Interest rate differentials and the dynamic asymmetry of exchange rates

J. Hambuckers$^{1,2,†}$ and M. Ulm$^2$

Abstract

In this paper, we revisit the predictive content of interest rates for daily exchange rate returns, taking into account dependencies of higher orders. In particular, we allow for time-varying asymmetries in the distribution of exchange rates. Then, we study the USD/EUR currency pair over the period 1999-2019. We find a dynamic asymmetry component to be significant and driven by interest rate differentials, but also by general uncertainty and past unexpected shocks. In line with recent currency crash theories, our study suggests that the high yield currency is more likely to appreciate, but is exposed to a higher likelihood of currency crashes. To assess the economic significance of our results, we introduce a directional forecasting approach derived from our model. We show that a trading rule based on these forecasts provides better in-sample and out-of-sample economic performance compared to benchmark models.

Keywords: Exchange rate, interest rate differential, asymmetry, currency crash, directional forecast, carry trade, GARCH, sinh-arcsinh.

JEL: F31, C53, C58.

$^1$ University of Liège, HEC Liège, Finance Department, 14 rue Louvrex, 4000 Liège, Belgium.
$^2$ Georg-August-Universität Göttingen, Chair of Econometrics, Humboldtallee 3, 37073 Göttingen, Germany.
$†$ Corresponding author: jhambuckers@uliege.be.
1 Introduction

The exact nature of the relationship between short term interest rate differential (IRD) and exchange rate returns is still an ongoing debate: on the one hand, economic theory postulates that IRD and foreign exchange rates are linked 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 carry trade (CT)\(^1\). Empirically, however, we observe the opposite: several studies (among others Gabaix and Maggiori [2015]) report significant returns of CT investment strategies and an appreciation of the high-yield currency over long periods of time, contradicting UIP. A potential explanation for this result is the existence of a currency crash premium.

This apparent contradiction, referred as the UIP or forward puzzle, is a longstanding question in the Finance literature. Early on, Meese and Rogoff [1983] notice that models based on IRD - but more generally macro-models - cannot beat a simple random walk in predicting future exchange rates, raising the question of its predictive content. 35 years later, few has changed in this regard, despite large progresses in term of data availability and econometrics techniques: 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 (an exception is Molodtsova and Papell [2009]). A similar conclusion is reached by Hsu et al. [2016] looking at profits made by trading strategies using various predictors, for a large panel of currencies.

Several 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, Ismailov and Rossi, 2018], but also misspecifications of the models traditionally used to conduct these forecasts [Cheung et al., 2005, Rossi, 2013, Ismailov and Rossi, 2018]. Indeed, whereas most models focus on conditional mean forecasts, exchange rates exhibit rather high-order dynamics and extremely weak or changing mean dynamics [Chung and Hong, 2007, Brunnermeier et al., 2008, Ismailov and Rossi, 2018]. In these conditions, tools like vector autoregressive models are ineffective [Herwartz, 2017] and other strategies should be considered, like modelling conditional skewness [Chung and Hong, 2007, Brunnermeier et al., 2008]. Moreover, most empirical studies focus on point forecasts of the conditional mean, and assess the predictive ability via mean-squared forecast errors criteria [Molodtsova and Papell, 2009]. However, from a financial perspective, directional forecasts and profit-based performance measures are more relevant [Christoffersen and Diebold, 2006, Chung and Hong, 2007, Hsu et al., 2010, Blaskowitz and Herwartz, 2011].

---

\(^1\)Carry trade is an investment strategy that consists in borrowing in a low interest rate currency and investing in another currency with a higher interest rate [Bakshi and Panayotov, 2013].
In light of these concerns, the purpose of the present paper is to revisit the link between IRD and daily exchange rate returns using an improved econometric strategy. In particular, we avoid establishing a causal or structural link between the level of future exchange rate returns and interest rates. Instead, we open the possibility for interest rates to have an effect on the density of future exchange rate returns, and more precisely on the asymmetry of this density. This set-up is therefore more parsimonious on its structural assumptions, and allows to test general assertions (“is a currency more likely to appreciate when its interest rate is relatively high?”) without too much restrictions. Thus, we can investigate if indeed IRD predicts the direction of change of exchange rates, if this direction is consistent with UIP or CT literature, and finally if IRD gives any indications on currency crashes.

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 skewness parameter varies over time according to a time series equation augmented with exogenous predictors. Therefore, our model answers to the critiques made in the empirical literature, since we account for a high-order dependence structures. We also allow for a time-varying risk premium via an in-mean component to match closely the empirical characteristics of the data.

We use this methodology to study into details the daily exchange rate returns of the Euro (EUR) vis-a-vis the US Dollar (USD) over the period 5th January 1999 - 28th March 2019. Our goal is thus to investigate if IRD is an important factor in the dynamics skewness of exchange rate returns, and to discuss the economic implications of this connection. We test various specifications of the skewness dynamics and consider additional control factors, like past innovations, past skewness parameter as well as the VIX. The inclusion of the latter allows us to discuss the potential effects of uncertainty on exchange rate returns, an important factor suggested in Brunnermeier et al. [2008], Ranaldo and Söderlind [2010], Menkhoff et al. [2012] and Ismailov and Rossi [2018]. To assess the significance of our findings, we look at the average return generated by a trading strategy based on our model, and compare it to several benchmarks. This approach is favored over statistical criteria like mean squared forecast error, since it focuses on economic significance. Last, we conduct a persistence analysis to assess if our results hold also out-of-sample. We use state-of-the-art tests to control for multiple testing and spurious correlation issues [Giacomini and White, 2006, Giacomini and Rossi, 2010, Hsu et al., 2016].

Although motivated primarily by empirical findings, the proposed econometric approach draws also on theoretical arguments recently put forward in the UIP and CT 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 [Rossi, 2013], we hypothesized instead that a large (absolute) IRD is indicative of a higher risk of a reverting mechanism, i.e. of a future depreciation of the funding currency: we assume the marginal effect of IRD to have an impact on the likelihood of a given move (appreciation, depreciation or crash), not on the move itself. This is achieved by allowing for local deviations due to a regression structure in the skewness dynamics, rather than at the mean level. Such a structure would be consistent with the absence of predictability for IRD in classical linear regression framework, but also with the existence of a predictive content of higher-orders. Moreover, Brunnermeier et al. [2008], Brunnermeier and Pedersen [2009] and Menkhoff et al. [2012] offer explanations for the financial channels connecting IRD, global uncertainty and exchange rate returns. They highlight that CT plays an active role on exchange rate markets and that the size of CT positions will be large when IRD is large. In this situation, one observe an appreciation of the investment currency instead of the depreciation predicted by UIP. In addition, funding constraints or an increase in global uncertainty (as measured, e.g. by the VIX) may lead to a currency crash due to the unwinding of those trading positions, or due to a flight-to-safety [Habib and Stracca, 2012]. This makes IRD and VIX variables of choices in a distributional model to capture these effects on the conditional distribution of exchange rate returns.

Our main conclusions are the following:

i. IRD, VIX and past unexpected shocks are important factors to model the dynamic skewness of USD/EUR exchange rate returns.

ii. An increase in IRD is associated with an increasing likelihood of appreciation of the investment currency, in line with the apparent success of CT strategies. However it comes at the price of an increasing risk in a large currency crash, a result consistent with the theoretical framework postulated by Fahri and Gabaix [2016].

iii. High values of the VIX are associated with a higher likelihood of appreciation of USD, but also with a bigger crash risk. We explain these results by the safe heaven characteristics of the US and the role played by a future laxer monetary policy.

iv. Past unexpected depreciation shocks are associated with a higher likelihood of appreciation, but also with a larger crash risk. This result suggests the existence of self-fulfilling mechanisms, as noticed in Habib and Stracca [2012].

v. The predictive content of these three factors is sufficiently strong to allow for significant economic gains when trading with a dynamic skewness model, both in- and out-of-sample.
As policy implications, our results suggest that, for USD/EUR, an economic policy favoring a persistent IRD exposes these economies to systemic issues like a brutal depreciation. It also highlights the role of self-fulfilling mechanisms and their interactions with exchange rate volatility in currency crashes, suggesting that the prevention of unexpected shocks in periods of high uncertainty reduces crash risk.

The rest of the paper is written 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 study the finite sample properties of the different procedures. In Section 4 we perform the empirical analysis and we conclude in Section 6.

2 Methodology

The fundamental feature of our econometric approach consists in the introduction of a time-varying asymmetry in the distribution of the stochastic shocks. In particular, we allow this asymmetry to be driven by IRD, as well as by a time series structure. Chung and Hong [2007] and Brunnermeier et al. [2008] emphasize the existence of conditional skewness in exchange rate returns, as well as its variation over time. Furthermore, Brunnermeier et al. [2008] find that IRD forecasts well future asymmetry.

To do so, we build upon a classical GARCH-type model, undoubtedly the most popular models for daily financial data [Engle, 1982, Bollerslev, 1986]. However, 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. The existing literature focuses on time-varying volatility [Engle, 1982, Bollerslev, 1986, Glosten et al., 1993, Hansen and Lunde, 2005, Francq and Zakoian, 2010], on the asymmetric response of volatility to the sign of stochastic shocks [Glosten et al., 1993], on leptokurtosis in the residuals [Engle and Gonzalez-Rivera, 1991, Hansen, 1994, Bai et al., 2003, Chen et al., 2008] and on the choice of a good error distribution [Klar et al., 2012, Stavroyiannis et al., 2012, Hambuckers and Heuchenne, 2017].

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] consider 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. [2009] 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]. Nevertheless, regarding exchange rates returns, the literature is 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 the present paper, we consider a different approach. Skewed-t, Pearson-type IV and Normal Inverse Gaussian, despite their flexibility, suffer from cumbersome constraints and numerical issues. For example, the skewed-t distribution is piecewise-defined and has restriction on its skewness and shape parameters that are difficult to handle in practice; the Pearson-type IV distribution is computationally expensive to use since it relies on complex Gamma and hypergeometric functions [Stavroyiannis et al., 2012]; whereas the Normal Inverse Gaussian distribution exhibits a flat likelihood function and has no closed-form solution for its quantile function [Barndorff-Nielsen and Blaesild, 1981]. In addition, these distributions are hardly 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, we consider instead 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 distributions previously cited, the standardized SH distribution has two parameters ($\epsilon$ and $\delta$) with interpretable meanings (asymmetry and shape), is centred on the Gaussian distribution (with $\epsilon = 0$ and $\delta = 1$) and has the single constraint $\delta > 0$. Moreover, it accounts for heavier and lighter tails than the normal, a feature not possible with the skewed-t distribution, and has all its moments that exist without additional restrictions. In the GARCH-SH model, the parameter $\epsilon$ is allowed to vary over time according to an ARMAX dynamic (i.e. an autoregressive-moving average structure with additional explanatory variables). The use of an ARMAX structures for the asymmetry parameter allows us to link the conditional distribution of exchange rate returns with relevant financial and economic factors. Thus, it enables to account for a dependence structure beyond the two first moments. Last, we allow the volatility level to 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 this quantity and the return of a currency. We detail the model and its essential features in the next subsections.

\footnote{Another constraint, although classical, is the finiteness of the parameters, needed to ensure that the distribution is proper.}
2.1 Model specification and interpretation

Our exchange rate model is described by the following set of equations: the log-return
\[ R_t = \log \left( \frac{S_t}{S_{t-1}} \right), \forall t \]
where \( S_t \) is the exchange rate at time \( t \), follows a multiplicative
heteroscedastic model of the form
\[
R_t = c + \lambda \sigma_t + r_t, \\
\sigma_t^2 = \omega + \alpha \sigma_{t-1}^2 + \beta r_{t-1}^2, \\
z_t[I_{t-1}] \overset{iid}{\sim} f(z_t; \epsilon_t, \delta_t|I_{t-1}), \\
\epsilon_t = g(I_{t-1}), \\
\delta_t = h(I_{t-1}),
\]

where \( c \) is a constant, \( \sigma_t \) the conditional variance and \( z_t \) the innovation at time \( t \) with
mean zero and unit variance. \( I_t \) denotes the information set up to time \( t \), composed of all
values of \( z_t \) and vectors of covariates \( x_t \) up to time \( t \). \( f(z_t; \epsilon_t, \delta_t|I_{t-1}) \) denotes the probability
density function (pdf) of the standardized sinh-arcsinh distribution with parameters \( \epsilon_t \)
and \( \delta_t \), conditional on \( I_{t-1} \). \( g(\cdot) \) and \( h(\cdot) \) are parametric functions linking asymmetry and
shape parameters to past information. Expressions for the pdf, the value of the location
and scale parameters in the standardized case as well as formula for the moments can be
found in Appendix A. Conditions stated in the same appendix ensure that \( \mathbb{E}(z_t) = 0 \) and
\( \mathbb{E}(z_t^2) = 1 \), so that \( \sigma_t^2 \) can be interpreted as the conditional variance of \( r_t \). The focus here
is on dynamic skewness.

As explained in Jones and Pewsey [2009], the SH distribution is conveniently built
around the Gaussian distribution such that, assuming a r.v. \( Y \sim N(0, 1) \), we can define
\( f(z; \epsilon, \delta) \) by the sinh-arcsinh transformation:
\[
Z = \sinh \left( \frac{\sinh^{-1}(Y) + \epsilon}{\delta} \right).
\]

Skewness increases with increasing \( \epsilon, \epsilon \in ]-\infty, +\infty[ \), positive skewness corresponding to
\( \epsilon > 0 \). Notice that positive (negative) skewness implies, for a standardized r.v., that
there is more probability mass below (above) zero. The kurtosis decreases with increasing
\( \delta, 0 < \delta < +\infty, \delta < 1 \) 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. One other 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. In our empirical application, we set \( \delta_t = \delta = \exp(b_0) \).
For $\epsilon_t$, we define eq. (5) as a function of past innovations $z_{t-1}$, past values of $\epsilon_{t-1}$ 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:

$$\epsilon_t = g(I_{t-1}) = a_0 + a_1 \epsilon_{t-1} + a_2 z_{t-1} + a_3 x_{t-1}.$$  

(8)

This equation can be modified or restricted in several ways. Assuming all parameters in (8) equal to zero, we are back to the symmetric case. Assuming that only $a_1 = a_2 = a_3 = 0$ leads to a model without dynamics but asymmetries, whereas assuming $a_1$ and $a_3$ to be zero leads to a model where only past innovations impact the asymmetry. On the other hand, 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, $\epsilon_t$ has converged towards its stationary value $\epsilon^* = a_0/(1 - a_1)$. Since we do not set an additional restriction linking $a_0$ and $a_1$, it exists an infinity of pairs $(a_0, a_1)$ that solve this equation. In practice, it results in an estimation that converges at random, depending on the starting value chosen for $\epsilon_0$. In our application, we assume that at least one of the other coefficients is always different from zero. In addition, stability of the process is fulfilled when $|a_1| < 1$.

From an economic perspective, eq. (8) can be used to study how explanatory variables and past stochastic shocks influence the distribution of exchange rate returns. In particular, we suggest to look at three quantities:

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

(9)

$$\rho^+_t = \mathbb{P}(z_t > q^+),$$  

(10)

$$\rho^-_t = \mathbb{P}(z_t < q^-),$$  

(11)

where $q^+ > 0$ is relatively large, $q^- < 0$ is small. Here, $\pi_t$ can be seen as the probability of a positive shock and is indicative of the likelihood of a depreciation of the home currency. Besides, $\rho^+_t$ and $\rho^-_t$ are measures of currency crash risk, and indicate if a sudden large depreciation (resp. appreciation) is likely. The effect of a marginal change in one component in eq. (8) is deduced from the sign of its regression coefficient. Table 1 summarizes the various scenarios. Notice here the important connection between $\epsilon_t$, $\pi_t$, $\rho^+_t$ and $\rho^-_t$: when $\epsilon_t$ is positive, the density is positively skewed (Figure 1, solid black line). It implies that $\pi_t < 0.5$ and that an appreciation is more likely than a depreciation. If $\epsilon_t$ becomes more positive (Figure 1, dashed red line), the density becomes more positively skewed, and $\pi_t$ becomes smaller (an appreciation is even more likely) but $\rho^+_t$ increases (the risk of a large depreciation increases). A similar reasoning holds for negative $\epsilon_t$ and $\rho^-_t$. 

8
\[
t_{x_{t-1}} - a_3 > 0 \quad t_{x_{t-1}} - a_3 < 0 \quad t_{x_{t-1}} - a_3 > 0 \quad t_{x_{t-1}} - a_3 < 0
\]

<table>
<thead>
<tr>
<th>[\epsilon_t]</th>
<th>[\pi_t]</th>
<th>[\rho_t^+]</th>
<th>[\rho_t^-]</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 1: Summary of the effect of a change in \(x_{t-1}\) on \(\epsilon_t\), \(\pi_t\), \(\rho_t^+\) and \(\rho_t^-\).

![Figure 1](image-url)  
Figure 1: Example of SH distribution 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.

### 2.2 Estimation procedure and inference

We estimate our model by means of maximum likelihood procedures. Denoting by \(\Theta\) the vector of all parameters in equations (1) to (6), \(y_T = \{R_t\}_{t=1,\ldots,T}\) the time series of observed log-returns and assuming conditional independence, the conditional log-likelihood function \(L(\Theta; y_T)\) is given by

\[
L(\Theta; y_T) = \frac{1}{T} \sum_{t=1}^{T} \log \left( \frac{1}{\sigma_t} f((R_t - c - \lambda \sigma_t)/\sigma_t; \epsilon_t, \delta_t|I_{t-1}) \right).
\]  

An estimator \(\hat{\Theta}\) of \(\Theta\) is obtained by maximizing numerically (12) with respect to \(\Theta\):

\[
\hat{\Theta} = \arg \max_\Theta L(\Theta; y_T),
\]
and subject to the usual GARCH constraints $\omega, \alpha, \beta > 0$ and $\alpha + \beta < 1$. In addition, we have $\delta_i > 0$. Regarding the other parameters, we do not set any constraints.  

Under correct specification of the model and usual stationarity conditions, the Fisher-Information matrix $H(\Theta)$ of (12) at $\hat{\Theta}$ can be used for testing the following null hypothesis

$$H_0: \theta_i = 0,$$

(14)

where $\theta_i$ is the $i^{th}$ element of $\Theta$. To do so, we use the Wald-type test statistic

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

(15)

where $\hat{\theta}_i$ is an estimator of $\theta_i$ and $\hat{\sigma}_{ii}^2$ is the $i^{th}$ diagonal element of $H^{-1}(\hat{\Theta})$. Under $H_0$, $w_i \sim N(0,1)$. In Section 3, we show that with time series of reasonable length, this approximation gives well-sized and respectably powerful tests. Similarly, restrictions in equations (8) can be tested using likelihood ratio (LR) test statistics of the type

$$LR = -2(\mathcal{L}(\Theta_0; y_T) - \mathcal{L}(\Theta_1; y_T)), $$

(16)

with $\Theta_0$ being a restricted version of $\Theta_1$. Under the null hypothesis of the restricted model being the true one, we have the usual result $LR \sim \chi^2_\nu$, $\nu$ being the number of restrictions.

### 2.3 Directional forecasts, performance measures and testing for superior ability

As noticed by Blaskowitz and Herwartz [2011], it is commonly accepted that information is helpful if it can be exploited to improve a decision making process. In economics, most decisions are based on forecasts of the direction of change of the variable of interest. In the specific context of exchange rate, monetary authorities and investors might be interested in the appreciation or depreciation of a currency. Therefore, we focus on daily directional forecasts (both in-sample and out-of-sample) of exchange rate returns, with the idea to assess the economic significance of a model with high-order dynamic.

Time-varying asymmetry is crucial for good directional forecasts [Liu, 2015]. In the framework of a multiplicative heteroscedastic models with a zero-mean like GARCH, if the distribution of the innovations is (dynamically) asymmetric, then tomorrow’s probability of a positive (resp. negative) return 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 determine an optimal forecasting strategy. An

---

3 Regarding the choice of a starting value $\epsilon_0$, 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 $\epsilon_t$ when $T \to +\infty$. 

10
easy analogy can be made: at each point in time, we are involved in a coin tossing bet, facing two choices - head or tail - whereas the time-varying probabilities of each result are not equal. It implies that if we knew these probabilities, we could choose the most likely outcome. At the contrary, if the conditional distribution is symmetric, there is an equal probability for each outcome, leaving us with no good forecasting strategy.

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

\begin{align}
  p_{t|t-1} &= 1 - \mathbb{P}(R_t < 0 | I_{t-1}), \\
  &= 1 - \mathbb{P}(c + \lambda \sigma_t + \sigma_t z_t < 0 | I_{t-1}), \\
  &= 1 - \mathbb{P}(z_t < -c/\sigma_t - \lambda | I_{t-1}), \\
  &= 1 - F(-c/\sigma_t - \lambda; \epsilon_t, \delta_t | I_{t-1}),
\end{align}

where $F(\cdot)$ denotes the SH cumulative distribution function (cdf), see Appendix A. Forecasts of $\hat{p}_{t|t-1}$, $t = 2, \ldots, T$ are obtained by plugging $\hat{\Theta}$ in (20). Then, the final directional forecast is obtained from the following indicator variable:

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

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

Notice, though, that in this framework, 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 his 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.

Translating the directional forecasts into a trading strategy, if $\hat{p}_{t|t-1} > .5$ (i.e. if the likelihood of depreciation of the home currency is above .5), one would take a short position i.e. own the foreign currency. At the contrary, if $\hat{p}_{t|t-1} < .5$ (i.e. if the likelihood of an appreciation is above .5), one would take a long position in the home currency, i.e. own USD. In case of a constant asymmetry parameter, the optimal trading strategy is to be either always in a short position (for a negative asymmetry) or always long (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

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

where $\hat{p}_j^*$ is given by equation (21), $1(\cdot)$ denotes an indicator function taking value 1 if the condition in parentheses is met, and $\text{sign}(\cdot)$ denotes the sign function. This criterion measures the raw performance of a model in a pure classification exercise. To assess the performance in term of CR, we use the independence test of Pesaran and Timmermann [2009] that accounts for serial correlation.

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

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

Diebold and Mariano [1995], Blaskowitz and Herwartz [2011] and Elliott and Timmermann [2016] argue that the use of 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 the 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. White [2000], Bajgrowicz and Scaillet [2012], Hambuckers and Heuchenne [2016] or Hsu et al. [2016].

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, whereas out-of-sample performances are compared with the test of Giacomini and White [2006] and the Fluctuation test of Giacomini and Rossi [2010]. We compare the performance of the GARCH-SH model to the random walk (RW), always-short (AS) and buy-and-hold (BH) approaches, as well as with various sub-specifications of the most complex model. The tests proposed by Giacomini and White [2006] and Giacomini and Rossi [2010] have the enormous advantage to cover explicitly loss functions that are based on direction-of-change, estimated parameters, and allow both the comparison of nested and non-nested

---

4Random walk directional predictions must be understood as predicting tomorrow’s direction of change using today’s sign of the return, whereas always-short and buy-and-hold strategies consist in always predicting a negative return or a positive one, respectively.
models. However, it is only valid in the cases of either a fixed estimation period or a rolling window estimation period, not in an expanding window context. Therefore we restrict our attention to this updating scheme of the parameters and conduct a persistence analysis in the idea of Bajgrowicz and Scaillet [2012].

Notice here that the performance measures are not used in any ways 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 goal to optimize directional forecasts. Thus, forecasting performance must be seen as a genuine by-product of the correctness of our model.

3 Simulation

We start by investigating the quality of the proposed maximum likelihood estimation procedure. Then, we study the size and power of the proposed hypothesis test given by (15), 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 2. For DGP1 to DGP3, we assume \( \lambda \) and \( c \) equal to zero, such that \( \mathbb{E}(R_t) = 0 \), whereas we introduce a mean structure in DGP4 to DGP6. When \( a_3 \neq 0 \), we assume that \( x_{t-1} \sim \text{iid } N(0, 1) \).

<table>
<thead>
<tr>
<th>Data generating processes</th>
<th>( \omega )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( c )</th>
<th>( \lambda )</th>
<th>( \delta )</th>
<th>( a_0 )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
<th>( a_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP1 ( 10^{-4} )</td>
<td>0.05</td>
<td>0.88</td>
<td>-</td>
<td>-</td>
<td>.8</td>
<td>-0.1</td>
<td>0.4</td>
<td>-0.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DGP2 ( 10^{-4} )</td>
<td>0.05</td>
<td>0.88</td>
<td>-</td>
<td>-</td>
<td>.8</td>
<td>-0.1</td>
<td>0.4</td>
<td>-0.9</td>
<td>0.85</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DGP3 ( 10^{-4} )</td>
<td>0.4</td>
<td>0.8</td>
<td>-</td>
<td>-</td>
<td>.85</td>
<td>-0.1</td>
<td>0.5</td>
<td>0.3</td>
<td>-0.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DGP4 ( 10^{-4} )</td>
<td>0.05</td>
<td>0.88</td>
<td>( 10^{-4} )</td>
<td>0.1</td>
<td>.8</td>
<td>-0.1</td>
<td>0.4</td>
<td>-0.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DGP5 ( 10^{-4} )</td>
<td>0.05</td>
<td>0.88</td>
<td>( 10^{-4} )</td>
<td>0.1</td>
<td>.8</td>
<td>-0.1</td>
<td>0.4</td>
<td>-0.9</td>
<td>0.85</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DGP6 ( 10^{-4} )</td>
<td>0.1</td>
<td>0.8</td>
<td>( 10^{-4} )</td>
<td>0.1</td>
<td>.85</td>
<td>-0.1</td>
<td>0.5</td>
<td>0.3</td>
<td>-0.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DGP7 ( 10^{-4} )</td>
<td>0.05</td>
<td>0.88</td>
<td>( 10^{-4} )</td>
<td>-0.1</td>
<td>.8</td>
<td>-0.1</td>
<td>0.5</td>
<td>0.1</td>
<td>0.15</td>
<td>0.3</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2: Values of the parameters considered in the different simulation set-ups.

Results are given in Tables 3 and 4. 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.
Table 3: Root-MSE divided by the value of the corresponding parameter, for DGP without a mean structure (DGP1 to DGP3).

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>No mean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T</td>
</tr>
<tr>
<td>DGP1</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td>3000</td>
</tr>
<tr>
<td>DGP2</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td>3000</td>
</tr>
<tr>
<td>DGP3</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td>3000</td>
</tr>
</tbody>
</table>

Table 4: Root-MSE divided by the value of the parameters, for DGP with a GARCH-in-Mean structure (DGP4 to DGP6).

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GARCH-in-Mean</td>
</tr>
<tr>
<td></td>
<td>T</td>
</tr>
<tr>
<td>DGP4</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td>3000</td>
</tr>
<tr>
<td>DGP5</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td>3000</td>
</tr>
<tr>
<td>DGP6</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td>3000</td>
</tr>
</tbody>
</table>

Now, we study the size and power of the suggested Wald-type tests for the skewness regression parameters and λ. We consider two explanatory variables $x_{t,1}$ and $x_{t,2}$ in eq. (8). 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 parameter of -0.4 (a value observed in our data). Baseline values of the parameters are given in Table 2.

5 We do not look at inference for the GARCH parameters since those are estimated under positivity constraints.
(DGP7). Afterwards, 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 2 summarizes our results, indicating good powers under the various scenarios. For $a_3$ and $a_4$, we obtain excellent sizes, whereas we reject a bit too often for $\lambda$, $a_1$ and $a_2$.

Figure 2: Rejection rates of Wald tests at the 5% test level, for various values of the parameters. Dashed: 5% threshold.
4 Empirical study

We start first by describing the data, then we conduct our regression analysis.

4.1 Data

The data are daily foreign exchange rates in U.S. dollar (USD) per unit of foreign currency for the Euro (EUR). 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 5 January, 1999 to 25 March 2019. We compute the log-return \( R_t = \log(S_t/S_{t-1}) \), where \( S_t \) is the nominal exchange rate at time \( t \). The final sample consists of 5,075 observations, where we removed dates for which the exchange rate data are missing.

Interest rates are 3-month London Inter-Bank Offered Rate (LIBOR) for the respective currencies. All Interest rates data have been retrieved from the website of the Federal Reserve of Saint-Louis. Missing LIBOR data are replaced by the previous observed rate. It concerned 71 and 115 days for EUR and USD, respectively. Ismailov and Rossi [2018] argue that the predictability of interest rates depends on 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 rate and the corresponding log-returns are plotted in Figure 3. The interest rates and the VIX are plotted in Figure 4. Several events such as negative LIBOR rates, soar of the VIX and the financial crisis might be sources of instabilities in the relationship between exchange rates and IRD (see the discussions in, e.g. Giacomini and Rossi [2010], Bacchetta and van Wincoop [2013] and Ismailov and Rossi [2018]). This preliminary observation motivates us to study the performance of the model over sub-periods of time in Section 5.

4.2 Interest rates, depreciation, and currency crashes

Our extended model enables us to investigate the relationship between IRD and exchange rate returns in higher orders. We start by fitting model (1)-(6) to the entire dataset and discussing the economic interpretation of our results. We use the following specifications of the general skewness equation given by (8):

\[
\text{CST: } \epsilon_t = a_0.
\]

\[
\text{ARX(IRD): } \epsilon_t = a_0 + a_1\epsilon_{t-1} + a_3\text{IRD}_{t-1}.
\]
Figure 3: Daily exchange (left) against USD and log-returns (right) over the period 5 January 1999 - 24 March 2019.

Figure 4: VIX (left) and 3-month LIBOR rates (right, EUR: solid, 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, hike of federal fund rate in 2016.
ARX(VIX): $\epsilon_t = a_0 + a_1 \epsilon_{t-1} + a_4 \text{VIX}_{t-1}$.

ARX(2): $\epsilon_t = a_0 + a_1 \epsilon_{t-1} + a_3 \text{IRD}_{t-1} + a_4 \text{VIX}_{t-1}$.

MA: $\epsilon_t = a_0 + a_2 z_{t-1}$.

MAX(IRD): $\epsilon_t = a_0 + a_2 z_{t-1} + a_3 \text{IRD}_{t-1}$.

MAX(VIX): $\epsilon_t = a_0 + a_2 z_{t-1} + a_4 \text{VIX}_{t-1}$.

MAX(2): $\epsilon_t = a_0 + a_2 z_{t-1} + a_3 \text{IRD}_{t-1} + a_4 \text{VIX}_{t-1}$.

ARMA: $\epsilon_t = a_0 + a_1 \epsilon_{t-1} + a_2 z_{t-1}$.

ARMAX(IRD): $\epsilon_t = a_0 + a_1 \epsilon_{t-1} + a_2 z_{t-1} + a_3 \text{IRD}_{t-1}$.

ARMAX(VIX): $\epsilon_t = a_0 + a_1 \epsilon_{t-1} + a_2 z_{t-1} + a_4 \text{IRD}_{t-1}$.

ARMAX(2): $\epsilon_t = a_0 + a_1 \epsilon_{t-1} + a_2 z_{t-1} + a_3 \text{IRD}_{t-1} + a_4 \text{VIX}_{t-1}$.

The variable IRD$_t$ is defined as LIBOR$_t^{USD} - \text{LIBOR}_t^{EUR}$, such that positive (resp. negative) values correspond to situations where USD is the investment (resp. funding) currency. The parameter $\delta_t$ is assumed constant over time. We assume either that $E_t(R_t) = 0$ (classical GARCH) or that $E_t(R_t) = c + \lambda_t \sigma_t$ (GARCH-in-Mean). For every specification, we report the estimated regression coefficients and the value of the likelihood function. We perform LR tests between the most complex model (ARMAX(2)) and nested alternatives, as well as Wald tests for the mean and skewness parameters. Then, we report CR and $\hat{m}$ as measures of the economic significance of our results. The results for the GARCH-in-Mean specifications are displayed in Table 5. In light of their poor performance, we do not discuss the results for GARCH specifications without mean components$^6$.

Using the LR test, we find the most complex specification (GARCH-in-Mean ARMAX with both IRD and VIX as predictors) to be superior to the considered alternatives. All LR tests reject the null of no differences with restricted specifications and the QQ-plot of the pseudo residuals indicates an excellent fit of this model (Figure 5). Therefore, we concentrate on the interpretation of that model. For the sake of clarity, we display the estimated coefficients in Table 6 a second time.

$^6$Those are available upon demand.
Table 5: The upper panel display the estimated parameters for the various specifications. Estimated standard errors are reported in parentheses whereas *, ** and *** indicate Wald's test statistics being significantly different from zero at the 10%, 5% and 1% test level, respectively. The bottom panel reports the correct classification rate (CR), mean return ($\hat{m}$, in equivalent yearly return), LR tests and the p-values of the associated tests (SSPA and PT09). Int. SSPA refers to SSPA test considering ARMAX(2), ARMAX(IRD), MAX(2), AS and RW$^-$ only. For buy-and-hold and random walk strategies, we report the performance of the best side of the strategies (here AS and RW$^+$ for the mean return criterion, RW$^+$ for the classification rate). Numbers in parentheses are p-values of the SSPA (Full and Int.) and PT09 tests.
The results 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 a constant in the mean that is close to zero. Moreover, we observe a negative but insignificant mean parameter for the volatility. This suggests that, all else equal, an increase in contemporaneous exchange rate volatility translates into more negative returns on average, i.e. an appreciation of USD against EUR. 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, like the last financial crisis [Habib and Stracca, 2012].

![Figure 5: QQ-plot of the residuals for ARMAX(2), fitted on the complete period.](image)

<table>
<thead>
<tr>
<th>GARCH-in-Mean ARMAX(2)</th>
<th>( \omega )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( c )</th>
<th>( \lambda )</th>
<th>( \delta )</th>
<th>( a_0 )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
<th>( a_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 0.000 )</td>
<td>( 0.029 )</td>
<td>( 0.968 )</td>
<td>( 0.000 )</td>
<td>( -0.032 )</td>
<td>( 0.772 )</td>
<td>( -0.043 )</td>
<td>( 0.631 )</td>
<td>( 0.042 )</td>
<td>( 1.521 )</td>
<td>( 0.171 )</td>
<td></td>
</tr>
<tr>
<td>( (0.414) )</td>
<td>( (0.108) )</td>
<td>( (0.004) )</td>
<td>( (0.000) )</td>
<td>( (0.068) )</td>
<td>( (0.183) )</td>
<td>( (0.060) )</td>
<td>( (0.153) )</td>
<td>( (0.013) )</td>
<td>( (0.451) )</td>
<td>( (0.061) )</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Estimated coefficient for GARCH-in-Mean ARMAX(2) model. \( a_0, a_1, a_2, a_3 \) and \( a_4 \) are the constant, AR, MA, IRD and VIX parameters in the skewness equation (8). Standard errors are put in parentheses.

In the following, we focus on the estimated coefficient of the regression equation for the skewness parameter. All parameters are found to be significant at the 5% test level.
4.2.1 Effect of IRD

We look first at the link between IRD and the distribution of the returns, captured by $a_3$. In particular, we look at the marginal effect of a change in IRD on the parameters $\pi_t$, $\rho^+_t$ and $\rho^-_t$. Whereas $\pi_t$ indicates the probability of a depreciation of USD, $\rho^+_t$ and $\rho^-_t$ are measures of currency crash risk. We find that an increase in USD interest rates is associated with an increase in the probability of appreciation of USD. This result suggests that a large and positive IRD opens the possibility for profitable carry trades, whereby USD is the investment currency and EUR the funding currency. As a consequence, more market participants borrow EUR at a small rate and buy USD to invest them at a higher rate, which leads to an appreciation of USD. This finding is in line with Brunnermeier et al. [2008], who report a positive connection between IRD and the activity of non-commercial traders on the futures market, which they use as a proxy for carry trade activities. Moreover, these results fit into the theoretical framework of Fahri and Gabaix [2016], where the currency of the high interest rate country appreciates, conditional on no disaster occurring.

Simultaneously, though, a larger positive IRD has an opposite effect on $\rho^+_t$: it becomes more likely to observe an extremely positive shock, synonym of a large depreciation of USD\(^7\). This is also in line with Brunnermeier et al. [2008], Fahri and Gabaix [2016] and Jurek [2014] who associate IRD with currency crash risk: the larger the IRD, the stronger are 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 CT: 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 highlight the potential endogeneity of the reverting mechanism, as suggested in Fahri and Gabaix [2016]: through their fear, investors turn themselves into a force of realignment that leads to a crash.

Figure 6 shows the empirical connection between IRD and the probabilities $\pi_t$, $\rho^+_t$ and $\rho^-_t$ as unraveled by the model: the larger IRD, the smaller the probability of an overall depreciation, but the larger the probability of an extreme depreciation.

\(^7\)A similar reasoning holds if EUR is the funding currency, leading to a increase in $\rho^-_t$ when IRD becomes more negative.
4.2.2 Effect of uncertainty

Besides IRD, the model includes the VIX as an additional predictor and proxy for global uncertainty. The corresponding coefficient $a_4$ is estimated to be positive. Hence, if the VIX increases, an appreciation of USD becomes more likely. Simultaneously, though, the likelihood of a currency crash increases as well. These results can be analyzed in light of the literature connecting global uncertainty, liquidity and exchange rates. In particular, Menkhoff et al. [2012] suggest the existence of a compensation mechanism for disaster risk related to an investment currency that leads low interest rate currencies to under-perform, except in periods of exceptionally high global exchange rate uncertainty. As a result, CT is a profitable strategy in low to moderate uncertainty periods. However, profitability arises from a compensation for exposure to currency crashes in extremely volatile periods. The proposed channel is that crashes are the consequences of the sudden unwinding of carry trade positions due to a decrease in risk appetite and funding liquidity [Bakshi and Panayotov, 2013], characteristics well captured by high values of the VIX [Collin-Dufresne et al., 2001, Adrian and Shin, 2010, Chung and Chuwonganant, 2014]. This is also the conclusion of Brunnermeier et al. [2008]. Hence, our model emphasizes a similar effect: with increasing uncertainty, an appreciation of USD is more likely, but the risk of a huge depreciation increases as well. This is emphasized in periods of high exchange rate volatility. A limitation of the present set-up is that we do not formally consider potential asymmetries that arise if there is a switch between funding and investment currencies: we might only capture an effect related to which currency is used for funding "on average". Nevertheless, an average appreciation of USD over EUR in time of financial stress 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

\[8\text{In the present case, USD is the investment currency most of the time}\]
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. Our results suggests that it 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, synonym of high risk taking in that economy.

4.2.3 Effect of past unexpected shocks

Finally, we look at the effect of past innovations on the asymmetry. We find $a_2$ to be positive and significant. 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 one day, the more likely the appreciation the next day on average but also the higher the likelihood of a very large depreciation. We suggest that this 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, positive unexpected depreciation are followed by appreciations. 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 even stronger crashes. Such a mechanism is consistent not only with clustered volatility, but also clustered signs or periods of time where several large crashes in the same direction take place.

4.2.4 Summary

We draw here four 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. Similarly, an increase in global uncertainty or risk aversion, as measured by the VIX, is positively associated to a higher likelihood of appreciation for USD. However, this "local" higher return of USD comes at the price of a larger currency crash risk. This is driven both by an increasing asymmetry of the shocks, and a higher multiplicative effect of the volatility on the shocks. Third, an increase in contemporaneous volatility increases the expected return in favor of USD. Last, we observe that a large unexpected depreciation (resp. appreciation) makes a large depreciation (resp. depreciation) more likely in the future, suggesting the existence of a self-fulfilling mechanism.
5 Economic performance

In the following, we investigate to which extend the statistical results translate into economic gains. We aim at answering the following questions: could an investor have guessed well 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 yield positive returns?

5.1 In-sample performance

We consider first the in-sample performance. Therefore we have a clear advantage over a realistic trader who would use less information. However, we do not use any profit measure, nor classification rate to fit our model.

Figure 7 panel (a) shows the one-step-ahead predicted probabilities of a positive return i.e. \( \{\hat{p}_{t|t-1}\}_{t=2}^{T} \). If the prediction is above .5, we conclude that the will observe a depreciation of USD in \( t \). For ARMAX(2), we observe a correct classification rate of 52.83% over the complete period (Table 5). Using the test of Pesaran and Timmermann [2009], we reject the null hypothesis of independence between the realized signs of the returns and our forecasts\(^9\). Looking at \( \hat{m} \), we are able to derive an average return\(^{10}\) of 5.45%. Figure 7 panel (b) shows 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, AS and BH whose performances in equivalent yearly log-return range between -.22% and .22%. Moreover, compared to the simpler specifications tested, we also achieve a higher performance on the \( \hat{m} \) criterion. Applying the SSPA test of Hsu et al. [2010] on ARMAX(2), ARMAX(IRD), MAX(2) and the benchmarks with a positive return (AS and RW\(^-\)), we find the ARMAX(2) model to have a performance significantly different from zero at the 5% test level (\( Int. \) SSPA in Table 5). Pooling all alternative models and benchmarks, the performance is significant at the 10% test level (p-val. = .082). Finally, testing if the difference with RW\(^-\) is zero, we reject again that hypothesis, at the 5% test level for ARMAX(2).

Now, we clearly see from the left panel of Figure 7 that the performance is especially good before the crisis and after the increase in USD rate 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 and 2016 IRD stays very close to 0. Thus, it is not surprising to observe such results,

---

\(^9\)Notice that the overall proportion of positive returns is 49.14%. Therefore we exhibit a correct classification rate, compared to a naive strategy, improved by 52.83/50.86 - 1 = 3.87%.

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

\(^{11}\)Exact dates are the following: 6/01/1999, 23/10/2008, 15/12/2015 and 25/03/2019.
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 direction in the exchange rate data (either appreciation or depreciation) during these periods, whereas the intermediate period is characterized by an absence of direction. To test for a potential structural breaks, we use a CUSUM test in the idea of Kulperger and Yu [2005] and Andreou and Ghysels [2002] 12. We are not able to detect a significant structural break, however a graphical inspection of the CUSUM process in Figure 11 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. The performance over the pre-crisis period increases drastically, reaching almost 15% in yearly percentages for \( \hat{m} \) and 55% for CR. At the contrary, we observe a change of sign for the mean component and smaller coefficients during the crisis period, as well as a small \( \hat{m} \). The last period is characterized by signs of the mean parameters similar to the ones of the first period. However, the effects of IRD and the VIX are small and not significant. The economic performance is positive (\( \hat{m} = 2.49\% \)) but not significantly different from zero.

5.2 Out-of-sample performance

So far, we have studied the performance of our model using information normally not available to an investor: we have used the complete data to estimate the parameters, not only past data. To restrict the information set in a more realistic way, we compute (pseudo) out-of-sample forecasts of the direction of change. We assess the performance of

---

12See Appendix B for technical details.
Table 7: 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. The reported performance marked by † are related to the BH strategy. Int. SSPA refers to SSPA tests including the best benchmarks, as well as MAX(2) and ARMAX(IRD) strategies.
Figure 8: Evolution over time of an initial investment of 1 USD with reinvestment of the proceed, for the three subperiods (a) 5/01/1999 - 1/12/2008, (b) 2/12/2008 - 14/12/2015 and (c) 15/12/2015 - 29/03/2019. In red, performance related to our model.
the ARMAX(2), ARMAX(IRD), ARMAX(VIX), MAX(2) and ARMA models in light of the benchmark forecasting strategies (AS/BH and RW−/RW+). To account for the effect of choosing the side of the benchmark (i.e. if we use the short/long or negative/positive RW strategy), we use the best performing benchmark of the in-sample rolling windows. To test for significant differences in forecasting abilities, we use a [Diebold and Mariano, 1995] test. Models are compared using the Conditional Predictive Ability test proposed by Giacomini and White [2006] and the Fluctuation test of Giacomini and Rossi [2010]. Those tests are denoted by DM, GW and GR, respectively. In particular, GR test allows to control for changes in forecasting performance over time, contrary to other tests that only take into account the average performance\(^{13}\).

We use 4000 observations as an initial training sample (from 5 January 1999 to 1 December 2014). Thus, we include both non-crisis and crisis data, as well as the period with low interest rates. 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.

Performance and tests results are displayed in Table 8. For ARMAX(2), we obtain an average performance of 5.19%, in equivalent yearly mean log-returns, over the forecast horizon. Testing if the profit is not different from zero, we reject this hypothesis at the 10% test level with the GW test, and at the 5% test level with GR for \(\tau \in \{.3,.5,.75\}\). DM is inconclusive. The performance of the random walk strategy is significantly negative at some point in time, as indicated by the GR test (\(\hat{m} = -1.26\%\)). Results are mostly inconclusive for the other specifications. Figure 9, panels (a) to (c) show the local performance computed with the GR test for ARMAX(2), ARMAX(VIX) and RW for \(\tau \in \{.3,.5,.75\}\). The model performs particularly well around 2017, during the hike of interest rates. Recent performance is more elusive (Figure 9, panel (d)). Looking at significant differences with respect to the RW strategy, we find ARMAX(2) to be significantly better over some periods of time, as indicated by the GR tests (\(\tau = .3\) and .5).

Thus, the performance observed in-sample is mostly preserved out-of-sample for the ARMAX(2) model. Again, it seems that the combination of IRD and VIX carries information about the future direction of change. It appears to be especially economically significant when IRD exhibits a changing intensity, as over the period 2016 - 2018. However, as soon as the difference is stabilized, we do not make any profit.

\(^{13}\)GR test can be seen as testing repeatedly for zero local differences in forecasting performance, using a rolling window of data containing a fraction \(\tau\) of the total, or as a sequence of DM tests.
USD/EUR Out-of-sample performance ($H_0 : |\hat{m}| \leq 0$)

<table>
<thead>
<tr>
<th></th>
<th>$\hat{m}$</th>
<th>GR($\tau = .3$)</th>
<th>GR($\tau = .25$)</th>
<th>GR($\tau = .75$)</th>
<th>GW</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMAX(2)</td>
<td>5.19%</td>
<td>2.67**</td>
<td>2.66**</td>
<td>2.51**</td>
<td>4.94*</td>
<td>1.22</td>
</tr>
<tr>
<td>-</td>
<td>(&lt; 0.05)</td>
<td>(&lt; 0.05)</td>
<td>(&lt; 0.05)</td>
<td>(0.085)</td>
<td>(0.112)</td>
<td></td>
</tr>
<tr>
<td>ARMAX(IRD)</td>
<td>-0.19%</td>
<td>1.69</td>
<td>1.59</td>
<td>1.29</td>
<td>0.00</td>
<td>-0.043</td>
</tr>
<tr>
<td>-</td>
<td>(&gt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(0.965)</td>
<td>(0.483)</td>
<td></td>
</tr>
<tr>
<td>ARMAX(VIX)</td>
<td>1.69%</td>
<td>1.73</td>
<td>1.97</td>
<td>1.36</td>
<td>0.15</td>
<td>0.38</td>
</tr>
<tr>
<td>-</td>
<td>(&gt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(0.695)</td>
<td>(0.352)</td>
<td></td>
</tr>
<tr>
<td>MAX(2)</td>
<td>-0.60%</td>
<td>2.21</td>
<td>2.28</td>
<td>1.19</td>
<td>0.02</td>
<td>-0.14</td>
</tr>
<tr>
<td>-</td>
<td>(&gt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(0.889)</td>
<td>(0.444)</td>
<td></td>
</tr>
<tr>
<td>ARMAX(VIX)</td>
<td>1.52%</td>
<td>2.18</td>
<td>2.05*</td>
<td>1.54</td>
<td>0.23</td>
<td>0.33</td>
</tr>
<tr>
<td>-</td>
<td>(&gt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(0.923)</td>
<td>(0.381)</td>
<td></td>
</tr>
<tr>
<td>RW</td>
<td>-1.26%</td>
<td>-2.47**</td>
<td>-2.04*</td>
<td>-0.68</td>
<td>-0.16</td>
<td>-0.30</td>
</tr>
<tr>
<td>-</td>
<td>(&lt; 0.10)</td>
<td>(&lt; 0.10)</td>
<td>(&lt; 0.10)</td>
<td>(0.893)</td>
<td>(0.369)</td>
<td></td>
</tr>
<tr>
<td>BH</td>
<td>-2.26%</td>
<td>-1.75</td>
<td>-1.27</td>
<td>-0.66</td>
<td>-0.31</td>
<td>-0.54</td>
</tr>
<tr>
<td>-</td>
<td>(&gt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(0.855)</td>
<td>(0.293)</td>
<td></td>
</tr>
</tbody>
</table>

Out-of-sample performance ($H_0 : |\Delta \hat{m}| \leq 0$)

<table>
<thead>
<tr>
<th></th>
<th>$\Delta \hat{m}$</th>
<th>GR($\tau = .3$)</th>
<th>GR($\tau = .25$)</th>
<th>GR($\tau = .75$)</th>
<th>GW</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMAX(2)</td>
<td>6.45%</td>
<td>2.44*</td>
<td>2.66**</td>
<td>1.60</td>
<td>2.12</td>
<td>0.92</td>
</tr>
<tr>
<td>-</td>
<td>(&lt; 0.10)</td>
<td>(&lt; 0.05)</td>
<td>(&lt; 0.10)</td>
<td>(0.346)</td>
<td>(0.179)</td>
<td></td>
</tr>
<tr>
<td>ARMAX(IRD)</td>
<td>1.07%</td>
<td>1.82</td>
<td>1.90</td>
<td>0.97</td>
<td>0.44</td>
<td>0.15</td>
</tr>
<tr>
<td>-</td>
<td>(&gt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(0.803)</td>
<td>(0.439)</td>
<td></td>
</tr>
<tr>
<td>ARMAX(VIX)</td>
<td>2.95%</td>
<td>2.10*</td>
<td>2.19</td>
<td>1.12</td>
<td>0.52</td>
<td>0.43</td>
</tr>
<tr>
<td>-</td>
<td>(&gt; 0.10)</td>
<td>(&lt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(0.769)</td>
<td>(0.334)</td>
<td></td>
</tr>
<tr>
<td>MAX(2)</td>
<td>0.66%</td>
<td>2.10</td>
<td>2.08*</td>
<td>0.82</td>
<td>0.14</td>
<td>0.10</td>
</tr>
<tr>
<td>-</td>
<td>(&gt; 0.10)</td>
<td>(&lt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(0.934)</td>
<td>(0.4622)</td>
<td></td>
</tr>
<tr>
<td>ARMA</td>
<td>2.78%</td>
<td>1.98</td>
<td>2.05*</td>
<td>1.15</td>
<td>0.14</td>
<td>0.39</td>
</tr>
<tr>
<td>-</td>
<td>(&gt; 0.10)</td>
<td>(&lt; 0.10)</td>
<td>(&gt; 0.10)</td>
<td>(0.931)</td>
<td>(0.384)</td>
<td></td>
</tr>
</tbody>
</table>

Table 8: The first column gives the out-of-sample performance, using a rolling window of 4,000 observations updated every 5 days. The other columns display the test statistics and associated p-values for GR, GW and DM tests. All tests are one-sided. Upper panel: tests of a positive (resp. negative) average profit. Lower panel: tests of a positive (resp. negative) difference in average profit with respect to a RW strategy. Positive values indicate a performance superior to RW. Notice that for the GR and DM tests, lines RW and BH, the reported p-values refer to testing $H_0 : \Delta \hat{m} \geq 0$ due to the negative sign of the test statistic. For DM, we use 8 lags in the computation of the HAC estimator.
Figure 9: Local performance measure of the GR tests using rolling windows with proportions (a) 30%, (b) 50% and (c) 75% of the total sample. The x-axis gives the dates of the central observation in the window. (d) Evolution over time of an initial investment of 1 USD with reinvestment of the proceeds. Red: ARMAX(2). Solid black: ARMAX(VIX). Solid dotted black: RW. Dashed: critical value of one-sided GR test with a 10% test level.
6 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 USD/EUR exchange rate, we obtain results consistent with the literature on currency crash risk, carry trade and safe heaven currencies. First, we find that the high yield currency is more likely to appreciate, but that 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], and suggests an 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. Relying on explanations advanced in Menkhoff et al. [2012], Bekaert et al. [2013] and Habib and Stracca [2012], we presume that liquidity shortage and increasing risk aversion lead investors towards USD, in the idea of a safe heaven currency. However, this increasing uncertainty leads also to future laxer monetary policies and risk-taking behaviours, increasing in turn the likelihood of currency crashes. Third, our results suggest the existence of a self-fulfilling mechanism 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 both in- and out-of-sample that the detected effects are sufficiently large to generate significant economic gains. This result highlights the importance of our findings. 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.

As policy implications, our results suggest that, for USD/EUR, an economic policy favoring an increase in IRD exposes these economies to systemic issues like a brutal depreciation. Moreover, they 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 reduces crash risk.

Finally, a fundamental innovative aspect of the present paper consists in investigating the high-order dependence structure of exchange rates. Therefore we are able to connect 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 promising 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 study a larger set of currencies and predictors, and see if our results generalize.

Acknowledgments

J. Hambuckers acknowledges the financial support of the DFG via the RTG 1644 Scaling problems in Statistics. The authors are grateful to P.-H. Hsu for providing them with A. Yu-Chin implementation of the SSPA test.

A Appendix: the sinh-arcsinh distribution

The pdf of the sinh-arcsinh distribution is given by

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

where

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

with $K$ being the modified Bessel function of the second kind. $\xi$ and $\eta$ are the location and the scale parameters, respectively, whose values are fixed to ensure zero mean and unit variance. The cumulative distribution function (cdf) $F(z; \epsilon, \delta)$ is obtained from the transformation given in (7) 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 $\Phi^{-1}$ is the quantile function of the Gaussian distribution.
Under equations (5) to (24), the skewness and kurtosis of $z_t$ are given by

$$SK_t = \frac{1}{4} \left\{ \sinh(3\epsilon_t/\delta_t)P_{3/\delta_t} - 3\sinh(\epsilon_t/\delta_t)P_{1/\delta_t} \right\},$$

$$KU_t = \frac{1}{8} \left\{ \cosh(4\epsilon_t/\delta_t)P_{4/\delta_t} - 4\cosh(2\epsilon_t/\delta_t)P_{2/\delta_t} + 3 \right\},$$

where $P_q$ is define by equation (24). Hence, one can observe that both quantities depend on both parameters. The response surfaces for skewness and kurtosis, for various values of $\epsilon$ and $\delta$, are displayed in Figure 10.

As shown on Figure 10 (left panel), $\delta$ 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 $\delta$. Similarly to what is done for $\epsilon_t$, the following equation can be used:

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

**B Appendix: Testing for structural breaks via CUSUM tests**

In this appendix, we provide technical details regarding the CUSUM tests used in Section 4. In particular, we discuss the necessary modifications to be done, 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 to work instead with Gaussian pseudo-residuals $\hat{u}_t$, based on the inverse of the sinh-arcsinh transform given by (7). 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 $\epsilon_t$ and $\delta_t$ are finite.

Then, we suggest to use the following test statistic, similar to the one provided in Kulperger and Yu [2005]:

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

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 (B) is beyond the scope of the paper. Based on the theoretical arguments enumerated previously, we will use the (approximated) results

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

i.e. that (B) converges to the supremum of a Brownian bridge. Size and power are studied in a simulation study in the next section.

Additionally, since we might face several breaks in the time series, we might 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 3, we briefly study the size of this test. Results are displayed in Table 9. CUSUM tests appear slightly under-sized.
Applying this test on the complete sample, we cannot reject the null of no structural breaks in any of the four first moments. However, a graphical inspection of the CUSUM process shows rather large instabilities (Figure 11 for $k = 3$).

![CUSUM test statistic](image)

Figure 11: CUSUM process over time, with $k = 3$.  

---

Table 9: 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).

<table>
<thead>
<tr>
<th>DGP</th>
<th>$T$</th>
<th>$k=2$</th>
<th>$k=3$</th>
<th>$k=4$</th>
<th>DGP</th>
<th>$T$</th>
<th>$k=2$</th>
<th>$k=3$</th>
<th>$k=4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP1</td>
<td>500</td>
<td>0.024</td>
<td>0.04</td>
<td>0.012</td>
<td>DGP4</td>
<td>500</td>
<td>0.04</td>
<td>0.048</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>1500</td>
<td>0.032</td>
<td>0.032</td>
<td>0.026</td>
<td></td>
<td>1500</td>
<td>0.044</td>
<td>0.038</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>0.044</td>
<td>0.052</td>
<td>0.05</td>
<td></td>
<td>3000</td>
<td>0.042</td>
<td>0.052</td>
<td>0.04</td>
</tr>
<tr>
<td>DGP2</td>
<td>500</td>
<td>0.024</td>
<td>0.042</td>
<td>0.034</td>
<td>DGP5</td>
<td>500</td>
<td>0.028</td>
<td>0.036</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>1500</td>
<td>0.03</td>
<td>0.046</td>
<td>0.032</td>
<td></td>
<td>1500</td>
<td>0.042</td>
<td>0.058</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>0.046</td>
<td>0.04</td>
<td>0.036</td>
<td></td>
<td>3000</td>
<td>0.042</td>
<td>0.034</td>
<td>0.036</td>
</tr>
<tr>
<td>DGP3</td>
<td>500</td>
<td>0.042</td>
<td>0.038</td>
<td>0.03</td>
<td>DGP6</td>
<td>500</td>
<td>0.04</td>
<td>0.036</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>1500</td>
<td>0.036</td>
<td>0.038</td>
<td>0.028</td>
<td></td>
<td>1500</td>
<td>0.07</td>
<td>0.044</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>0.044</td>
<td>0.042</td>
<td>0.032</td>
<td></td>
<td>3000</td>
<td>0.024</td>
<td>0.042</td>
<td>0.022</td>
</tr>
</tbody>
</table>
References


