Interest rate differentials and the dynamic asymmetry of exchange rates

J. Hambuckers[1,†] and M. Ulm[2]

Abstract

We propose a unified econometric strategy to revisit the informational content of interest rate differentials (IRD) for predicting exchange rates. The novelty of our approach consists in allowing for a time-varying asymmetry component in the conditional distribution of the depreciation rate, therefore explicitly modeling the link between interest rates and the likelihood of a depreciation. To assess the economic significance of IRD as a predictor, we derive a directional forecasting procedure from our model and apply this technique to daily exchange rates of the Euro and the Swiss Franc. We document in-sample and out-of-sample performances significantly superior to benchmark models, both in terms of sign forecasts and trading profits. Overall, we find the dynamic asymmetry component to be driven by interest rate differentials, but also by general uncertainty and past unexpected shocks. These findings empirically confirm currency crash theories for recent time periods, suggesting that the larger the difference between interest rates, the more likely the high yield currency appreciates but also exhibit larger depreciation risks.

Keywords: Exchange rate, interest rate differential, GARCH, dynamic asymmetry.

JEL: C53, C58, C22, F31.

[1] HEC Liège, Finance Department, Liège, Belgium.

[2] Georg-August-Universität Göttingen, Chair of Statistics and Econometrics, Göttingen, Germany.

[†] Corresponding author: jhambuckers@uliege.be.

1 Introduction

The exact nature of the relationship between short term interest rate differential (IRD) and currency depreciation remains an ongoing debate: on the one hand, economic theory postulates that IRD and foreign exchange rates are linked over time via the uncovered interest rate parity (UIP) rule. In this framework, the currency of the high yield economy is expected to depreciate, offsetting possible gains derived from a carry trade strategy[1]. Empirically, however, we observe the opposite: several studies (among others Gabaix and Maggiori [2015]) report an appreciation of the high yield currency over long periods of time, which contradicts UIP.

This apparent contradiction is a long-standing question in the finance literature. Early on, Meese and Rogoff [1983] notice that models based on IRD cannot beat a simple random walk in predicting future exchange rates, raising the question of its predictive content. Despite considerable progress in terms of data availability and econometrics techniques for the past 35 years, few has changed in this regard. In a recent review, Rossi [2013] concludes that even if many predictors and models provide economically significant in-sample forecasts for the mean, few produce significant out-of-sample forecasts, especially at short time horizon. A similar conclusion is reached by Hsu et al. [2016], who look at profits made by trading strategies using various predictors for a large panel of currencies. They observe that, in recent times, currencies of developed economies have been particularly hard to predict.

Several theoretical reasons are advanced for this lack of performance: most notably the time-varying predictive content of the fundamentals like IRD [Bacchetta and van Wincoop, 2013, Berge, 2014, Ismailov and Rossi, 2018], but also the misspecifications of the models traditionally used to conduct these forecasts [Cheung et al., 2005, Rossi, 2013, Ismailov and Rossi, 2018, Amat et al., 2018]. Indeed, whereas most models focus on conditional mean forecasts [Husted et al., 2018], exchange rates exhibit high-order dynamics and extremely weak mean dynamics [Chung and Hong, 2007, Brunnermeier et al., 2008, Ismailov and Rossi, 2018]. To illustrate this relation between exchange rates and IRD, we conduct the following analysis: we fit a simple GARCH(1,1) model on the daily log-rate of change of the USD/EUR exchange rate between 1999 and 2019, and look at the relationship between empirical skewness of the residuals and IRD. That is, for a defined IRD threshold (e.g. 1.5%), we pool all residuals whose associated IRD is larger (resp. smaller) than this threshold, and compute the empirical skewness of this subsample. We repeat this operation for a sequence of thresholds. If the true model has its random part independent from IRD, then conditional skewness must be zero. Figure 1 shows the obtained results considering negative and positive thresholds (solid red line). Using residual-bootstrap confidence intervals where we enforce the independence with IRD, we conclude that the independence hypothesis has to be rejected. On the contrary, we observe some intriguing

[1] *Carry trade* here refers to an investment strategy that consists in borrowing in a low interest rate currency and investing in another with a higher interest rate over a certain period of time, without hedging exchange rate risk.

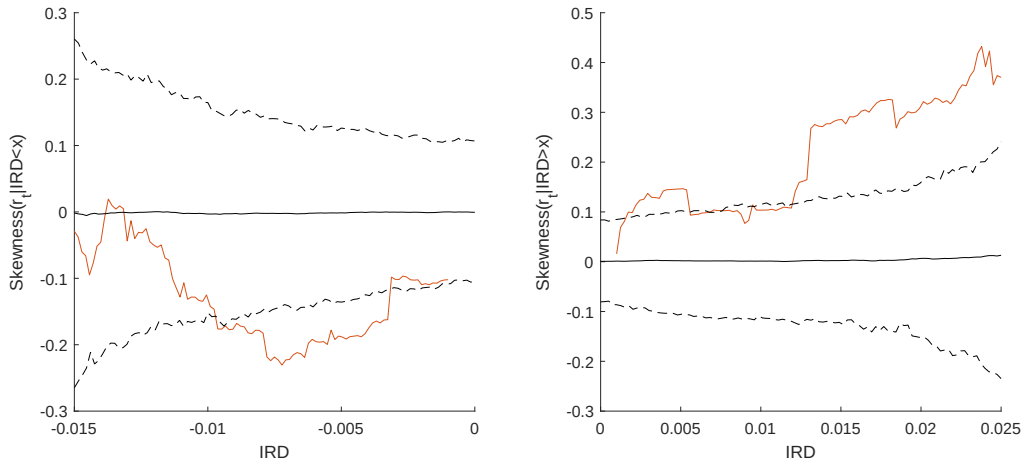


Figure 1: Solid red: Empirical skewness for residuals of a GARCH(1,1) computed on daily returns of USD/EUR, conditional on smaller or larger IRD (left: smaller values, right: larger values, as given by the x-axis). Dashed black: 95% confidence intervals for the empirical skewness of GARCH(1,1) data simulated independently of the IRD process.

patterns between skewness of the residuals and IRD. In these conditions, tools like vector autoregressive models are ineffective [Herwartz, 2017] and other strategies such as modelling conditional skewness should be considered [Chung and Hong, 2007, Brunnermeier et al., 2008, Anatolyev and Gospodinov, 2010, Liu, 2015].

In light of these concerns, the purpose of the present paper is to detail an improved econometric strategy to revisit the predictive content of IRD for the daily exchange rate depreciation rate of developed economies. In particular, we avoid establishing a structural link between the *level* of future (log) depreciation rates and interest rates. Instead, motivated by Figure 1, we allow interest rates to convey information on the *density* of future depreciation rates, and more precisely on its *asymmetry*. We focus on the latter component since it is the major factor in the *likelihood* of a depreciation. Therefore, this set-up is more parsimonious in its structural assumptions, and allows to test empirically general assertions such as "is a currency more likely to appreciate when its interest rate is relatively high?" without too many restrictions. Thus, we can investigate if IRD predicts the direction of change of exchange rates and if this direction is consistent with UIP literature. In particular, we are among the first to answer this last question with a unified and time series approach, instead of relying on "static" or empirical measures of skewness as in Brunnermeier et al. [2008].

The main feature of our model is a GARCH structure of the variance, associated with a dynamic non-Gaussian distribution for the innovations. In this model, the asymmetry parameter varies over time according to a time series equation augmented with exogenous predictors. It allows for time-varying skewness (and kurtosis), thus addressing the critiques of neglecting high-order dependence structures. We detail our approach in Section 2.

We use this methodology to study the depreciation rate of the US Dollar (USD) vis-a-vis

two major currencies: the Euro (EUR) and the Swiss Franc (CHF). Our choice of currencies is motivated by the findings of Hsu et al. [2016], who could not find meaningful trading strategies for these economies. Our goal is thus to investigate if IRD is an important factor in the dynamic skewness of these currencies, and if economically meaningful forecasts can be derived from it. We test various specifications of the skewness dynamics, considering also additional control factors like past innovations, past skewness parameter as well as the VIX, an important factor suggested in Ranaldo and Söderlind [2010], Menkhoff et al. [2012] and Ismailov and Rossi [2018]. To assess the economic significance of our findings, we look at the average return generated by a simple directional trading strategy based on our model, and compare it to several benchmarks. We also conduct a persistence analysis to assess if our results hold out-of-sample. A novelty of our analysis is that we do not use our economic performance criteria (directional accuracy or trading profit) to estimate and select our model, but only statistical ones (i.e. the log-likelihood). This approach makes our findings a by-product of genuine forecasting ability.

Although motivated primarily by empirical findings, the proposed approach also takes its roots in theoretical arguments recently put forward in the exchange rate literature. In particular, Fahri and Gabaix [2016] link the time-varying probability of rare disasters and the exposure of a country to such disasters to the risk of a depreciation. They argue that relatively risky countries feature high interest rates because investors need to be compensated for a potential depreciation in case of a disaster. This suggests that IRD is informative about the likelihood of a future depreciation, as well as about the anticipation of a currency crash. Thus, whereas UIP postulates instantaneous realignment pressures when IRD increases, we hypothesize that a large (absolute) IRD is indicative of a higher risk of a reverting mechanism, i.e. of a future depreciation of the low-yield currency. We thereby assume the marginal effect of IRD to convey information on the likelihood of an appreciation or depreciation, instead of on the move itself. This is implemented by means of a regression structure in the skewness dynamics, rather than at the mean level, allowing for local deviations in the likelihood of a depreciation. Such a structure is consistent with the empirical findings of an absence of predictability for IRD in the classical regression framework, but with the existence of a more general predictive content.

Our main conclusions are the following:

i. IRD, past unexpected shocks and VIX are important factors to model the dynamic skewness of the daily depreciation rate for EUR and CHF.

ii. An increase in IRD is associated with an increasing likelihood of appreciation of the high-yield currency, but at the price of an increasing risk in a large currency crash.

iii. In terms of sign forecasts, we show a statistically significant predictive performance of our model over some periods of time for both currencies.

iv. The predictive content of the three mentioned factors is sufficiently strong to generate significant economic gains when trading with a dynamic skewness model, both in- and

out-of-sample.

The rest of the paper is structured as follows: in Section 2, we detail the features of the statistical model, the estimation approach and the interpretation of the model. In Section 3, we perform the empirical analysis and investigate the economic significance of our results in Section 3.3. We conclude in Section 4.

2 Methodology

The fundamental feature of our econometric approach is a time-varying asymmetry in the distribution of the stochastic component, depending on IRD. To do so, we build upon a classical GARCH model, undoubtedly the work-horse model for daily financial data [Engle, 1982, Bollerslev, 1986]. In this model, the distribution of the innovations is usually assumed to be symmetric and time-constant (e.g. Gaussian or t-distributed). Surprisingly, and despite the consensus on the non-Gaussian, time-varying nature of financial time series, few studies are concerned with dynamic conditional asymmetry. On the contrary, the existing literature focuses on time-varying volatility [Hansen and Lunde, 2005, Francq and Zakoian, 2010], on the asymmetric response of volatility [Glosten et al., 1993] or on leptokurtosis in the error distribution [Bai et al., 2003, Klar et al., 2012].

The idea of time-varying asymmetry of GARCH innovations can be traced back to Hansen [1994], who introduces the autoregressive conditional distribution (ACD) model. In this model, the GARCH structure of the volatility is combined with skewed-t innovations where the skewness parameter varies over time. Harvey and Siddique [1999] as well as Jondeau and Rockinger [2003] build upon this work to introduce variants where the skewness itself varies over time. More recently, Grigoletto and Lisi [2009] have considered a similar approach with the Pearson-type IV distribution instead of the skewed-t distribution. Wilhelmsson [2009] proposes a variant based on the Normal Inverse Gaussian distribution whereas Bali et al. [2008] rely on the skewed generalized-t distribution, with time-varying kurtosis. For stock indices, dynamic asymmetry has been studied, e.g. in Hansen [1994], Harvey and Siddique [1999], Jondeau and Rockinger [2003], Wilhelmsson [2009] and Grigoletto and Lisi [2009][2]. On exchange rates, the literature is especially limited. Looking at time-varying skewness for several currency pairs, Jondeau and Rockinger [2003] find that its dynamic can be explained by an autoregressive process.

In general, skewed-t, Pearson-type IV and Normal Inverse Gaussian, despite their flexibility, suffer from cumbersome constraints and numerical issues. In addition, these distributions are not always tractable and some of their parameters are difficult to interpret or need to be constrained to ensure the existence of the first four moments. To avoid these shortcomings,

[2]More recently, Bekaert et al. [2015] have relied on two gamma-distributed innovations to account for non-Gaussianity in GARCH-type models. Although in effect their approach models time-varying asymmetry, they focus instead on differencing the effects of negative and positive shocks on the stochastic component of stock returns.

we instead consider a GARCH-type model combined with a sinh-arcsinh distribution for the innovations (SH, Jones and Pewsey [2009]), abbreviated GARCH-SH in the later. Contrary to the aforementioned distributions, the standardized SH distribution has two parameters (ϵ and δ) with interpretable meanings (asymmetry and shape), is centered on the Gaussian distribution (with $\epsilon = 0$ and $\delta = 1$) and has the single constraint[3] $\delta > 0$. Moreover, it accounts for heavier and lighter tails than the normal distribution, a feature not possible with the skewed-t distribution, and has all its moments that exist without additional restrictions. This last feature is particularly appealing, as the existence of high-order moments is often a needed requirement for inference. In our suggested GARCH-SH approach, we specify the parameter ϵ to evolve according to an ARMAX structure, i.e. an autoregressive-moving average structure complemented by explanatory variables. Thus, we can link the conditional distribution of exchange rate returns with relevant financial and economic factors, and account for a dependence structure beyond the first and second moments. Furthermore, we let the volatility level enter the mean equation, defining a GARCH-in-Mean model as in Glosten et al. [1993]. Empirically, the use of the contemporaneous volatility in the mean equation is motivated by Ranaldo and Söderlind [2010] and Menkhoff et al. [2012], who find a significant relation between the volatility and the expected depreciation of a currency. We detail the model and its essential features in the following subsections.

2.1 Model specification and interpretation

We specify the exchange rate model according to the following set of equations: denoting by S_t the nominal exchange rate at time t, the log-rate of change $R_t = \log\left(S_t/S_{t-1}\right)$ follows a multiplicative heteroscedastic process of the form

$$
\begin{aligned}
R_t &= c + \lambda\sigma_t + r_t, && (1)\\
r_t &= \sigma_t z_t, && (2)\\
\sigma_t^2 &= \omega + \alpha\sigma_{t-1}^2 + \beta r_{t-1}^2, && (3)\\
z_t|\mathcal{I}_{t-1} &\overset{iid}{\sim} f(z_t; \epsilon_t, \delta|\mathcal{I}_{t-1}), && (4)\\
\epsilon_t &= g(\mathcal{I}_{t-1}), && (5)
\end{aligned}
$$

where c is a constant, σ_t^2 the conditional variance of r_t, and z_t the innovation at time t with mean zero and unit variance. $\mathcal{I}_t$ denotes the information set up to time t, composed of all values of z_t and vectors of covariates $\mathbf{x}_t$ up to time t. The probability density function (pdf) of the standardized sinh-arcsinh distribution with parameters ϵ_t and δ, conditional on $\mathcal{I}_{t-1}$ is denoted by $f(z; \epsilon_t, \delta|\mathcal{I}_{t-1})$. Moreover, $g(\cdot)$ is a parametric functions linking the asymmetry parameter to past information. Expressions for the pdf, the value of the location and scale parameters in the standardized case as well as formulas for the moments can be found in Appendix A.

[3]Another constraint, although classical, is the finiteness of the parameters, which is needed to ensure that the distribution is proper.

Notice that we assume the parameter δ constant over time. Without loss of generality, we can easily relax this assumption to obtain a more flexible model, but prefer keeping a low number of parameters. In addition, since the kurtosis depends on ϵ_t as well (see Appendix A), dynamic kurtosis is automatically implied from our specification. Conditions stated in the same appendix ensure that $\mathbb{E}(z_t) = 0$ and $\mathbb{E}(z_t^2) = 1$, so that σ_t^2 can be interpreted as the conditional variance of r_t.

As explained in Jones and Pewsey [2009], the SH distribution is conveniently built around the Gaussian distribution such that, assuming a random variable $Y \sim N(0,1)$, we can define $f(z;\epsilon,\delta)$ by the sinh-arcsinh transformation:

$$Z = \sinh\left(\frac{\sinh^{-1}(Y) + \epsilon}{\delta}\right). \tag{6}$$

Skewness increases with increasing ϵ for $\epsilon \in]-\infty, +\infty[$, where $\epsilon > 0$ corresponds to positive skewness. Notice that positive (negative) skewness, for a standardized random variable, implies that there is more probability mass below (above) zero. The kurtosis decreases with increasing δ, $0 < \delta < +\infty$, $\delta < 1$ yielding heavier tails than the normal distribution. Thus, the Gaussian distribution has a central position in the SH distribution. This is an advantage compared to other distributions, for which the Gaussian distribution is usually a limiting case. Another advantage of the SH distribution is the existence of all its moments for finite values of the parameters. This is particularly useful for inference and residuals-based tests.

For ϵ_t, we define eq. (5) as a function of past innovations z_{t-1}, lagged values ϵ_{t-1} as well as past values of explanatory variables x_{t-1} (for the sake of exposition, we assume x_{t-1} to be a scalar here, but one can easily generalize to x_{t-1} being a vector). Eq. (5) is expressed in the following way:

$$\epsilon_t = g(\mathcal{I}_{t-1}) = a_0 + a_1\epsilon_{t-1} + a_2 z_{t-1} + a_3 x_{t-1}. \tag{7}$$

This equation can be modified or restricted in several ways. For instance, assuming that all parameters in (7) take value zero, we are back to the symmetric case. Setting $a_1 = a_2 = a_3 = 0$ leads to a model without dynamics but including asymmetry, whereas assuming a_1 and a_3 to be zero leads to a model where only past innovations impact on the asymmetry. Furthermore, as explained in Jondeau and Rockinger [2003], a model where $a_2 = a_3 = 0$ and $a_1 \neq 0$ is not properly identified: for t sufficiently far from zero, ϵ_t equals its stationary value $\epsilon^* = a_0/(1-a_1)$. Since we do not set an additional restriction linking a_0 and a_1, there exists an infinity of pairs (a_0, a_1) solving this equation. In practice, the estimation will converge at random, depending on the starting value chosen for ϵ_0. Thus, in our application, we assume that at least one of the other coefficients is always different from zero. In addition, the stability of the process is fulfilled when $|a_1| < 1$.

From an economic perspective, eq. (7) can be used to study how explanatory variables and past stochastic components influence the distribution of exchange rate returns. In particular, we suggest looking at three quantities: the probability of a positive shock (π_t), indicative of

the likelihood of a depreciation of the home currency[4]. Then, two measures of *currency crash risk*, indicating the likelihood of a sudden large depreciation (resp. appreciation) of the home currency. These measures are denoted ρ_t^+ and ρ^-, respectively. Mathematically, these quantities are defined by

$$\pi_t = \mathbb{P}(z_t > 0), \tag{8}$$

$$\rho_t^+ = \mathbb{P}(z_t > q^+), \tag{9}$$

$$\rho_t^- = \mathbb{P}(z_t < q^-), \tag{10}$$

where $q^+ > 0$ is large and $q^- < 0$ is small (e.g. an empirical quantile far in the tail of the distribution). In our empirical study, we use $q^+ = 2$ and $q^- = -2$, values roughly corresponding to the 98% and 2% quantile of a random shock following a standardized Gaussian distribution.

The effect of a marginal change in one component in eq. (7) on these quantities is deduced from the sign of the regression coefficients. Table 1 summarizes the various scenarios and highlights the important connections between the asymmetry parameter (ϵ_t), the likelihood of a depreciation (π_t) and the crash risks (ρ_t^+ and ρ_t^-): if the density is positively skewed (i.e. if ϵ_t is positive, Figure 2, solid black line), then an appreciation is more likely than a depreciation (i.e. $\pi_t < 0.5$). If the asymmetry parameter becomes more positive (Figure 2, dashed red line), the density becomes more positively skewed, and an appreciation is even more likely (i.e. π_t becomes smaller). However, the risk of a large depreciation increases (i.e. ρ_t^+ increases). A similar reasoning holds for negative asymmetry parameter ϵ_t and the risk of a large appreciation ρ_t^-.

	$x_{t-1} \nearrow$, $a_3 > 0$	$x_{t-1} \nearrow$, $a_3 < 0$	$x_{t-1} \searrow$, $a_3 > 0$	$x_{t-1} \searrow$, $a_3 < 0$
ϵ_t	+	-	-	+
π_t	-	+	+	-
ρ_t^+	+	-	-	+
ρ_t^-	-	+	+	-

Table 1: Summary of the effect of a change in x_{t-1} on ϵ_t, π_t, ρ_t^+ and ρ_t^-.

[4]Throughout the paper, we use the direct quotation for exchange rates, i.e. we express one unit of the foreign currency in terms of the home currency.

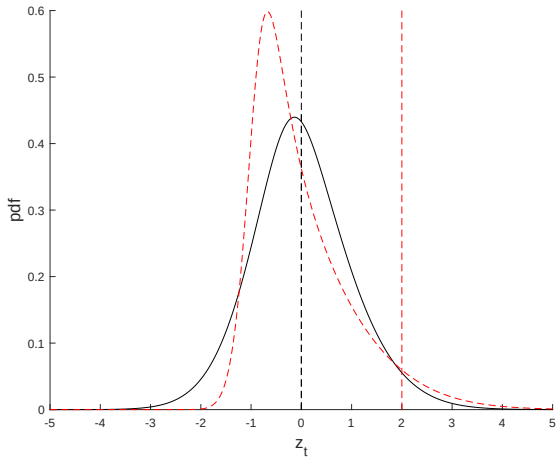


Figure 2: Example of SH distributions with $\delta = 0.85$ and $\epsilon_t = 0.1$ (black) or $\epsilon_t = 0.7$ (dashed red). Those values imply a skewness around 0.25 and 1.3, respectively. For $\epsilon_t = 0.7$, we observe $\pi_t = 0.395$ and $\rho_t^+ = 0.05$. For $\epsilon_t = 0.1$, we observe 0.48 and 0.031 for these two quantities, respectively. Vertical black (resp. red) line denotes the mean (resp. the 99% quantile).

2.2 Estimation procedure and inference

We estimate the model by means of maximum likelihood procedures. Denoting by Θ the vector of all parameters in equations (1) to (5), by $\mathbf{y_T} = \{R_t\}_{t=1,\dots,T}$ the time series of observed log-rate of change and assuming conditional independence, the conditional log-likelihood function $\mathcal{L}(\Theta; \mathbf{y}_T)$ is given by

$$\mathcal{L}(\Theta; \mathbf{y}_T) = \frac{1}{T} \sum_{t=1}^{T} \log \left(\frac{1}{\sigma_t} f((R_t - c - \lambda\sigma_t)/\sigma_t; \epsilon_t, \delta | \mathcal{I}_{t-1}) \right). \tag{11}$$

An estimator $\hat{\Theta}$ of Θ is obtained by maximizing numerically (11) with respect to Θ:

$$\hat{\Theta} = \arg\max_{\Theta} \mathcal{L}(\Theta; \mathbf{y}_T), \tag{12}$$

and subject to the the constraints $\omega, \alpha, \beta > 0$, $\alpha + \beta < 1$ and $\delta > 0$. We do not set constraints on the other parameters[5]. A simulation study of the finite-sample performance of the proposed estimation method can be found in Appendix B.

Under correct specification of the model and usual stationarity conditions, the Fisher-Information matrix $H(\Theta)$ of (11) at $\hat{\Theta}$ can be used for testing the following null hypothesis:

$$H_0 : \theta_i = 0, \tag{13}$$

where θ_i is the i^{th} element of Θ. To do so, we use the Wald-type test statistic

$$w_i = \hat{\theta}_i / \hat{\sigma}_{ii}, \tag{14}$$

[5]Regarding the choice of a starting value ϵ_0, we use $(a_0 + \sum_{j>2} a_j \bar{x}_j)/(1 - a_1)$. We check also *a posteriori* if the estimated parameters ensure finite values of ϵ_t when $T \to +\infty$

where $\hat{\theta}_i$ is an estimator of θ_i and $\hat{\sigma}_{ii}^2$ is the i^{th} diagonal element of $H^{-1}(\hat{\Theta})$. Under H_0, $w_i \overset{as.}{\sim} N(0,1)$. In Appendix B, we perform a simulation study showing that with time series of reasonable lengths, this approximation gives well-sized and respectably powerful tests. Similarly, restrictions in eq. (7) can be tested using likelihood ratio (LR) test statistics of the type

$$LR = -2(\mathcal{L}(\Theta_0; \mathbf{y}_T) - \mathcal{L}(\Theta_1; \mathbf{y}_T)), \tag{15}$$

with Θ_0 being a restricted version of Θ_1. Under the null hypothesis of the restricted model being the true one, we have the usual result $LR \overset{as.}{\sim} \chi^2_\nu$, ν being the number of restrictions.

2.3 Directional forecasts, performance measures and testing for superior ability

As noticed by Blaskowitz and Herwartz [2011], in the specific context of exchange rates, monetary authorities and investors are particularly interested in the direction of change of the market: for monetary authorities, a good anticipation of the direction of exchange rate movements is important for policy implementation, whereas for investors this knowledge helps to hedge currency risk or devise an investment strategy. With the goal of assessing the economic significance of a model of exchange rate with high-order dynamic, we focus on producing daily directional forecasts using the proposed model (both in-sample and out-of-sample), and on measuring its directional accuracy.

This task is particularly appropriate for the considered model, since time-varying asymmetry is crucial for good directional forecasts [Liu, 2015]. In the framework of a multiplicative heteroscedastic model with a zero-mean like GARCH, if the distribution of the innovations is (dynamically) asymmetric, then tomorrow's probability of a positive (resp. negative) variation would be lower (resp. larger) than a negative one. Consequently, knowing the level and sign of asymmetry enables us to compute a probability of appreciation or depreciation, and to derive a forecasting strategy. An easy analogy can be made: at each point in time, we are involved in a coin tossing bet - facing heads or tails - whereas the time-varying probabilities of each result are not equal. This implies that if we knew these probabilities, we could choose the most likely outcome. On the contrary, if the conditional distribution is symmetric, there is an equal probability for each outcome, leaving us with no dominant forecasting strategy.

Under correct specification, we can easily compute, at each point in time, the probability that the foreign currency appreciates (i.e. that $R_t > 0$), given the information set at time $t-1$. This probability is denoted $p_{t|t-1}$ and is obtained from

$$\begin{aligned}
p_{t|t-1} &= 1 - \mathbb{P}(R_t < 0 | \mathcal{I}_{t-1}), && (16)\\
&= 1 - \mathbb{P}(c + \lambda\sigma_t + \sigma_t z_t < 0 | \mathcal{I}_{t-1}), && (17)\\
&= 1 - \mathbb{P}(z_t < -c/\sigma_t - \lambda | \mathcal{I}_{t-1}), && (18)\\
&= 1 - F(-c/\sigma_t - \lambda; \epsilon_t, \delta | \mathcal{I}_{t-1}), && (19)
\end{aligned}$$

where $F(\cdot)$ denotes the SH cumulative distribution function (cdf), see Appendix A. Estimates of $\hat{p}_{t|t-1}$, for $t = 2, \ldots, T$, are obtained by plugging $\hat{\Theta}$ in (19). Then, the final directional forecast is obtained from the following indicator variable:

$$\hat{p}_t^* = \begin{cases} 1 \text{ if } \hat{p}_{t|t-1} > 0.5, \\ -1 \text{ otherwise.} \end{cases} \tag{20}$$

where 1 indicates an appreciation of the foreign currency (or a positive variation) and -1 an appreciation of the home currency (or a negative variation). If $\hat{p}_{t|t-1}$ is larger than .5, we forecast an appreciation of the foreign currency. Then, the rational strategy consists in buying the foreign currency (resp. borrowing the domestic currency) at the beginning of the period, and in closing the position at the end of the day.

In this framework, though, the direction of change cannot be perfectly forecast except if $|\epsilon_t|$ is very large. In that case, the density function is degenerate with almost all its mass above or below 0. As a result, the sign of the return will be either positive or negative with certainty: the stronger the asymmetry, the better our ability to make a correct directional forecast.

To translate the directional forecasts into a trading strategy, we use the following rule: if the likelihood of a depreciation of the home currency is above .5 (i.e. if $\hat{p}_{t|t-1} > .5$), the investor takes a short position or own the foreign currency. On the contrary, if the likelihood of an appreciation is above .5 (i.e. if $\hat{p}_{t|t-1} < .5$), the investor takes a long position in the home currency, i.e. own USD. In the case of a constant asymmetry and negligible mean dynamics of the conditional density, the best trading strategy is to be either always in a short position (for a negative asymmetry) or always in a long one (for a positive asymmetry).

To assess the quality of these forecasts, we use several measures. First, we use the correct classification rate over h time periods (starting in $t+1$), given by

$$\text{CR} = \frac{1}{h} \sum_{j=t+1}^{t+h} \mathbb{1}(sign(R_j) = \hat{p}_j^*), \tag{21}$$

where $\hat{p}_j^*$ is given by equation (20), $\mathbb{1}(\cdot)$ denotes an indicator function taking value 1 if the condition in parentheses is met, and $sign(.)$ denotes the sign function. This criterion measures the raw performance of a model in a pure classification exercise. To assess the in-sample performance in term of CR, we use the independence test of Pesaran and Timmermann [2009] which accounts for serial correlation.

Second, we use the mean return obtained with our directional forecasts over the same period, and given by

$$\hat{m} = \frac{1}{h} \sum_{j=t+1}^{t+h} \hat{p}_j^* R_j. \tag{22}$$

Diebold and Mariano [1995], Blaskowitz and Herwartz [2011] and Elliott and Timmermann [2016] argue that employing a realized economic value is often more sensible than a statistical

value in evaluating the usefulness of a forecast. In particular, $\hat{m}$ measures the economic usefulness of "being right", i.e. it combines the correct prediction with the timing of a success [Blaskowitz and Herwartz, 2011, 2014]. Hence, if we predict the correct direction of change, we make a gross profit of R_j, whereas a loss of the same amount is suffered if the prediction is wrong. Such a criterion is used throughout the financial literature to assess trading rules, see e.g. Bajgrowicz and Scaillet [2012]. The significance of in-sample performance is assessed with the stepwise-superior predictive ability (SSPA) test of Hsu et al. [2010] to control for data snooping issues.

We compare the performance of the GARCH-SH model to the random walk (RW^+), inverted random walk (RW^-), always-short (AS) and buy-and-hold (BH) approaches, as well as with various sub-specifications of the dynamic skewness model. Random walk (resp. inverted random walk) directional predictions must be understood as predicting tomorrow's direction of change using today's sign (resp. opposite sign) of the return, whereas always-short and buy-and-hold strategies consist in always predicting an appreciation or a depreciation of the home currency, respectively.

In Section 3, we also conduct an out-of-sample analysis of the proposed model. In that cas, the out-of-sample performance is assessed with the Diebold and Mariano [1995] test, the conditional predictive ability test of Giacomini and White [2006] and the fluctuation test of Giacomini and Rossi [2010]. The tests proposed by Giacomini and White [2006] and Giacomini and Rossi [2010] have the advantage of explicitly covering loss functions that are based on direction-of-change, involve estimated parameters, and allow both the comparison of nested and non-nested models. Thus, we can apply them on out-of-sample versions of (21) and (22). The latter test can be seen as testing repeatedly for zero local differences in forecasting performance, using a rolling window of data containing a fraction τ of the total, or as a sequence of DM tests. Both tests, however, are only valid in the cases of either a fixed estimation period or a rolling window estimation period, not in an expanding window context. Therefore, for the evaluation of the out-of-sample performance, we restrict our attention to the rolling window updating scheme of the parameters, and conduct a persistence analysis inspired from Bajgrowicz and Scaillet [2012].

Finally, we also report a series of in-sample and out-of-sample performance measures traditionally used in the exchange rate and trading rule literature: area under the correct classification frontier (AUC) and its return-weighted version (AUC*) of Jordà and Taylor [2012], gain-loss (G/L) ratio of Bernardo and Ledoit [2000], Sharpe ratio[6] and skewness of the daily profit, maximum drawdown on the compounded profits.

Notice here that these performance measures are not used in any way in the estimation procedure. Our model is entirely based on either theoretical or empirical considerations regarding the structure of exchange rate dynamics, but not with the purpose of optimizing directional

[6]For simplicity, we assume a 0% risk-free rate, making the Sharpe ratio equivalent to the coefficient of variation.

forecasts. Thus, the forecasting performance must be seen as a genuine by-product of the correctness of our model.

3 Empirical study

In the present section, we turn to the study of exchange rate dynamics. We first describe the data and discuss the specifications of our models in Sections 3.1 and 3.2, respectively. Before turning to the forecasting exercise in Section 3.3, we discuss the economic interpretation of the fitted models.

3.1 Data

The data are daily foreign exchange rates in U.S. dollar (USD) per unit of foreign currency for EUR and CHF. This choice is motivated by the fact that EUR and CHF are two of the most traded currencies in the world. Moreover, since Hsu et al. [2016]have not been able to find profitable trading rules for developed countries in recent time periods, it seems interesting to challenge this conclusion by using more advanced procedure.

Exchange rates are noon buying rates in New York for cable transfers payable and available from the Board of Governors of the Federal Reserve System. The considered time period ranges from 6 January 1999 to 25 March 2019. We compute the log-rate of change $R_t = \log(S_t/S_{t-1})$, where S_t is the nominal exchange rate at time t. The final samples consist of 5,074 observations.We removed the dates for which exchange rate data were missing.

Interest rates are 3-month London Inter-Bank Offered Rate (LIBOR) for the respective currencies, following Jordà and Taylor [2012], Ismailov and Rossi [2018] and Du et al. [2018]. All interest rates data have been retrieved from the website of the Federal Reserve of Saint-Louis. Missing LIBOR data are replaced by the previously observed rate. Ismailov and Rossi [2018] argue that the predictability of interest rates also depends on the uncertainty prevailing on financial markets. Therefore, as a robustness check, we consider the VIX as an additional predictor. VIX data are daily closing prices and are provided by the CBOE. Missing data have been replaced by the first prior price available (54 occurrences).

The exchange rates time series and the corresponding log-rate of change are plotted in Figures 3 and 4. The interest rates and the VIX are plotted in Figure 5. Several events such as negative LIBOR rates, the soar of the VIX, the financial crisis or the removal of the CHF floor rate might indicate instabilities in the relationship between exchange rates and IRD (see, e.g., the discussions in Giacomini and Rossi [2010], Bacchetta and van Wincoop [2013] and Ismailov and Rossi [2018]). This first observation motivates us to study the performance of the model over sub-periods of time in Section 3.3.

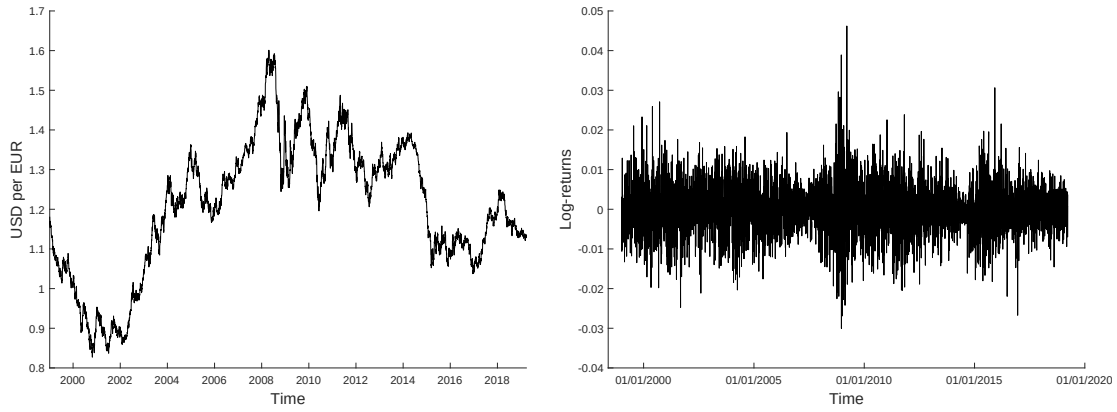


Figure 3: (i) Daily exchange rate of EUR against USD, and (ii) log-returns (right) over the period 6 January 1999 - 25 March 2019.

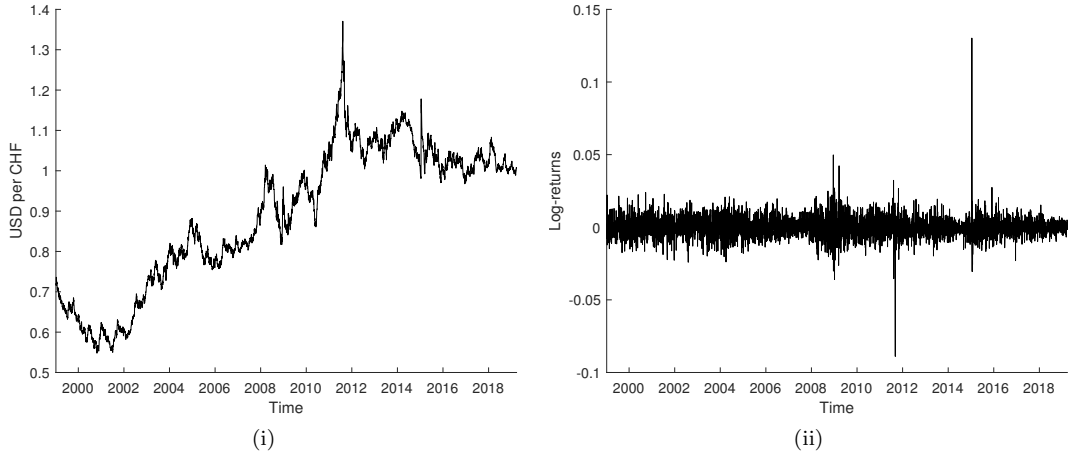


Figure 4: (i) Daily exchange rate of CHF against USD, and (ii) log-returns over the period 6 January 1999 - 25 March 2019.

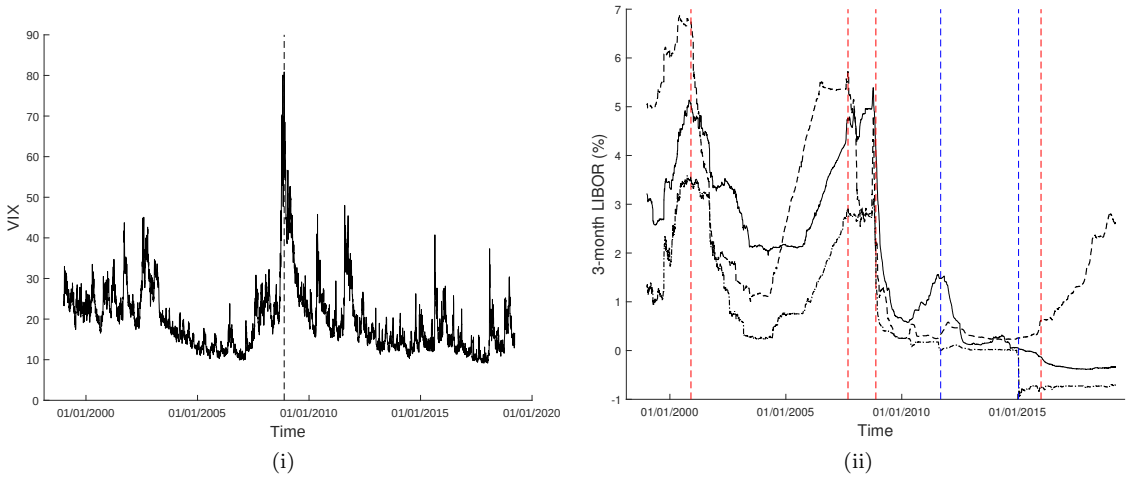


Figure 5: (i) VIX and (ii) 3-month LIBOR rates (EUR: solid, CHF: dashed-dotted, USD: dashed). The dashed vertical lines indicate remarkable events: drop of the Dow Jones index by 445 basis points, dotcom bubble crash, liquidity crisis of 2007, banks bailout of 2008, and the hike of federal fund rate in 2016 (red), establishment and removal of CHF capping (blue).

3.2 Model estimation and economic interpretation

We start by fitting model (1)-(5) to the entire dataset and discussing the economic interpretation of our results. To this end, we consider up to 12 specifications of the general skewness equation given by (7), and focus on the 6 most complex ones (specifications 7 to 12 in Table 2).

Specification number	Name	Equation
(1)	**CST**	$\epsilon_t = a_0.$
(2)	**ARX(VIX)**	$\epsilon_t = a_0 + a_1\epsilon_{t-1} + a_3\text{VIX}_{t-1}.$
(3)	**ARX(IRD)**	$\epsilon_t = a_0 + a_1\epsilon_{t-1} + a_4\text{IRD}_{t-1}.$
(4)	**ARX(2)**	$\epsilon_t = a_0 + a_1\epsilon_{t-1} + a_3\text{IRD}_{t-1} + a_4\text{VIX}_{t-1}.$
(5)	**MA**	$\epsilon_t = a_0 + a_2 z_{t-1}.$
(6)	**ARMA**	$\epsilon_t = a_0 + a_1\epsilon_{t-1} + a_2 z_{t-1}.$
(7)	**MAX(VIX)**	$\epsilon_t = a_0 + a_2 z_{t-1} + a_3\text{VIX}_{t-1}.$
(8)	**MAX(IRD)**	$\epsilon_t = a_0 + a_2 z_{t-1} + a_4\text{IRD}_{t-1}.$
(9)	**MAX(2)**	$\epsilon_t = a_0 + a_2 z_{t-1} + a_3\text{IRD}_{t-1} + a_4\text{VIX}_{t-1}.$
(10)	**ARMAX(VIX)**	$\epsilon_t = a_0 + a_1\epsilon_{t-1} + a_2 z_{t-1} + a_3\text{VIX}_{t-1}.$
(11)	**ARMAX(IRD)**	$\epsilon_t = a_0 + a_1\epsilon_{t-1} + a_2 z_{t-1} + a_4\text{IRD}_{t-1}.$
(12)	**ARMAX(2)**	$\epsilon_t = a_0 + a_1\epsilon_{t-1} + a_2 z_{t-1} + a_3\text{IRD}_{t-1} + a_4\text{VIX}_{t-1}.$

Table 2: Tested specifications of the skewness equation.

The variable IRD_t is defined as $\text{LIBOR}_t^{Home} - \text{LIBOR}_t^{Foreign}$, so that positive (resp. negative) values correspond to situations where the home currency is the investment (resp. funding) currency. We always choose USD as the home currency. For the mean equation, we assume a GARCH-in-Mean process such that

$$\mathbb{E}_t(R_t) = c + \lambda_t \sigma_t.$$

In Tables 3 and 4, we report the estimated coefficients of specifications 7 to 12, for the two currencies. For EUR, using LR tests against simpler nested specifications (Table 5), we find ARMAX(2) to be superior to all considered alternatives. AIC and BIC also suggest that a model with IRD as the only predictor is preferable to a model solely driven by the VIX, and that the autoregressive component is probably not necessary since MAX(IRD) is the best model on the BIC criterion. Thus, the presence of IRD in the skewness specification seems necessary to provide a good fit, but it is well complemented by the informational content of the VIX. Looking at the QQ-plots of the pseudo-residuals for ARMAX(2), we observe an excellent fit (Figure 6, left panel). Following Rossi and Sekhposyan [2014], we also report the results of Berkowitz [2001], Doornik and Hansen [2008] and Anderson-Darling specification tests (BK, DH and AD hereafter). Whereas the test of Berkowitz [2001] focuses on detecting jointly departures from the expected mean, variance and independence properties of the residuals, Doornik and Hansen

[2008] and Anderson-Darling tests target skewness, kurtosis and extreme values of the residuals' distribution. The three tests indicate a good fit of the various SH-GARCH models. Applying the same tests to the residuals of a Gaussian GARCH-in-mean model, DH and AD tests reject the hypothesis of a correct specification (Appendix D). Lastly, we repeat the analysis conducted in the introduction: we compute the empirical skewness of subgroups of residuals obtained from ARMAX(2) and pooled by IRD levels (Figure 7). We find skewness levels close to zero for every IRD level. These results suggest that a dynamic asymmetry component is needed to correctly model the data. In the next section we concentrate on the economic interpretation of the ARMAX(2) model. Estimated conditional volatility, skewness and kurtosis for this model are displayed in Figure 9, upper panel.

For CHF, ARMAX(2) is found to be significantly superior to simpler alternatives using likelihood ratio test, with the exceptions of ARMAX(IRD), MAX(2) and MAX(IRD). However, in terms of AIC and BIC, MAX(IRD) is found to be superior to all models tested, whereas the models based solely on the VIX are dominated by their IRD alternatives. These results again suggest the importance of IRD to model the skewness, but also that, for CHF, the VIX and an autoregressive component are likely superfluous predictors. QQ-plots of the residuals for MAX(IRD) are satisfactory (Figure 6, right panel). Repeating the preliminary analysis based on empirical skewness of the residuals for MAX(IRD), we find skewness levels close to zero for every IRD level and much smaller than those obtained with a GARCH(1,1) model (Figure 8). We also conduct BK and AD tests, and we do not reject the null hypothesis. On the contrary, DH test rejects a correct specification. After a graphical inspection, this results appears to be strongly driven by a single date, namely the end of CHF ceiling on 15 January 2015. Estimated conditional volatility, skewness and kurtosis for MAX(IRD) can be found in Figure 9, bottom panel. On 15 January 2015, we observe a dramatic surge of the conditional skewness and kurtosis, taking respectively values of 476 and 8,679 at that date (for clarity, the scale of the y-axis is not adjusted for these points). To test for a structural break, we adapt the CUSUM test of Kulperger and Yu [2005] and Andreou and Ghysels [2002] to our model [7], and endogenously date a structural break in the second moment on 6 September 2011, corresponding to the introduction of the CHF ceiling by the Swiss national bank (Figure 10). Thus, although the fit of the MAX(IRD) model appears reasonably good along most considered metrics, these results serve as a motivation for studying subperiods of time in our forecasting exercise.

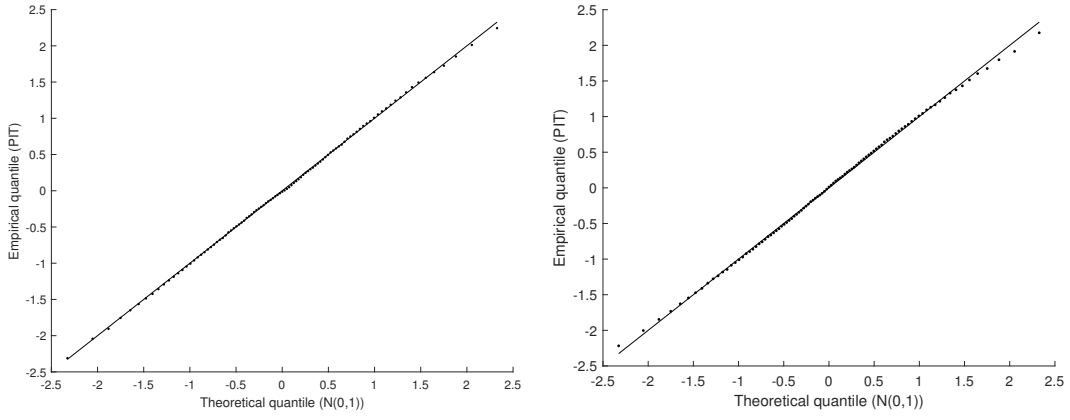
Regarding the volatility process, the estimated coefficients are in line with common features of GARCH-type models: a high persistence of the volatility process (close to an integrated process); stochastic shocks exhibiting excess kurtosis as indicated by $\delta < 1$, and the constant in the mean equation that is close to zero. Moreover, we observe a negative but insignificant mean parameter λ for the volatility component. This suggests that, all else equal, an increase

[7] See Appendix C for technical details.

in contemporaneous exchange rate volatility, on average, translates into an appreciation of USD against the foreign currency. These results suggest the existence of a small positive time-varying premium for investors in dollars and is in line with the observed appreciation of USD during volatile periods, such as the last financial crisis [Habib and Stracca, 2012].

USD/EUR (1999M1 - 2019M3)											
Model	ω	α	β	c	λ	δ	a_0	a_1	a_2	IRD	VIX
ARMAX(2)	0.000	0.029	0.968	0.000	-0.032	0.772	-0.043 *	0.631 ***	0.042 ***	1.520 *	0.171 *
	(0.699)	(0.177)	(0.007)	(0.000)	(0.052)	(0.022)	(0.025)	(0.184)	(0.014)	(0.813)	(0.101)
ARMAX(IRD)	0.000	0.029	0.968	0.000	-0.054	0.769	-0.007	0.589***	0.042***	1.205*	-
	(0.615)	(0.166)	(0.006)	(0.000)	(0.052)	(0.022)	(0.007)	(0.199)	(0.014)	(0.679)	-
ARMAX(VIX)	0.000	0.03	0.968	0.000	-0.025	0.771	-0.025	0.543***	0.045***	-	0.10
	(0.628)	(0.164)	(0.006)	(0.000)	(0.052)	(0.022)	(0.019)	(0.198)	(0.014)	-	(0.079)
MAX(2)	0.000	0.029	0.968	0.000	-0.027	0.772	-0.103***	-	0.045***	3.821***	0.401**
	(0.435)	(0.132)	(0.006)	(0.000)	(0.054)	(0.022)	(0.038)	-	(0.014)	(1.058)	(0.160)
MAX(IRD)	0.000	0.029	0.968	0.000	-0.047	0.770	-0.016	-	0.043***	2.805***	-
	(0.435)	(0.132)	(0.004)	(0.000)	(0.054)	(0.022)	(0.015)	-	(0.014)	(0.978)	-
MAX(VIX)	.000	0.029	0.968	0.000	-0.023	0.772	-0.045	-	0.046***	-	0.184
	(0.435)	(0.132)	(0.004)	(0.000)	(.053)	(.022)	(0.034)	-	(0.014)	-	(0.147)

Table 3: For EUR: estimated coefficients for different skewness models. a_0, a_1, a_2 are the constant, AR and MA parameters in the skewness equation (7), respectively. IRD and VIX refer to the estimated parameters of the corresponding predictors. Standard errors are put in parentheses. *, ** and *** denote Wald tests significant at the 10%, 5% and 1% test level.



(i) (ii)

Figure 6: QQ-plot of the residuals for (i) ARMAX(2) (EUR), and (ii) MAX(IRD) (CHF), fitted on the complete period.

3.2.1 Effect of IRD

First, we examine the link between IRD and the conditional distribution of the depreciation rate, captured by a_3. In particular, we look at the marginal effect of a change in IRD on the

USD/CHF (1999M1 - 2019M3)											
Model	ω	α	β	c	λ	δ	a_0	a_1	a_2	IRD	VIX
ARMAX(2)	0.000	0.037	0.959	0.000	-0.045	0.687	-0.021	-0.227	0.041***	4.304***	0.094
	(0.253)	(0.098)	(0.004)	(0.000)	(0.048)	(0.017)	(0.043)	(0.218)	(0.013)	(1.341)	(0.161)
ARMAX(IRD)	0.000	0.037	0.960	0.000	-0.049	0.687	0.000	-0.220	0.041***	4.180***	-
	(0.252)	(0.098)	(0.04)	(0.000)	(0.048)	(0.017)	(0.022)	(0.216)	(0.013)	(1.304)	-
ARMAX(VIX)	0.000	0.036	0.960	0.000	-0.032	0.700	0.061	-0.235	0.040***	-	-0.008
	(0.252)	(0.098)	(0.04)	(0.000)	(0.047)	(0.017)	(0.039)	(0.211)	(0.013)	-	(0.158)
MAX(2)	0.000	0.037	0.959	0.000	-0.046	0.686	-0.017	-	0.040***	3.524***	0.072
	(0.253)	(0.098)	(0.004)	(0.000)	(0.048)	(0.017)	(0.035)	-	(0.013)	(0.901)	(0.131)
MAX(IRD)	0.000	0.037	0.959	0.000	-0.050	0.686	-0.000	-	0.040***	3.448***	-
	(0.253)	(0.098)	(0.004)	(0.000)	(0.047)	(0.017)	(0.018)	-	(0.013)	(0.890)	-
MAX(VIX)	0.000	0.037	0.959	0.000	-0.033	0.686	0.050	-	0.039***	-	-0.012
	(0.253)	(0.098)	(0.004)	(0.000)	(0.047)	(0.017)	(0.031)	-	(0.013)	-	(0.129)

Table 4: For CHF: estimated coefficients for different skewness models. a_0, a_1, a_2 are the constant, AR and MA parameters in the skewness equation (7), respectively. IRD and VIX refer to the estimated parameters of the corresponding predictors. Standard errors are put in parentheses. *, ** and *** denote Wald tests significant at the 10%, 5% and 1% test level.

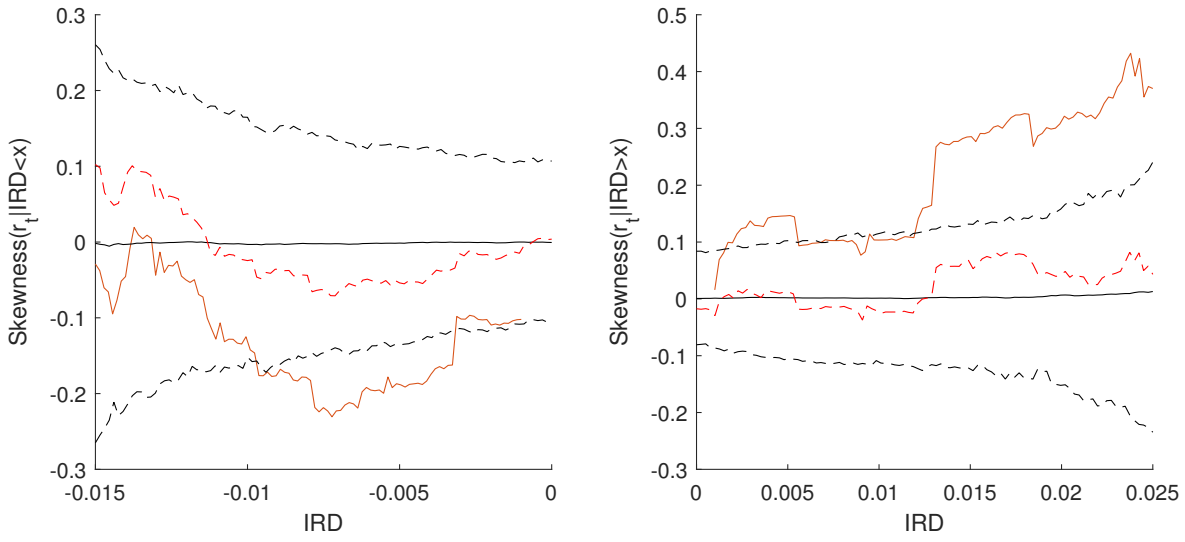


Figure 7: Empirical skewness of USD/EUR residuals, conditional on observing IRD smaller (resp. larger) than a given IRD level. Solid red: residuals of a GARCH(1,1) model. Dashed red: residuals of ARMAX(2).

EUR	ARMAX(2)	ARMAX(IRD)	ARMAX(VIX)	MAX(2)	MAX(IRD)	MAX(VIX)
LLF	-19,026.94	-19,022.82	-19,019.48	-19,024.26	-19,020.47	-19,017.16
AIC	-38,031.88	-38,025.64	-38,018.96	-38,028.52	-38,022.94	-38,016.32
BIC	-37,960.02	-37,960.32	-37,953.64	-37,963.20	-37,964.14	-37,957.53
LR	-	8.24***	14.92***	5.36**	12.94***	19.56***
BK	0.702	0.758	0.840	0.252	0.735	0.770
DH	2.885	3.030	3.217	2.758	2.895	3.019
AD	0.353	0.371	0.328	0.369	0.339	0.357

CHF	ARMAX(2)	ARMAX(IRD)	ARMAX(VIX)	MAX(2)	MAX(IRD)	MAX(VIX)
LLF	-18,670.06	-18,669.88	-18,662.59	-18,669.52	-18,669.38	-18,661.98
AIC	-37,318.11	-37,319.77	-37,305.18	-37,319.05	-37,320.75	-37,305.97
BIC	-37,246.26	-37,254.44	-37,239.86	-37,253.73	-37,261.96	-37,247.18
LR	-	0.35	14.94***	1.07	1.37	16.15***
BK	3.32	3.36	3.06	3.39	3.41	3.13
DH	18.63***	18.74***	16.57***	18.95***	19.04***	16.90***
AD	1.08	1.09	1.07	1.12	1.12	1.11

Table 5: Model selection and specification criteria. LLF denotes the value of the negative log-likelihood function. The line *LR* displays the likelihood ratio test statistics between ARMAX(2) and the competing models. The lines labelled BK, DH and AD report the test statistics for Berkowitz [2001], Doornik and Hansen [2008] and Anderson-Darling tests, respectively. *** denote tests significant at the 1% level.

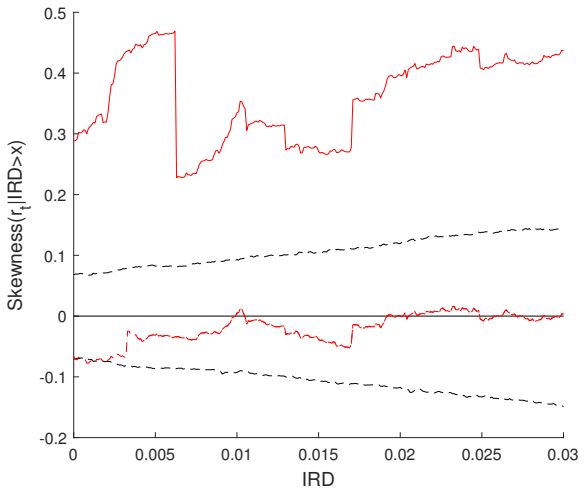
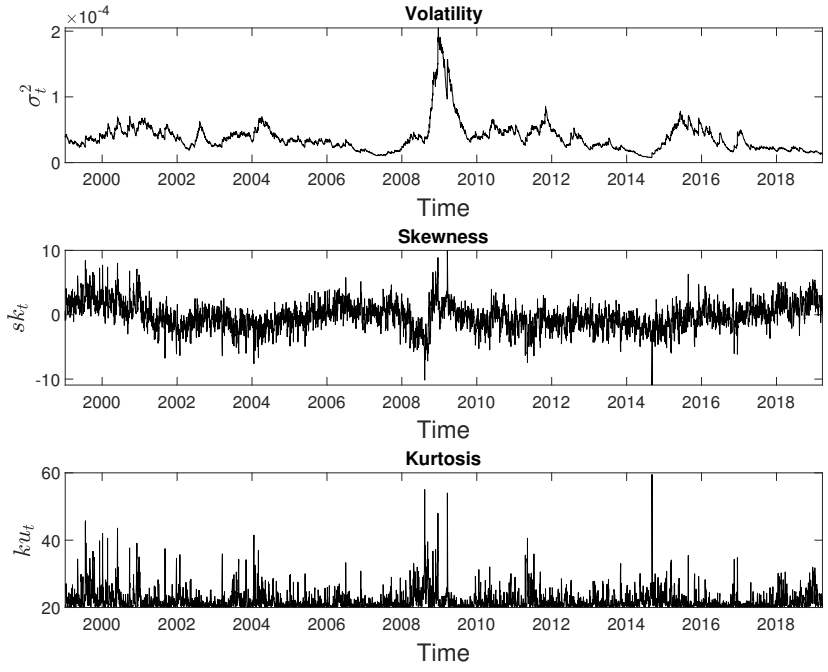
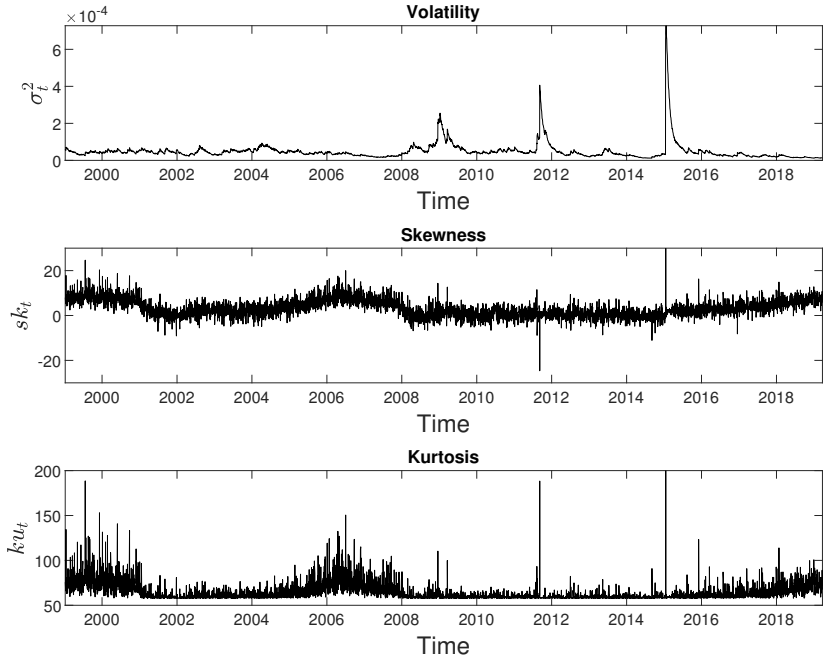


Figure 8: Empirical skewness of USD/CHF residuals, conditional on observing IRD larger than a given IRD level. Solid red: residuals of a GARCH(1,1) model. Dashed red: residuals of ARMAX(2). Dashed black: 95% bootstrap confidence intervals for the GARCH(1,1) model. Contrary to USD/EUR, we observe mainly positive IRD values, thus we do not measure the empirical skewness of the residuals for IRD values *smaller* than a given IRD level.



(i) USD/EUR (ARMAX(2))



(ii) USD/CHF (MAX(IRD))

Figure 9: Conditional variance, skewness and kurtosis for EUR (top) and CHF (bottom).

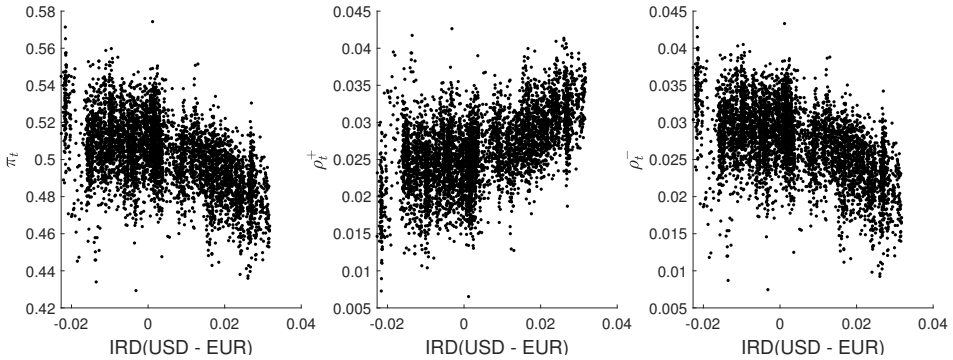
Figure 10: CUSUM process over time, CHF, with $k = 2$ (solid). Dashed red: break date (6 September 2011) identified with the algorithm of Inclan and Tiao [1994]. Dashed black: rejection threshold of the test.

probability of a depreciation of USD (measured by π_t), and on the crash risk (measured by ρ_t^+ and ρ_t^-, ρ_t^+ measuring a sudden depreciation and ρ_t^- a sudden appreciation of USD, see Table 1). We find a positive and significant effect (at the 10% level for EUR) of IRD for both currencies. This result implies that an increase in USD (resp. foreign) interest rates is associated with an increase in the probability of an appreciation (resp. depreciation) of USD. It suggests that a large and positive IRD opens the possibility for profitable carry trades, whereby USD is the investment currency. In this set-up, market participants could borrow the foreign currency at a small rate, buy USD, invest them at a higher rate and still expect an appreciation of USD. These results fit into the theoretical framework of Fahri and Gabaix [2016], where the currency of the country with high interest rates appreciates, conditional on no disaster occurring.

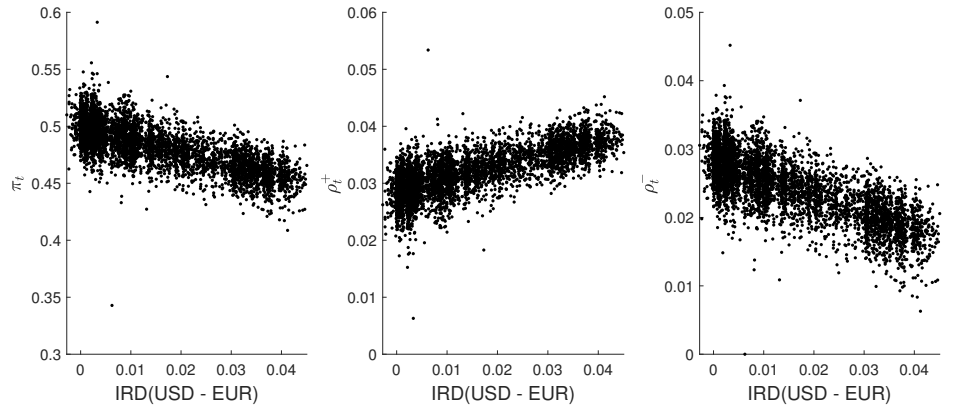
At the same time, though, an increase in (positive) IRD has an opposite effect on ρ_t^+: it becomes more likely to observe an extremely positive shock, synonymous with a large depreciation of USD[8]. This observation is in line with Fahri and Gabaix [2016] and Jurek [2014] who associate IRD with currency crash risk: the larger the IRD, the stronger the realignment pressures. Hence, we are more likely to observe a reverting move or a crash on the exchange rate market. A potential explanation for this effect is the increasing share of market participants involved in carry trades when IRD are large [Brunnermeier et al., 2008]: the larger the IRD, the more carry trade investors fear realignments of the exchange rate. As a consequence, they might unwind their positions in the investment currency, leading to abrupt appreciations of the funding currency. This result also highlights the potential endogeneity of the reverting mechanism, as suggested in Fahri and Gabaix [2016]: out of fear, investors turn themselves into a force of realignment that leads to a crash.

[8]A similar reasoning holds if the foreign currency is the funding currency, leading to a increase in ρ_t^- when IRD becomes more negative.

Figure 11 shows the empirical connection between IRD and the probabilities π_t, ρ_t^+ and ρ_t^- for EUR and CHF, as unraveled by the models: the larger IRD, the smaller the probability of an overall depreciation, but the larger the probability of an extreme depreciation.



(i) USD/EUR (ARMAX(2))



(ii) USD/CHF (MAX(IRD))

Figure 11: Standardized effects of IRD on depreciation and currency crash risks of EUR, measured by π_t, ρ_t^+ and ρ_t^- with $q^+ = 2$ and $q^- = -2$ (from left to right). Estimates are obtained from the ARMAX(2) model for EUR (top) and MAX(IRD) for CHF (bottom).

3.2.2 Effect of past unexpected shocks

We now look at the effect of past innovations on the asymmetry. We find a_2 to be positive and significant for all specifications and both currencies, with similar magnitudes. Hence, past positive shocks have a positive effect on the likelihood of an appreciation of USD, but also a positive effect on large depreciation. In other words, the larger an unexpected depreciation on one day, the more likely the appreciation on the next day *on average* but also the higher the likelihood *of a very large depreciation*. We suggest that this effect is connected to the existence of self-fulfilling mechanisms, as found by Habib and Stracca [2012]. According to these authors, exchange rates fluctuate around some equilibrium value. As a result, unexpected depreciation is followed by appreciation periods. However, large unexpected depreciation may lead more and more economic agents to believe into a future depreciation and to short USD, thus increasing

the risk of a sudden USD crash. If this phenomenon takes place at a time of high volatility, shocks will be amplified, leading to stronger crashes. Such a mechanism is consistent not only with *clustered volatility*, but also with *clustered signs* or periods of time where several large crashes *in the same direction* (depreciation or appreciation) take place.

3.2.3 Effect of uncertainty

Besides IRD, we include the VIX in the model as an additional predictor and proxy for global uncertainty. The corresponding coefficient a_4 is found to be positive for the two currencies, although much larger for EUR and not significant for CHF. Thus, for high values of the VIX, an appreciation of USD against EUR becomes more likely. Simultaneously, though, the likelihood of a currency crash increases as well. An appreciation of USD over other currencies in time of financial stress (as captured by the VIX) is in line with several findings related to safe heaven currencies and funding liquidity constraints: Habib and Stracca [2012] find that the larger the size of the economy, relative to world GDP, the higher the currency excess returns in times of financial stress. Hui et al. [2011] highlight the role played by the USD funding needs of European banks during the crisis. However, our results suggest that this appreciation comes at the price of a larger reversal risk. This is consistent with e.g. Bekaert et al. [2013] who show that high values of the VIX (in its uncertainty component) forecast short-term laxer monetary policy in the US, synonymous with high risk taking in that economy. For CHF, the interpretation is similar, although likelihood ratio tests do not suggest that the VIX is a significant predictor.

3.2.4 Autoregressive dynamics

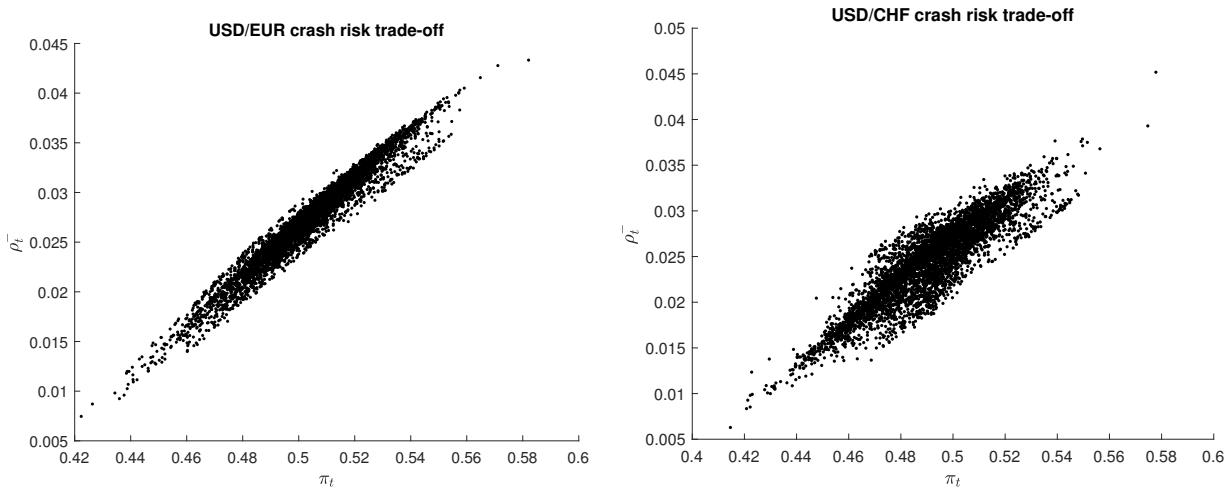
Finally, we find a positive and significant autoregressive for EUR. This implies a high persistence of asymmetries for this currency: when an increase (decrease) of the asymmetry takes place, a long period of time is needed to return to pre-shock levels. This persistence can be a source of sign correlation of the returns since, ceteris paribus, it implies time periods of clustered asymmetry. For CHF, we observe a negative, smaller and insignificant estimated coefficient, whereas likelihood-ratio tests do not suggest the existence of an autoregressive dynamic. This result implies less persistence of the asymmetry.

3.2.5 Summary

From our analysis, we draw three main conclusions: first, the larger the difference between interest rates, the more likely the high-yield currency is to appreciate but also to experience a currency crash. However, this "local" higher appreciation of the high-yield currency comes at the price of a larger likelihood of a currency crash risk (i.e. a large depreciation) for the same currency. Second, we observe that a large unexpected depreciation (resp. appreciation) makes an appreciation (resp. depreciation) more likely the next day, but is also associated

with an increase in the likelihood of an extreme depreciation (resp. appreciation), suggesting the existence of a self-fulfilling mechanism. Finally, an increase in global uncertainty or risk aversion, as measured by the VIX, is positively associated to a higher likelihood of appreciation for USD against the EUR, although, again, it comes at the price of an increasing risk of an extreme depreciation. These three effects combined lead to a crash-risk trade-off: the larger the probability of an appreciation, the more likely we are to suffer an extreme depreciation (ceteris paribus). Figure 12 illustrates this result, displaying the risk of an extreme depreciation shock (ρ_t^-) suffered by the foreign currency as a function of the probability of appreciation of the foreign currency (π_t) vis-a-vis the US Dollar. We observe a clear, upward relationship for both currencies.

Notice that the USD/EUR pair displays the strongest dynamics, with all components of the skewness equation found to be large and significant. The results of the likelihood-ratio tests hint at a more parsimonious specification (MAX(IRD)) for CHF. In the next section, we conduct our forecasting exercise focusing on the ARMAX(2) model for EUR, and on the MAX(IRD) model for CHF.



(i) USD/EUR (ARMAX(2)) (ii) USD/CHF (MAX(IRD))

Figure 12: Crash-risk trade-off between the probability of an appreciation (π_t) and the probability of an extreme depreciation (ρ_t^-) for the foreign currency.

3.3 Forecasting with dynamic asymmetry

We now investigate the extend to which the proposed modeling approach translates into economically and statistically significant forecasts. We aim at answering the following questions: could an investor have correctly guessed the direction of change of exchange rates, using the suggested framework? Is the dynamic skewness, and in particular the effect of IRD, sufficiently strong to be exploited and to yield an investment strategy with positive returns?

3.3.1 In-sample performance

We first consider the in-sample performance. That is, we fit the different models on the complete dataset, and use the estimated parameters to generate forecasts of the one-step-ahead predicted probabilities of an increase in exchange rate i.e. $\{\hat{p}_{t|t-1}\}_{t=2,\dots,T}$. If $\hat{p}_{t|t-1}$ is above .5, we forecast a depreciation of USD in t (thus, an appreciation for EUR or CHF).

EUR In-sample performance

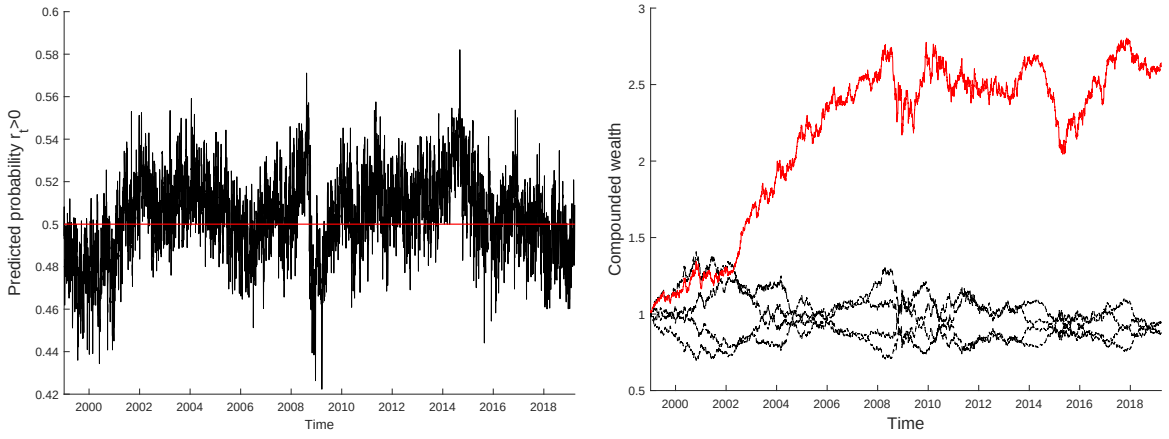
We first consider the in-sample performance for EUR. Figure 13, left panel shows the one-step-ahead predicted probabilities $\{\hat{p}_{t|t-1}\}_{t=2,\dots,T}$ obtained with ARMAX(2) (specification (12)). For ARMAX(2), we observe a correct classification rate of 52.83% over the complete period (Table 6). Using the test of Pesaran and Timmermann [2009], we reject the null hypothesis of independence between the realized signs of the variations and our forecasts[9]. This is also the case for the ARMAX(IRD), MAX(2), MAX(IRD) and MAX(VIX). Looking at $\hat{m}$, we are able to derive an average return[10] of 5.45%. On Figure 13, right panel we display the cumulative wealth obtained by using our directional forecasts. We achieve a performance which is substantially superior to the naive benchmark strategies, i.e. RW^+, RW^-, AS and BH whose performances in equivalent yearly log-returns range between -.22% and .22%. Moreover, compared to the simpler specifications tested, we also achieve a higher performance on the $\hat{m}$ criterion. In particular, a model with no dynamic skewness (speficiation (1)) only achieves an average return of .42%. Applying the SSPA test of Hsu et al. [2010] to test simultaneously the performance of all alternative models and benchmarks, we find $\hat{m}$ to be significantly different from 0 at the 10% test level (p-val. = .083, column *Full SSPA*, Table 6). ARMAX(2) is the only model that passes the test (however, at a high confidence level). Since predictive ability tests are known for their rapid loss of power when the number of alternative models is large, we repeat the procedure only retaining specifications 9 to 12. ARMAX(2) is now found having $\hat{m}$ different from 0 at the 5% test level (column *Int. SSPA*), whereas no changes are observed for the other models. Finally, we test one by one if the average difference in profit of a given model is significantly different from 0 or from the random walk benchmarks. We reject this hypothesis for ARMAX(2) (columns labeled *p-val.*). P-values are obtained from the same bootstrap techniques as the one used in the SSPA test. In Table 7, we report additional performance indicators. They confirm the good performance of ARMAX(2) in terms of Sharpe ratio, AUC, gain-loss ratio and maximum drawdown. The main conclusions, therefore, are that IRD conveys valuable information for forecasting exchange rate depreciation periods. However, one needs to take into account financial uncertainty to obtain a superior profitability.

[9]Notice that the overall proportion of positive returns is 49.14%. Therefore, we exhibit a correct classification rate, compared to a naive strategy that always predicts an appreciation of USD, improved by $52.83/50.86 - 1 = 3.87\%$.

[10]The reported number here is the equivalent yearly rate $\hat{m}^y = (1 + \hat{m}^d)^{252} - 1$.

USD/EUR (1999M1 - 2019M3)										
Specification	CR	PT09	$\hat{m}$	*Full. SSPA*	*Int. SSPA*	*p-val.*	$\Delta\hat{m}$	*Full. SSPA*	*Int. SSPA*	*p-val.*
(12)	52.83%	6.52**	5.45%	0.083	0.039	0.023	5.22%	0.124	0.074	0.022
(11)	52.49%	3.02*	3.49%	0.401	0.257	0.074	3.27%	0.347	0.252	0.094
(10)	51.57%	2.10	1.36%	0.898	0.739	0.295	1.14%	0.746	0.615	0.287
(9)	52.04%	4.15**	3.40%	0.424	0.275	0.099	3.18%	0.391	0.286	0.110
(8)	52.3%	7.20***	2.17%	0.752	0.551	0.181	1.95%	0.581	0.455	0.193
(7)	51.47%	3.21*	2.69%	0.614	0.419	0.093	2.47%	0.34	0.248	0.087
(6)	50.5%	1.92	-1.30%	1.000	-	0.295	-1.51%	0.973	-	0.287
(5)	51.68%	5.01**	2.52%	0.668	-	0.068	2.29%	0.294	-	0.06
(4)	51.31%	0.18	2.12%	0.763	-	0.261	1.9%	0.700	-	0.259
(3)	51%	0.06	1.39%	0.892	-	0.348	1.17%	0.789	-	0.363
(2)	50.25%	0.01	0.57%	0.973	-	0.459	0.35%	0.881	-	0.464
(1)	49.79%	0.02	0.42%	0.978	-	0.458	0.2%	0.898	-	0.473
RW^+/RW^-	50.7%	0.00	0.22%	0.983	-	0.467	-	-	-	-
BH/AS	50.1%	-	0.19%	0.985	-	0.470	-	-	-	-

Table 6: In-sample forecasting performance for EUR. $*$, $**$ and $***$ indicate tests significant at the 10%, 5% or 1% levels, respectively. $\Delta\hat{m}$ refers to the average *excess performance* over the random walk benchmark.



(i) (ii)

Figure 13: (i) Predicted probability of an appreciation of EUR over time and (ii) Evolution over time of an initial investment of 1 USD with reinvestment of the proceed.

Now, we clearly see from the right panel of Figure 13 that the performance is particularly good before the crisis and after the increase in USD interest rates of December 2016: we average 10.19% and 5.73% over these two periods, respectively, whereas we average -.89% in the intermediate period[11]. These two periods exhibit rather large IRD, whereas between 2009

[11]Exact dates for the computation of the these numbers are the following: 6/01/1999, 23/10/2008, 15/12/2015

USD/EUR (1999M1 - 2019M3)						
Spec.	Sharpe	Skew.	AUC	AUC*	G/L	Max Draw.
(12)	0.56	-0.26	0.54	0.51	1.10	0.26
(11)	0.36	-0.20	0.54	0.51	1.06	0.34
(10)	0.14	-0.24	0.52	0.50	1.02	0.33
(9)	0.35	-0.24	0.53	0.51	1.06	0.34
(8)	0.22	-0.27	0.53	0.51	1.04	0.42
(7)	0.28	-0.16	0.52	0.50	1.05	0.36
(6)	-0.13	-0.05	0.52	0.49	0.98	0.43
(5)	0.26	-0.15	0.52	0.50	1.04	0.34
(4)	0.22	-0.30	0.52	0.51	1.04	0.51
(3)	0.14	-0.22	0.53	0.52	1.02	0.53
(2)	0.06	-0.11	0.51	0.50	1.01	0.34
(1)	0.04	-0.14	0.51	0.51	1.01	0.38
RW^+/RW^-	0.02	-0.18	0.52	0.50	1.00	0.30
BH/AS	0.02	-0.11	-	-	1.00	0.50

Table 7: In-sample performance measures for the trading rules derived from the different models.

and 2016, IRD stays very close to 0. Thus, it is not surprising to observe such results since an absence of differences in interest rates leads to a skewness close to 0, and the exchange rate behaves more like a random walk. Furthermore, we observe a clear trend in the exchange rate data (either appreciation or depreciation) during these periods, whereas the intermediate period is characterized by an absence of clear directionality. We are not able to detect a significant structural break using the CUSUM test described in Appendix C, but a graphical inspection of the CUSUM process in Figure 14 shows rather large instabilities. This motivates us to fit the model to the following three subperiods: 6/01/1999 to 23/10/2008, 24/10/2008 to 15/12/2015, and 16/12/2015 to 25/03/2019. Estimated parameters can be found in Appendix, Table 20. Performance indicators are displayed in Table 8. First, we observe that the performance of ARMAX(2) over the pre-crisis period increases, reaching almost 15% in yearly percentages for $\hat{m}$ and 55% for CR. Signs of the regression coefficients are alike, but the marginal effect of IRD is estimated to be much larger compared to the one obtained on the full sample (3.54 instead of 1.52). Inspecting the other models (Table 8, left panel), we find a similar pattern: models based on IRD exhibit a better performance than those without IRD, and better results compared to those obtained on the full sample. During the crisis period (Period 2), on the contrary, we observe smaller values for $\hat{m}$. Two exceptions are models (8) and (5) (MAX(IRD) and MA), exhibiting a profit of around 8% annually (Table 8, middle panel). The last period is characterized by smaller coefficients of IRD and the VIX, which are not found significant.

and 25/03/2019.

The economic performance is positive ($\hat{m} = 1.14\%$) but not significantly different from zero (Figure 15). Looking at the performance of alternative models (Table 8, right panel), we reach a similar conclusion: no model delivers a profit significantly different from 0, although most of them exhibit a positive one. Model (2) (ARX(VIX)) reaches 6% annually, but we cannot reject the possibility that this result is due to luck. Surely, these differences can be partially explained by smaller samples sizes for periods 2 and 3 compared to period 1. Overall, these results point to the direction of a time-varying predictability of IRD, and of a change in its relationship with exchange rate log-returns over time. This motivates us to use rolling-window estimation in an out-of-sample forecasting exercise in Section 3.3.2.

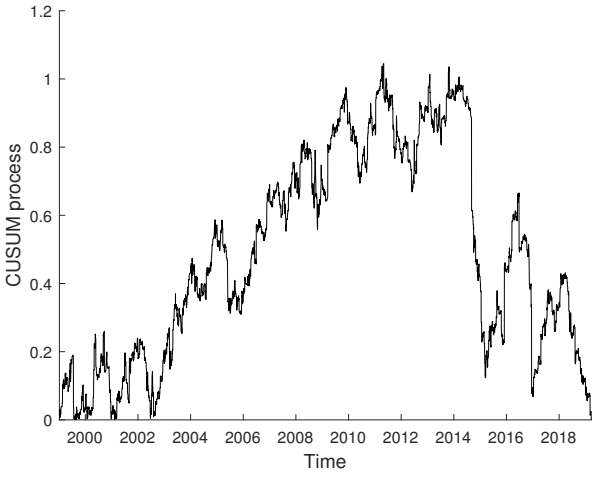


Figure 14: CUSUM process over time, for EUR, with $k = 3$.

USD/EUR (1999M1 - 2019M3)												
	1999M1 - 2008M10 ($n = 2,470$)				2008M10 - 2015M12 ($n = 1,792$)				2015M12 - 2019M3 ($n = 814$)			
Spec.	CR	PT09	$\hat{m}$	*Full SSPA*	CR	PT09	$\hat{m}$	*Full SSPA*	CR	PT09	$\hat{m}$	*Full SSPA*
(12)	54.98%	12.31***	14.50%	0.000	51.25%	0.56	0.64%	0.989	52.09%	0.10	1.14%	0.977
(11)	54.17%	5.75**	12.36%	0.000	50.87%	0.00	0.98%	0.984	53.07%	0.02	2.61%	0.903
(10)	52.72%	3.28*	6.14%	0.121	50.48%	0.00	-1.08%	1.000	52.21%	0.33	2.55%	0.908
(9)	54.66%	10.89***	14.67%	0.000	51.31%	0.27	0.89%	0.986	48.40%	0.00	-0.35%	0.999
(8)	53.44%	3.12*	9.88%	0.007	52.15%	2.48	8.54%	0.180	48.89%	0.66	1.25%	0.975
(7)	52.51%	1.96	4.78%	0.269	51.31%	1.65	0.51%	0.992	49.51%	0.00	1.85%	0.951
(6)	53.04%	12.84***	5.89%	0.142	50.92%	0.00	0.95%	0.985	52.58%	0.09	2.94%	0.874
(5)	52.55%	2.15	4.98%	0.246	52.09%	2.05	7.95%	0.235	49.51%	0.15	1.42%	0.972
(4)	54.17%	4.85**	11.30%	0.001	48.80%	0.20	-1.94%	1.000	48.77%	0.15	2.48%	0.910
(3)	53.32%	3.92**	8.97%	0.015	50.53%	0.88	4.71%	0.670	47.91%	1.02	-2.21%	1.000
(2)	52.31%	0.90	8.11%	0.034	49.81%	0.00	1.04%	0.983	50.25%	0.58	6.00%	0.449
(1)	52.19%	0.50	7.74%	0.041	49.08%	0.74	3.30%	0.842	49.51%	0.15	-0.13%	0.998
RW^+/RW^-	50.65%	0.00	0.01%	0.945	49.97%	0.00	0.16%	0.928	49.88%	0.00	1.84%	0.969
BH/AS	49.19%	-	0.74%	0.886	51.14%	-	2.31%	0.744	48.77%	-	0.83%	0.988

Table 8: Forecasting performance for all models during the 3 subperiods defined by the following dates: 6/01/1999, 23/10/2008, 15/12/2015 and 25/03/2019. PT09 refers to the independence test of Pesaran and Timmermann [2009]

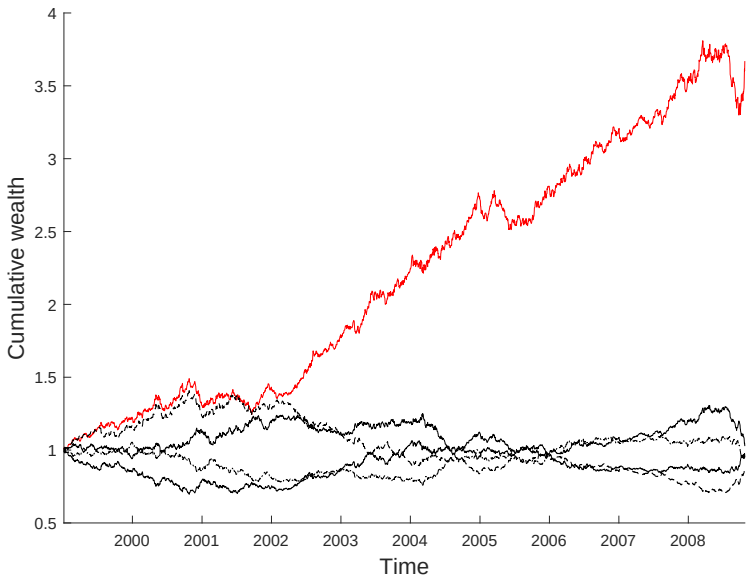


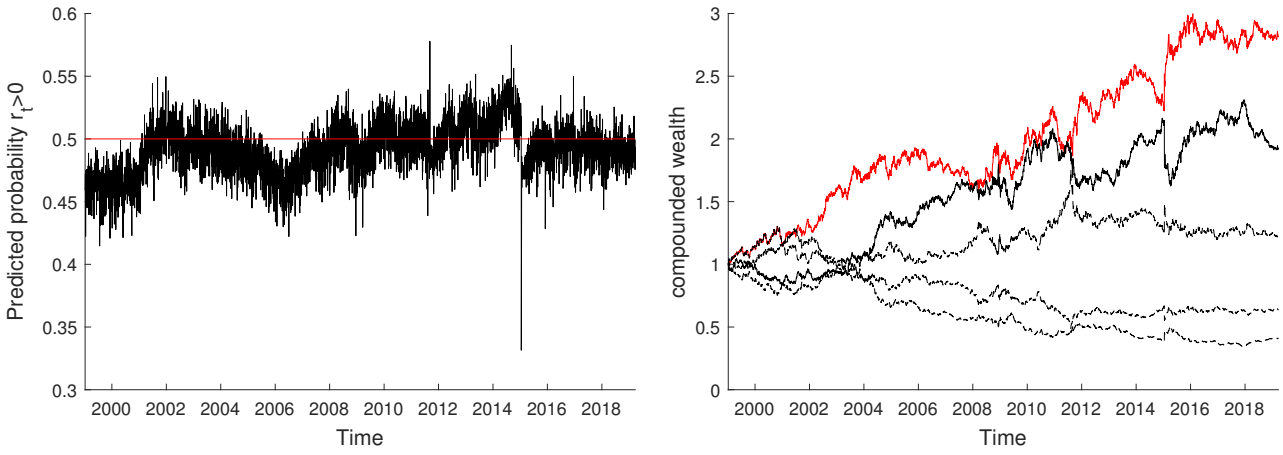
Figure 15: Evolution over time of an initial investment of 1 USD in EUR, with reinvestment of the proceed, for the subperiod 6/01/1999 - 1/12/2008.

CHF In-sample performance

We conduct a similar in-sample analysis for CHF. Figure 16 shows the predicted probability of a CHF appreciation (panel (i)) and cumulated profit (panel (ii)) derived from the MAX(IRD) model. On 16 January 2015, the day after the removal of the CHF capping, we observe a huge drop in our estimate of the appreciation probability for CHF. Correct classification rates and economic performance can be found in Table 9. For the most complex models based on IRD (specification 7 to 12), we systematically reject the independence hypothesis between predicted and realized signs. In particular, for MAX(2) and MAX(IRD) (specification 9 and 8), we register a correct classification rate of 52.32% and 52.21% that translates into average profits $\hat{m}$ of 6.2% and 5.89%, respectively. These results are superior to all the other models, and found to be significantly different from 0 at the 5% and 10% test levels using the most restrictive test (*Full SSPA*). The performance is particularly good before 2006 and after 2008. Another model worth mentioning is ARX(IRD) (specification (3)), recording a performance of 4.64%. Rather surprisingly, both the random walk benchmark RW^- and the simple Gaussian GARCH-in-mean model perform well, reaching average profits of 3.92% and 4.29% despite poor classification rates (see Appendix). Although our reference models have a superior performance, we cannot conclude that they generate significantly more profits than the benchmarks. Looking at alternative performance indicators (Table 10), we find MAX(2), MAX(IRD) and ARX(IRD) to perform remarkably well: they simultaneously exhibit a positive Sharpe ratio and a positive skewness, indicating that extreme trading losses are limited. This is confirmed by a gain/loss ratio superior to 1 and the smallest maximum drawdown among all models.

As discussed in Section 3.2, we probably face a structural break in the data due to the introduction and removal of a CHF capping between September 2011 and January 2015. To

neutralize the effect of these events, we fit the various models on three subperiods: before, during and after the capping period. Performance measures can be found in Table 11. We find a significant predictive content of MAX(2) and MAX(IRD) before the capping period, both in terms of directional forecast and average profit. The average profits of these models reach 11% in yearly equivalent rate. During the capping period, though, no directional forecasting ability is found for any of the models considered. In terms of average profits, the models based on the VIX perform better, with ARMAX(2) and ARMAX(VIX) reaching 10.7% and 12.6%. No clear pattern is observable for the post-capping period, although most models exhibit a positive performance, with ARX(2) registering 6.34% on an annual basis. ARX(IRD) is worth mentioning: it registers a positive performance during the three periods, ranging between 10.52% during the pre-capping period, and 3.22% during the post-capping period. A significant predictive content is found for this model during the more recent period. Overall, these changes in performance over time suggest instabilities in the dynamics, which motivates us to study the out-of-sample performance obtained with rolling-window estimates.



(i) (ii)

Figure 16: (i) One-step ahead predicted probability of an appreciation for CHF. (ii) Evolution over time of an initial investment of 1 USD with reinvestment of the proceed. Red: MAX(IRD). Solid black: RW^-. Dashed black: BH, AS, RW^+.

3.3.2 Out-of-sample performance

To restrict the information set in a more realistic way, we produce out-of-sample directional forecasts using rolling-window estimates of the parameters. We use the period ranging from 6 January 1999 to 1 December 2014 (3999 observations) as initial training sample. Thus, we include both non-crisis and crisis data, as well as parts of the more recent period with low interest rates. For CHF, we mix pre-capping and capping periods. We re-estimate the parameters of the model every 5 days, and predict the direction of change up to March 2019. Overall, we perform one-step-ahead predictions for 1,075 days.

USD/CHF (1999M1 - 2019M3)										
Specification	CR	PT09	$\hat{m}$	*Full. SSPA*	*Int. SSPA*	*p-val.*	$\Delta\hat{m}$	*Full. SSPA*	*Int. SSPA*	*p-val.*
(12)	52.52%	11.52***	3.91%	0.314	0.175	0.059	-0.01%	0.859	0.736	1.000
(11)	52.19%	7.70***	3.27%	0.444	0.264	0.085	-0.64%	0.912	0.815	1.000
(10)	51.16%	1.26	0.84%	0.939	0.713	0.355	-3.07%	0.994	0.973	1.000
(9)	52.33%	9.72***	6.20%	0.043	0.023	0.005	2.28%	0.555	0.421	0.223
(8)	52.21%	7.37***	5.89%	0.060	0.032	0.007	1.97%	0.606	0.456	0.256
(7)	50.99%	0.79	-0.23%	1.000	1.000	1.000	-4.14%	0.999	0.990	1.000
(6)	51.28%	1.81	1.11%	0.913	-	0.302	-2.81%	0.992	-	1.000
(5)	50.87%	0.38	-0.45%	1.000	-	1.000	-4.36%	1.000	-	1.000
(4)	50.91%	0.09	0.19%	0.985	-	0.458	-3.73%	0.993	-	1.000
(3)	51.58%	0.32	4.64%	0.177	-	0.037	0.73%	0.804	-	0.428
(2)	50.06%	1.12	-1.67%	1.000	-	1.000	-5.59%	1.000	-	1.000
(1)	50.12%	1.27	-1.58%	1.000	-	1.000	-5.49%	1.000	-	1.000
RW^+/RW^-	50.87%	0.00	3.92%	0.309	-	0.055	-	-	-	-
BH/AS	49.96%	-	3.72%	0.855	-	0.239	-	-	-	-

Table 9: In-sample forecasting performance for CHF. *, ** and *** indicate tests significant at the 10%, 5% and 1% levels, respectively. $\Delta\hat{m}$ refers to the average *excess performance* over the random walk benchmark.

USD/CHF (1999M1 - 2019M3)						
Spec.	Sharpe	Skew.	AUC	AUC*	G/L	Max Draw.
(12)	0.36	-1.20	0.53	0.53	1.06	0.31
(11)	0.30	-1.24	0.53	0.53	1.05	0.31
(10)	0.08	-1.18	0.52	0.52	1.01	0.31
(9)	0.57	1.53	0.53	0.53	1.10	0.21
(8)	0.54	1.54	0.53	0.53	1.10	0.18
(7)	-0.02	-1.15	0.52	0.52	1.00	0.35
(6)	0.10	-1.18	0.52	0.52	1.02	0.29
(5)	-0.04	-1.15	0.52	0.52	0.99	0.36
(4)	0.02	-1.13	0.53	0.53	1.00	0.45
(3)	0.43	1.59	0.53	0.53	1.08	0.36
(2)	-0.15	-1.11	0.51	0.51	0.97	0.62
(1)	-0.14	-1.11	0.51	0.51	0.97	0.61
RW^+/RW^-	0.36	-1.15	0.52	0.51	1.07	0.25
BH/AS	0.15	1.11	-	-	1.03	0.32

Table 10: In-sample performance measures for the profit of the trading rules derived from the different models.

USD/CHF (1999M1 - 2019M3)												
	1999M1 - 2011M09 ($n = 3,188$)				2011M09 - 2015M01 ($n = 840$)				2015M01 - 2019M3 ($n = 1,045$)			
Spec.	CR	PT09	$\hat{m}$	*Full SSPA*	CR	PT09	$\hat{m}$	*Full SSPA*	CR	PT09	$\hat{m}$	*Full SSPA*
(12)	53.98%	7.61^{***}	11.20%	0.002	51.73%	1.13	10.73%	0.145	51.39%	0.00	-0.88%	0.923
(11)	53.48%	2.71^{*}	9.93%	0.008	51.85%	0.21	5.91%	0.609	51.96%	0.64	2.94%	0.597
(10)	51.41%	0.32	4.49%	0.312	52.80%	1.97	12.63%	0.056	52.82%	0.01	2.99%	0.593
(9)	53.83%	5.45^{***}	11.18%	0.002	50.54%	0.00	3.69%	0.824	52.34%	0.26	3.22%	0.562
(8)	53.76%	6.95^{***}	10.63%	0.003	51.85%	0.49	4.87%	0.720	52.34%	0.10	3.15%	0.574
(7)	51.35%	0.22	4.50%	0.311	52.32%	0.94	8.67%	0.308	52.73%	0.55	2.62%	0.631
(6)	51.51%	0.66	5.43%	0.197	51.97%	0.89	9.01%	0.275	51.87%	1.54	2.08%	0.685
(5)	51.57%	0.73	5.48%	0.191	51.49%	0.26	3.33%	0.847	52.92%	0.15	2.54%	0.640
(4)	53.26%	0.02	9.84%	0.008	52.56%	2.39	5.12%	0.693	53.59%	1.67	6.34%	0.225
(3)	53.42%	0.02	10.52%	0.004	51.49%	1.16	7.85%	0.384	52.54%	5.36^{**}	3.22%	0.562
(2)	50.60%	0.01	0.51%	0.899	51.73%	0.69	7.90%	0.382	53.40%	0.19	4.60%	0.410
(1)	50.53%	0.08	0.83%	0.867	50.77%	0.53	1.76%	0.951	53.59%	0.58	4.58%	0.411
RW^{+}/RW^{-}	51.19%	0.00	5.34%	0.193	51.01%	0.00	4.84%	0.696	49.86%	0.00	1.10%	0.788
BH/AS	49.75%	-	3.85%	0.434	49.70%	-	4.92%	0.687	51.58%	-	2.50%	0.626

Table 11: Forecasting performance for all models during the 3 subperiods defined by the following dates: 6/01/1999, 6/09/2011, 15/01/2015.

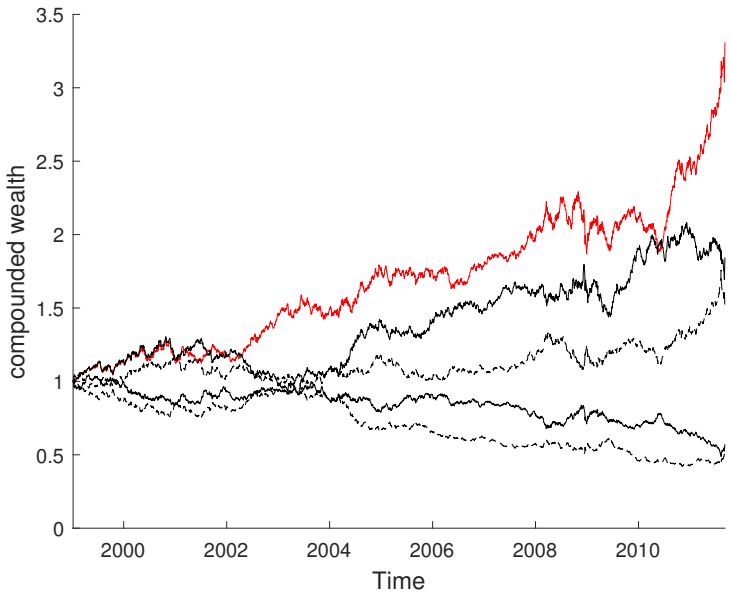


Figure 17: Evolution, over time, of an initial investment of 1 USD in CHF (red), with reinvestment of the proceed, for the subperiod 6/01/1999 - 6/09/2011 (before capping). Black: benchmark strategies (solid: BH and RW^{+}, dashed: AS and RW^{-}).

To account for the effect of choosing the side of the naive benchmark (i.e. if we use the long or short side of the BH/AS and RW^+/RW^- strategies), we use the best performing benchmark during the in-sample period to conduct out-of-sample forecasts. For the BH and AS benchmarks, this is similar to forecasting exchange rates with a momentum strategy using the training data to determine the direction of the momentum. For RW^+/RW^-, this strategy can be seen also as a 1-day momentum strategy.

As described in Section 2.3, we use several tests to analyze the out-of-sample forecasts. To test for significant differences in forecasting abilities, we use a Diebold and Mariano [1995] test. The models are compared using the conditional predictive ability test proposed by Giacomini and White [2006] and the fluctuation test of Giacomini and Rossi [2010]. These tests are denoted by DM, GW and GR, respectively. In particular, the GR test enables us to control for changes in forecasting performance over time, contrary to other tests that only take the average performance into account. We perform the GR test for a grid of value for τ ranging between .1 and .85. Details of the loss functions are given in the next section. We also report out-of-sample values for AUC, AUC*, G/L and Sharpe ratios, skewness and maximum drawdown.

EUR out-of-sample performance

The out-of-sample predicted probabilities of appreciation for EUR, obtained with ARMAX(2) and ARMAX(IRD) are displayed in Figure 18. For ARMAX(2), we obtain an average performance $\hat{m}$ of 4.94% in equivalent yearly rate over the forecast horizon. This is the best result across all tested models (Table 12). The second best performing model is ARMAX(IRD), with a profit of 3.01%. In Figure 19, panel (i), we display the compounded value over time of investing 1 USD at the beginning of the forecasting period trading rules defined by either ARMAX(2) or ARMAX(IRD). Although we register losses at the outset, we rapidly make profit between 2017 and 2018. On the contrary, both benchmark models exhibit a negative performance (-1.26% for the random walk strategy, -2.26% for the momentum strategy). In term of CR, ARMAX(IRD) obtains a correct classification rate of 52.37%, and is the only model with a sign forecasting ability found significant by the test of Pesaran and Timmermann [2009]. Figure 19, panel (ii), illustrates this result by reporting the cumulative profit of a trading rule earning 1 USD if the sign is correctly guessed and losing 1 USD otherwise. Additional performance measures are reported in Table 13. The highest values for AUC, AUC*, the Sharpe and G/L ratios are obtained with ARMAX(2). We also obtain a positive skewness and one of the smallest maximum drawdowns.

Is the out-of-sample profit $\hat{m}$ obtained with ARMAX(2) and ARMAX(IRD) significantly different from 0? To test this hypothesis, we use the GR test for several $\tau \in [.1, .85]$. The loss function used for the test is given by eq. (22). The null hypothesis is rejected at the 5% test level for ARMAX(2), and at the 10% for ARMAX(IRD)(Figure 20). GW and DM tests are inconclusive (Table 12). Looking at significant differences with respect to the RW benchmark,

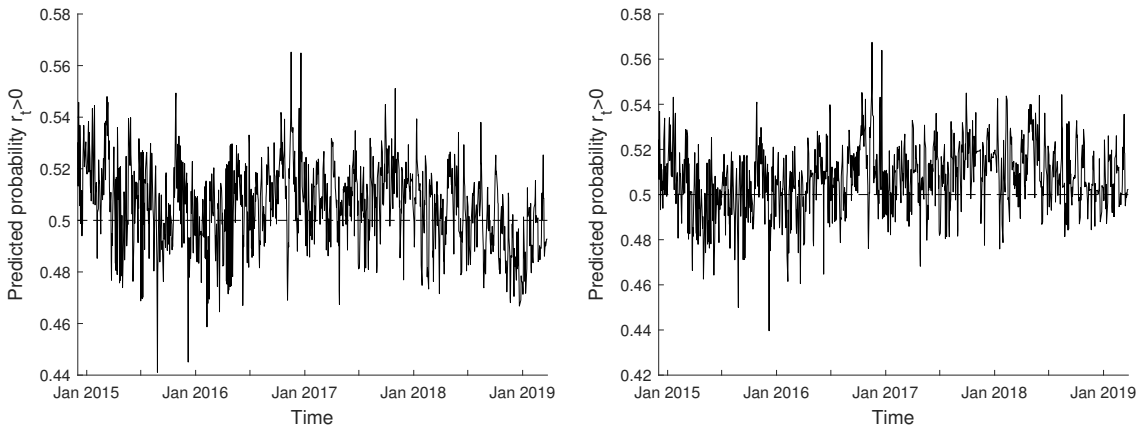


Figure 18: Out-of-sample predicted probability of appreciation for EUR over the period 2014M12 - 2019M3, obtained from (i) ARMAX(2) and (ii) ARMAX(IRD).

USD/EUR (2014M12 - 2019M3)								
Spec.	CR	PT09	$\hat{m}$	DM	GW	$\Delta\hat{m}$	DM	GW
(12)	50.98%	0.15	4.94%	1.15	4.21	6.2%	0.90	2.53
(11)	52.37%	5.11**	3.01%	0.70	0.74	4.27%	0.61	0.54
(10)	50.23%	1.99	1.62%	0.37	0.62	2.88%	0.44	0.97
(9)	49.40%	0.15	-0.60%	-0.14	0.08	0.66%	0.09	0.14
(8)	49.30%	0.02	-2.77%	-0.64	0.53	-1.51%	-0.21	0.05
(7)	47.72%	1.00	-2.89%	-0.67	0.73	-1.63%	-0.24	0.07
(6)	51.16%	2.32	2.27%	0.50	0.25	3.53%	0.49	0.23
(5)	48.47%	0.22	-2.70%	-0.61	0.63	-1.44%	-0.20	0.05
(4)	49.21%	0.18	-1.59%	-0.36	0.16	-0.33%	-0.05	1.43
(3)	48.28%	0.29	-4.18%	-0.99	1.05	-2.92%	-0.50	0.38
(2)	47.72%	0.11	0.37%	0.08	0.00	-1.63%	0.29	0.09
(1)	47.44%	1.63	-3.05%	-0.74	0.85	-1.79%	-0.31	0.23
RW^+/RW^-	48.00%	0.00	-1.26%	-0.34	0.16	-	-	-
BH/AS	47.81%	-	-2.26%	-0.55	0.31	-	-	-

Table 12: Out-of-sample forecasting performance for EUR. ** indicates tests significant at the 5% test levels. $\Delta\hat{m}$ refers to the average *excess performance* over the random walk benchmark.

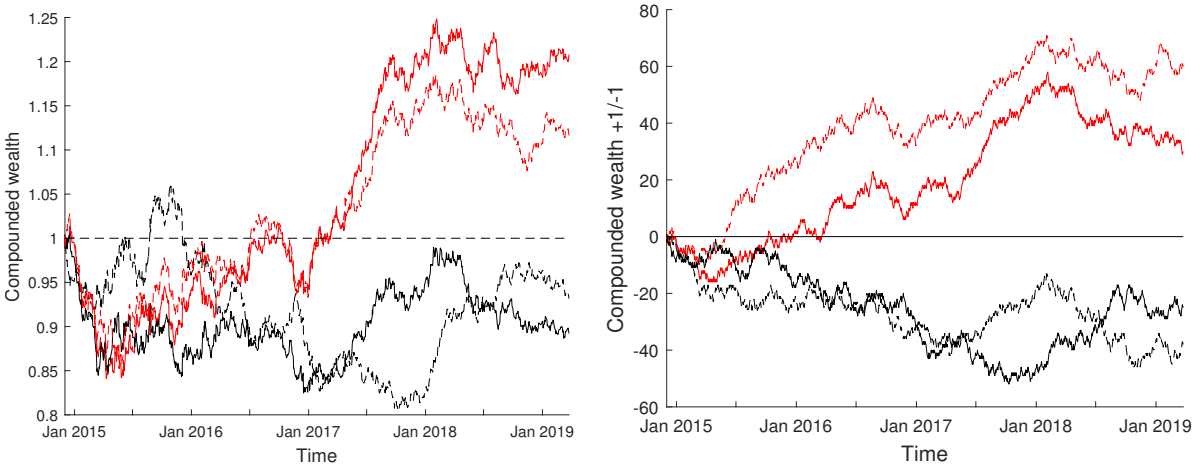


Figure 19: (i) Compounded value of an initial investment of 1 USD in the trading rules derived from ARMAX(2) (solid red), ARMAX(IRD) (dashed red), random walk (dashed black) and momentum (solid black) approaches. (ii) Cumulative profit, over time, of a trading rule based on the different models that earns 1 if the sign is correctly forecast, -1 otherwise.

USD/EUR (2014M12 - 2019M3)						
Spec.	Sharpe	Skew	AUC	AUC*	G/L	Max. Draw.
(12)	0.56	0.04	0.53	0.53	1.10	0.17
(11)	0.34	-0.05	0.53	0.50	1.06	0.17
(10)	0.18	0.10	0.53	0.52	1.03	0.16
(6)	0.26	0.04	0.53	0.51	1.04	0.20
(9)	-0.07	0.05	0.51	0.52	0.99	0.20
(8)	-0.31	0.08	0.52	0.51	0.95	0.23
(7)	-0.33	0.06	0.51	0.51	0.95	0.22
(5)	-0.31	0.05	0.52	0.51	0.95	0.23
(4)	-0.18	-0.10	0.50	0.50	0.97	0.29
(3)	-0.47	-0.10	0.50	0.50	0.92	0.24
(2)	0.04	0.20	0.49	0.50	1.01	0.13
(1)	-0.35	0.09	0.50	0.51	0.94	0.21
RW^+/RW^-	-0.14	-0.28	0.48	0.49	0.98	0.24
BH/AS	-0.26	0.09	-	-	0.96	0.18

Table 13: Out-of-sample performance measures for the profit of the trading rules derived from the different models (USD/EUR).

our loss function becomes

$$L_{\Delta}^{(1)} = \frac{1}{h} \sum_{j=t+1}^{t+h} (\hat{p}_j^* - \hat{p}_j^{*,RW}) R_j,$$

where $\hat{p}_j^{*,RW}$ is the sign forecast obtained from the random walk benchmark. We also find both model to be significantly better over some periods of time (Figure 21). The results for the other specifications (available upon demand) are mostly inconclusive. Finally, repeating the same procedure with the following loss function:

$$L_{\Delta}^{(2)} = \frac{1}{h} \sum_{j=t+1}^{t+h} (\hat{p}_j^* - \hat{p}_j^{*,RW}) \text{sign}(R_j),$$

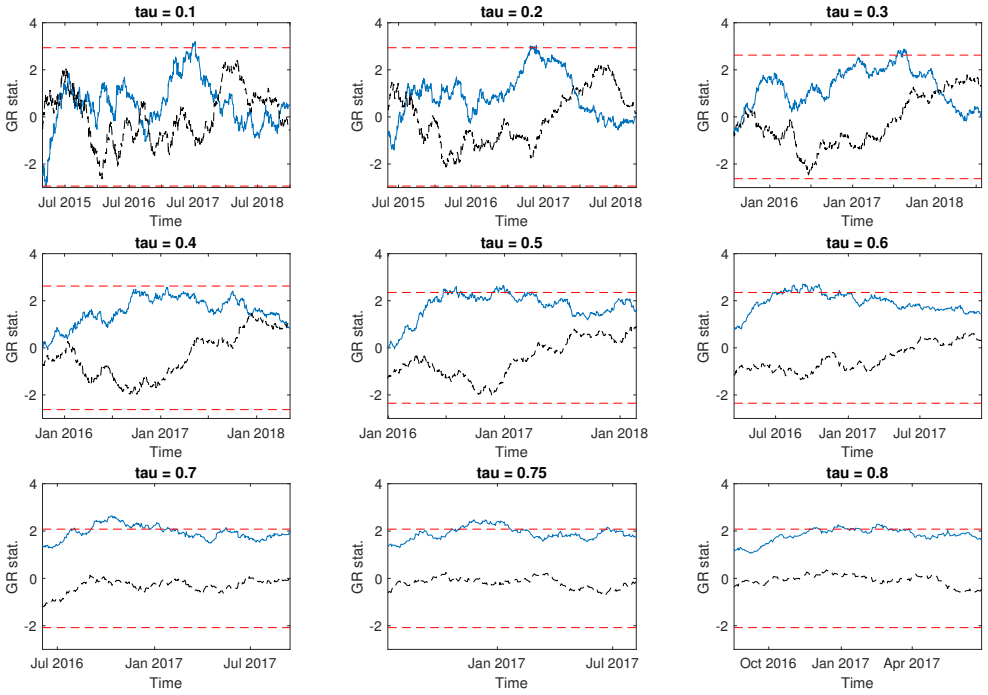
to test for superior sign predictability, we obtain the same results (Figure 22). Being more stringent and using the best (ex-post) BS/AS benchmark[12], i.e. without accounting for in-sample benchmark selection, the results stay unchanged for ARMAX(2) (see Figure 30 in Appendix).

To summarize this section, we conclude that the performance observed in-sample is mostly preserved out-of-sample for the ARMAX(2) model (and less evidently for ARMAX(IRD)). Evidence is less obvious after 2018. Again, it seems that the combination of IRD and VIX carries information about the future direction of change of USD/EUR exchange rate. It also appears that the predictive power is economically significant when IRD exhibits a changing intensity, as over the period 2016 - 2018.

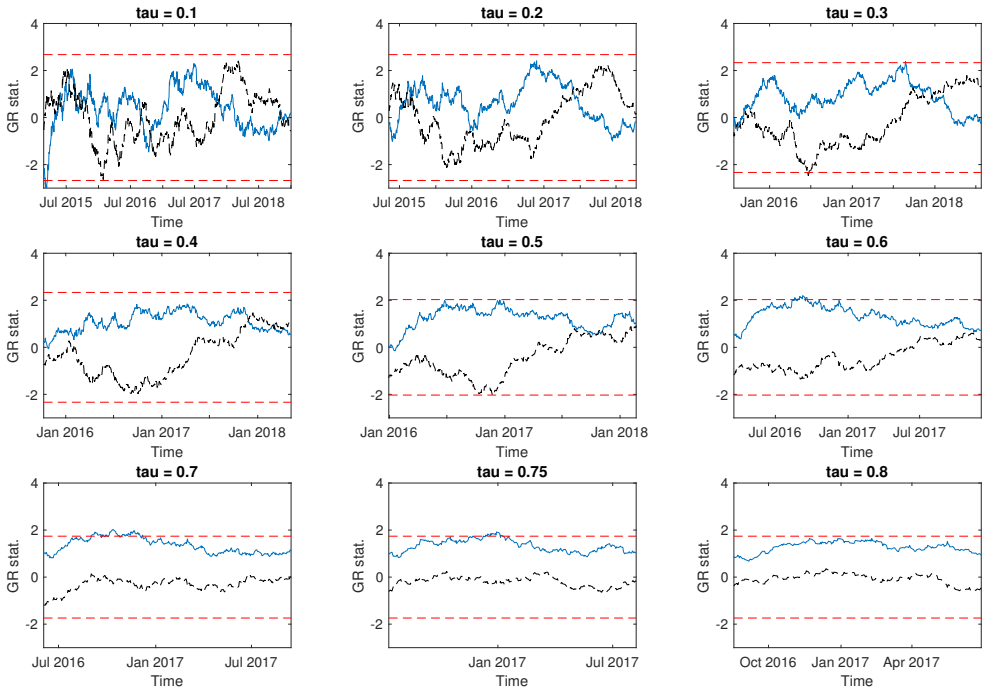
CHF out-of-sample performance

We produce similar out-of-sample forecasts for CHF. The various indicators are reported in Table 14 and 15. As for the in-sample analysis, the models based on IRD or mixing both IRD and VIX seem to perform better than the others. In terms of $\hat{m}$, MAX(2), MAX(IRD), ARX(2) and AR(IRD) exhibit an average profit ranging between 8.37% and 6.22% in equivalent yearly rate. Only AR(IRD) (specification (3)) is found to have a significant $\hat{m} > 0$. ARMAX(2) and ARMAX(IRD) are also found to be significantly better than the random walk benchmark. In terms of sign forecasts, only MAX(VIX), ARMA and MA are found to have a significant directional ability, although without translating into superior economic forecasts. Looking at the additional indicators (Table 15), the trading strategies derived from ARMAX(2), ARMAX(IRD), MAX(2), MAX(IRD), ARX(2) and ARX(IRD) all exhibit positive skewness, G/L ratio above one, good AUC* and small maximum drawdown. These results highlight the predictive content of IRD. Surprisingly, the model with constant asymmetry (specification (1)) performs quite well, too, although its classification rate is below 50%. To illustrate these results, we display in Figure 23 the compounded value over time of investing 1 USD in our trading rules in December 2014. The RW benchmark and the model with constant asymmetry (specification (1)) are displayed in black, exhibiting average returns of 0.95% and 4.68%, respectively. Most

[12]In this case, this is the AS benchmark, with a correct classification rate of 51.35%

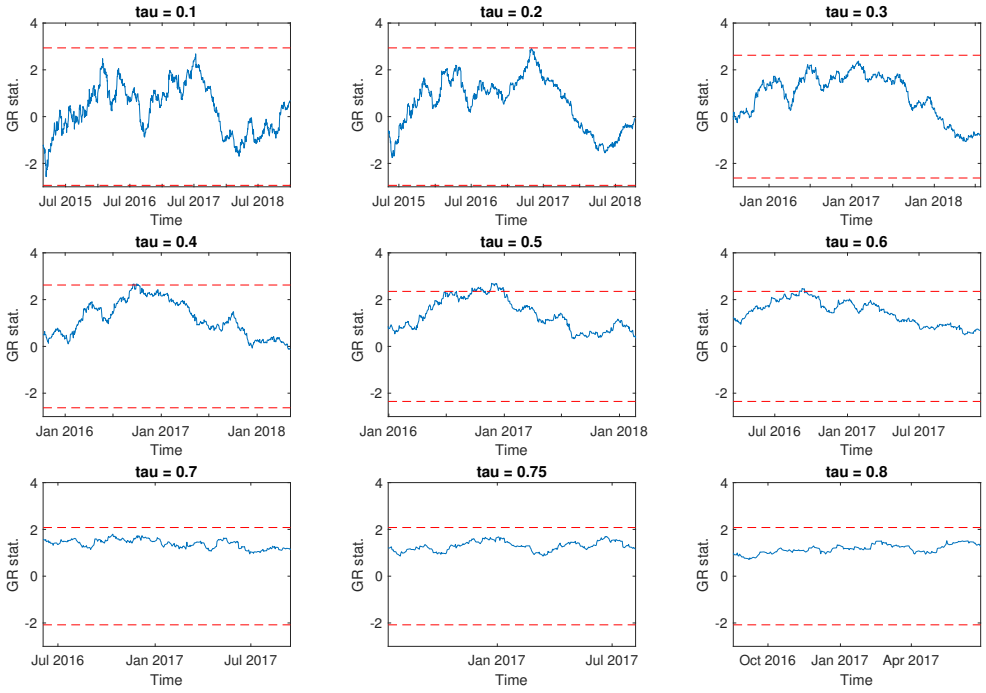


(i) GR test for ARMAX(2) ($H_0 : |\hat{m}| \leq 0$)

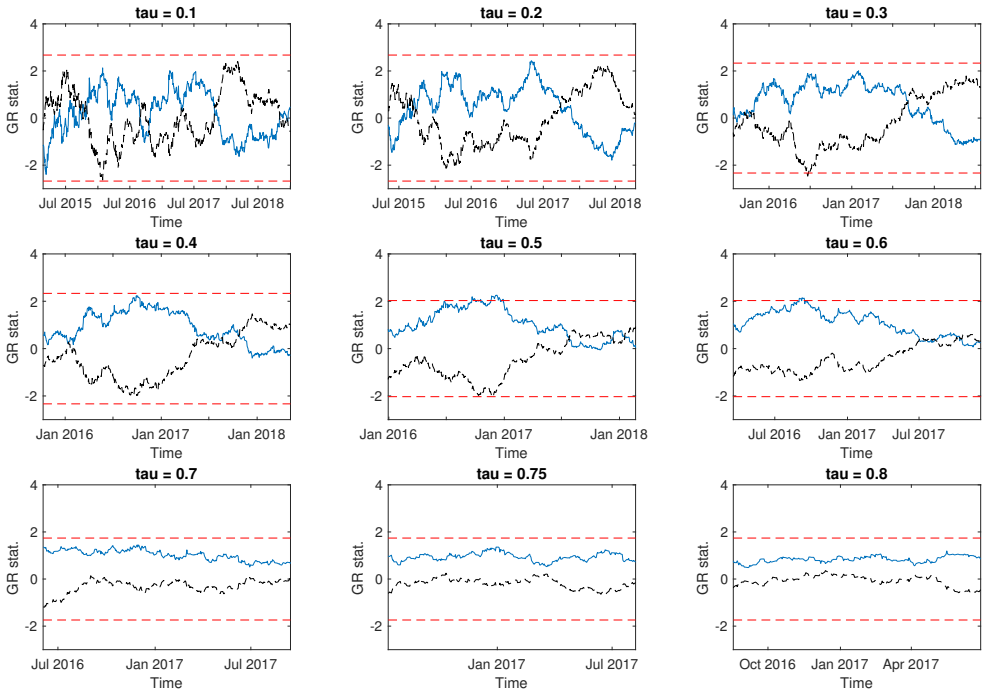


(ii) GR test for ARMAX(IRD) ($H_0 : |\hat{m}| \leq 0$)

Figure 20: GR test statistic (blue) with $\tau \in [.1, .85]$ using $\hat{m}$ as loss function. If the statistic is above the rejection threshold (dashed red), we reject the null hypothesis $H_0 : |\hat{m}| \leq 0$. Dashed black: test statistic for the RW strategy. (i) ARMAX(2) and rejection threshold at the 5% test level. (ii) ARMAX(IRD) and rejection threshold at the 10% test level.

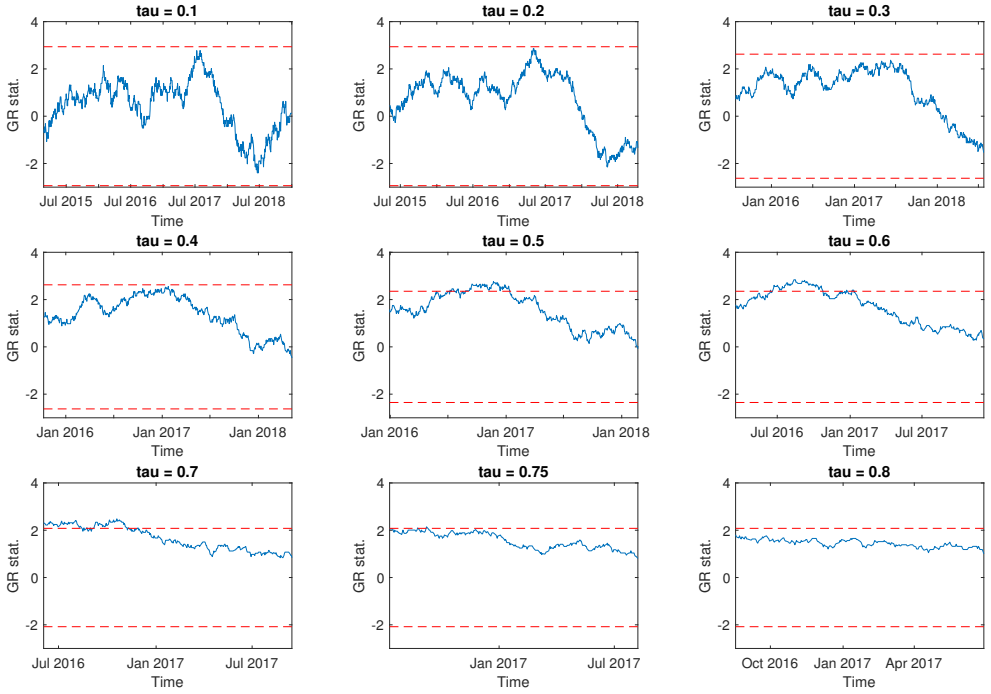


(i) GR test for ARMAX(2) ($H_0 : |\Delta\hat{m}| \leq 0$)

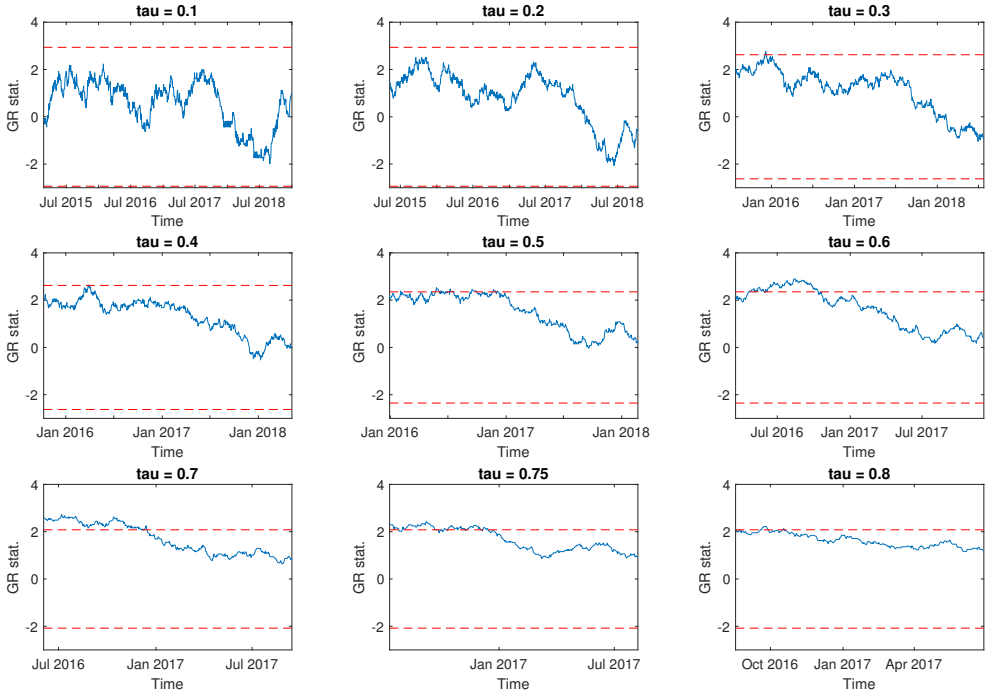


(ii) GR test for ARMAX(IRD) ($H_0 : |\Delta\hat{m}| \leq 0$)

Figure 21: GR test statistic (blue) with $\tau \in [.1, .85]$ using $L_{\Delta}^{(1)}$ as loss function. If the statistic is above the rejection threshold (dashed red), we reject the null hypothesis $H_0 : |\Delta\hat{m}| \leq 0$ where $\Delta\hat{m}$ is the average difference in profit with respect to the random walk benchmark. (i) ARMAX(2) and rejection threshold at the 5% test level. (ii) ARMAX(IRD) and rejection threshold at the 10% test level.



(i) GR test for ARMAX(2) ($H_0 : |L_\Delta^{(2)}| \leq 0$)



(ii) GR test for ARMAX(IRD) ($H_0 : |L_\Delta^{(2)}| \leq 0$)

Figure 22: GR test statistic (blue) with $\tau \in [.1, .85]$ using $L_\Delta^{(2)}$ as loss function. If the statistic is above the rejection threshold (dashed red), we reject the null hypothesis $H_0 : |L_\Delta^{(2)}| \leq 0$. (i) ARMAX(2) and (ii) ARMAX(IRD). Both rejection thresholds are at the 5% test level.

of the performance seems to be built at the beginning of the period, i.e. between 2015 and mid 2016. In particular, the models performing well are the ones that correctly predicted the surge in CHF value on 15 January 2015.

To test if the out-of-sample profit $\hat{m}$ obtained with IRD-based models is truly superior, we report the results of the GR test for the following specifications: MAX(2), MAX(IRD) and ARX(IRD). Figures 24 to 26, panels (i), display the GR statistics used to test the null hypothesis $H_0 : |\hat{m}| \leq 0$. We reject this hypothesis at the 5% test level for the three models, with various values of τ. As suggested by Figure 23, the significant performance always takes place in 2015-2016. Testing now $H_0 : |\Delta\hat{m}| \leq 0$ (i.e. using $L_\delta^{(1)}$ as loss function), we obtain similar results (Figures 24 to 26, panels (ii)). Using $L_\delta^{(2)}$, we do not find significant results, in line with the low performance on the pure classification rate (graphs are available upon demand).

Overall, these results point towards an out-of-sample predictive ability of our models, although the performance seems driven partly by the sharp appreciation of the CHF at the beginning of 2015 and not much significant after 2017.

USD/CHF (2014M12 - 2019M3)								
Spec.	CR	PT09	$\hat{m}$	DM	GW	$\Delta\hat{m}$	DM	GW
(12)	51.16%	1.27	5.16%	1.07	1.21	4.21%	0.54	4.95*
(11)	51.53%	0.92	5.10%	1.06	1.12	4.15%	0.52	5.70*
(10)	51.26%	2.26	3.32%	0.59	1.87	2.37%	0.24	1.18
(9)	50.70%	0.24	7.62%	1.32	2.96	6.67%	0.85	1.17
(8)	50.42%	0.02	6.22%	1.17	2.46	5.27%	0.68	0.65
(7)	51.26%	3.55*	1.71%	0.32	1.18	0.76%	0.08	0.91
(6)	52.47%	3.20*	3.21%	0.55	1.77	2.26%	0.23	1.18
(5)	51.72%	4.71**	2.60%	0.48	1.51	1.65%	0.17	1.03
(4)	51.16%	1.47	6.85%	1.31	3.76	5.90%	1.01	1.26
(3)	50.79%	0.19	8.37%	1.62	4.78*	7.42%	1.28	1.43
(2)	50.70%	0.15	3.19%	0.62	2.22	2.24%	0.39	0.57
(1)	48.74%	0.61	4.68%	0.92	2.37	3.73%	0.64	0.47
RW^+/RW^-	47.72%	0.00	0.95%	0.18	1.22	-	-	-
BH/AS	47.07%	-	-0.69%	-0.13	0.80	-	-	-

Table 14: Out-of-sample forecasting performance for CHF. * and ** indicate tests significant at the 10% and 5% test levels. $\Delta\hat{m}$ refers to the average *excess performance* over the random walk benchmark.

USD/CHF (2014M12 - 2019M3)						
Spec.	Sharpe	Skew	AUC	AUC*	G/L	Max. Draw.
(12)	0.48	6.02	0.50	0.51	1.10	0.11
(11)	0.47	6.04	0.51	0.52	1.10	0.11
(10)	0.31	-6.79	0.50	0.50	1.07	0.18
(9)	0.71	6.20	0.51	0.52	1.15	0.13
(8)	0.58	6.19	0.51	0.53	1.13	0.12
(7)	0.16	-6.82	0.51	0.50	1.03	0.18
(6)	0.30	-6.89	0.53	0.50	1.06	0.18
(5)	0.24	-6.84	0.51	0.50	1.05	0.18
(4)	0.63	6.25	0.52	0.55	1.14	0.16
(3)	0.78	6.30	0.51	0.55	1.17	0.10
(2)	0.30	6.21	0.53	0.52	1.06	0.19
(1)	0.43	6.32	0.52	0.51	1.09	0.10
RW^+/RW^-	0.09	6.87	0.49	0.51	1.02	0.31
BH/AS	-0.06	6.86	-	-	0.99	0.19

Table 15: Out-of-sample performance measures for the profit of the trading rules derived from the different models (USD/CHF).

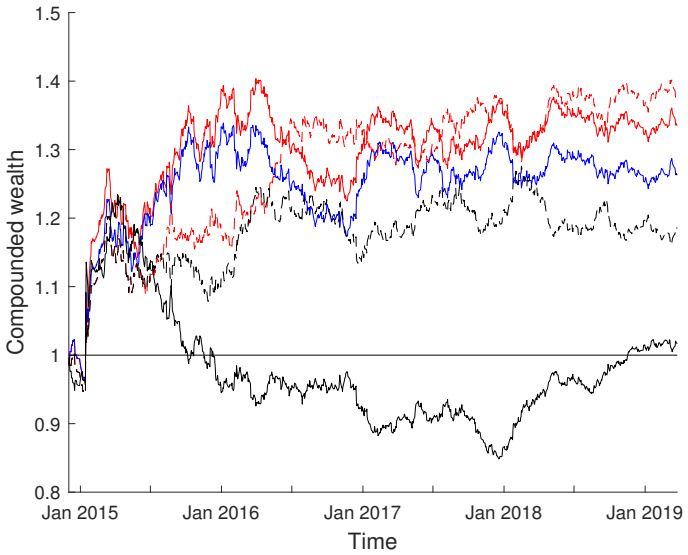
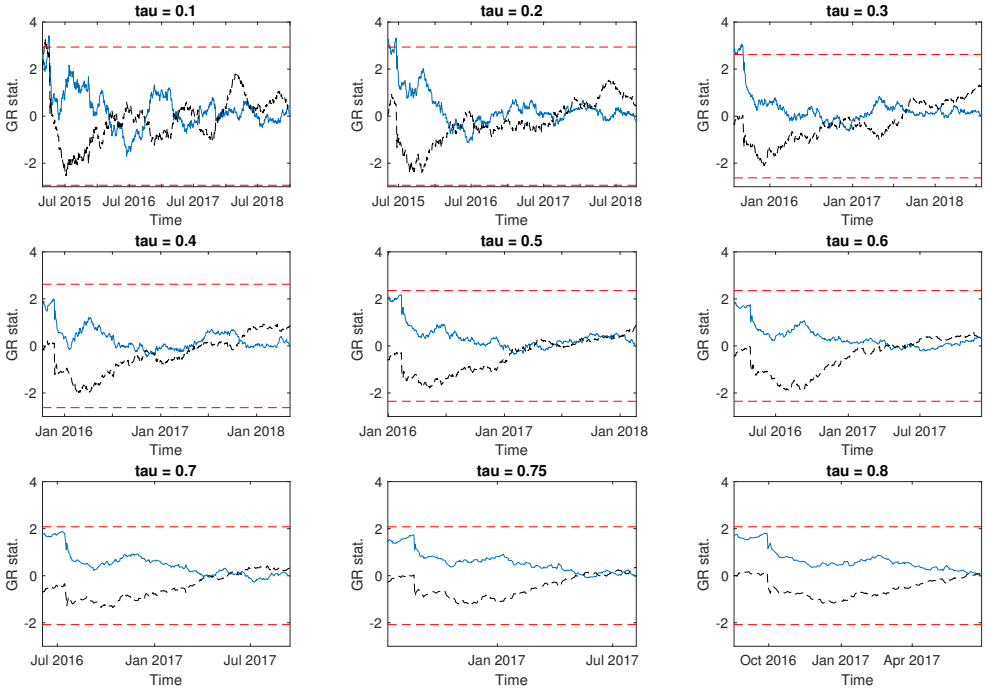
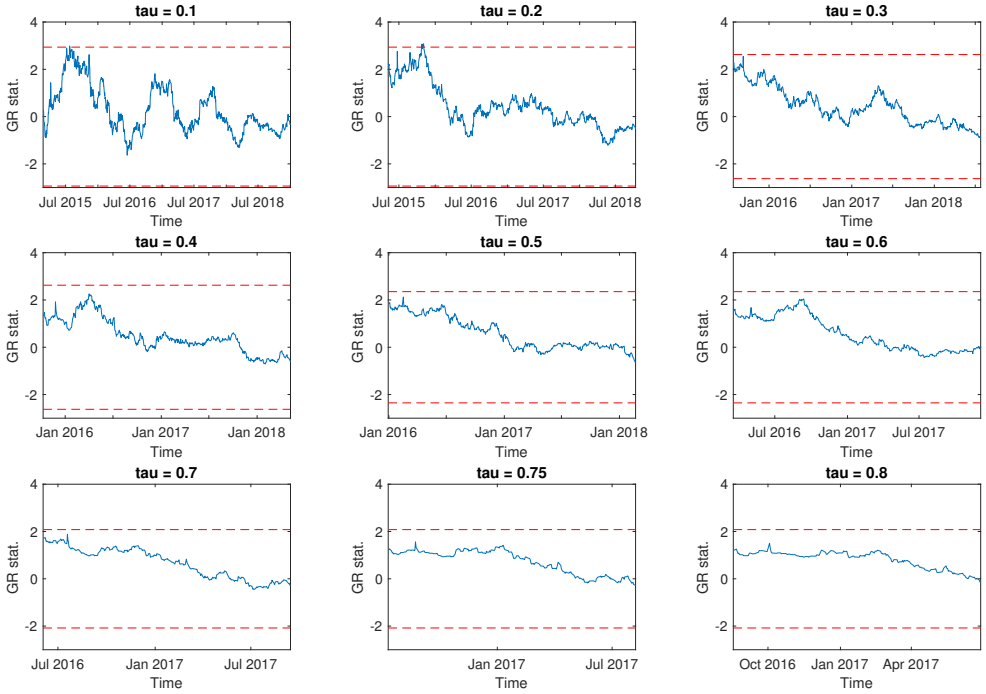


Figure 23: For CHF, compounded value of an initial investment of 1 USD in the trading rules derived from MAX(2) (solid red), MAX(IRD) (blue), ARX(IRD) (dashed red), random walk (solid black) and constant asymmetry (specification (1), dashed black) models.

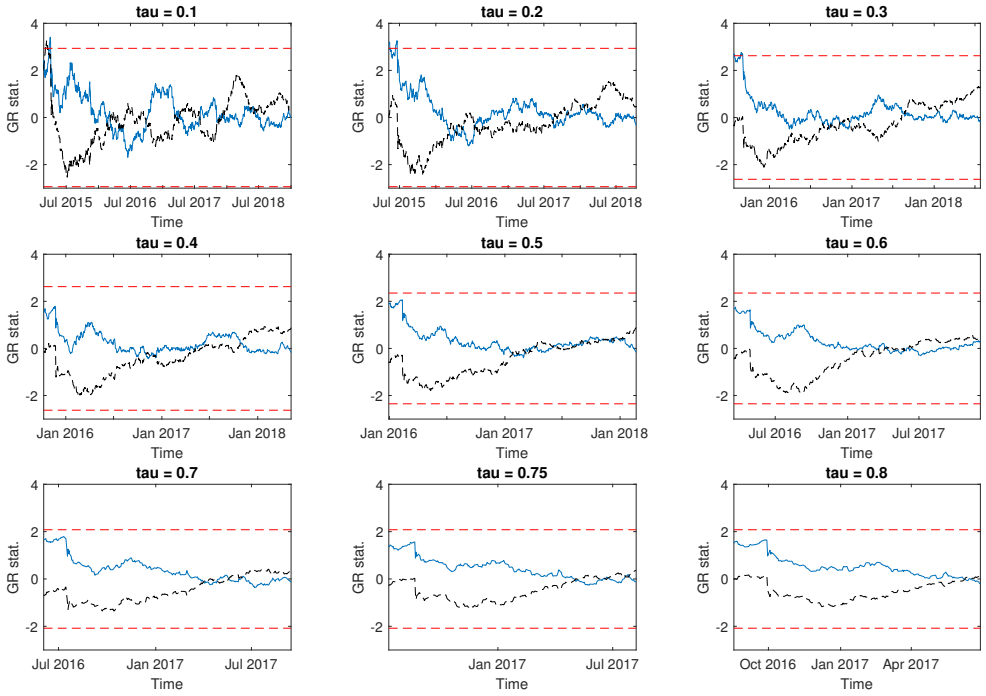


(i) GR test for MAX(2) ($H_0 : |\hat{m}| \leq 0$)

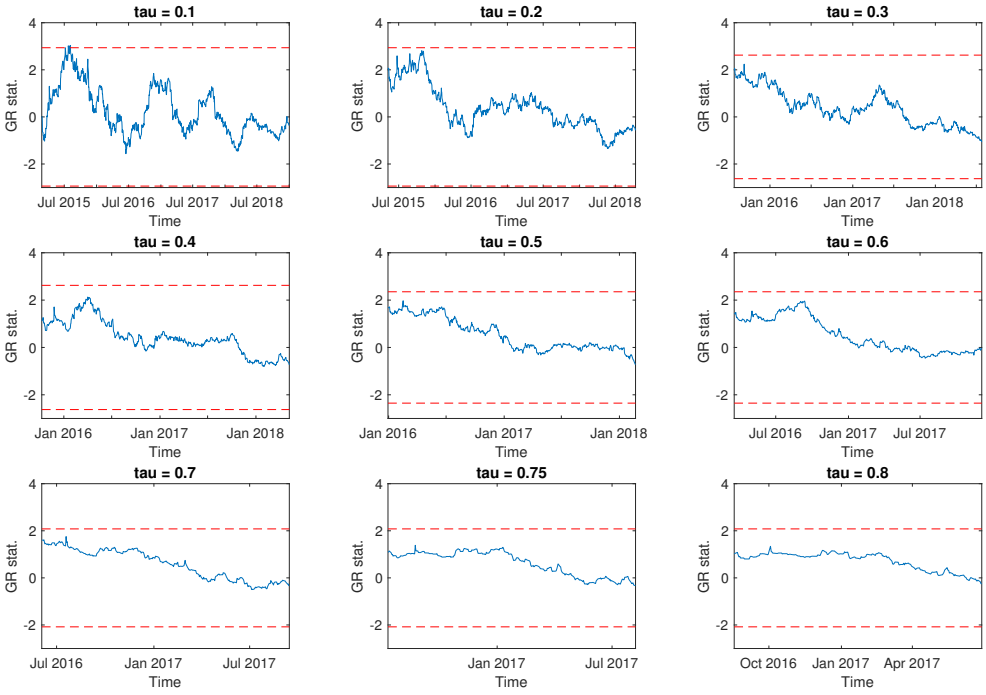


(ii) GR test for MAX(2) ($H_0 : |\Delta\hat{m}| \leq 0$)

Figure 24: GR test statistic (blue) with $\tau \in [.1, .85]$ using (i) $\hat{m}$ and (ii) $\Delta\hat{m}$ as loss functions. If the statistic is above the rejection threshold (dashed red), we reject the null hypothesis at the 5% test level. Dashed black: test statistics for the RW benchmark.

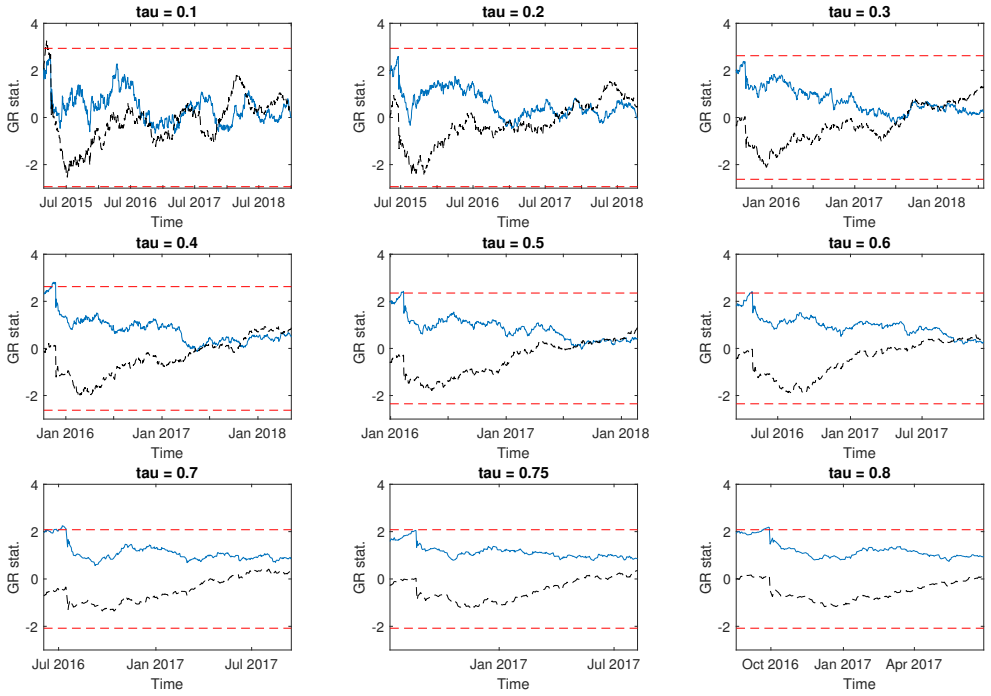


(i) GR test for MAX(IRD) ($H_0 : |\hat{m}| \leq 0$)

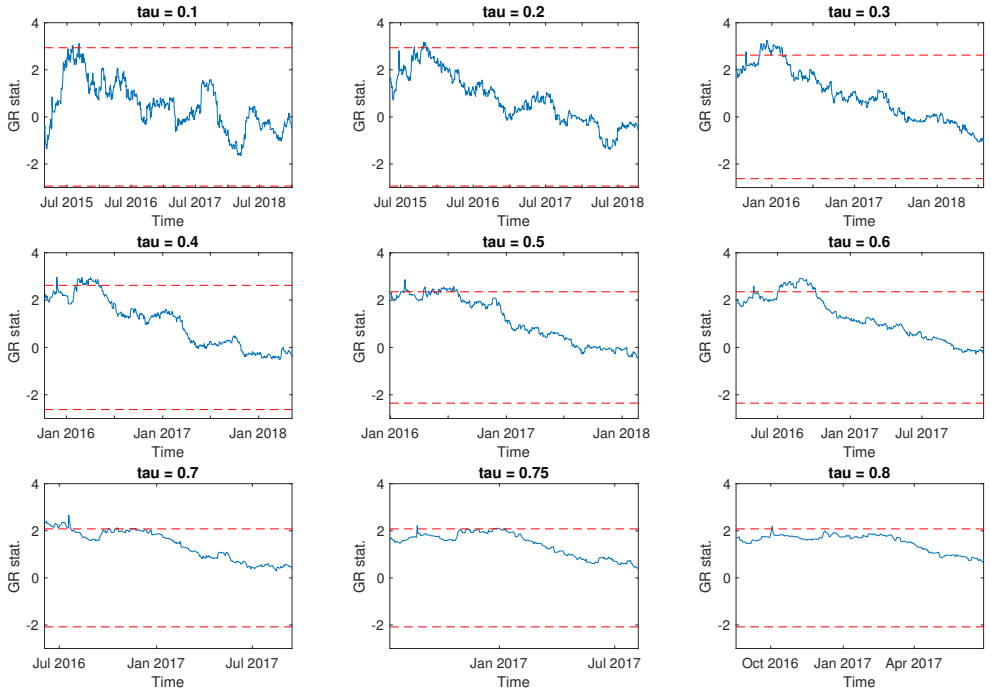


(ii) GR test for MAX(IRD) ($H_0 : |\Delta\hat{m}| \leq 0$)

Figure 25: GR test statistic (blue) with $\tau \in [.1, .85]$ using (i) $\hat{m}$ and (ii) $\Delta\hat{m}$ as loss functions. If the statistic is above the rejection threshold (dashed red), we reject the null hypothesis at the 5% test level. Dashed black: test statistics for the RW benchmark.



(i) GR test for ARX(IRD) ($H_0 : |\hat{m}| \leq 0$)



(ii) GR test for ARX(IRD) ($H_0 : |\Delta\hat{m}| \leq 0$)

Figure 26: GR test statistic (blue) with $\tau \in [.1, .85]$ using (i) $\hat{m}$ and (ii) $\Delta\hat{m}$ as loss functions. If the statistic is above the rejection threshold (dashed red), we reject the null hypothesis at the 5% test level. Dashed black: test statistics for the RW benchmark.

4 Conclusion

Using a model that allows for conditional dynamic asymmetry, we revisit the link between interest rate differentials and exchange rate returns. We also account for the effect of financial uncertainty by the inclusion of the VIX in our analysis. Applying this approach to the study of EUR and CHF exchange rates vis-a-vis the US Dollar, we find that the larger the difference between interest rates, the more likely the high-yield currency is to appreciate but it comes at the cost of a higher likelihood of a very large depreciation (i.e. crash risk). This result is in line with the theoretical framework of Fahri and Gabaix [2016] and Brunnermeier et al. [2008] who suggest the influence of carry trades through the brutal unwinding of those positions. Second, we find that USD is more likely to appreciate with respect to EUR when the VIX increases, but also that it is exposed to a higher risk of currency crashes. These results are in line with Menkhoff et al. [2012], Bekaert et al. [2013] and Habib and Stracca [2012] who suggest that liquidity shortage and increasing risk aversion lead investors towards buying USD, in the idea of a safe haven currency. However, this increasing uncertainty also leads to an increasing likelihood of a USD crash. Third, our results suggest the existence of self-fulfilling mechanisms as in Habib and Stracca [2012], where past unexpected shocks generate an increase in future crash risk.

Relying on the proposed model, we predict the direction of change of exchange rates and use these forecasts to build a trading strategy. We show that the detected effects are sufficiently large to generate significant economic gains both in- and out-of-sample, similarly to the findings of Amat et al. [2018]. However, we could show the existence of this effect at a daily frequency, whereas previous literature focused only on a monthly frequency. In addition, we obtain these results with a transparent and theoretically motivated econometric model.

Notice, though, that we do not account for the selection of the model itself. Therefore nothing guarantees that one could have obtained a profit *ex ante*, as assessed, e.g. in Bajgrowicz and Scaillet [2012]. This is a limit of the present analysis. However, as discussed in Inoue and Kilian [2005], in-sample results typically exhibit a higher power in performance tests. Hence, the consistence between in-sample and out-of-sample tests as well as the small number of tested specifications point towards a limited risk of spurious findings.

From a policy standpoint, our results suggest that favoring an increase in IRD correlates with systemic issues like a brutal depreciation. They also highlight the importance of self-fulfilling mechanisms and interactions with exchange rate volatility in currency crashes, suggesting that the prevention of unexpected shocks in periods of high uncertainty would reduce crash risk.

Finally, a last innovative aspect of the present paper consists in connecting the *likelihood* of a depreciation and of a currency crash with economic fundamentals, rather than the *level* of exchange rate returns. We believe that this change of perspective has interesting applications and could reconcile some of the apparent contradictions found in a literature mostly focused on mean effects. Future research could extend the present approach to studying a larger set of

currencies and predictors, and see if our results can be generalized.

Acknowledgments

J. Hambuckers acknowledges the financial support of the DFG via the RTG 1644 *Scaling problems in Statistics* and of the National Bank of Belgium (project REFEX). The authors are grateful to P.-H. Hsu for providing them with A. Yu-Chin implementation of the SSPA test. Parts of the research for this paper were conducted while JH was affiliated with the University of Goettingen (Germany), Chair of Statistics.

References

C. Amat, T. Michalski, and G. Stoltz. Fundamentals and exchange rate forecastability with simple machine learning methods. *Journal of International Money and Finance*, 88:1–24, 2018.

S. Anatolyev and N. Gospodinov. Modeling financial return dynamics via decomposition. *Journal of Business & Economic Statistics*, 28(2):232–245, 2010.

E. Andreou and E. Ghysels. Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics*, 17(5):579–600, 2002.

P. Bacchetta and E. van Wincoop. On the unstable relationship between exchange rates and macroeconomic fundamentals. *Journal of International Economics*, 91(1):18 – 26, 2013.

X. Bai, J.R. Russel, and G.C. Tiao. Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics*, 114(2):349–360, 2003.

P. Bajgrowicz and O. Scaillet. Technical trading revisited: False discoveries, persistence tests, and transaction costs. *Journal of Financial Economics*, 106(3):473–491, 2012.

T. Bali, H. Mo, and Y. Tang. The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR. *Journal of Banking & Finance*, 32(2):269–282, 2008.

G. Bekaert, M. Hoerova, and Duca M.L. Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, 60(7):771–788, 2013.

G. Bekaert, E. Engstrom, and A. Ermolov. Bad environments, good environments: A non-gaussian asymmetric volatility model. *Journal of Econometrics*, 186(1):258–275, 2015.

T.J. Berge. Forecasting disconnected exchange rates. *Journal of Applied Econometrics*, 29(5): 713–735, 2014.

J. Berkowitz. Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474, 2001.

A.E. Bernardo and O. Ledoit. Gain, loss, and asset pricing. *Journal of Political Economy*, 108 (1):144–172, 2000.

O. Blaskowitz and H. Herwartz. On economic evaluation of directional forecasts. *International Journal of Forecasting*, 27(4):1058–1065, 2011.

O. Blaskowitz and H. Herwartz. Testing the value of directional forecasts in the presence of serial correlation. *International Journal of Forecasting*, 30(1):30–42, 2014.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

M. Brunnermeier, S. Nagel, and L. Pedersen. Carry trades and currency crashes. *NBER Macroeconomics Annual*, 23:313–347, 2008.

Y-W. Cheung, M.D. Chinn, and A.G. Pascual. Empirical exchange rate models of the nineties: Are any fit to survive? *Journal of International Money and Finance*, 24(7):1150–1175, 2005.

J. Chung and Y. Hong. Model-free evaluation of directional predictability in foreign exchange markets. *Journal of Applied Econometrics*, 22(5):855–889, 2007.

F.X. Diebold and R.S. Mariano. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, 1995.

J.A. Doornik and H. Hansen. An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70(s1):927–939, 2008.

W. Du, A. Tepper, and A. Verdelhan. Deviations from covered interest rate parity. *The Journal of Finance*, 73(3):915–957, 2018.

G. Elliott and A. Timmermann. *Economic Forecasting*. Princeton University Press, 2016. ISBN 978-0-691-14013-1.

R.F. Engle. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1007, 1982.

M. Fahri and X. Gabaix. Rare disasters and exchange rates. *Quarterly Journal of Economics*, 131(1):1–52, 2016.

C. Francq and J.M. Zakoian. *GARCH Models : Structure, Statistical Inference and Financial Applications*. John Wiley, 2010. ISBN 978-0-470-68391-0.

X. Gabaix and M. Maggiori. International liquidity and exchange rate dynamics. *Quarterly Journal of Economics*, 130(3):1369–1420, 2015.

R. Giacomini and B. Rossi. Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4):595–620, 2010.

R. Giacomini and H. White. Tests of Conditional Predictive Ability. *Econometrica*, 74(6): 1545–1578, 2006.

L. Glosten, R. Jagannathan, and D. Runkle. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, 48(5):1779–1801, 1993.

M. Grigoletto and F. Lisi. Looking for skewness in financial time series. *Econometrics Journal*, 12(2):310–323, 2009.

M. Habib and L. Stracca. Getting beyond carry trade: What makes a safe haven currency? *Journal of International Economics*, 87(1):50–64, 2012.

B. Hansen. Autoregressive conditional density estimation. *International Economic Review*, 35 (3):705–730, 1994.

P. Hansen and A. Lunde. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7):873–889, 2005.

C. Harvey and A. Siddique. Autoregressive Conditional Skewness. *The Journal of Financial and Quantitative Analysis*, 34(4):465–487, 1999.

H. Herwartz. Stock return prediction under GARCH - an empirical assessment. *International Journal of Forecasting*, 33(3):569–580, 2017.

P.-H. Hsu, Y. Hsu, and C. Kuan. Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias. *Journal of Empirical Finance*, 17(3):471–484, 2010.

P.-H. Hsu, M. Taylor, and Z. Wang. Technical trading: Is it still beating the foreign exchange market? *Journal of International Economics*, 102:188–208, 2016.

C.-H. Hui, H. Genberg, and T.-K. Chung. Funding liquidity risk and deviations from interest–rate parity during the financial crisis of 2007–2009. *International Journal of Finance and Economics*, 16(4):307–323, 2011.

L. Husted, J. Rogers, and B. Sun. Uncertainty, currency excess returns, and risk reversals. *Journal of International Money and Finance*, 88:228–241, 2018.

C. Inclan and G. Tiao. Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance. *Journal of American Statistical Association*, 89(427):913–923, 1994.

A. Inoue and L. Kilian. In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, 23(4):371–402, 2005.

A Ismailov and B. Rossi. Uncertainty and deviations from uncovered interest rate parity. *Journal of International Money and Finance*, 88:242–259, 2018.

E. Jondeau and M. Rockinger. Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, 27(10):1699–1737, 2003.

M.C. Jones and A. Pewsey. Sinh-arcsinh distributions. *Biometrika*, 96(4):761–780, 2009.

O. Jordà and A.M. Taylor. The carry trade and fundamentals: Nothing to fear but FEER itself. *Journal of International Economics*, 88(1):74 – 90, 2012.

J.W. Jurek. Crash-neutral currency carry trades. *Journal of Financial Economics*, 113(3):325 – 347, 2014.

B. Klar, F. Lindner, and S.G. Meintanis. Specification tests for the error distribution in GARCH models. *Computational Statistics and Data Analysis*, 56(11):3587–3598, 2012.

R. Kulperger and H. Yu. High moment partial sum processes of residuals in garch models and their applications. *The Annals of Statistics*, 33(5):2395–2422, 2005.

X. Liu. Modeling time-varying skewness via decomposition for out-of-sample forecast. *International Journal of Forecasting*, 31(1):30–42, 2015.

R. Meese and K. Rogoff. Empirical exchange rate models of the seventies. Do they fit out of sample? *Journal of International Economics*, 14(1-2):3–24, 1983.

L. Menkhoff, L. Sarno, M. Schmeling, and A. Schrimpf. Carry Trades and Global Foreign Exchange Volatility. *Journal of Finance*, 67(2):681–717, 2012.

H. Pesaran and A. Timmermann. Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, 104(485):325–337, 2009.

A. Ranaldo and P. Söderlind. Safe haven currencies. *Review of Finance*, 14:385–407, 2010.

B. Rossi. Exchange rate predictability. *Journal of Economic Literature*, 51(4):1063–1119, 2013.

B. Rossi and T. Sekhposyan. Evaluating predictive densities of us output growth and inflation in a large macroeconomic data set. *International Journal of Forecasting*, 30(3):662 – 682, 2014.

A. Wilhelmsson. Value-at-Risk with time-varying variance, skewness and kurtosis: the NIG-ACD model. *Econometrics Journal*, 12(1):82–104, 2009.

A Appendix: the sinh-arcsinh distribution

The pdf of the sinh-arcsinh distribution is given by

$$f(z;\epsilon,\delta) = \eta^{-1} Z_{\xi,\eta}(x)^{-1/2} \delta C_{\epsilon,\delta}((x-\xi)\eta) \exp(-S^2_{\epsilon,\delta}((x-\xi)/\eta)/2),$$

where

$$\begin{aligned}
Z_{\xi,\eta}(x) &= (2\pi(1+((x-\xi)/\eta)^2)), \\
C_{\epsilon,\delta}(x) &= \cosh(\delta \sinh^{-1}(x) - \epsilon) = (1+S^2_{\epsilon,\delta}(x))^{1/2}, \\
S_{\epsilon,\delta}(x) &= \sinh(\delta \sinh^{-1}(x) - \epsilon), \\
\xi &= -\eta \sinh(\epsilon_t/\delta) P_{1/\delta} \\
\eta &= \sqrt{1/(0.5(\cosh(2\epsilon_t/\delta_t)P_{2/\delta} - 1) - \sinh(\epsilon_t/\delta)P_{1/\delta})^2}, \\
P_q &= \frac{\exp(1/4)}{8\pi^{1/2}} \left(K_{(q+1)/2}(1/4) + K_{(q-1)/2}(1/4) \right).
\end{aligned}$$

with K being the modified Bessel function of the second kind. ξ and η are the location and the scale parameters, respectively, whose values are fixed to ensure zero mean and unit variance. The cumulative distribution function $F(z;\epsilon,\delta)$ is obtained from the transformation given in (6) and is simply:

$$F(z;\epsilon,\delta) = \Phi(\sinh(\delta \sinh^{-1}((z-\xi)/\eta) - \epsilon).$$

where $\Phi(\cdot)$ is the cdf of the standardized Gaussian distribution. The quantile function F^{-1} is easily derived in the same way; and is given by

$$F^{-1}(p;\epsilon,\delta) = \sinh\left((1/\delta) * \sinh^{-1}(\Phi^{-1}(p)) + (\epsilon/\delta) \right) * \eta + \xi,$$

where Φ^{-1} is the quantile function of the Gaussian distribution.

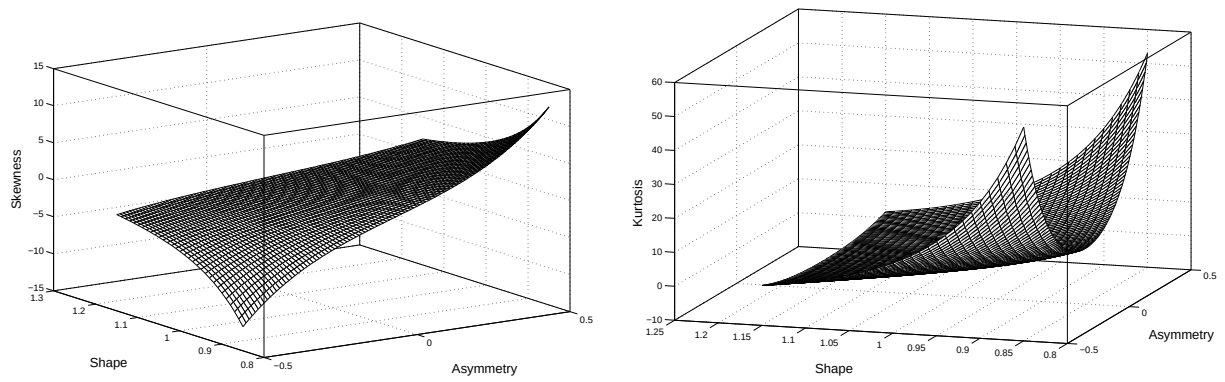
The skewness and kurtosis of z_t are given by

$$\begin{aligned}
SK_t &= \frac{1}{4}\{\sinh(3\epsilon_t/\delta)P_{3/\delta_t} - 3\sinh(\epsilon_t/\delta_t)P_{1/\delta}\}, \\
KU_t &= \frac{1}{8}\{\cosh(4\epsilon_t/\delta)P_{4/\delta} - 4\cosh(2\epsilon_t/\delta_t)P_{2/\delta} + 3\}.
\end{aligned}$$

where P_q is define by equation (23). Hence, one can observe that both quantities depend on both parameters. The response surfaces for skewness and kurtosis, for various values of ϵ and δ, are displayed in Figure 27.

As shown on Figure 27 (left panel), δ seems to have a limited impact on the skewness (the response surface is quite flat on this dimension). Greater flexibility can be introduced by specifying a more complicated equation for δ. Similarly to what is done for ϵ_t, the following equation can be used:

$$\delta_t = h(\mathcal{I}_{t-1}) = \exp(b_0 + b_1\delta_{t-1} + b_2 z_{t-1} + b_3 x_{t-1}).$$



(a) (b)

Figure 27: (a) Skewness and (b) Kurtosis

B Appendix: Simulation study

In this appendix, we study the finite sample properties of our econometric approach. We start by investigating the quality of the proposed maximum likelihood estimation procedure. Then, we study the size and power of the hypothesis test based on (14), focusing on the regression parameters in the skewness equation. We consider three sample sizes: $T \in \{500, 1500, 3000\}$. The parameters of the various data generating processes (DGP) are displayed in Table 16. For DGP1 to DGP3, we assume λ and c equal to zero, such that $\mathbb{E}(R_t) = 0$, whereas we introduce a mean structure in DGP4 to DGP6. If $a_3 \neq 0$, we assume that $x_{t-1} \overset{iid}{\sim} N(0,1)$. For each DGP, we simulate $1,000$ time series.

Data generating processes											
	ω	α	β	c	λ	δ	a_0	a_1	a_2	a_3	a_4
DGP1	10^{-4}	0.05	0.88	-	-	.8	-0.1	0.4	-0.9	-	-
DGP2	10^{-4}	0.05	0.88	-	-	.8	-0.1	0.4	-0.9	0.85	-
DGP3	10^{-4}	0.1	0.8	-	-	.85	-0.1	0.5	0.3	-0.5	-
DGP4	10^{-4}	0.05	0.88	10^{-4}	0.1	.8	-0.1	0.4	-0.9	-	-
DGP5	10^{-4}	0.05	0.88	10^{-4}	0.1	.8	-0.1	0.4	-0.9	0.85	-
DGP6	10^{-4}	0.1	0.8	10^{-4}	0.1	.85	-0.1	0.5	0.3	-0.5	-
DGP7	10^{-4}	0.05	0.88	10^{-4}	-0.1	.8	-0.1	0.5	0.1	0.15	0.3

Table 16: Values of the parameters considered in the different simulation set-ups.

Results are given in Tables 17 and 18. Overall, we observe a decreasing mean squared error in the estimated parameters when the sample size increases, and no differences across DGP. As for most time-series models, the simulations highlight the need for large samples (i.e. several thousands observations) for a high level of precision.

Now, we study the size and power of the suggested Wald-type tests for the skewness regres-

RMSE									
No mean	T	ω	α	β	a_0	a_1	a_2	a_3	δ
DGP1	500	1.07	0.32	0.1	0.52	0.15	0.77	-	0.23
	1500	0.26	0.16	0.03	0.23	0.08	0.38	-	0.08
	3000	0.17	0.11	0.02	0.16	0.06	0.27	-	0.06
DGP2	500	0.53	0.27	0.05	0.64	0.12	0.17	0.18	0.6
	1500	0.21	0.13	0.02	0.3	0.06	0.08	0.09	0.33
	3000	0.13	0.09	0.01	0.23	0.04	0.06	0.06	0.23
DGP3	500	0.43	0.23	0.07	0.61	0.15	0.27	0.22	0.44
°	1500	0.18	0.12	0.03	0.29	0.07	0.13	0.11	0.22
	3000	0.11	0.08	0.02	0.2	0.05	0.09	0.07	0.16

Table 17: Root-MSE divided by the value of the corresponding parameter, for DGP without a mean structure (DGP1 to DGP3).

RMSE											
GARCH-in-Mean	T	ω	α	β	c	λ	a_0	a_1	a_2	a_3	δ
DGP4	500	1.02	0.31	0.1	275.55	7.82	0.61	0.16	0.71	-	0.21
	1500	0.25	0.17	0.03	42.19	1.13	0.34	0.09	0.43	-	0.1
	3000	0.16	0.11	0.02	26.76	0.71	0.21	0.062	0.28	-	0.07
DGP5	500	0.76	0.28	0.07	88.54	2.79	0.77	0.13	0.2	0.21	0.65
	1500	0.22	0.14	0.02	33.31	1.39	0.37	0.07	0.09	0.09	0.32
	3000	0.13	0.09	0.01	22.89	1.17	0.24	0.04	0.05	0.06	0.23
DGP6	500	0.41	0.3	0.07	97.2	3.21	0.63	0.14	0.3	0.23	0.43
	1500	0.19	0.15	0.03	35.06	1.38	0.31	0.07	0.15	0.11	0.24
	3000	0.13	0.12	0.02	23.99	1.19	0.20	0.05	0.09	0.07	0.17

Table 18: Root-MSE divided by the value of the parameters, for DGP with a GARCH-in-Mean structure (DGP4 to DGP6).

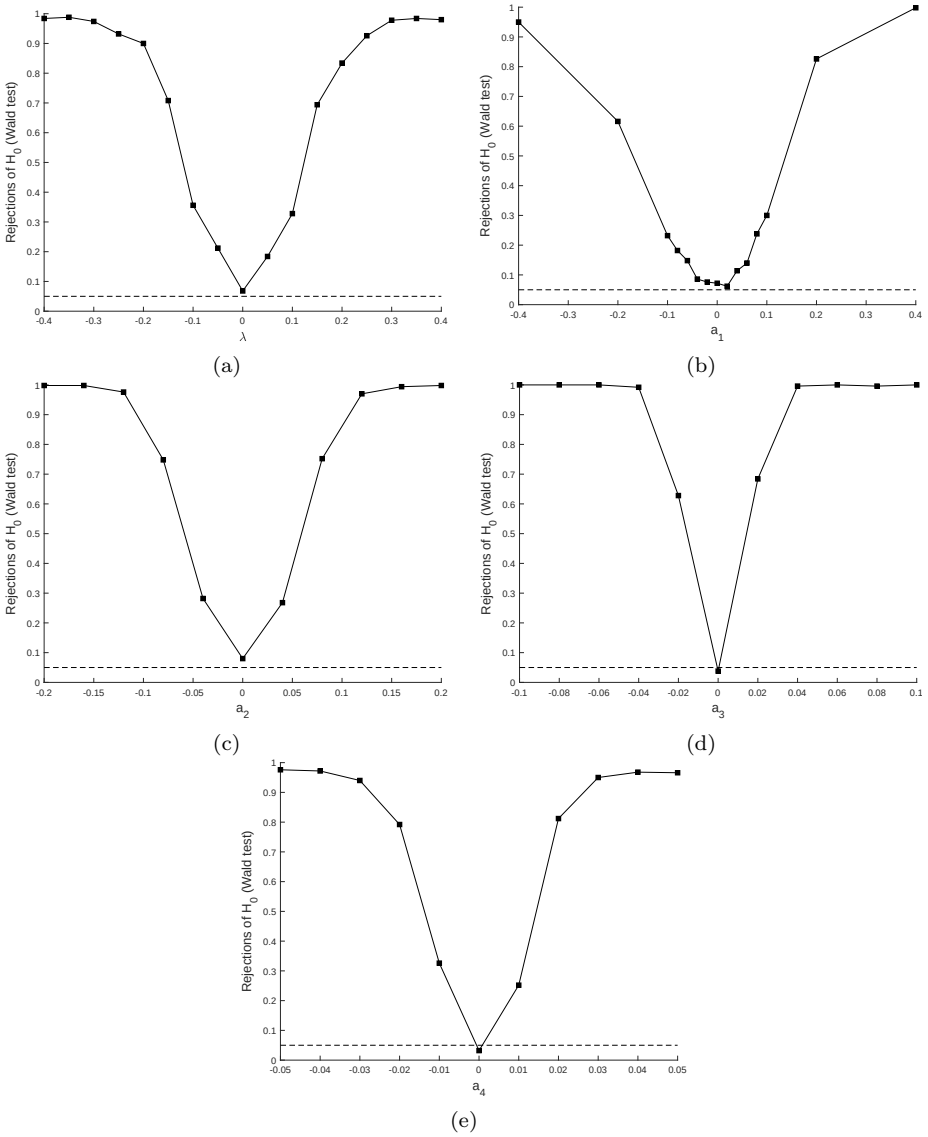


Figure 28: Rejection rates of Wald tests at the 5% test level, for various values of the parameters. Dashed: 5% threshold.

sion parameters and λ. We consider two explanatory variables $x_{t-1,1}$ and $x_{t-1,2}$ in eq. (7)[13]. We generate $x_{t,1}$ and $x_{t,2}$ from two AR(1) processes where the AR parameters are equal to 0.9 and the error terms follow a bivariate normal distribution with a correlation parameter of -0.4 (a value observed in our data). Baseline values of the parameters are given in Table 16 (DGP7). Then, we sequentially replace one of the parameters of interest by a range of values (including 0), keeping the others at their baseline value. In line with our empirical study, we set $n = 1500$. Figure 28 summarizes our results, indicating good powers under the various scenarios. For a_3 and a_4, we also obtain excellent sizes, whereas we reject a bit too often for λ, a_1 and a_2.

[13]We do not look at inference for the GARCH parameters since these parameters are estimated under positivity constraints.

C Appendix: Testing for structural breaks via CUSUM tests

In this appendix, we provide technical details regarding the CUSUM tests used in Section 3. In particular, we discuss the necessary modifications to be made, in order to account for the specifics of our model.

In Kulperger and Yu [2005], the authors derived the asymptotic properties of partial sum processes constructed on k^{th} power of GARCH residuals, showing that it converges toward a Brownian process plus a correction term. Such CUSUM statistics can be used to test for a change in conditional (potentially high-order) moments over time. As implied by their Theorem 1.1 and 1.2, the partial sum process behaves as if the residuals $\hat{z}_t = r_t/\hat{\sigma}_t$ were asymptotically the same as the innovations z_t. However, in usual GARCH models, z_t are assumed (unconditionally) i.i.d, whereas in our GARCH-SH model, it is not the case under the null hypothesis of no breaks. To circumvent this issue, we suggest working instead with Gaussian pseudo-residuals $\hat{u}_t$, based on the inverse of the sinh-arcsinh transform given by (6). Thus, we defined these pseudo-residuals as

$$\hat{u}_t = \sinh\left(\hat{\delta}_t \sinh^{-1}(\hat{z}_t) - \hat{\epsilon}_t\right),$$

where $\hat{z}_t = (R_t - \hat{c} - \hat{\lambda}\hat{\sigma}_t)/\hat{\sigma}_t$. Under a correct specification of the sinh-arcsinh distribution, $\hat{u}_t$ is asymptotically N(0,1) distributed and fulfills the main assumptions of Kulperger and Yu [2005]. It also fulfills the assumptions of zero-mean and unit-variance. Two additional requirements are the finiteness of the k^{th} moment of z_t and that u_0 is a non-degenerate random variable. These conditions are fulfilled when we assume that ϵ_t and δ_t are finite. Then, we suggest using the following test statistic, similar to the one proposed in Kulperger and Yu [2005]:

$$CUSUM^{(k)} = \max_{1 \leq i \leq T} \frac{|\sum_{t=1}^{i} \hat{u}_t^k - i\hat{\mu}_k|}{\hat{s}_k \sqrt{T}}, \tag{23}$$

where $\hat{\mu}_k$ is the empirical moment of order k of the residuals, and $\hat{s}_k$ an estimate of $E(u_0^k - \mu_k)^2$. A formal proof of the asymptotic properties of (23) is beyond the scope of the paper. On the basis of the theoretical arguments enumerated previously, we use the (approximated) results that (23) converges to the supremum of a Brownian bridge:

$$CUSUM^{(k)} \xrightarrow{a.s.} \sup_{0 \leq u \leq 1} |B_0(u)|.$$

Additionally, since we might face several breaks in the time series, we need an algorithm to sequentially identify the dates of the breaks. We simply apply the procedure detailed in Inclan and Tiao [1994], consisting in repeatedly partitioning our time series, until no more breaks are found.

Relying on the simulation set-up described in Section B, we briefly study the size of this test. Results are displayed in Table 19. CUSUM tests appear slightly under-sized.

Rejections of H_0: no breaks

No mean					GARCH-in-Mean				
DGP	T	k=2	k=3	k=4	DGP	T	k=2	k=3	k=4
DGP1	500	0.024	0.04	0.012	DGP4	500	0.04	0.048	0.018
	1500	0.032	0.032	0.026		1500	0.044	0.038	0.026
	3000	0.044	0.052	0.05		3000	0.042	0.052	0.04
DGP2	500	0.024	0.042	0.034	DGP5	500	0.028	0.036	0.028
	1500	0.03	0.046	0.032		1500	0.042	0.058	0.038
	3000	0.046	0.04	0.036		3000	0.042	0.034	0.036
DGP3	500	0.042	0.038	0.03	DGP6	500	0.04	0.036	0.042
	1500	0.036	0.038	0.028		1500	0.07	0.044	0.05
	3000	0.044	0.042	0.032		3000	0.024	0.042	0.022

Table 19: Type-I error for testing the null hypothesis of no structural breaks, for DGP either with no mean structure (left) or with a GARCH-in-Mean structure (right).

D Appendix: Additional results

D.1 EUR: Estimated parameters - subperiods

D.2 Results for a Gaussian GARCH-in-mean model

In this section, we provide several additional results for a benchmark Gaussian GARCH-in-mean model. In Table 21, we report the estimated parameters as well as various specification and selection statistics. Figure 29 displays the empirical distribution of the PIT residuals used to conduct the specification tests for both the ARMAX(2) and the Gaussian models. We see clear departure from uniformity in the Gaussian case.

D.3 Out-of-sample test EUR

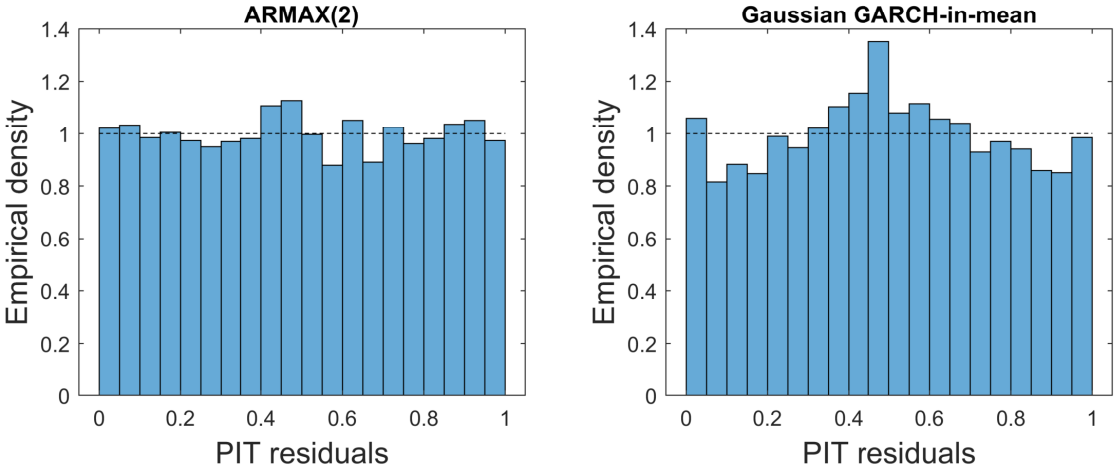
In this section, we provide additional results for the out-of-sample analysis related to EUR. Figure 30 displays the results of the GR test using the best ex-post benchmark (i.e. AS with a classification rate of 51.38%).

USD/EUR	ARMAX(2) - GARCH-in-Mean			
	Full sample	Period 1	Period 2	Period 3
ω	0.000	0.000	0.000	0.000
α	0.029	0.025	0.034	0.03
β	0.968	0.973	0.966	0.931
c	0.000	0.002***	-0.001*	0.001
	(0.000)	(0.001)	(0.000)	(0.001)
λ	-0.032	-0.273***	0.093	-0.151
	(0.052)	(0.095)	(0.057)	(0.267)
δ	0.772	0.788	0.752	0.77
	(0.022)	(0.032)	(0.038)	(0.053)
$a0$	-0.043*	-0.055	-0.047**	-0.011
	(0.025)	(0.039)	(0.021)	(0.011)
$a1$	0.631***	0.285	0.728***	0.951***
	(0.184)	(0.231)	(0.114)	(0.029)
$a2$	0.042***	0.060***	0.047**	0.034**
	(0.014)	(0.012)	(0.019)	(0.016)
$a3$	1.520*	3.54***	2.556*	0.100
	(0.813)	(0.972)	(1.54)	(0.295)
$a4$	0.171*	0.237	0.208**	0.055
	(0.101)	(0.174)	(0.099)	(0.058)
CR	52.83%	54.98%	51.26%	52.09%
(PT09)	(6.52***)	(12.31**)	(0.56)	(0.10)
CR(RW^+/RW^-)	50.68%	50.65%	49.97%	49.88%
CR(BH/AS)	50.05%	49.19%	51.14%	48.77%†
$\hat{m}$	5.45%	14.71%	0.64%	1.14%
(Full SSPA)	(0.03)	(0.000)	(0.896)	(0.977)
$\hat{m}(RW^+/RW^-)$	0.22%	0.01%	0.16%	1.84%
(Full SSPA)	(0.984)	(0.945)	(0.928)	(0.969)
$\hat{m}$(BH/AS)	0.19%	0.74%	2.31%	0.83%
(Full SSPA)	(0.985)	(0.886)	(0.744)	(0.988)

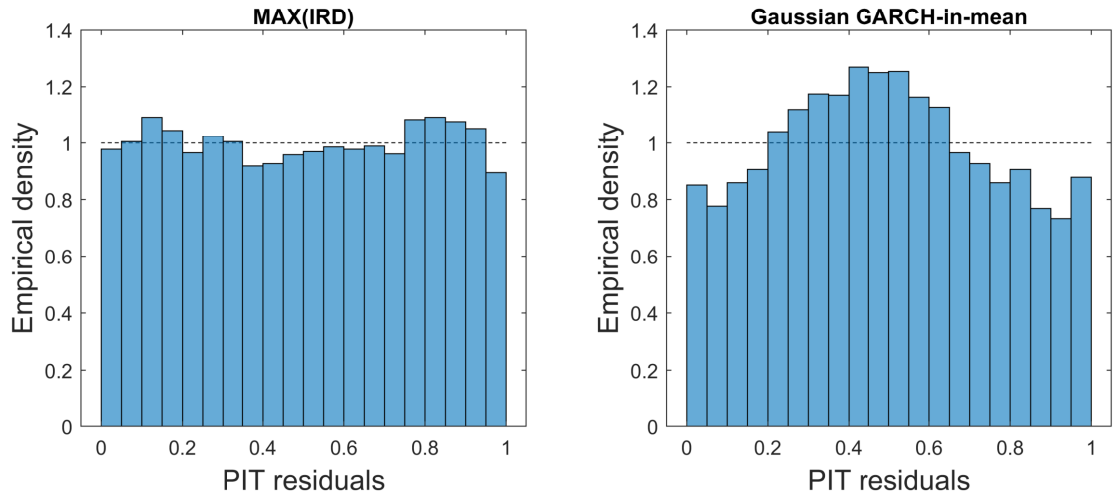
Table 20: Estimated parameters for ARMAX(2) and performance measures on three subperiods defined by the following dates: 6/01/1999, 23/10/2008, 15/12/2015 and 25/03/2019. *Full SSPA* refers to p-values of the SSPA tests obtained when testing simultaneously all models. For the benchmark strategies, we report the results for the best side (long or short).

Gaussian GARCH-in-mean													
Currency	ω	α	β	c	λ	AIC	BIC	LR	$\hat{m}$	CR	BK	DH	AD
EUR	0.000	0.028	0.969	0.000	-0.001	-37,897.08	-37,864.42	146.8^{***}	-.19%	49.14%	2.5	32.02^{***}	6.8^{***}
CHF	0.000	0.042	0.954	0.000	-0.033	-367.1318	-366.8052	616.9^{***}	4.29%	49.79%	45.4^{***}	67.04^{***}	22.4^{***}

Table 21: Estimated parameters, model selection and specification criteria for a classical Gaussian GARCH-in-Mean model, for both currencies. *, ** and *** denote tests significant at the 10%, 5% and 1% level.

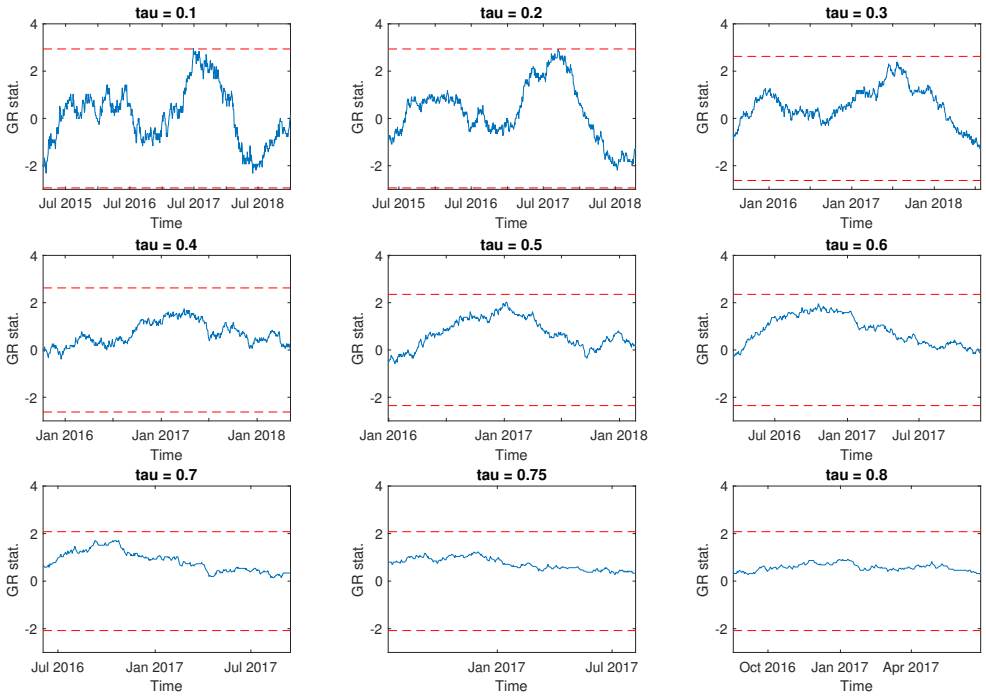


(i) USD/EUR

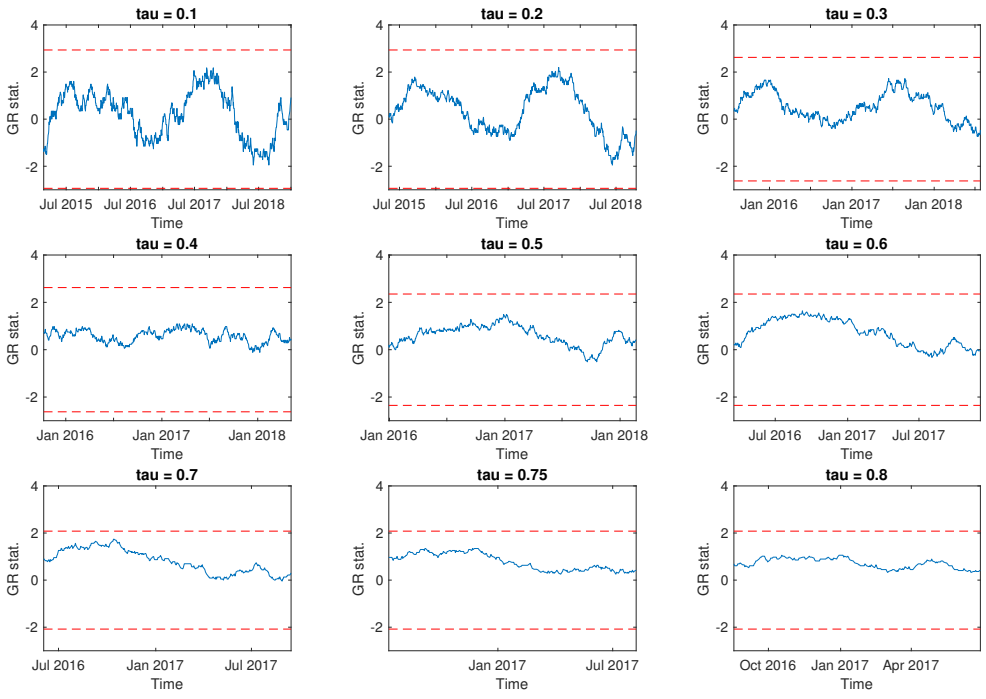


(ii) USD/CHF

Figure 29: Histogram of the PIT residuals for ARMAX(2) and MAX(IRD)(left) versus Gaussian GARCH-in-mean (right). Upper panel: EUR. Lower panel: CHF.



(i) GR test for ARMAX(2) ($H_0 : |L_\Delta^{(3)}| \leq 0$)



(ii) GR test for ARMAX(IRD) ($H_0 : |L_\Delta^{(3)}| \leq 0$)

Figure 30: GR test statistic (blue) with $\tau \in [.1, .85]$ using $L_\Delta^{(3)} = \frac{1}{h} \sum_{j=t+1}^{t+h} (\hat{p}_j^* - \hat{p}_j^{*,AS}) \text{sign}(R_j)$ as loss function. If the statistic is above the rejection threshold (dashed red), we reject the null hypothesis $H_0 : |L_\Delta^{(3)}| \leq 0$. (i) ARMAX(2) and (ii) ARMAX(IRD). Both rejection thresholds are at the 5% test level.