A Monte Carlo Analysis of the VAR-Based Indirect Inference Estimation of DSGE Models[*]

David Dubois[†]

April, 2011

Abstract

In this paper we study estimation of DSGE models. More specifically, in the indirect inference framework, we analyze how critical is the choice of the reduced form model for estimation purposes. As it turns out, simple VAR parameters performs better than commonly used impulse response functions. This can be attributed to the fact that IRF worsen identification issues for models that are already plagued by that phenomenon.

JEL Classification: C53, E13, E17
Keywords: DSGE, Identification, Indirect inference, VAR, Impulse response functions

1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models are now common in central banks and academic institutions. Although the recent crisis has shed some doubts on the relevance of DSGE models, or at least on the credibility of some of their assumptions, studying the way they are estimated is still of interest for future development. It seems that so much effort has been put out to try to make these models as reliable as possible that the quality of their estimation has somehow been neglected. By their structural nature, DSGE are very sensitive to identification issues. However this problem is often neglected in practice, researchers preferring to calibrate the unidentified parameters. This strategy is not without risk and can severely bias estimations as shown by Iskrev (2010) and Canova and Sala (2009). Among the various methods of estimation that can be used to estimate this type of models we propose to study indirect inference estimation procedure developed by Gourieroux, Monfort, and Renault (1993). This estimation procedure which is a general simulated method of moments (SMM) makes use of an auxiliary model and we wonder how critical is the choice of that auxiliary model. A common choice is to use impulse response functions (IRF). Although they do have a nice economic interpretation, it is not necessarily a good criterion to select an auxiliary model, which by only serves for estimation purposes. More relevant is the ability of the auxiliary model to correctly summerize the data. In that perspective, IRF do not seem so attractive because they are non linear transformations of the VAR parameters, thus

adding complexity to an already complex problem. Non linearity can have undesirable consequences in terms of identification and more generally estimation. An alternative to IRF is to merely use the VAR parameters themselves which can be seen as a representation of the second moments of the data. In the next section we briefly discuss two articles that study eestimation issues of DSGE. In section 3 we present a Real Business Cycle model that we linearize and solve. Section 4 is devoted to a presentation of indirect inference estimation procedure. The Monte Carlo experiment is presented in section 5. Finally, section 6 concludes the paper.

2   Literature Review

Iskrev (2010) studies identification for a state-of-the-art DSGE model in the line of the model developped by Smets and Wouters (2007). While this model has been recognized as a great achievement in DSGE literature, results on identification are a bit concerning. Often, identification issues are not treated explicitely. Whenever a parameter is unidentified, researchers usually merely fix its value. Its value is chosen among previous studies thought to be relevant. We can already see the problem of such an approach, by doing that researchers dangerously lean toward calibration. As emphasized by Iskrev, this issue is amplified by the used of Bayesian techniques. Indeed these allow one to hide (although not necessarily consciously) unidentifiability of a parameter by specifying a prior distribution with very small variance. But everything comes at a cost, and although it makes the identification issue disappear, one can wonder whether the results obtained come more from the priors imposed than from the information actually contained in the data. The analysis leads to the conclusion that the model is poorly identified, essentially structurally (i.e. from the model itself, not from the data). Not surprisingly the lack of identifiability, leads to overweight the prior distribution in a Baysian estimation.

Canova and Sala (2009) study identification in DSGE models as well, however their approach is relatively different. They have a more practical approach and emphasize the consequences of identification issues on the behaviour of objective function and thus, on the importance of the choice of this function. Different problems can affect the objective function. If the objective function does not have a unique minimum, it means that different values for the structural parameters have the same reduced form (observational equivalence). Too small curvature of the objective function leads to obvious problems of finding the minimum (weak identification). Some parameters may also disappear from the objective function (under-identification). They share the same doubts than Iskrev about calibrated unidentified parameters and show that wrongly calibrated parameters even make other parameters estimates biased. Also, using a similar model than Iskrev they find it poorly identified too. Along with practical advices for researchers they point out the importance of the choice of the objective function. The less informative the objective function the more likely identification issues. IRF might contain less information than VAR thus exarcerbating identification issues. We are quite in line with Canova and Sala but our goal is different in the sense that we are comparing two auxiliary models and we are trying to assess which objective function is the best for estimation.

3 Model

For our study, we use a very simple real business cycles (RBC) model. It is based on the stochastic Ramsey model. The production function is a one-factor Cobb-Douglas. As there is no labour, the representative household does not receive any wage, but it receives dividends for the capital it owns (bought with the savings) so that the household is both consumer and producer. It maximizes its intertemporal utility which is a function of the consumption. So the household has to choose between consumption now or later which influences the savings (at each period, the agent must allocate is revenue between either consumption or investment), themselves influencing possible future consumption (via the capital revenue). Its horizon is infinite so we assume the transversality condition holds. The utility function of the household $u(c)$ is a function of consumption. It is an increasing concave function. As a consequence, the agent will tend to have a relatively smooth consumption path rather than very high consumption at some periods and very low at others. The only stochastic shock in the economy, z, is a total factor productivity (TFP) shock. The shock is generated by an autoregressive process wihtout intercept for which the shocks are drawn from a normal distribution of mean zero. As z can be negative it enters the production function as argument of an exponential function to avoid negative production values. Because of this shock the household does not know what will be its future consumption so we need to think in term of expected utility.
The representative agent maximizes expected lifetime utility defined by:

$$\max_{\{c_t\}_{t=1}^{\infty}} E_t \quad \sum_{s=t}^{\infty} \quad \beta^{s-t} u(c_t) \tag{1}$$

$$s.t.$$

$$u(c_t) \quad = \quad \frac{c_t^{1-\theta}}{1-\theta} \tag{2}$$

$$k_{t+1} \quad = \quad f(k_t, z_t) - c_t + (1-\delta)k_t, \quad \forall t \tag{3}$$

$$z_t \quad = \quad \rho z_{t-1} + \varepsilon_t, \quad \forall t \tag{4}$$

$$f(k_t, z_t) \quad = \quad A \exp(z_t) k_t^{\alpha}, \quad \forall t \tag{5}$$

where $\rho \in (-1, 1)$ and ε_t is iid $N(0, \sigma^2)$
Finally the good market clearing requires

$$y_t = c_t + i_t$$

The utility function is a constant relative risk aversion (CRRA) function. Unlike the original model, we do not consider the agent to be totally indifferent between present and future consumption, this is traduced by the factor $\beta \in [0; 1)$. The larger β the lower discounter rate. One extreme case, when $\beta = 1$ means we do not discount future consumption and thus consider it as useful as today's consumption. In the other extreme case when $\beta = 0$ we totally discount future consumption, so that we consume everything today. For any other β (different than the two extremes), the farther the period the higher the discount rate. The law of motion of capital, equation (3), merely states that the stock of capital is equal to the depreciated stock of capital of the previous period plus the investment of the previous period (the production minus consumption). The production function, equation (5), is a one-factor Cobb-Douglas that depends on the stock of capital and on the total factor productivity level A. The lower the

risk aversion the more the agent will be willing to substitute present consumption for future consumption (stated differently, $1/\theta$ represents the propension of the household to deviate from a smooth consumption path).

3.1 Non-linear system and Steady State

We need to solve the intertemporal maximisation. This can be done using for instance dynamic programming, the derivation is given in appendix.
We end up whith the following three-equation non-linear system:

$$
\begin{aligned}
c_t^{-\theta} &= \beta E_t[Aexp(z_{t+1})\alpha k_{t+1}^{\alpha-1} + (1-\delta)c_{t+1}^{-\theta}] \\
k_{t+1} &= A\exp(z_t)k_t^{\alpha} - c_t + (1-\delta)k_t \\
z_t &= \rho z_{t-1} + \varepsilon_t
\end{aligned}
$$

From this system we can compute the steady state value for each variable (that we designate by the subscript ss). We decide to fix the steady state value of capital (k_{ss}) at 10 and we let the TFP factor A be determined by k_{ss}.

$$
\begin{aligned}
k_{ss} &= 10 \\
A &= \frac{\delta}{\alpha\beta k_{ss}^{\alpha-1}} \\
c_{ss} &= Ak_{ss}^{\alpha} - \delta k_{ss} \\
y_{ss} &= Ak_{ss}^{\alpha} \\
i_{ss} &= \delta k_{ss} \\
z_{ss} &= 0;
\end{aligned}
$$

So at the steady state, the value of the exogenous variable is zero. The investment is equal to the depreciation so that the stock of capital remains constant. Of course it makes the production and the consumption constant as well.

3.2 Linearization

A closed form analytical solution of a non-linear rational expectations model is typically unavailable. So the next step is to linearize the model. A common and easy way to do so is to use the first-order Taylor series expansion. As we are dealing with a multivariate system of difference equations, we need to compute the Jacobian of the system. Also, we are intrested in studying the model and its variations around the steady state, so we evaluate the Jacobian at the steady state of the model.
Mathematically we can express the system in a compact notation:

$$
\Psi(x_{t+1}, x_t) = 0
$$

where x and 0 are vectors containing the variables of the system.
If $\bar{x}$ denotes the steady state value, the first order Taylor Series approximation around the steady state is

$$
0 \approx \frac{\partial \Psi}{\partial x_t}(\bar{x})(x_t - \bar{x}) + \frac{\partial \Psi}{\partial x_{t+1}}(\bar{x})(x_{t+1} - \bar{x})
$$

Let's denote A and $-B$ the part of the Jacobian for which we take the derivatives with respect to x_{t+1} and x_t respectively. The variable in deviation from the steady state is denoted $\tilde{x}$. Moreover we know that $\Psi(\bar{x}) = 0$ so eventually,

$$A\tilde{x}_{t+1} = B\tilde{x}_t \tag{6}$$

Note that during the linearization step we completely disregard the expectation operator in the Euler equation. Next, assuming A is invertible, we can premultiply B by A^{-1} to obtain:[1]

$$\tilde{x}_{t+1} = \widetilde{A}\tilde{x}_t$$

However the model is still not empirically amenable. A last step need to be performed. Indeed the Euler's equation still contains an expectation operator, so we have to solve for it.

3.3   Solution of the linearized model

We want to end up with a system of the following form:

$$x_{t+1} = Fx_t + Gv_t$$

where x is the vector of endogenous variables and v a vector of structural shocks. Different methods has been proposed to solve linear rational expectations models, see for instance Blanchard and Kahn (1980), Klein (2000) or Sims (2002). The idea behind these methodologies is always the same: decoupling the system into the predetermined part and the non-predetermined part and then solve them separetely. They differ essentially in the way the linearized system must be written, the type of matrix decomposition applied and level of generality possible.[2] However, for our purpose, the choice of the solution methods is not critical and we decided to go with the widely used method developed by Blanchard and Kahn.

Three conditions are required to apply the Blanchard-Kahn method.

First the model must be written in the following form:

$$\begin{bmatrix} x_{1t+1} \\ E_t(x_{2t+1}) \end{bmatrix} = \widetilde{A} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} + Ez_t \tag{7}$$

where x_1 is a vector of predetermined variables (variables for which $E_t(x_{1,t+1}) = x_{1,t+1}$), x_2 a vector of non-predetermined variables and z a vector of exogenous variables.

Second, we assume rational expectations:

$$E_t(x_{2t+1}) = E(x_{2t+1}|\Omega_t)$$

where $E_t(x_{2t+1})$ is the mathematical expected value of x_{2t+1} at time t and Ω_t the information set at time t containing at least past and current values of x_1, x_2 and z. This condition ensures that once we know the value of exogenous variables, we necessarily know the value of the endogenous variables.

---

[1]The $\tilde{A}$ matrix we found for our model is given in appendix.

[2]For instance the methodology proposed by Klein allows to overcome singularity issue that might arise with some of the partionned matrices.

Third, we impose a technical condition on z to prevent it from exploding. It can be defined relatively broadly (Blanchard and Kahn 1980) but in our case it is merely ensured by a stable autoregressive process, i.e. with $\rho < 1$.

The first step is to partition the matrix $\widetilde{A}$ using the Jordan decomposition. The Jordan method decomposes a matrix into a diagonal matrix of the eigen values of the original matrix and a matrix with its associated eigen vectors.

Formally,

$$\widetilde{A} = \Lambda^{-1} J \, \Lambda$$

where J is the diagonal matrix containing the eigen values of $\widetilde{A}$ and Λ a matrix containing the associated eigen vectors.

We sort the eigen values by increasing absolute value and the associated vectors are sorted similarly. The system is partioned and ordered decoupling stable eigen values (within or on the unit circle in absolute value) and explosive ones (outside the unit circle in absolute value).

$$J = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}$$

where J_1 contains the stable eigen values and J_2 the explosive eigen values. The matrices Λ and E are partioned accordingly,

$$\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}, \quad E = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}$$

We can rewrite (7) as follows:

$$\begin{bmatrix} x_{1t+1} \\ E_t(x_{2t+1}) \end{bmatrix} = \Lambda^{-1} J \, \Lambda \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} + \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} f_t$$

Next we premultiply the system by Λ:

$$\begin{bmatrix} \acute{x}_{1t+1} \\ E_t(\acute{x}_{2t+1}) \end{bmatrix} = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix} \begin{bmatrix} \acute{x}_{1t} \\ \acute{x}_{2t} \end{bmatrix} + \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} f_t \tag{8}$$

where

$$\begin{bmatrix} \acute{x}_{1t+1} \\ \acute{x}_{2t+1} \end{bmatrix} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \begin{bmatrix} x_{1t+1} \\ x_{2t+1} \end{bmatrix}$$

$$\begin{bmatrix} D_1 \\ D_2 \end{bmatrix} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}$$

We see that the system (8) is actually decoupled with non-predetermined variables depending on explosive eigen values and the predetermined variables depending on stable eigen values. Using the rational expectations, it can be shown that the solution of lower portion of (8) is the following:

$$x_{2t} = -\Lambda_{22}^{-1} \Lambda_{21} x_{1t} - \Lambda_{22}^{-1} J_2^{-1} (I - \rho J_2^{-1} D_2)^{-1} z_t \tag{9}$$

The solution for x_1 is now straightforward, we expand the upper portion of (8):

$$x_{1t+1} = \widetilde{A}_{11} x_{1t} + \widetilde{A}_{22} x_{2t} + E_1 f_t$$

where $\widetilde{A}_{11}$ and $\widetilde{A}_{22}$ are partitions of $\Lambda^{-1} J \Lambda$ conformable with x_{1t} and x_{2t} respectively. Then we can substitute for x_{2t} with the expression found in (9):

$$x_{1t+1} = \widetilde{A}_{11} x_{1t} + \widetilde{A}_{22}(-\Lambda_{22}^{-1}\Lambda_{21} x_{1t} - \Lambda_{22}^{-1} J_2^{-1} (I - \rho J_2^{-1} D_2)^{-1} z_t) + E_1 f_t$$

So far, implicitely we have made the assumption that the number of explosive eigen values ($\bar{m}$) is equal to the number of non-predetermined variables (m). This assumption is actually a necessary condition to find a unique solution to the system. It is called the Blanchard-Kahn condition, named after (Blanchard and Kahn 1980). Two other situations may arise.
First if $\bar{m} > m$, no solution that respects the other conditions exist. Second if $\bar{m} < m$, there is an infinity of solutions. Economically speaking, the Blanchard-Kahn condition requires that the model respects the saddle point property.

4 Indirect Inference

Indirect inference has been developed by Gourieroux, Monfort, and Renault (1993) and can be viewed as a generalization of the simulated method of moments (SMM). Indeed, Gourrieroux et al. showed that SMM is a particular case of the indirect inference procedure. They extend simulated methods to a very general framework. Instead of talking about moments, they rather use an auxiliary (reduced form) model. So now we are dealing with two kinds of models, an auxiliary model and a structural model that is supposed to be the data generating process. What we do is to fit the auxiliary model to the observed variables, simulate the structural model (with some initial value for the structural parameters), fit the auxiliary model to the simulated variables and compare the two sets of auxiliary parameters. Then the value of the structural parameters is updated in order to reduce the distance between the two sets of auxiliary parameters. The whole procedure is repeated until the distance between the two sets of parameters is as small as possible.

4.1 Structural Model

The structural model contains three kinds of variables: endogenous, strongly exogenous and exogenous forcing variable. The model we propose to study does not have any strongly exogenous variables so we won't insist on that aspect, however the crucial assumption regarding that variable is that it is an homogeneous Markov Process. We also assume to know the distribution G of ε. However this is not a necessary condition, if we consider ε to be a white noise process of a known distribution, we can estimate its parametrization by including it in θ.[3] The important thing here is that the model can be simulated with some initial values for θ by drawing simulated values of ε.

4.2 Auxiliary Parameter

We need to define how can be obtained the auxiliary parameters β. It is broadly defined as the argument that maximizes a criterion function depending on the

---

[3]Note that in our empirical study, we consider it given.

observed data:

$$\max_{\beta \in B} Q_T(y_T, \beta) \tag{10}$$

so

$$\hat{\beta}_T = \arg\max_{\beta \in B} Q_T(y_T, \beta)$$

In the limit, when the sample size goes to infinity, the criterion function is assumed to be non-stochastic and to have a unique maximum at β_0. Note that y_T is a function of θ and G (which is known)[4], so β_0 is the reduced form parametrization corresponding to the true structural parameters values θ_0. Of course, as θ_0 is unknown, so is β_0.

When $T \to \infty$ the criterion function can be express as:

$$Q_\infty = (\theta_0, \beta)$$

and we have

$$\beta_0 = \arg\max_{\beta \in B} Q_\infty(\theta_0, \beta)$$

From these concepts, we can derive the "binding" function which is defined as

$$b(\theta) = \arg\max_{\beta \in B} Q_\infty(\theta, \beta)$$

so it is the function that maps the structural parameters to the auxiliary parameters. The binding function is assumed to be one-to-one. A one-to-one function (also called injective function) is such that every element of its codomain is mapped to by at most one element of its domain. Applied to indirect inference, it means that to one auxiliary parameters corresponds at most one structural parameter. This is a crucial assumption for the identification of the structural parameters.

Finally, we also require $\frac{\partial b}{\partial \theta'}(\theta_0)$ to be full-column rank.

4.3 Estimation of the structural parameters

We can apply the estimation procedure of the previous section to simulated paths. We can simulate H paths (each of one denoted h) of lenght T by drawing ε (independently) $H \times T$ times and by setting some initial values for the variables and the parameters. In the previous notation:

$$\tilde{\beta}_T^h(\theta, z_0) = \arg\max_{\beta \in B} Q_t((\tilde{y}^h)_T^1, \beta)$$

As previously we have,

$$\lim_{T \to \infty} \tilde{\beta}_T^h(\theta, z_0) = b(\theta)$$

therefore, $\tilde{\beta}_T^h$ is a consistent estimator of $b(.)$

The indirect estimator will try to make the average of $\tilde{\beta}$ over the H simulations

---

[4]G being known, we disregard it hereafter in order to alleviate the notation.

$(\frac{1}{H} \sum_{h=1}^{H} \tilde{\beta}_T^h(\theta, z_0))$ as close as possible to $\hat{\beta}$ (i.e. the β estimated from the observed data). Overall to estimate θ, we end-up with the following minimization problem:

$$\min_{\theta \in \Theta} [\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^{H} \tilde{\beta}_T^h(\theta, z_0)]' \hat{\Omega}(\hat{\beta}_T - \frac{1}{H} \sum_{h=1}^{H} \tilde{\beta}_T^h(\theta, z_0)) \tag{11}$$

Or using the notation of the previous sections, the sample analogs of the moments conditions are

$$g(X, \theta) = \hat{\beta}(x_{obs}) - \tilde{\beta}(X, \theta)$$

and estimation is given by

$$\hat{\theta} = \min_{\theta} \Gamma(\theta) = g(X, \theta)' \; \Omega \; g(X, \theta)$$

From the objective function (11), we see clearly why it is called an indirect estimator as the estimation of θ is done through the estimation of β.

To summerize, for one minimization (i.e. one estimation of θ) we apply the maximization procedure (to get β) one time to the observed data and H times to the H simulated paths. In practice, algorithms used to reach the minimum proceed by trial and error so that the minimization is computed several times. It can lead to a huge amount of computations.

Gourrieroux et al. (1993) propose a second version of the indirect estimators less computationally intensive. Instead of applying H times the maximisation procedure to a sample of length T, it is asymptotically equivalent to do the maximization one time to a path of length HT.

Then the minimization is

$$\min_{\theta \in \Theta} [\hat{\beta}_T - \tilde{\beta}_{HT}(\theta, z_0)]' \hat{\Omega}(\hat{\beta}_T - \tilde{\beta}_{HT}(\theta, z_0))$$

Like the GMM estimator, indirect inference estimators have nice asymptotic properties. It can be shown that

Lemma 1 $\sqrt{T}(\tilde{\theta}_T^H(\Omega) - \theta_0) \xrightarrow{d} N(0, W(H, \Omega))$

where $W(H, \Omega) = (1 + \frac{1}{H})(\frac{\partial b'}{\partial \theta}(G_0, \theta_0)\Omega \frac{\partial b}{\partial \theta'}(G_0, \theta_0))^{-1} \frac{\partial b'}{\partial \theta}(G_0, \theta_0)\Omega J_0^{-1}$
$(I_0 - K_0)^{-1} J_0^{-1}\Omega \frac{\partial b}{\partial \theta'}(G_0, \theta_0)(\frac{\partial b'}{\partial \theta}(G_0, \theta_0)\Omega \frac{\partial b}{\partial \theta'}(G_0, \theta_0))^{-1}$

W has 2 arguments, H the number of simulations and Ω the weighting matrix.

Regarding Ω, the optimal choice is

$$\Omega^* = J_0(I_0 - K_0)^{-1} J_0$$

and thus $W_H^* = (1 + \frac{1}{H})(\frac{\partial b'}{\partial \theta}(G_0, \theta_0)J_0(I_0 - K_0)J_0^{-1} \frac{\partial b}{\partial \theta'}(G_0, \theta_0))$

Also the greater H the more precise the estimates.

4.4 Reduced Form Model

We now turn to the choice of the auxiliary model. The estimation procedure defined in (10) is indeed very general and allows to choose almost any kind

of specifications. A very convenient tool to summerize data is a VAR model. Indeed its coefficients are merely functions of the second moments, so as long as we consider we don't need higher moments, VAR is a good choice. The idea is to fit a VAR model to both observed and simulated data and compare the sets of parameters. $\hat{\theta}$ is then chosen is such a way that it minimizes the distance between the 2 sets of estimated VAR coefficients. However, in practice, a lot of researchers match the impulse-response functions of the VAR instead of the VAR parameters themselves.

Let us first look at what are exactly these functions. Recall that any weakly stationary VAR model has an infinite MA representation:

$$
\begin{aligned}
y_t &= \Phi(L)y_t + \varepsilon_t \\
y_t &= \frac{1}{1 - \Phi(L)}y_t + \varepsilon_t \\
y_t &= \sum_{i=1}^{\infty} \Phi(L)^t \varepsilon_t \\
y_t &= \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + ...
\end{aligned}
$$

Ψ is a matrix whose rows indicate the equation (the variable shocked) and whose columns indicate the origin of the shock. The (i, j) element of Ψ_s show the effect on the ith variable of a shock of the jth variable after s periods *ceteris paribus*. This interpretation can be expressed as a partial derivative:

$$
\frac{\partial y_{t+s}}{\partial \varepsilon_t'} = \Psi_s
$$

or taking each equation separetely,

$$
\frac{\partial y_{i,t+s}}{\partial \varepsilon_{jt}}
$$

We can compute the value of the (i, j) element for different s, starting at the period after the shock and for the subsequent periods. If we express these values as a function of time (s) we obtain the impulse response function of the variable i to a shock of the variable j. Typically, for stationary variables, after a bump at the moment of the shock the IRF decreases and tends to zero (more or less rapidly depending on the variables under consideration). So IRF have a nice economic interpretation, they allow one to see the effect that a shock in an economic variable has on other economic variables. Economically speaking it is not meaningless to seek a specification that generates the same impulse response functions for observed data and simulated data. However from an econometric point of view it is not necessarily the case, it might be that two different structural models generate the same IRF. Of course this is true for VAR parameters as well, but by their nature impulse response functions might be worse in that respect. To see that more clearly we derive the IRF of a simple 2-variable VAR(1). IRF can be computed by simulating the VAR model: we set all the variables to zero then we apply a unit shock to one variable and simulate the system for the subsequent periods. We end up with the IRF of all the variables of the system with respect to the variable we choosed to shock.

The model is as follows:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$$

Initially both variables are set to 0 but in $t = 0$ we apply a unit shock to y_1 thus we have:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

In $t = 1$: $\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix}$

In $t = 2$: $\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11}^2 + a_{12}a_{21} \\ a_{21}a_{11} + a_{22}a_{21} \end{bmatrix}$

In $t = 3$: $\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11}^3 + a_{12}a_{21}a_{11} + \\ a_{12}(a_{21}a_{11} + a_{22}a_{21}) \\ a_{11}^2 a_{21} + a_{12}a_{21}^2 + a_{22}a_{21}a_{11} + a_{22}^2 a_{21} \end{bmatrix}$

and so on...

So basically, impulse-response functions are just non-linear transformations of the VAR parameters. Then one might think that impulse-response functions contain the same information than the VAR parameters and that matching one or the other is equivalent. As a matter of fact it is not the case and the former procedure might contain less information than the latter.
If the 2 approaches give similar results in some cases, on the contrary in others problems may arise.
The first problem is that when matching two non-linear transformations, algorithms used for the computation may fail (due to the complexity added by non-linearity) and thus not giving any result. The other issues that may arise are essentially about identification. In the matching of impulse response functions exercise, we have to focus on a certain range (a certain period of time), it can happen that because of that limitation different values of $\hat{\theta}$ matches the impulse-response function of the observed data. Another identification issue is that some parameters might not appear in the impulse response function while they are present in the VAR. Because identification is an important issue we analyse it in more detail in the next section.

5   Monte Carlo

The first step is to generate observed data. To avoid wasting computing time, we input directly the linearized form of the model (which is merely function of the parameter values).[5] The model as we specified it allow us to compute the series for capital, consumption and TFP shock. For the TFP shock we draw values for ε in a N(0,0.001) distribution. From these series we compute the production y, the investment i and the interest rate r. Production is computed using the function we specified, investment is merely the production minus the consumption and the interest rate is the return on capital so it is the deriative

---

[5]The matrix $\widetilde{A}$ is given in appendix.

of the production with respect to capital.

$$y_t = Aexp(z_t)k_t^{\alpha}$$

$$r_t = Aexp(z_t)\alpha k_t^{\alpha-1}$$

In practice the TFP shock is unobserved and the stock of capital is hard to measure so we do not use these two series for the estimation.

The true parameter values we use are fairly standard. The share of capital in production $\alpha = 0.35$, the discount factor $\beta = 0.97$ (very close to one, so agents do not discount too much future consumption), the relative risk aversion $\theta = 2$ and the autoregressive coefficient (of the TFP shock) $\rho = 0.85$ so that the shock is relatively persistent. For the simulated data, we burn the first 100 observations out of 200. So we end up with 100 observations for each variable, it can be seen as 25 years of data if you consider quarterly data (wich seems realistic given the variation of the data during one period).

We use 2 auxiliary models:

- VAR: several VAR(1) of 2 variables are fitted to the observed data. We cannot use VAR of more than 2 variables because we have only one stochastic shock for several variables. If we use more variables, then there would exist deterministic relations between some variables. This issue is called stochastic singularity. (Ruge-Murcia 2007) showed that it is more problematic for estimation with Maximum Likelihood (ML) methods than for methods of moments (MM). This is so because stochastic singularity limits the number of variables one can use for estimation with ML while for MM it only limits the number of moments that can be used (and you can find linearly independant moments that involve linearly dependent variables). So we estimate VAR of GDP with consumption, GDP with investment, GDP with the interest rate, consumption with investment and investment with the interest rate.[6] The VAR coefficients are estimated by simple OLS.

- IRF: we also compute several impulse response functions of these VAR. To do so we use the method presented previously, i.e. by simulating the system for a unit shock. Note that we do not limit ourselves to structural shocks (i.e. shocks that have an economic interpretation in the framework of the model under investigation). Indeed we just need them to appropriately summerize the data (as we use them as empirical targets). We found that a span of 7 periods gave the best results for the IRF.[7]

We specify upper and lower bounds for every parameter that has to be estimated. In practice the use of bounds is necessary to attain proper estimation. Without any constraints, the algorithm will most likely diverge. Then one might think that what we do is as unsatisfactory as calibration. However it is not the case and we ensure to let the data speak by continuously enlarging the bounds until the estimates of the parameter do not hit them (whenever possible). After a lot of trials and errors, we end up using 2 sets of bounds, one with relatively tight bounds and another one with larger bounds for some parameters. Of course it

---

[6] For every VAR we add a constant to the regressors.

[7] It is a bit surprising as we generally observed that the IRF completely die out around the 12th lags.

would be better to use the large one but it turned out to be possible only under certain conditions (typically if we fix another parameter to its true value).

- α: [0.15;0.45] and [0.15;0.7]

- β: [0.9;0.999] and [0.9;0.999]

- δ: [0.05;0.2] and [0.05;0.2]

- ρ: [0.65;0.99] and [0.65;0.99]

- θ: [1.5;3.5] and [1.1;3.8]

Regarding the optimal weighting matrix W, we already know is the inverse of the variance-covariance matrix of the estimator, so $W = [Cov(g(X,\theta)]^{-1}$. In practice, we approximate it from the observed data. We define the covariance matrix with j lags as

$$\Gamma_j = \frac{1}{N} \sum_{t=j+1}^{N} g_t(\hat{X}) g_{t-j}(\hat{X})'$$

It can be approximated by the Newey-West method:

$$Cov(g(X,\theta)) = \Gamma_0 + \sum_{j=1}^{J} w_j (\Gamma_j + \Gamma_j')$$

where J are the number of lags (the bandwidth) and w_j the weight. We use the Bartlett weights defined as $w_i = \frac{1-j}{J+1}$. We construct the different observations of $g(\hat{X})$ by estimating the empirical targets for an increasing sample size starting at half of the full sample size. In our case T = 100, so we estimate VAR and IRF for a sample size of 50, 51, and so on till 100 (so we got 51 estimations of the targets).

The weighting matrix allows to give less weight to the less informative conditions, it turned out to be crucial to obtain convergence. Note that we compute this matrix from the observed data so it is computed only once.

5.1 Analysis

5.1.1 Observed Data

We briefly describe the data generated. On figure 1 we plot GDP, consumption, investment and interest rate. Of course any pattern that might be observed is particular to one simulation and another simulation would show a different one. The series look well-behave, stationnary and more importantly, the data seem to vary around its steady state. This is confirmed by the Table 1 where we report various statistics on the observed data. We can see that the mean of the series is very close to the steady state. Also the maximum variation (the variation between the minimum and the maximum value) is the largest for investment which is typically a very volatile variable. GDP and interest rate are relatively stable while consumption is more volatile (but less than investment).
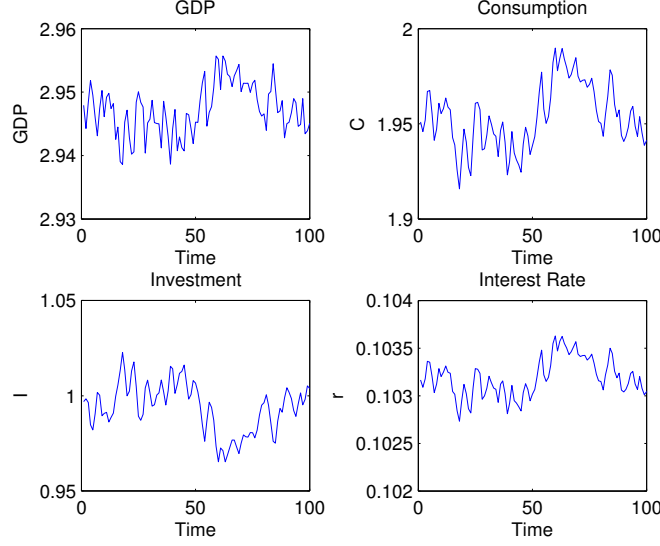
Figure 1: Observed Data

Table 1: Observed Data

	Steady State	Mean	Max	Min	Max Variation in %
GDP	2.9455	2.9553	2.9664	2.9364	0.64
Consumption	1.9455	1.9494	1.9889	1.9180	3.70
Investment	1.0000	0.9969	1.0226	0.9653	5.93
Interest rate	0.1031	0.1031	0.1036	0.1028	0.78

5.1.2 Monte Carlo Results

Estimation with Identity Matrix and small bounds   Now we shall compare the estimation results between VAR estimation and IRF estimation. We start by analyzing the results obtained when we use the identity matrix as weighting matrix. For both VAR and IRF estimation, the estimates tends to accumulate on the bounds for some parameters, but other parameters are well estimated (see Figures 2 and 3 and Table 2a). Regarding α mean and median are relatively similar for VAR and IRF and close to the true value, but VAR estimation presents a lower variance and positive excess kurtosis while IRF dislplays a negative excess kurtosis. So VAR has more observation at the mean value than IRF. Estimation of β is similar for both procedures, it is well centered but there is a lot of accumulation on the upper bound and the distribution is relatively spread and flat. δ on the other hand is very well estimated in both case but VAR is better with a lot of estimates close to the true value (positive excess kurtosis). ρ is problematic for VAR estimation, it accumulates a lot on both bounds (around half of the estimates) and the distribution is flat elswhere. In

14

this symetric configuration the central moments are not affected and are relatively good but the standard deviation is quite large. IRF estimation is much better even if ther is still a little bit of accumulation. Estimation of θ is flatter for IRF, but VAR estimates accumulates a lot on the lower bound and IRF accumulates less but on both bounds. The central moments are slightly downward biased for VAR and more largely upward biased for IRF.
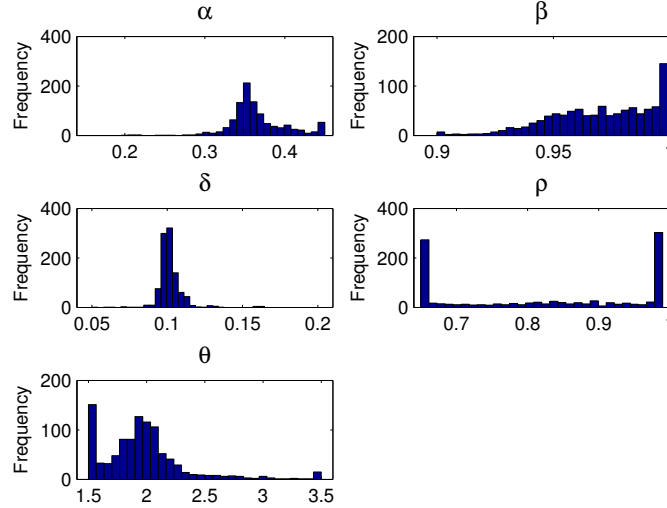

Figure 2: Histogram of VAR estimates with Identity Weighting matrix and small bounds

Overall the results are mixed. VAR estimation is a bit better for most parameters but is way worse for ρ. With such estimations (both for IRF and VAR), it seems that enlarging the bounds might solve the problem. However using the second set of bounds often led to some combinations of parameters that makes the Blanchard-Kahn solution method fail (the number of explosive eigen values was larger than the number of non-predermined variables so that no solution exists for the system). One way that might help to avoid accumulating on the bound without enlarging them is the use of the optimal weighting matrix.

Estimation with Optimal Weighting Matrix and small bounds The use of this matrix with the IRF as auxiliary model turned out to worsen the estimation. We tried to play with the number of lags but it did not improve the results. From Figures 4 and 5 we can already see that estimation with a VAR reduced form is much better than with the identity matrix while it is much worse with the IRF. The VAR estimates do not hit the bound, are less spread and are around the true value. On the other hand, IRF estimates systematically hit the bounds and accumulate on them, moreover they are biased. Table 2b shows more complete results with usual descriptive statistics. Regarding α, estimation is very precise for the VAR, maximum and minimum values would be themselves very good estimates and define a very tight interval. Both the mean and the

Table 2: Estimation Results

(a) Identity weighting matrix and small bounds

	α (0.35)		β (0.97)		δ (0.1)		ρ (0.85)		θ (2)	
	VAR	**IRF**	**VAR**	**IRF**	**VAR**	**IRF**	**VAR**	**IRF**	**VAR**	**IRF**
Min.	0.19541	0.17354	0.90000	0.90000	0.05006	0.05207	0.65000	0.65000	1.50000	1.50000
Max.	0.45000	0.45000	0.99900	0.99900	0.16458	0.17753	0.99000	0.99000	3.50000	3.50000
Median	0.35688	0.36408	0.97237	0.97142	0.10045	0.10565	0.84207	0.85945	1.94476	2.25429
Mean	0.36475	0.36274	0.97027	0.96887	0.10188	0.10686	0.82838	0.84501	1.96336	2.34735
Standard deviation	0.03538	0.05916	0.02196	0.02580	0.00871	0.01722	0.14117	0.07335	0.36706	0.51245
Skewness	0.25062	-0.42714	-0.59734	-0.74611	2.49997	0.42677	-0.11843	-0.74844	1.64147	0.52775
Kurtosis	2.15294	-0.13748	-0.15631	-0.00550	20.00730	1.38735	-1.66027	0.41904	4.39701	-0.41499

(b) Optimal weighting matrix and small bounds

	α (0.35)		β (0.97)		δ (0.1)		ρ (0.85)		θ (2)	
	VAR	**IRF**	**VAR**	**IRF**	**VAR**	**IRF**	**VAR**	**IRF**	**VAR**	**IRF**
Min	0.34805	0.15000	0.96933	0.90000	0.09985	0.05000	0.82152	0.65000	1.98936	1.50000
Max	0.35243	0.45000	0.97080	0.99900	0.10020	0.20000	0.94758	0.99000	2.01455	3.50000
Mean	0.35000	0.28655	0.97000	0.98626	0.10000	0.10035	0.85082	0.93873	2.00023	2.15802
Median	0.35000	0.27172	0.97000	0.99843	0.10000	0.06629	0.84999	0.99000	2.00018	1.56382
Standard deviation	0.00032	0.12305	0.00011	0.02418	0.00003	0.05975	0.00730	0.08770	0.00268	0.83642
Skewness	0.69871	0.13928	0.13851	-2.40939	0.60154	0.91686	8.81499	-1.99785	0.25171	0.75181
Kurtosis	11.16882	-1.75128	4.52345	5.35871	7.89234	-0.93188	93.79817	3.30789	1.91573	-1.23368

(c) Identity weighting matrix, large bounds and fixed β

	α (0.35)		β (0.97)		δ (0.1)		ρ (0.85)		θ (2)	
	VAR	**IRF**	**VAR**	**IRF**	**VAR**	**IRF**	**VAR**	**IRF**	**VAR**	**IRF**
Min.	0.19234	0.15000	-	-	0.06396	0.05000	0.65000	0.65000	1.10000	1.10000
Max.	0.53892	0.70000	-	-	0.12909	0.19684	0.99000	0.99000	3.80000	3.78504
Median	0.35224	0.40227	-	-	0.10030	0.11540	0.85438	0.85903	1.94248	2.06489
Mean	0.36410	0.42112	-	-	0.10060	0.11786	0.84460	0.84614	1.91026	2.14260
Standard deviation	0.05513	0.10137	-	-	0.00623	0.02249	0.11536	0.06781	0.32319	0.49708
Skewness	0.71832	0.76170	-	-	-0.07700	0.51956	-0.25574	-0.71008	0.31461	0.54982
Kurtosis	0.23783	0.50720	-	-	2.01263	0.47263	-1.17858	0.23893	1.64429	0.25830

The true value of parameter is given in parenthesis.

Figure 3: Histogram of IRF estimates with Identity Weighting matrix and small bounds
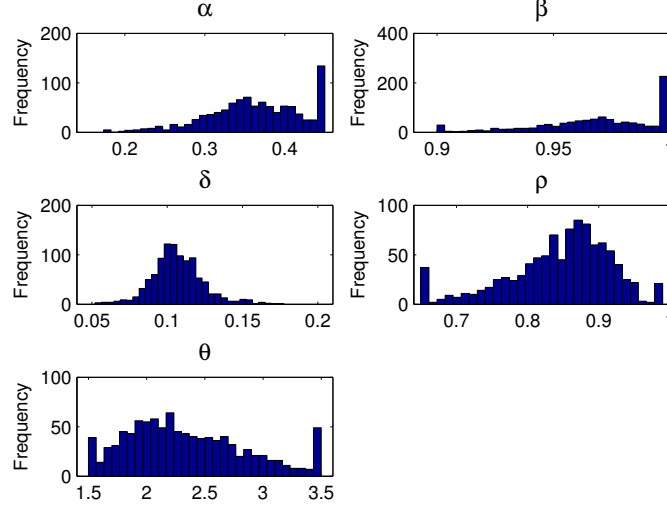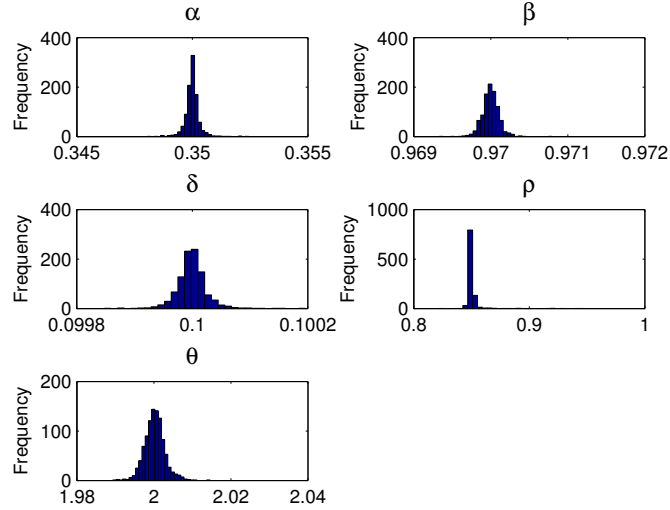

Figure 4: Histogram of VAR estimates with Optimal Weighting matrix and small bounds

median are at the true value.[8] The distribution is slightly skewed to the right,

---

[8]If we consider, like for the algorithm settings, that we are only interested in the parameters value up to 3 decimals.

but skewness remains close to zero, confirming similarity of median and mean. The high peak of observations around the mean and also the fact that extreme observations are relatively frequent is traduced by a very high positive kurtosis. However in this case, given the value of maximum and minimum, extreme values are not problematic. Estimation with IRF is not so good. First maximum and minimum are the bounds we set for the estimation. Second both the mean and the median are way too small with respect to the true value and the standard deviation is very large. The distribution is relatively symetric (slightly skewed to the right). The estimation of β is also very good for VAR (again both mean and median are at the true value and minimum and maximum values are close to the true value), not skewed (skewness close to 0) and presents an excess kurtosis close to 5. β is overestimated with IRF (there is a large accumulation on the upper bound).

δ has very tight bounds between minimum and maximum. The standard deviation is the lowest of all estimations and the distribution is slightly skewed to the right. Again, but it is the case for all VAR estimates, we observed a large positive excess kurtosis. IRF estimation is at the true value if consider the mean but displays a large negative bias if we consider the median.

ρ presents the bigger difference between maximum and minimum. But it is also characterized by a huge excess kurtosis, with a massive amount of estimations around the mean (which is almost at the true value), however the distribution is skewed to the right. IRF completely overestimate ρ, the median is even situated at the upper bound.

Finally, VAR estimates of θ are very precise with low standard deviation, tight

Figure 5: Histogram of IRF estimates with Optimal Weighting matrix and small bounds

18
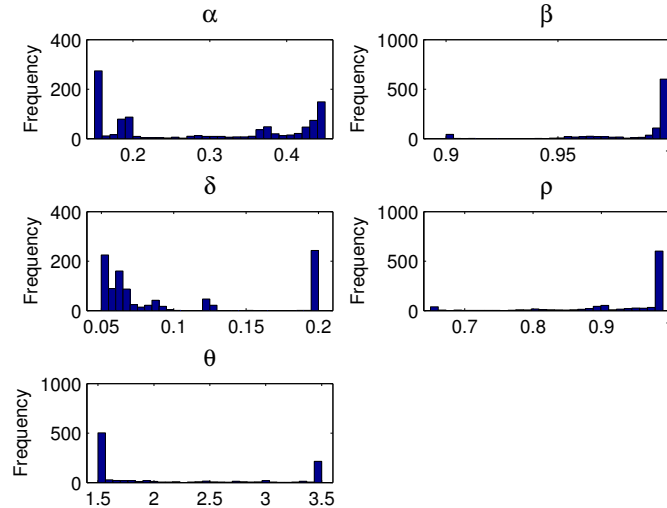
bounds and mean and median equal to the true value. IRF estimates are again less precise and biased, the median is very small (almost at the lower bound) while the mean is a bit to high with respect to the true value.

Estimation with Identity Weighting Matrix, large bounds and fixed β
The IRF estimates being very unsatisfactory with the optimal weighting matrix we try another approach to estimate the model. It still seems that enlarging the bounds could improve estimation. However, we saw that with the large bounds the solution method fails for some combination of parameters, so the idea is to use these large bounds except for β that we fix to its true value. Indeed, from the first estimation (i.e. with identity weighting matrix and small bounds, see Figure 3) β seemed to be the less well identified parameter, so fixing it might improve estimation of the other parameters. In order to be able to compare the results we apply the same procedure to the VAR model. From the histograms (Figures 6 and 7) we see that IRF estimation is better, on the other VAR estimation is better than the first estimation but worse than the second where we used the optimal weighting matrix.[9]   Enlarging the bounds had exactly

Figure 6: Histogram of VAR estimates with Identity Weighting matrix, fixed β and large bounds
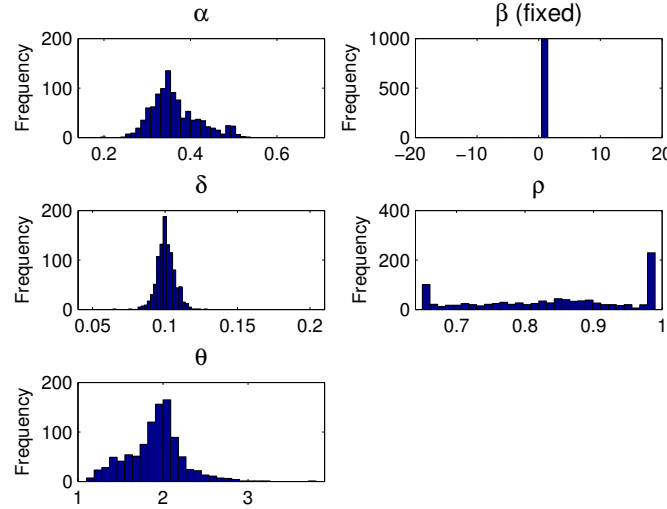
the effect expected. There is no accumulation on the bounds anymore and it increased the number of observations around the mean. Table 2c is very similar to Table 2a and tells us more or less the same things. Note that with the

---

[9]We also made the same estimation with the optimal weighting matrix but results for IRF were not so good (see Figure 11 in appendix). With large bounds, free β and optimal weighting matrix results for the VAR were very similar than with the small bounds (the histogram is given in the appendix, Figure 12) but the Blanchard-Kahn procedure failed for the IRF estimation.

Figure 7: Histogram of IRF estimates with Identity Weighting matrix, fixed β and large bounds
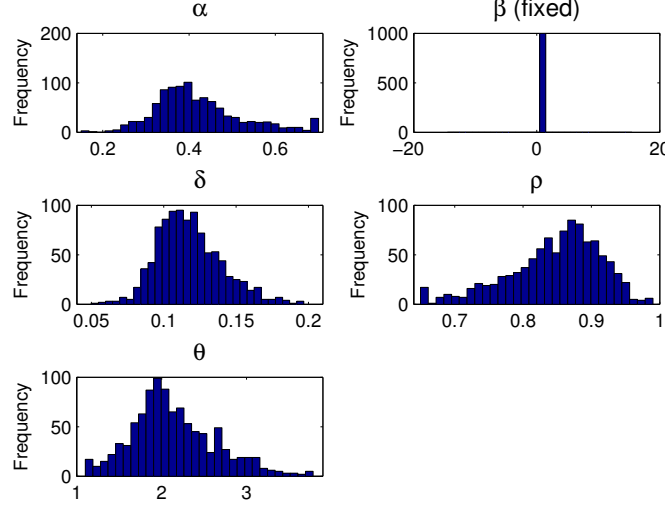
IRF α is now largely biased while it was not the case in the first regression. θ estimation on the other hand is now less biased. Regarding VAR estimates, the essential differences are the skewness and kurtosis which are closer to those of a Normal distribution for most parameters. Again VAR estimation is better than IRF if we except ρ whose distribution is similar to the one obtained with the first estimation. So it turns out that to properly estimate ρ with a VAR specification we need to use the optimal weighting matrix.

5.2 Sensitivity analysis

In previous sections we pointed out that it might come from identification issues, we also emphasized the relation between identification and the behaviour of the objective function. In Figures 8 and 9 we plot the criterion function for VAR and IRF when we vary one parameter at a time holding the others constant (at their true value).[10]

The most striking feature is the similarity of the criterion function for IRF and VAR. For most parameters the shape is indeed very similar (but not necessarily in value). α seems relatively well identified although the minimum not seem to be at the left of the true value (VAR looks slightly closer to the true value). The same comments can be made for β *mutatis mutandi*. δ has its minimum on the right of the true value for both auxiliary models. ρ and θ have more or less the same shape and does not seem identifiable. This similarity between VAR and IRF is a bit surprising, given the results of the previous section

---

[10]For every parameter, we actually run a Monte Carlo simulation with 250 simulations, we compute the criterion function at each loop, then we average it over the simulations.
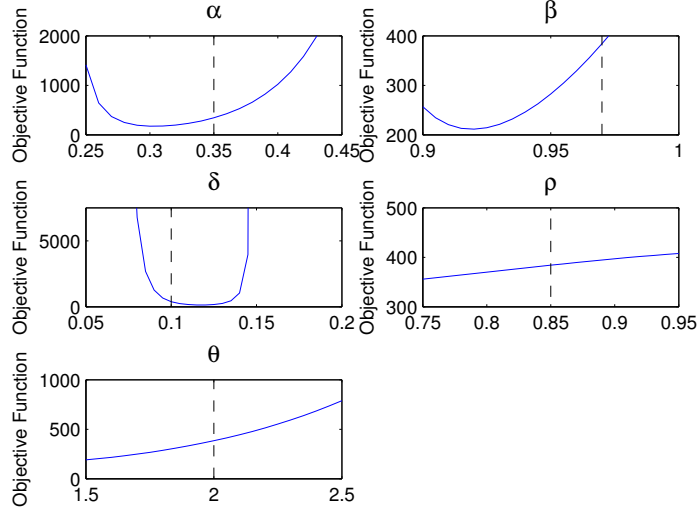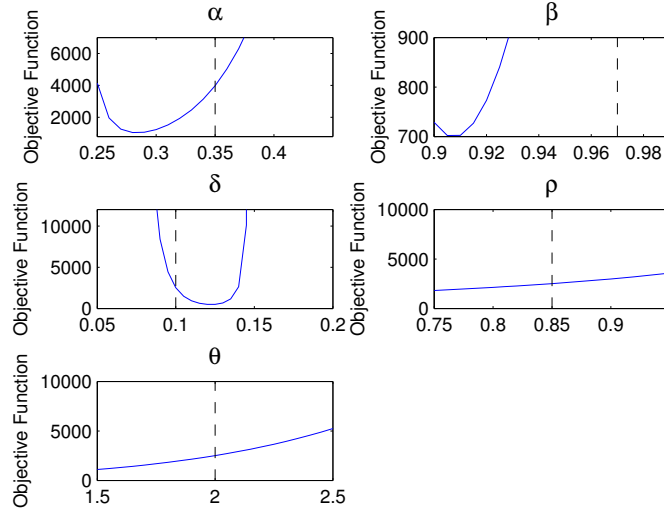
Figure 8: Objective Functions - VAR

Figure 9: Objective Function - IRF

we would expect the objectif functions to behave better for VAR than for IRF. However to compute the criterion function we had to use the identity weighting matrix rather than the optimal weighting matrix because otherwise the function can take very different values from loop to loop and average it would not make sense. This point is important because we have seen in the previous section how

21

the use of the optimal weighting matrix dramatically improves the results for VAR estimation. Another problem is that the plot of the objective function is very sensitive to scaling. Depending on the interval on which we represent it, a same objective function can exhibit very different behaviours. To avoid this problem (Canova and Sala 2009) use elasticity of the criterion function with respect to every parameter instead. The elasticity representing a relative variation, the scale problem disappears. Like any other elasticity we can define it as

$$\frac{\Delta Q}{\Delta \theta_i} \frac{\theta_i}{Q}$$

where Q is the objective function and θ_i one structural parameter.

We compute the elasticity by continuously increasing the value of the parameter, so we know that $\Delta \theta_i > 0$.[11] Q being a quadratic function, it is necessarily positive as well and none of our parameters can take negative value. The only element that can be negative is the variation of the objective function. So we

Figure 10: Elasticities of the Objective Function
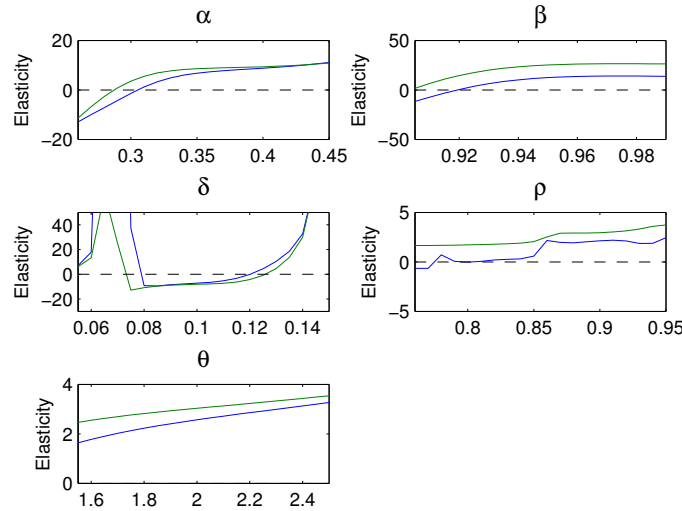
Blue lines represent VAR elasticities and red lines represent IRF elasticities.

would expect the elasticity to be negative for parameter values smaller than the true value (so that increasing parameter values would decrease the criterion function) and, following the same reasoning, positive for parameter values larger than the true value. At the true value it should be around zero. Results are consistent with what we found with the plots of the objective function. Again they look similar for most parameters. Elasticity for α is relatively well behaved despite the fact that it reaches zero at a lower value than the true value, but

---

[11] The procedure we use to compute the elasticites is similar to the one we used to compute the objective function, we run 250 Monte Carlo simulations and we average the result over the simulations.

VAR elasticity crosses zero a bit closer to the true value than IRF elasticity. δ presents the same characteristics, it is negative at first and turn positive aftewards but not at the true value, but very close to it (0.12) and again VAR is slightly closer to the true value. β is again better for VAR (negative at first and then positive) than for IRF (always positive) even if again elasticity is not zero at true value. θ and ρ are more problematic. Their elasticities are always positive, except the VAR elasticity for ρ which is negative for a few values but goes positive very quickly. It is a bit puzzling that VAR and IRF elasticity for ρ behave so similarly while their estimation differed greatly (although we may recall that objective function is plot when every other parameters are at their true value which is not the case the during estimation). From an identification perspective it is also interesting to note that for every parameter elasticity is relatively large, in the sense that a modification of the value of a parameter has a sufficient impact on the value of the objective function (for instance it would not be the case if the parameter is under identified). The conclusions are the same than with the plot of the objective function. Identification issues look similar for VAR and IRF with the identity matrix. The critical issue seems to be the optimal weighting matrix. Indeed the use of the optimal weighting matrix largely improves the estimation for VAR but not for IRF.

6 Conclusions

From our empirical investigation we found that identification was a relevant issue when we use the identity matrix as weighting matrix for the estimator. Intrestingly, with that specification results obtained were quite similar with both reduced forms if we except the autoregressive coefficient of the technology shock. However with the use of the optimal weighting matrix, estimates were very good and precise for all parameters with a VAR auxiliary model while it worsened estimates with the IRF auxiliary model. We were able to improve IRF estimation by fixing the parameter β to its true value (which is of not possible as we do not know the true value of β). In that case, with the identity matrix, estimation was reasonably good but not as good as VAR estimates and again the use of the optimal weighting matrix did not help. This fact is in itself intresting and one can wonder why the weighting matrix performs so poorly with IRF.

As both auxiliary models are derived from the same specification (thus not requiring any effort for using one rather than the other) it seems that it is preferable to use a VAR reduced form rather than an IRF reduced form. Of course, this result is particular to the RBC model we used and it would be intresting to see if it remains true for other specifications. However, note that we deliberately chose a simple specification, supposed to be less subject to identification and estimation problems, any more complex models should worsen estimation issues we encountered.

References

BLANCHARD, O. J., AND C. M. KAHN (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48(5), 1305–1311.

CANOVA, F., AND L. SALA (2009): "Back to square one: Identification issues in DSGE models," *Journal of Monetary Economics*, 56(4), 431–449.

GOURIEROUX, C., A. MONFORT, AND E. RENAULT (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8(S), S85–118.

ISKREV, N. (2010): "Local identification in DSGE models," *Journal of Monetary Economics*, 57(2), 189–202.

KLEIN, P. (2000): "Using the generalized Schur form to solve a multivariate linear rational expectations model," *Journal of Economic Dynamics and Control*, 24(10), 1405–1423.

RUGE-MURCIA, F. J. (2007): "Methods to estimate dynamic stochastic general equilibrium models," *Journal of Economic Dynamics and Control*, 31(8), 2599–2636.

SIMS, C. A. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20(1-2), 1–20.

SMETS, F., AND R. WOUTERS (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97(3), 586–606.

A   Solution of the inter-temporal maximisation

The problem is the following:

$$\max_{\{c_t\}_{t=1}^{\infty}} E_t \quad \sum_{s=t}^{\infty} \quad \beta^{s-t}u(c_t) \quad s.t. \quad (2) \ (3) \ (4) \ and \ (5) \tag{12}$$

If we define $U_t = E_t \sum_{s=t}^{\infty} \beta^{s-t}u(c_t)$ the optimal program can be expressed as follows

$$V(k_t, z_t) = \max_{c_t} U_t \quad s.t. \quad k_{t+1} \quad = \quad g(k_t, c_t, z_t)$$

where $k_{t+1} = g(k_t, c_t, z_t)$ represents the constraint imposed by the equation for capital. Actually it is the only constraint as the production function is included in it and the z is considered as exogeneously given.

The first step is to specify to so-called Bellman equation and the first order conditions (FOC). For that purpose, we need to distinguish state variables and control variables. A state variable is defined as a variable that the agents do not choose directly as opposed to a control variable which is chosen by the agent. However, unless the state variable is totally exogenous, the choice of the level of the control variable indirectly influences the level of the state variable. Thus in our case, consumption is the control variable and capital the state variable. Moreover we consider z_t as a state variable, as it is exogeneously determined we do not need to take into account the specific form that generates z_t. We just need to know that it is a source of uncertainty and thus requires the expectation sign. Also, rather than keeping an equation for the production we specify it directly in the equation for the evolution of capital.

- Bellman Equation and FOC
  Using the fact that the utility function is additively separable we can write $U_t = u(c_t) + \beta U_{t+1}$. Moreover in order to focus on today's optimization problem only, we consider future path as optimal, so we replace U_{t+1} by $V(k_{t+1}, z_{t+1})$.
  This gives us the value function:

$$V(k_t, z_t) = \max_{c_t} u(c_t) + \beta E_t V(k_{t+1}, z_{t+1})$$

  The FOC can be easily derived:

$$u'(c_t) \quad + \quad \beta E_t \frac{\partial V(k_{t+1}, z_{t+1})}{\partial k_{t+1}} \frac{\partial k_{t+1}}{\partial c_t} = 0$$

$$\Leftrightarrow u'(c_t) \quad = \quad \beta E_t V'(k_{t+1} z_{t+1}) \tag{13}$$

- Evolution of the costate variable
  The costate variable is the derivative of the value function with respect to the state variable.
  With the envelope theorem we can immediately compute the derivative of the maximized Bellman equation with respect to the state variable:

$$V'(k_t, z_t) \quad = \quad \beta E_t \frac{\partial V(k_{t+1}, z_{t+1})}{\partial k_{t+1}}$$

$$= \quad \beta E_t \frac{\partial k_{t+1}}{\partial k_t} \frac{\partial V(k_{t+1}, z_{t+1})}{\partial k_{t+1}}$$

$$= \quad \beta E_t [1 - \delta + \frac{\partial y_t}{\partial k_t}] V'(k_{t+1}, z_{t+1}) \tag{14}$$

- Inserting FOC

  We merely insert the FOC (equation 13) in (14) and we obtain the Euler equation

  $$u'(c_t) = \beta E_t[1 - \delta + \frac{\partial y_{t+1}}{\partial k_{t+1}} u'(c_{t+1})]$$

  If we apply it to our model specification:

  $$c_t^{-\theta} = \beta E_t[1 - \delta + \alpha k_t^{\alpha-1} A exp(z_t) c_{t+1}^{-\theta}] \tag{15}$$

B   Matrix $\tilde{A}$ for the linearized system

$$\tilde{A} = \begin{pmatrix} A\alpha k_{ss}^{\alpha-1}+1-\delta & -1 & Ak_{ss}^{\alpha} \\[2ex] \dfrac{A\alpha c_{ss} k_{ss}^{\alpha-1}(\alpha-1)(A\alpha k_{ss}^{\alpha-1}+A\alpha k_{ss}^{\alpha})}{\theta(k_{ss}-\delta k_{ss}+A\alpha k_{ss}^{\alpha})} & \dfrac{k_{ss}}{\beta(k_{ss}-\delta k_{ss}+A\alpha k_{ss}^{\alpha})} - \dfrac{A\alpha c_{ss} k_{ss}^{\alpha-1}(\alpha-1)}{\theta(k_{ss}-\delta k_{ss}+A\alpha k_{ss}^{\alpha})} & \dfrac{A\alpha c_{ss} k_{ss}^{\alpha}\rho}{\theta(k_{ss}-\delta k_{ss}+A\alpha k_{ss}^{\alpha})} + \dfrac{A^2\alpha c_{ss} k_{ss}^{\alpha} k_{ss}^{\alpha-1}(\alpha-1)}{\theta(k_{ss}-\delta k_{ss}+A\alpha k_{ss}^{\alpha})} \\[2ex] 0 & 0 & \rho \end{pmatrix}$$

C Histograms for estimation with large bounds

Figure 11: Histogram of IRF estimates with Optimal Weighting matrix, large bounds and fixed β
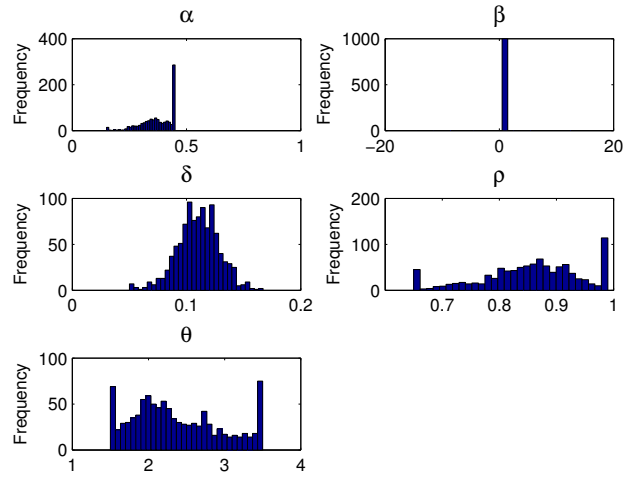

Figure 12: Histogram of VAR estimates with Optimal Weighting matrix, large bounds and free β