterns are effectively present in both the individual-specific and general error second order moments of the two models. In all cases, heteroscedasticity seems related to input ratios : more capitalistic and/or skilled labor intensive firms appear more heterogeneous both in the between and within dimensions than more laboristic and/or unskilled labor intensive firms.

The captured heteroscedasticity however does not seem to be notably related to size. Figures 1 and 2 portray this latter point. In these figures, estimated general error and individual error variances are graphed against the observations sorted in ascending order according to individual means of the fitted dependent variable and, within each individual, according to the values of the fitted dependent variable itself.

Figure 1 : Estimated variances versus size in Model I

General error term variances : $\hat{\sigma}^2_{\nu_{it}}$ Individual error term variances : $\hat{\sigma}^2_{\mu_i}$

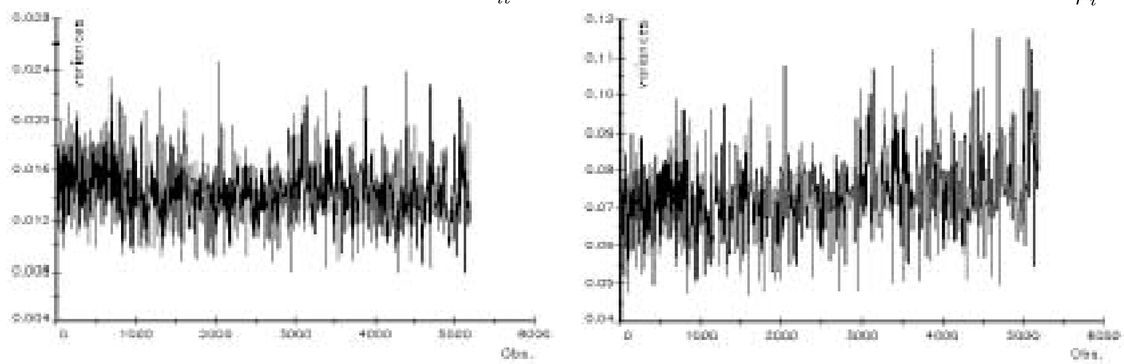
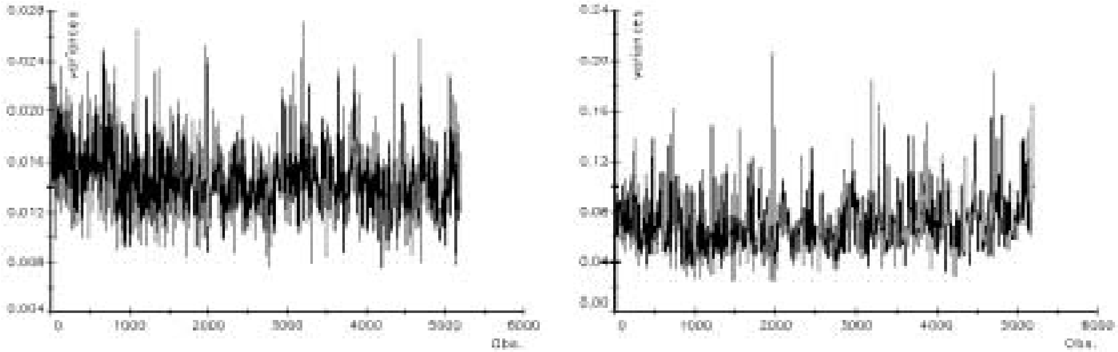


Figure 2 : Estimated variances versus size in Model II

General error term variances : $\hat{\sigma}^2_{\nu_{it}}$ Individual error term variances : $\hat{\sigma}^2_{\mu_i}$



None of these figures reveals notable links between variances and size. They however outline two other points. First, variations in the observed inputs ratios imply variations in the estimated variances — identified by the difference between the lower and upper levels of the estimated variances — of more than a factor 2. Second, the estimated individual-specific variances are roughly 5 - 6 times higher than the estimated general error variances.

4.3.3. Specification testing

Following Section 3.4, we check the conditional mean and the conditional variance specifications of model I and II through Hausman type tests, information matrix type tests and tests against auxiliary nested alternatives. The performed tests are described hereafter. They are all implemented as outlined in Section 3.4.1 and 3.4.2.

- Conditional mean diagnostic tests :
  - Test (1) : Hausman type test based on comparing the GPML2 and OLS estimators of all mean parameters (including the dummies).
  - Test (2) : idem than (1) except that it concentrates on the non-intercept mean parameters (thus excluding the dummies).
  - Test (3) : information matrix type test based on checking the nullity of the sub-block of the hessian corresponding to the cross-derivatives between the non-intercept mean parameters and all variance parameters (excepted the intercept of the individual-specific variance function[4]).
  - Test (4) : test against an auxiliary nested alternative including as additional variables the interactions between a trend and the first order terms of the tranlog function. This tests for non-neutral technical progress[5].
  - Test (5) : test against an auxiliary nested alternative including as additional variables to the null translog specification terms of third power[6].

[4] To avoid singularity (see Section 3.4.1).

[5] Non-neutral technical progress is typically modelled by considering a trend as an additional input. The trend and trend-squared terms being already captured by the set of dummies, it thus remains to test for the interaction terms between the trend and the first order terms of the translog function.

This tests the functional form.

- Test (6) : test against an auxiliary nested alternative allowing for the non-intercept mean parameters to be time-period specific. This tests for time heterogeneity.
- Test (7) : test against an auxiliary nested alternative allowing for the non-intercept mean parameters to be sector-specific. This tests for sectorial heterogeneity.

- Conditional variance diagnostic tests :
  - Test (8) : Hausman type test based on comparing the GPML2 and (unweighted) MNLS[7] estimators of all variance parameters.
  - Test (9) : information matrix type test based on checking the non-redundant elements of the sub-block of the information matrix equality associated with the non-intercept mean parameters.
  - Test (10) : idem than (9) except that it considers the sum of the indicators on which (9) is based.
  - Test (11) : test against an auxiliary nested alternative specifying both the individual-specific and general error variances as (the exponential of) translog functions instead of Cobb-Douglas like functions. This tests the functional forms.
  - Test (12) : test against an auxiliary nested alternative allowing for all variance parameters to be sector-specific. This tests for sectorial heterogeneity.

For the record, all conditional mean tests are robust to distributional and conditional variance misspecification while all conditional variance tests are robust to distributional misspecification. Note further that none of the above diagnostic tests against auxiliary alternatives resort to variables which are not functions of the original sets of conditioning variables (CV_{I} for model I and CV_{II} for model II). The null hypothesis of these tests is thus never more than the null models themselves.

Table 9 reports the results obtained from the computation of the above diagnostic tests of model I and II.

Let us first consider the conditional mean specification tests.

As it may be seen, the conditional mean tests confirm what was already felt from simply comparing the OLS and WOLS estimators of the models : model II appears seriously misspecified while model I does not appear to exhibit patent misspecification, the observed differences between the alternative estimators of model I seeming to be well and truly attributable to randomness. Regarding in particular the very large sample size, model I appears very surprisingly well specified : it does not seem to patently suffer of endogeneity of the inputs, functional misspecification, sectorial heterogeneity or temporal instability, to mention some of the "best-sellers" of the misspecification catalogue. The only statistic which indicates some possible deviation from the null is the test (4). Its p-value is however not really worrying : from a formal point of view, according a standard Bonferroni approach, for rejecting at 5 % the null hypothesis that the conditional mean is correctly specified, we "need"

[6] i.e., for model I, K^3, L^3, KL^2 and K^2L, and for model II, K^3, L_s^3, L_{us}^3, KL_s^2, KL_{us}^2, K^2L_s, K^2L_{us}, $L_sL_{us}^2$, $L_s^2L_{us}$, KL_sL_{us}. This corresponds to the third order terms of a (multivariate) Taylor expansion, the translog function being itself a second order Taylor expansion.

[7] See Section 3.4.2.

that at least one of the 7 separate tests rejects the null at 0.71 % ($0.05/7 \simeq 0.0071$). Viewed in a less formal way, it is normal to find out some statistics which (moderately) deviate when multiplying the number of diagnostic tests.

Table 9 : Specification tests of model I and II (GPML2 estimation)

	Model I			Model II		
	Stat.	D.f.	p-value	Stat.	D.f.	p-value
Conditional mean tests						
(1) Hm (all ind.)	85.0	95	0.7602	190.1	99	0.0000
(2) Hm (all select. ind.)	5.9	5	0.3180	114.8	9	0.0000
(3) Im (all select. ind.)	33.7	25	0.1141	192.0	63	0.0000
(4) H_1: non-neutral TP	8.4	2	0.0146	10.4	3	0.0151
(5) H_1: third power	2.8	4	0.5961	36.7	10	0.0001
(6) H_1: time heterogeneity	57.1	45	0.1064	100.58	81	0.0693
(7) H_1: sectorial heterogeneity	41.0	40	0.4249	104.2	72	0.0078
Conditional variance tests						
(8) Hm (all ind.)	18.4	6	0.0052	27.6	8	0.0006
(9) Im (all select. ind.)	45.6	15	0.0001	79.8	45	0.0011
(10) Im (sum of all select. ind.)	5.66	1	0.0173	9.1	1	0.0025
(11) H_1: second power	2.2	6	0.9015	11.3	12	0.5003
(12) H_1: sectorial heterogeneity	98.6	48	0.0000	98.6	64	0.0036

Identifying the plausible source(s) of a detected misspecification as in model II is in essence a perilous task. In this respect, diagnostic tests without explicit alternatives are not really helpful. Diagnostic tests against explicit alternatives are more informative. In the present case, they suggest that sectorial heterogeneity and misspecified functional form — possibly for the same kind of underlying reasons — are involved. On the other hand, time heterogeneity does not appear to be a major issue.

Table 10 and 11 provide further evidence of the absence of patent misspecification of model I and more information about the possible source(s) of misspecification of model II. These tables report the same set of conditional mean diagnostic tests than above, but computed at the sectorial level, based on separate GPML2 estimation of model I and II sector by sector[8]. For conciseness, we only report the p-values of the test statistics.

Before looking at the obtained results, note that the intercept of all estimated sector-specific version of model I and II is time-specific : each model thus contains 10 dummies (= 10 periods). Note further that whenever the dimension of the misspecification indicator underlying a test is larger than the number of firms in the sector, a nonsingular estimates of its covariance matrix cannot be obtained, so that the test statistic is not computed. The abbreviation 'sing.' appearing in Table 10 and 11 refers to these situations.

[8] Note that given the very small number of observations available in sector 03 and 09, in these sectors, the conditional variance specification has been confined to the standard homoscedastic one-way error components structure.

Table 10: P-values of sector-specific conditional mean specification tests of model I (sector-specific GPML2 estimation)

	Sector								
Test	02	03	04	05	06	07	08	09	10
(1)	0.7831	sing.	0.4878	0.0830	0.0422	0.3660	0.4776	0.3351	0.3255
(2)	0.8764	0.2966	0.2105	0.2076	0.0182	0.1986	0.6013	0.5653	0.0098
(3)	0.2944	0.2851	0.0371	0.0937	0.4959	0.3684	0.5544	0.5125	0.0698
(4)	0.2401	0.1825	0.0134	0.6231	0.0493	0.7778	0.1284	0.1222	0.7402
(5)	0.3932	0.2232	0.6022	0.4298	0.4168	0.0507	0.2996	0.1095	0.0611
(6)	0.3185	sing.	0.3487	0.1279	0.8169	0.7221	0.2372	sing.	0.5743

Table 11: P-values of sector-specific conditional mean specification tests of model II (sector-specific GPML2 estimation)

	Sector								
Test	02	03	04	05	06	07	08	09	10
(1)	0.5102	sing.	0.1400	0.0030	0.0002	0.3960	0.0047	0.4353	0.1136
(2)	0.2955	0.4939	0.0569	0.0013	0.0000	0.1538	0.0007	0.1837	0.0053
(3)	sing.	0.5089	0.1378	0.1503	0.1601	sing.	0.0287	0.3895	0.4583
(4)	0.4441	0.3668	0.0338	0.4740	0.0403	0.8857	0.1834	0.1031	0.5626
(5)	0.3514	0.2511	0.1293	0.0330	0.3258	0.2733	0.0079	0.1112	0.1913
(6)	sing.	sing.	0.5076	0.2409	0.7238	sing.	0.6144	sing.	sing.

As it may be seen, added to the absence of detected sectorial heterogeneity, Table 10 leads us to come to the same conclusion than above: model I does not appear to exhibit patent misspecification. It may be viewed as a satisfactory statistical representation of the available data.

The picture drawn by Table 11 is more ambiguous: only three sectors (05, 06 and 08) — to a smaller extent also sector 10 — exhibit imprecise but quite firm misspecification. It is however worth recalling that these three sectors are precisely three out of the four main sectors represented in the sample, i.e., three out of the four sectors for which a (very) large number of observations is available. In other words, the test statistics might lack power in the smaller sectors. Be that as it may, it appears that sectorial heterogeneity is certainly not the only source of the model II misspecification detected at the inter-sectorial level. Something deeper seems involved. Note by the way that this is congruent with the outlined fact that model I and II are essentially incompatible. On the other hand, it turns out that the conditional mean diagnostic tests prove to be unable to discriminate between model I and II in the smallest sectors. Finally, it seems that time heterogeneity is not a major issue[9].

Taking correct conditional mean specification of model I for granted, we may examine the results of the diagnostic tests of its conditional variance. This of course

[9] To complete Table 11, note in this respect that time heterogeneity tests of model II against an auxiliary nested alternative allowing for the non-intercept mean parameters to be specific in the periods 78-81, 82-85 and 86-88 (considering only three periods reduces the size of the indicator) yield for sector 02, 07 and 10 p-values respectively equal to 0.6801, 0.8091 and 0.2756.

does not really make sense for model II since the null of correct conditional variance specification embodies the null of correct conditional mean specification, which proved to be violated. Note by the way that this implies that the (conservative) Wald test for heteroscedasticity in model II reported in Table 8 is not valid. It is nevertheless quite likely that a heteroscedasticity-like pattern is indeed both present in the data and, as suggested in Table 9, misspecified.

The test for heteroscedasticity in model I reported in Table 8 is well and truly valid, and it clearly indicates the actual presence of a heteroscedasticity-like pattern in the second order conditional moments of the observations. According to Table 9, the assumed specification however turns out to be seriously misspecified. Test (11) suggests that relaxing the functional form would not really help. On the other hand, test (12) points out that a problem of sectorial heterogeneity might be involved.

To shed light on the latter point as well as to gauge the sensibility of the conditional mean estimates and diagnostic tests to the specification of the conditional variance, Table 12, 13 and 14 respectively report GPML2 estimation, Wald testing and diagnostic testing — the same tests than above — of a variant of model I (entitled model Ib) letting both the individual-specific and general error variance parameters to be sector-specific[10].

Table 12 : GPML2 estimation of model Ib

Variable	Coefficient	Std. error	t-ratio	p-value
K	0.2455	0.0169	14.54	0.0000
L	0.7519	0.0210	35.77	0.0000
K^2	0.0557	0.0062	9.03	0.0000
L^2	0.0639	0.0101	6.29	0.0000
KL	-0.1165	0.0148	-7.87	0.0000

Standard errors computed according to (3.22)

Table 13 : Wald tests of model Ib (GPML2 estimation)

	Stat.	D.f.	p-value
H_0: Cobb-Douglas	82.9	3	0.0000
H_0: no sector × time effects	3 027.2	89	0.0000
H_0: no heteroscedasticity	309.3	48	0.0000
H_0: no sector-specific heteroscedasticity	244.7	44	0.0000

For conciseness, Table 12 only reports the mean parameter estimates. Compared to GPML2 estimation of model I, the mean parameters estimates does not sensibly change. On the other hand, Table 13 confirms that both the simple Cobb-Douglas formulation and the hypothesis of no 'sector × time' effects are strongly rejected[11].

The conservative Wald tests reported in Table 13 also confirm the presence of a heteroscedasticity-like pattern in the second order conditional moments of the

[10] Again because of the very small number of observations available in sector 03 and 09, note that the non-intercept parameters of, on the one hand, the individual-specific variance function, and on the other hand, the general error variance function, have been constrained to be equal across these two sectors.

[11] As for model I, note that separate tests show that sector-specific, time-specific and sector-time interactions are all responsible of the rejection of the joint 'sector × time' effects test.

observations and the fact that the heteroscedasticity-like patterns are sector-specific. Note however that, although somewhat constrasted across sectors, the estimated sector-specific variances still mainly vary with input ratios, and that in the way outlined above.

Table 14 : Specification tests of model Ib (GPML2 estimation)

	Stat.	D.f.	p-value
Conditional mean tests			
(1) Hm (all ind.)	93.7	95	0.5175
(2) Hm (all select. ind.)	6.5	5	0.2579
(3) Im (all select. ind.)[12]	38.7	25	0.0396
(5) H_1: non-neutral TP	3.9	2	0.1446
(6) H_1: third power	3.3	4	0.5061
(7) H_1: time heterogeneity	55.6	45	0.1341
(8) H_1: sectorial heterogeneity	36.0	40	0.6505
Conditional variance tests			
(9) Hm (all ind.)	72.1	50	0.0221
(10) Im (all select. ind.)	52.8	15	0.0000
(11) Im (sum of all select. ind.)	4.5	1	0.0348

Finally, Table 14 further corroborates our finding that the conditional mean specification of model I does not exhibitit patent misspecification. Further, it shows that allowing for sector-specific variance functions does not solve our misspecification problem in the conditional variance. It is nevertheless not useless. Comparing the standard errors of the mean parameters reported in Table 6 and 12, it may indeed be seen that allowing for a more flexible conditional variance specification has entailed (moderate) efficiency gains : the reduction of the standard errors ranges from -10.1 % to -23.4 %. This shows that, as argued, besides providing some (possibly misleading) insights about the within and between heterogeneity, a misspecified conditional variance may also get efficiency benefits — for estimation but also testing of the conditional mean — from approximately taking into account the actual scedastic structure of the data.

4.4. Concluding comments

The primary purpose of this chapter was to exemplify the potential usefulness of the proposed full heteroscedastic one-way error components model and its accompanying robust inferential methods.

This illustration provides some support to the key points which motivated the theoretical developments undertaken in this dissertation : the need for some extension of the standard model which allows, in an intuitively appealing way and at least approximately, to take into account and to account for phenomenons of variable het-

[12] Test (3) is here based on checking the nullity of the sub-block of the hessian corresponding to the cross-derivatives between the non-intercept mean parameters and the variance parameters (excepted the intercept of the individual-specific variance) of the 'sector of reference' (taken as the largest sector : sector 05) by reference to which the other sector variance functions are defined (through sets of dummies).

erogeneity, i.e., heteroscedasticity, and the desirability to have at one's disposal a reasonably computationally convenient integrated inferential framework for both estimation and testing, explicitly taking into account the possibility of second order misspecification and allowing to easily handle incomplete panels. In this latter respect, note that in the present case, dropping from the original panel the individuals for which the observations are not complete (observed over the ten years 1979-1988) would have meant discarding almost one half of the available observations.

Regarding the obtained empirical results, the absence of detected misspecification in model I is a good news but also a surprising one, in particular given the very large sample size. On the other hand, the fact that model I appears more appropriate than model II is not really surprising. Considering a model with an enlarged set of conditioning variables may indeed be viewed as to go through the data into more details. But, going into details, the actual production activities — ranging from car manufacturing to hosiery — of the individual firms contained in the sample have almost nothing in common. It is thus not really surprising that, in this kind of modelling problem, going through the data into more details reveals additional heterogeneity which proves to be difficult to capture through a simple common parametrization. This of course does not discard other explanations but seems to us to be one of the relevant ones.

Conclusion

Starting from the acknowledgment that there is some need for generalized versions of the standard one-way error components model which take into account and account for phenomenons of variable heterogeneity, this dissertation pursued two main objectives: on the one hand, to propose and discuss such an extension of the standard model, and on the other hand, to provide an as comprehensive as possible statistical tool-box for its estimation and specification testing.

Chapter 3 exposed both the proposed extension of the standard model and a relevant statistical tool-box — following from the general results derived in Chapters 1 and 2 — to deal with.

The basic idea underlying the proposed extension is very simple. It amounts to letting both the individual-specific and the general error terms variances change by parametrically specifying these variances as functions of some set of explanatory variables. Doing this means adopting an economically and statistically appealing quite flexible parametrization allowing for variable heterogeneity both in the between and within dimensions.

For the estimation of the model, we argued for using second order pseudo-maximum likelihood methods. Chapter 1 provided the theoretical developments propping up this assertion. In a much more comprehensive framework than actually needed, in this chapter we outlined sufficient and necessary conditions for a second order pseudo-maximum likelihood estimator to be robust to conditional variance misspecification, and described the limiting distribution properties of such a nicely behaved estimator. It provided us with a potentially efficient and computationally convenient estimator of the model of interest, explicitly managing a possible misspecification of the assumed form of heterogeneity and further allowing to easily handle incomplete panels.

On the other hand, for specification testing of the model, we argued for taking advantage of the very powerful and flexible m-testing / Wooldridge's modified m-testing framework. Chapter 2 underpinned this claim. Remaining in the same comprehensive framework than in Chapter 1, it described how to check the specification of second order semi-parametric models. It put at our's disposal a large spectrum of m-type diagnostic tests for the model under scrutiny, whose prominent characteristic is that their validity requires no more than just the null hypothesis of interest. Combined with the nested nature of the null hypotheses of correct conditional mean and correct conditional variance specification, this provides ways to hopefully unambiguously identify eventual departures from the prominent aspects — mean and variance — of the proposed model specification.

Finally, Chapter 4 was intended to exemplifying through an empirical illustration the potential usefulness of the proposed full heteroscedastic one-way error components model and its accompanying robust inferential methods.

As suggested by Chapter 4, we believe that the proposed full heteroscedastic model and its accompanying robust inferential methods should be useful for analyzing short, possibly unbalanced, microeconomic panel datasets. On the one hand, from an economic point of view, it offers an intuitively appealing way for modelling variable heterogeneity in both the between and within dimensions, so that if it actually proves to be correctly specified, it may provide very interesting information about the heterogeneity of the economic relationship under consideration. On the other hand, from a more statistical point of view, this specification embodies the scedastic characteristics which are the most likely to be observed when dealing with microeconomic panel data : autocorrelation in the time-series dimension and heteroscedasticity in the cross-section dimension. So, even if it actually proves to be second order misspecified, besides also providing some (possibly misleading) insights about the heterogeneity of the observations, it nevertheless allows to (eventually) get efficiency gains — both for estimation and testing of the conditional mean — from approximately taking into account the scedastic structure of the data. At this level, the robustness to conditional variance property of the outlined estimation and testing procedures is of course essential.

Treating estimation and specification testing in a much more comprehensive framework than actually needed for the model under scrutiny, if somewhat more cumbersome, has the obvious by-product advantage that the obtained results may be used in a large spectrum of situations. From the panel data point of view, virtually all models, linear or nonlinear, assuming strictly exogenous explanatory variables and letting mean and variance parameters to vary independently, may actually be treated along the same lines than in Chapter 3. This includes numerous extensions of the standard one-way error components model : models with autocorrelation in the general error term, random coefficient models, seemingly unrelated regressions (SUR) models or any other similar extensions. In all cases, if computationally convenient, the model may be estimated by gaussian pseudo-maximum likelihood of order 2 without worrying about possible misspecification of the second order moments and may likewise be extensively tested through m-type diagnostic tests, including tests for competing non-nested specifications of the second order moments.

Although the generalized method of moments (GMM) offers a much more flexible framework to deal with, the same methodology may also be used when it is felt that the lack of strict exogeneity of the explanatory variables is only due to correlation between the individual effects and the regressors. In this case, we may put the observations in (first) differences and, taking into account the (possibly heteroscedastic) moving average process induced by this transformation, then proceed in the same way.

Throughout this work, we tried to stress the crucial role, regarding both estimation and specification testing, of the choice of the set of conditioning variables in an econometric modelling exercise. We emphasized the facts that this choice is ultimately up to the researcher, depending on what is of interest to him, and that different choices of conditioning variables actually yield different models. Obviously, we will usually not be able to get correctly specified models for any set of conditioning variables. Sometimes, as it happened in our empirical illustration, correctly specified models will prove to be easier to obtain for limited sets of conditioning variables, while in other cases, it may turn out to be the opposite. This kind of considerations seems to be overlooked by virtually all authors. We hope that this work will encourage to consider this issue more deeply.

To conclude, it is worth pointing out that all the results derived in this dissertation are asymptotic results. Whether or not they yield accurate and reliable approximations for finite samples remains an open question, in particular regarding the test statistics. Because the outlined tests take advantage of the generalized residuals structure of the problem at hand, we may hopefully expect that it does not too much suffer from the sometimes very poor finite sample properties exhibited by the standard Newey's (1985) outer-product gradient implementation of m-tests, as argued by Wooldridge (1990).

References

Amemiya T. (1973), "Regression analysis when the variance of the dependent variable is proportional to the square of its expectation", *Journal of the American Statistical Association*, 68, 928-934.

Amemiya T. (1985), *Advanced Econometrics*, Basil Blackwell, Oxford.

Amemiya T. and MaCurdy T.E. (1986), "Instrumental variable estimation of an error components model", *Econometrica*, 54 (4), 869-880.

Ahn S.C. and Low S. (1996), "A reformulation of the Hausman test for regression models with pooled cross-section-time-series data", *Journal of Econometrics*, 71, 309-319.

Arellano M. (1987), "Computing robust standard errors for within-groups estimators", *Oxford Bulletin of Economics and Statistics*, 49, 431-434.

Arellano M. (1993), "On the testing of correlated effects with panel data", *Journal of Econometrics*, 59, 87-97.

Balestra P. (1996), "Fixed effect models and fixed coefficient models", Ch. 3 in *The Econometrics of Panel Data*, eds. Mátyás L. and Sevestre P., Kluwer Academic Publishers, Dordrecht.

Baltagi B.H. (1985), "Pooling cross-sections with unequal time-series lengths", *Economics Letters*, 37, 39-44.

Baltagi B.H. and Griffin J.M. (1988), "A generalized error component model with heteroscedastic disturbances", *International Economic Review*, 29 (4), 745-753.

Baltagi B.H. (1988), "An alternative heteroscedastic error components model", Problem 88.2.2, *Econometric Theory*, 4, 349-350.

Baltagi B.H. and Li Q. (1990), "A Lagrange multiplier test for error components model with incomplete panels", *Econometric Reviews*, 9 (1), 103-107.

Baltagi B.H. and Li Q. (1991), "A joint test for serial correlation and random individual effects", *Statistics and Probability Letters*, 11, 277-280.

Baltagi B.H., Chang Y. and Li Q. (1992a), "Monte-Carlo results on several new and existing tests for the error component model", *Journal of Econometrics*, 54, 95-120.

Baltagi B.H. and Ray B. (1992b), "A survey of recent theoretical developments in the econometrics of panel data", *Empirical Economics*, 17, 85-109

Baltagi B.H., Chang Y. (1994), "Incomplete panels : a comparative study of alternative estimators for the unbalanced one-way error component regression model", *Journal of Econometrics*, 62, 67-89.

Baltagi B.H. (1995), *Econometric Analysis of Panel Data*, John Wiley & Sons, New York.

Baltagi B.H. (1996), "Specification issues", Ch. 12 in *The Econometrics of Panel Data*, eds. Mátyás L. and Sevestre P., Kluwer Academic Publishers, Dordrecht.

Baltagi B.H. (1997), "Specification tests in panel data models using artificial regressions", paper presented at the Seventh International Conference on Panel Data held at La Sorbonne, Paris, 19-20 June 1997.

Bates C. and White H. (1985), "Determination of estimators with minimum asymptotic covariance matrices", *Econometric Theory*, 9, 633-648.

Bates C. and White H. (1993), "A unified theory of consistent estimation for parametric models", *Econometric Theory*, 1, 151-178.

Bera A.K. and Jarque C.M. (1982), "Model specification tests. A simultaneous approach", *Journal of Econometrics*, 20, 59-82.

Bera A.K. and McKenzie C.R. (1986), "Alternative forms and properties of the score test", *Journal of Applied Statistics*, 13, 13-25.

Bierens H.J. (1994), *Topics in Advanced Econometrics: Estimation, Testing and Specification of Cross-section and Time series Models*, Cambridge University Press, Cambridge.

Biørn E. (1981), "Estimating economic relations from incomplete cross-section/time-series data", *Journal of Econometrics*, 16, 221-236.

Blanchard P. and Mátyás L. (1994), "Robustness of tests for error components models to non-normality", paper presented at the Fifth Conference on Panel Data held at Université Paris XII - Val de Marne, Paris, 16-17 June 1994.

Blanchard P., Bresson G., Cayatte J-L., Pirotte A., Sevestre P. and Teurlai J.C. (1996), "Coûts d'ajustement de l'emploi, structure de marché et gestion de la main d'oeuvre: analyses théoriques et estimations économétriques sur données d'entreprises", Rapport au Commissariat Général du Plan, Université Paris - Val de Marne, ERUDITE, Paris.

Bollerslev T. and Wooldridge J. (1992), "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances", *Econometric Reviews*, 11 (2), 143-172.

Bollerslev T., Engle R. and Nelson B. (1994), "Arch models", Ch. 49 in *Handbook of Econometrics*, Vol. 4, eds. Engle R. and McFadden D., North Holland, Amsterdam.

Breush T.S. and Pagan A.R. (1979), "A simple test for heteroscedasticity and random coefficient variation", *Econometrica*, 47 (5), 1287-1297.

Breush T.S. and Pagan A.R. (1980), "The Lagrange multiplier test and its applications to model specification in econometrics", *Review of Economic Studies*, 47, 239-253.

Breush T.S. (1987), "Maximum likelihood estimation of random effects models", *Journal of Econometrics*, 36, 383-389.

Breush T.S., Mizon G.E. and Schmidt P. (1989), "Efficient estimation using panel data", *Econometrica*, 57 (3), 695-700.

Broze L. and Gourieroux C. (1993), "Pseudo-maximum likelihood method, ajusted pseudo-maximum likelihood method and covariance estimators", CORE Discussion Paper 9313, Université Catholique de Louvain, forthcoming in *Journal of Econometrics*.

Burguete J., Gallant R. and Souza G. (1982), "On unification of the asymptotic theory of nonlinear econometric models", *Econometric Reviews*, 1, 151-190.

Chamberlain G. (1984), "Panel data", Ch. 22 in *Handbook of Econometrics*, Vol. 2, eds. Griliches Z. and Intriligator M.D., North Holland, Amsterdam.

Chamberlain G. (1987), "Asymptotic efficiency in estimation with conditional moment restrictions", *Journal of Econometrics*, 34, 305-334.

Chow G.C., (1960), "Testing for equality between sets of coefficients in two linear regressions", *Econometrica*, 28, 591-605.

Christensen L.R., Jorgenson D.W. and Lau L.J. (1973), "Transcendental logarithmic production frontiers", *Review of Economics and Statistics*, 25, 28-45.

Cobb C.W. and Douglas P.H. (1928), "A theory of production", *American Economic Review*, 18 (Supplement), 139-165.

Cox D.R., (1961), "Tests of separate families of hypotheses", in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley, 105-123.

Cox D.R., (1962), "Further results on tests of separate families of hypotheses", *Journal of the Royal Statistical Society*, series B, 406-424.

Cragg J. (1983), "More efficient estimation in the presence of heteroscedasticity of unknown form", *Econometrica*, 51 (3), 751-763.

Dastoor K.N. (1981), "A note on the interpretation of the Cox procedure for nonnested hypotheses", *Economics Letters*, 8, 113-119.

Douglas P.H. (1948), "Are there laws of production ?", *American Economic Review*, 38, 1-41.

Davidson R. and MacKinnon J.G. (1981), "Several tests for model specification in the presence of alternative hypotheses", *Econometrica*, 49 (3), 781-793.

Davidson R. and MacKinnon J.G. (1984), "Model specification tests based on artificial linear regressions", *International Economic Review*, 25 (2), 485-502.

Davidson R. and MacKinnon J.G. (1985a), "The interpretation of test statistics", *Canadian Journal of Economics*, 18, 38-57.

Davidson R. and MacKinnon J.G. (1985b), "Heteroskedasticity-robust tests in regression directions", *Annales de l'INSEE*, 59/60, 183-218.

Davidson R. and MacKinnon J.G. (1993), *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

Engle R.F. (1982), "A general approach to Lagrange multiplier model diagnostics", *Journal of Econometrics*, 20, 83-104.

Engle R.F. (1984), "Wald, likelihood ratio and Lagrange multiplier tests in econometrics", Ch. 13 in *Handbook of Econometrics*, Vol. 2, eds. Griliches Z. and Intriligator M.D., North Holland, Amsterdam.

Fomby T.B., Carter H.R. and Johnson S.R. (1984), *Advanced Econometric Methods*, Springer-Verlag, New York.

Graag J.G. (1993), "A bilinear model for heteroscedastic panel data", *Annales d'Économie et de Statistique*, 30, 121-141.

Gallant R.G. (1987), *Nonlinear Statistical Models*, John Wiley & Sons, New-York.

Greenwald B.C. (1983), "A general analysis of bias in the estimated standard errors of least squares coefficients", *Journal of Econometrics*, 22, 323-338.

Godfrey L.G. (1988), *Misspecification Tests in Econometrics*, Cambridge University Press, Cambridge.

Goldberger A. (1968), "The interpretation and estimation of Cobb-Douglas functions", *Econometrica*, 35 (3), 464-472.

Gourieroux C.and Monfort A. (1983), "Testing nested or non-nested hypotheses", *Journal of Econometrics*, 21, 83-115.

Gourieroux C., Monfort A. and Trognon A. (1984a), "Pseudo-maximum likelihood methods : theory", *Econometrica*, 52 (3), 681-700.

Gourieroux C., Monfort A. and Trognon A. (1984b), "Pseudo-maximum likelihood methods : applications to poisson models", *Econometrica*, 52 (3), 701-720.

Gourieroux C. and Monfort A. (1989), *Statistique et Modèles Econométriques*, Economica, Paris. Translated in 1995, *Statistics and Econometric Models*, Cambridge University Press, Cambridge.

Gourieroux C. (1992), *Modèles ARCH et applications financières*, Economica, Paris.

Gourieroux C. and Monfort A. (1993), "Pseudo-likelihood methods", Ch. 12 in *Handbook of Statistics*, Vol. 11, eds. Maddala G.S., Rao C.R. and Vinod H.D., North Holland, Amsterdam.

Griliches Z. and Mairesse J. (1990), "Heterogeneity in panel data : are there stable production functions ?", in *Essays in Honour of Edmond Malivaud*, eds. Champsaur P. and al., 3, Empirical Economics, MIT Press, Cambridge.

Griliches Z. and Mairesse J. (1995), "Production functions : the search for identification", NBER Working Paper n° 5067, National Bureau of Economic Research, Cambridge.

Harvey A.C. (1976), "Estimating regression models with multiplicative heteroscedasticity", *Econometrica*, 44 (3), 461-465.

Harvey A.C. (1981), *The Econometric Analysis of Time Series*, Philip Allan, Oxford.

Hausman J. (1978), "Specification tests in econometrics", *Econometrica*, 46 (6), 1251-1272.

Hausman J. and Taylor W.E. (1981), "Panel data and unobservable individual effects", *Econometrica*, 49 (6), 1377-1398.

Heijmans R.D.H. and Magnus J.R. (1986a), "Consistent maximum likelihood estimation with dependent observations. The general (non-normal) case and the normal case", *Journal of Econometrics*, 32, 253-285. Corrigenda, *Journal of Econometrics*, 35, 395.

Heijmans R.D.H. and Magnus J.R. (1986b), "Asymptotic normality of maximum likelihood estimators obtained from normally distributed but dependent observations", *Econometric Theory*, 2, 374-397.

Hildreth C. and Houck J.P. (1968), "Some estimators for a linear model with random coefficients", *Journal of the American Statistical Association*, 63, 584-595.

Hoch I. and Mundlak Y. (1965), "Consequences of alternatives specifications in estimation of Cobb-Douglas production functions", *Econometrica*, 33 (4), 814-828.

Holly A. (1987), "Specification tests : an overview", Ch. 2 in *Advances in Econo-*

metrics - Fifth World Congress, Vol. 1, eds. Bewley T., Cambridge University Press, New-York.

Honda Y. (1985), "Testing the error components model with non-normal disturbances", *Review of Economic Studies*, 52, 681-690.

Honda Y. (1988), "A size correction to the Lagrange Multiplier test for heteroscedasticity", *Journal of Econometrics*, 38, 375-386.

Hsiao C. (1986), *Analysis of Panel data*, Cambridge University Press, Cambridge.

Jarque C.M. and Bera A.K. (1980), "Efficient tests for normality, homoscedasticity and serial independence of regression residuals", *Economics Letters*, 6, 255-259.

Jorgenson D.W. (1986), "Econometric methods for modelling producer behavior", Ch. 31 in *Handbook of Econometrics*, Vol. 3, eds. Griliches Z. and Intriligator M.D., North Holland, Amsterdam.

Koenker R. (1981), "A note on studentizing a test for heteroscedasticity", *Journal of Econometrics*, 17, 107-112.

Lejeune B. (1996a), "A full heteroscedastic one-way error components model for incomplete panel : Maximum likelihood estimation and Lagrange Multiplier testing", CORE Discussion Paper n° 9606, Université Catholique de Louvain, Louvain-la-Neuve.

Lejeune B. (1996b), "A full heteroscedastic one-way error components model for incomplete panel : Pseudo-maximum likelihood estimation and specification testing", paper presented at the Econometric Society European Meeting, 25-29 August 1996, Istanbul.

Lejeune B. (1997), "Second order pseudo-maximum likelihood estimation and conditional variance misspecification", CORE Discussion Paper n° 9779, Université Catholique de Louvain, Louvain-la-Neuve.

Li Q. and Stengos T. (1994), "Adaptive estimation on the panel data error component model with heteroscedasticity of unknown form", *International Economic Review*, 35, 981-1000.

MacKinnon J.G (1983a), "Model specification tests against non-nested alternatives", *Econometric Reviews*, 2, 85-110.

MacKinnon J.G., White H. and Davidson R. (1983b), "Tests for model specification in the presence of alternative hypotheses : Some further results", *Journal of Econometrics*, 21, 53-70.

MacKinnon J.G. and White H. (1985), "Some heteroscedasticity consistent covariance matrix estimators with improved finite sample properties", *Journal of Econometrics*, 29, 305-325.

MacKinnon J.G. and Magee L. (1990), "Transforming the dependent variable in regression models", *International Economic Review*, 31 (2), 315-339.

MacKinnon J.G. (1992), "Model specification tests and artificial regressions", *Journal of Economic Literature*, XXX, 102-146.

Maddala G.S. (1971), "The use of variance components models in pooling cross section and time series data", *Econometrica*, 39 (2), 341-358.

Magnus J.R. (1978), "Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix", *Journal of Econometrics*, 7, 281-312. Corrigenda, *Journal of Econometrics*, 10, 261.

Magnus J.R. (1982), "Multivariate error components analysis of linear and non-linear regression models by maximum likelihood", *Journal of Econometrics*, 19, 239-285.

Magnus J.R. and Neudecker H. (1986), "Symmetry, 0-1 matrices and jacobians: a review", *Econometric Theory*, 2, 157-190.

Magnus J.R. and Neudecker H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons, New York.

Mairesse J. (1988), "Les lois de la production ne sont plus ce qu'elles étaient : une introduction à l'économétrie des panels", *Revue Economique*, 39, 225-271.

Marschak J. and Andrews W.H. (1944), "Random simultaneous equations and the theory of production", *Econometrica*, 12 (3), 143-205.

Mátyás L. and Lovrics L. (1991), "Missing observations and panel data : a Monte-Carlo analysis", *Economics Letters*, 37, 39-44.

Mátyás L. and Sevestre P. (1996), *The Econometrics of Panel Data*, Kluwer Academic Publishers, Dordrecht.

Mazodier P. and Trognon A. (1978), "Heteroscedasticity and stratification in error components models", *Annales de l'INSEE*, 30-31, 451-482.

Mazodier P. and Trognon A. (1993), *Econométrie II*, Collection ENSAE, Malakoff.

Monfort A. (1995), "A reappraisal of misspecified econometric models", Working Paper 9545, CREST-INSEE, Paris.

Moulton B.R. (1986), "Random group effects and the precision of regression estimates", *Journal of Econometrics*, 32, 385-397.

Moulton B.R. and Randolf W.C. (1989), "Alternative tests of the error components model", *Econometrica*, 57 (3), 685-693.

Mundlak Y. (1961), "Empirical production function free of management bias", *Journal of Farm Economics*, 43 (1), 44-56.

Mundlak Y. (1978), "On the pooling of time series and cross section data", *Econometrica*, 46 (1), 69-85.

Muus L. and Wansbeek T. (1994), "Diagnostic tests for heteroscedasticity in panel data models", paper presented at the Fifth Conference on Panel Data held at E.R.U.D.I.T.E. (Paris), 16-17 June 1994.

Newey W. (1985), "Maximum likelihood specification testing and conditional moment tests", *Econometrica*, 53 (5), 1047-1070.

Newey W. (1990), "Semiparametric efficiency bounds", *Journal of Applied Econometrics*, 5, 99-135.

Newey W. (1993), "Efficient estimation of models with conditional moment restrictions", Ch. 16 in *Handbook of Statistics*, Vol. 11, eds. Maddala G.S., Rao C.R. and Vinod H.D., North Holland, Amsterdam.

Newey W. and McFadden D. (1994), "Large sample estimation and hypothesis testing", Ch. 36 in *Handbook of Econometrics*, Vol. 4, eds. Engle R. and McFadden D., North Holland, Amsterdam.

Newey W. and Steigerwald D. (1997), "Asymptotic bias for quasi-maximum likelihood estimators in conditional heteroscedasticity models", *Econometrica*, 63 (3), 587-599.

Pagan A. and Sabau H. (1991), "On the inconsistency of the MLE in certain heteroscedastic regression models", *Estudios Económicos*, 6 (2), 159-172.

Pötscher B.M. and Prucha I.R. (1997), *Dynamic Nonlinear Econometric Models*, Springer-Verlag, Berlin.

Quandt R.E. (1983), "Computational problems and methods", Ch. 12 in *Handbook of Econometrics*, Vol. 1, eds. Griliches Z. and Intriligator M.D., North Holland, Amsterdam.

Ramsey J.B. (1969), "Tests for specification errors in classical linear least-squares regression analysis", *Journal of the Royal Statistical Society*, series B, 2, 350-371.

Ramsey J.B. and Schmidt P. (1976), "Some further results on the use of OLS and BLUS residuals in specification error tests", *Journal of the American Statistical Association*, 71, 389-390.

Randolf W.C. (1988a), "A transformation for heteroscedastic error components regression models", *Economics Letters*, 27, 349-354.

Randolf W.C. (1988b), "Housing depreciation and aging bias in the consumer price index", *Journal of Business and Economic Statistics*, 6 (3), 359-371.

Rao S.R.S., Kaplan J. and Cochran W.G. (1981), "Estimators for the one-way random effects model with unequal error variances", *Journal of the American Statistical Association*, 76, 89-97.

Ruud P.A. (1984), "Tests of specification in econometrics", *Econometric Reviews*, 3, 211-242.

Savin N.E. (1980), "The Bonferroni and the Scheffé multiple comparison procedures", *Review of Economic Studies*, 47, 255-273.

Savin N.E. (1984), "Multiple hypothesis testing", Ch. 14 in *Handbook of Econometrics*, Vol. 2, eds. Griliches Z. and Intriligator M.D., North Holland, Amsterdam.

Swamy P. (1970), "Efficient inference in a random coefficient regression model", *Econometrica*, 38 (2), 311-322.

Tauchen G. (1985), "Diagnostic testing and evaluation of maximum likelihood models", *Journal of Econometrics*, 30, 415-443.

Theil H. (1971), *Principles of econometrics*, John Wiley and Sons, Amsterdam.

Trognon A. (1987), "Les méthodes du pseudo-maximum de vraisemblance", *Annales d'Économie et de Statistique*, 8, 117-134.

Verbeek M. and Nijmans T. (1996), "Incomplete panels and selection bias", Ch. 18 in *The Econometrics of Panel Data*, eds. Mátyás L. And Sevestre P., Kluwer Academic Publishers, Dordrecht.

Verbon H.A.A. (1980), "Testing for heteroscedasticity in a model of seemingly unrelated regression equations with variance components", *Economics Letters*, 5, 149-153.

Wansbeek T. and Kapteyn A. (1989), "Estimation of the error components model with incomplete panels", *Journal of Econometrics*, 41, 341-361.

Wansbeek T. (1989), "An alternative heteroscedastic error components model", Solution 88.1.1, *Econometric Theory*, 5, 326.

White H. (1980a), "Using least squares to approximate unknown regression functions", *International Economic Review*, 21 (1), 149-170.

White H. (1980b), “Nonlinear regression on cross-section data”, *Econometrica*, 48 (3), 721-746.

White H. (1980c), “A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity”, *Econometrica*, 48 (4), 817-839.

White H. (1981), “Consequences and detection of misspecified nonlinear regression models”, *Journal of the American Statistical Association*, 76, 419-433.

White H. (1982), “Maximum likelihood estimation of misspecified models”, *Econometrica*, 50 (1), 1-25.

White H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, San Diego.

White H. (1987), “Specification testing in dynamic models”, Ch. 1 in *Advances in Econometrics - Fifth World Congress*, Vol. 1, eds. Bewley T., Cambridge University Press, New-York.

White H. (1990), “A consistent model selection procedure based on m-testing”, Ch. 16 in *Modelling Economic Series: Reading in Econometrics Methodology*, eds. Granger C.W.J., Oxford University Press, Oxford.

White H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press, Cambridge.

Wooldridge J. (1990), “A unified approach to robust, regression-based specification tests”, *Econometric Theory*, 6, 17-43.

Wooldridge J. (1991a), “On the application of robust, regression-based diagnostics to models of conditional means and conditional variances”, *Journal of Econometrics*, 47, 5-46.

Wooldridge J. (1991b), “Specification testing and quasi-maximum likelihood estimation”, *Journal of Econometrics*, 48, 29-55.

Wooldridge J. (1994), “Estimation and inference for dependent processes”, Ch. 45 in *Handbook of Econometrics*, Vol. 4, eds. Engle R. and McFadden D., North Holland, Amsterdam.

Zellner A., Kmenta L. and Drèze J. (1966), “Specification and estimation of Cobb-Douglas production function models”, *Econometrica*, 34 (4), 784-795.