Testing Uncovered Interest Parity: A Continuous-Time Approach*

Antonio Diez de los Rios
Bank of Canada
diez@bankofcanada.ca

Enrique Sentana
CEMFI
sentana@cemfi.es

June 2010

Abstract

Nowadays researchers can choose the sampling frequency of exchange rates and interest rates. If the degree of overlap is large relative to the sample size, standard GMM asymptotic theory provides unreliable inferences in UIP regression tests. We specify a continuous-time model for exchange rates and forward premia robust to temporal aggregation, unlike existing discrete-time models. We test the UIP restrictions on the continuous-time model parameters and propose a novel specification test that compares estimators at different frequencies. Our results based on correctly specified models provide little support for UIP at both short and long horizons.

JEL Classification: F31, G15.

Keywords: Exchange Rates, Forward Premium Puzzle, Hausman Test, Interest Rates, Orstein-Uhlenbeck Process, Temporal Aggregation.

*We would like to thank Jason Allen, Manuel Arellano, Marcus Chambers, Marcos Dal Bianco, Lars Hansen, Bob Hodrick, Javier Gardeazabal, Nour Meddahi, Eric Renault, Frank Schorfheide (the editor), and three anonymous referees, as well as audiences at the European Summer Meeting of the Econometric Society (Stockholm, 2003), European Winter Meeting of the Econometric Society (Madrid, 2003), Finance Forum (Alicante, 2003), Symposium on Economic Analysis (Murcia, 2005), Bank of Canada, Graduate Institute of International Studies (Geneva), Universidad de Alicante, Universidad Autónoma de Barcelona, Universidad Carlos III (Madrid), Université de Montréal and Universidad de Valencia for useful comments and suggestions. We are also grateful to Sebastian Schich, Andy Haldane and especially Mike Joyce for their help with the data. Special thanks are also due to Angel León who co-wrote the first draft of this paper with us. Of course, we remain responsible for any remaining errors. Financial support from the Spanish Ministry of Science and Innovation through grant ECO 2008-00280 (Sentana) is gratefully acknowledged. *Address for correspondence:* Enrique Sentana, CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Canada.

1 Introduction

Over the last thirty years the majority of studies have rejected the hypothesis of uncovered interest parity, which in its basic form implies that the (nominal) expected return to speculation in the forward foreign exchange market conditioned on available information should be zero. Many of these studies have regressed ex post rates of depreciation on a constant and the forward premium, rejecting the null hypothesis that the slope coefficient is one. In fact, a robust result is that the slope is negative. This phenomenon, known as the "forward premium puzzle", implies that, contrary to the theory, high domestic interest rates relative to those in the foreign country predict a future appreciation of the home currency. In fact, the so-called "carry trade", which involves borrowing low-interest-rate currencies and investing in high-interest-rate ones, constitutes a very popular currency speculation strategy developed by financial market practitioners to exploit this "anomaly" (see Burnside et al. 2006). However, this is not by any means a risk-free strategy: Julian Robertson's Tiger Fund lost $2 billion in 1998 on the unraveling of US$/Yen carry trade positions that followed the Russian default and the subsequent LTCM crisis.

While some authors have argued that the empirical rejections found could be due to the existence of a rational risk premium in the foreign exchange rate market, "peso problems", or even violations of the rational expectations assumption, the focus of our paper is different.[1] We are interested in assessing whether existing tests of uncovered interest parity provide reliable inferences. In this sense, it is interesting to emphasize that the empirical evidence against uncovered interest parity has been lessened in more recent studies. In particular, Flood and Rose (2002) find that this hypothesis works better in the 1990's, Bekaert and Hodrick (2001) find that the evidence against uncovered interest parity is much less strong under finite sample inference than under standard asymptotic theory, while Baillie and Bollerslev (2000) and Maynard and Phillips (2001) cast some doubt on the econometric validity of the forward premium puzzle on account of the highly persistent behaviour of the forward premium.

In this paper, we focus instead on the impact of temporal aggregation on the statistical properties of traditional tests of uncovered interest parity, where by temporal aggregation we mean the fact that exchange rates evolve on a much finer time-scale than the frequency of observations typically employed by empirical researchers. While in many areas of

[1]See Lewis (1989) for details of the "peso problem approach", and Mark and Wu (1998) for a model that adapts the overlapping-generation noise-trader model of De Long et al. (1990) to a foreign exchange context.

economics the sampling frequency is given because collecting data is very expensive in terms of time and money (e.g. output or labor force statistics), this is not the case for financial prices any more. For exchange rates and interest rates in particular, nowadays the sampling frequency is to a large extent chosen by the researcher.

Two important problems arise when we consider the impact of the choice of sampling frequency on traditional uncovered interest parity tests. The first one affects the usual regression approach in which one estimates a single equation that linearly relates the increment of the spot exchange rate over the contract period to the forward premia at the beginning of the period. As is well known, if the period of the forward contract is longer than the sampling interval, then there will be overlapping observations and, thereby, serially correlated regression errors. For that reason, Hansen and Hodrick (1980) use Hansen's (1982) Generalized Method of Moments (GMM) to obtain standard errors that are robust to autocorrelation. Unfortunately, if the degree of overlap (number of observations per contract period) is large relative to the sample size (which in terms of test power should be a good thing), standard GMM asymptotic theory no longer provides a good approximation to the finite sample distribution of overlapping regression tests (see e.g. Richardson and Stock, 1989 or Valkanov, 2003). For example, imagine that we are interested in testing a long-horizon version of uncovered interest parity using yearly data on five-year interest rates, as in one of the robustness tests in Chinn and Meredith (2004). Since the degree of overlap is only 5 periods, we may expect the usual asymptotic results to be reliable if the number of years in the sample is reasonably large. But if we decide to use monthly (weekly) data instead, then we will have an overlap of 60 (260) periods, which is likely to render standard GMM asymptotics useless. Therefore, by choosing the sampling frequency, we are in effect taking a stand on the degree of overlap and, inadvertently, on the finite-sample size and power properties of the test.

The second problem affects the alternative approach that first specifies the joint stochastic process driving the forward premia and the increment on the spot exchange rate over the sampling interval, and then tests the constraints that uncovered interest parity implies on the dynamic evolution of both variables. In this second approach, one usually specifies a vector autoregressive (VAR) model in which the variation of the spot exchange rate is measured over the sampling interval in order to avoid serially correlated residuals. However, the election of the sampling frequency also has implications in this context because VAR models are not usually invariant to temporal aggregation. For instance, if daily observations of the forward premia and the rate of depreciation follow a VAR model, then monthly observations of the same variables will typically satisfy a more complex vector autoregressive moving average (VARMA) model (see e.g. McCrorie and Chambers, 2006). Therefore, having a model that is invariant to temporal aggregation or, in other words, a model that is "sampling-frequency-proof", will eliminate the misspecification problems that may arise from mechanically equating the data generating interval to the sampling interval when the former is in fact finer. This is important because testing uncovered interest parity in a multivariate framework is a joint test of the uncovered interest parity hypothesis and the dynamic specification of the model and, like in many other contexts, having a misspecified model will often result in misleading tests.

Motivated by these two problems, we use a continuous-time approach to derive a new test of uncovered interest parity. In the spirit of Renault et al. (1998), who consider a multivariate continuous-time VAR process to address temporal aggregation problems that arise in testing for causality between exchange rates, we assume that there is an underlying joint continuous-time process for exchange rates and interest rate differentials and then derive UIP conditions that are valid for any observation frequency. We then estimate the parameters of the underlying continuous process on the basis of discretely sampled data, and test the implied uncovered interest parity restrictions. In this way, we can accommodate situations with a large ratio of observations per contract period, with the corresponding gains in asymptotic power. At the same time, though, the model that we estimate is the same irrespective of the sampling frequency.

An alternative approach would be to assume that the data is generated at some specific discrete-time frequency (e.g. daily), which is finer than the sampling interval (e.g. weekly). Then, one could use the results in Marcellino (1999) to obtain the model that the observed data follows. However, such an approach requires knowledge of the data generating frequency, which seems arbitrary. In this paper, we effectively take this approach to its logical limit by assuming that exchange rate and interest rate data are generated on a continuous-time basis.

Previous papers that jointly model exchange and interest rates in continuous-time include Kyu Moh (2006) and Mark and Kyu Moh (2007), who propose non-Gaussian continuous-time uncovered interest parity models, as well as Brandt and Santa-Clara (2002), Brennan and Xia (2006), and Diez de los Rios (2009), who propose continuous-time arbitrage-free models of the international term structure. However, these studies do not test the validity of the uncovered interest parity hypothesis, which is the main objective of our paper.

We begin our analysis by deriving the conditions that uncovered interest parity imposes on the Wold decomposition of continuous-time processes. However, given that working directly with this decomposition is difficult in practice, we follow the discrete-time literature on uncovered interest parity tests and translate these restrictions into testable hypotheses on the continuous-time analogue of a state-space model.[2] Then, we explain how to evaluate the Gaussian pseudo-likelihood function of data observed at arbitrary discrete intervals via the prediction error decomposition using Kalman filtering techniques, which, under certain assumptions, allow us to obtain asymptotically efficient estimators of the parameters characterizing the continuous-time specification. We also assess the usefulness of our proposed methodology by comparing it to existing methods. In particular, we provide a detailed Monte Carlo study which suggests that: (i) in situations where traditional tests of the uncovered interest parity hypothesis have size distortions, the test based on our continuous-time approach has the right size, and (ii) in situations where existing tests have the right size, our proposed test is more powerful.

Importantly, we also propose a specification test that exploits the fact that discrete-time observations generated by a correctly specified continuous-time model will satisfy a valid discrete-time representation regardless of the sampling frequency. The idea is the following: if the model is well-specified, then the estimators of the model parameters obtained at different frequencies converge to their common true values. However, if the model is misspecified then the probability limit of the coefficients estimated at different frequencies will diverge. Although we concentrate on continuous-time models for the exchange rate and interest rate differentials, our testing principle has much wider applicability.

Finally, we apply our continuous time approach to test uncovered interest parity at both short and long horizons on the basis of weekly data on U.S. dollar bilateral exchange rates against the British pound, the German DM-Euro and the Canadian dollar. We use Eurocurrency interest rates of maturities one, three, six-months and one-year to test uncovered interest parity at short horizons, while we use zero-coupon bond yields of maturities one, two and five-years to test it at long horizons. Note that our methodology is especially useful to handle the large degree of overlap (relative to the sample size) that characterizes uncovered interest parity at long-horizons. Importantly, we also use our proposed specification test to check the validity of the continuous-time processes that we estimate. The results that we obtain with correctly specified models continue to reject the

[2] See chapter 9 in Harvey (1989) for a discussion of state-space models with a transition equation in continuous time.

uncovered interest parity hypothesis at short horizons even after taking care of temporal aggregation problems. We also find little support for uncovered interest parity at long horizons. This is in line with Bekaert et al. (2007), and in contrast to Chinn and Meredith (2004) who cannot reject the validity of uncovered interest parity at long horizons on the basis of quarterly data.

The paper is organized as follows. Section 2 details our dynamic framework, the testable restrictions that uncovered interest parity imposes on continuous-time models, our estimation method, and the Monte Carlo evidence on size and power. In Section 3, we introduce our specification test, while Section 4 contains our empirical results. Finally, we provide some concluding remarks and future lines of research in Section 5. Auxiliary results are gathered in an appendix.

2 A continuous-time framework

2.1 Conditions for uncovered interest parity

The most common version of uncovered interest parity (UIP) states that the (nominal) expected return to speculation in the forward foreign exchange market conditioned on available information is zero. Typically, this hypothesis is formally written as:

$$E_t\left(s_{t+\tau} - s_t\right) = p_{t,\tau}, \tag{1}$$

where $E_t(\cdot)$ denotes expectations conditional on the information available up to time t, s_t is the logarithm of the spot exchange rate S_t (e.g. dollar per euro), $p_{t,\tau} = f_{t,\tau} - s_t$ is the τ-period forward premium,[3] and $f_{t,\tau}$ is the logarithm of the forward rate $F_{t,\tau}$ contracted at t that matures at $t+\tau$. As a consequence, if (1) holds then the (log) forward exchange rate will be an unbiased predictor of the τ-period ahead (log) spot exchange rate. For this reason, UIP is also known as the "Unbiasedness Hypothesis". A frequent criticism of this version of UIP is that it pays no attention to issues of risk aversion and intertemporal allocation of wealth. However, Hansen and Hodrick (1983) show that with an additional constant term, equation (1) is consistent with a model of rational maximizing behaviour in which assets are priced by a no-arbitrage restriction. In what follows, we shall refer to this "Modified Unbiasedness Hypothesis" as UIP. To simplify our notation, we will

[3] Most often, UIP is stated in terms of the interest rate differential between two countries. In particular, the covered interest parity hypothesis states that the forward premium is equal to the interest rate differential between two countries: $f_{t,\tau} - s_t = r_{t,\tau} - r^*_{t,\tau}$, where $r_{t,\tau}$ and $r^*_{t,\tau}$ are the (continuously-compounded) τ-period interest rates on a deposit denominated in domestic and foreign currency, respectively.

also understand $p_{t,\tau}$ and Δs_t as the demeaned values of forward premium and the first difference of the spot exchange rate, respectively.

As mentioned before, we could simply specify a joint covariance stationary process for Δs_t and $p_{t,\tau}$ in discrete-time, and test the constraints that UIP implies on the dynamic evolution of both variables. In typical discrete-time models, both the τ-period forward and spot exchange rates have a unit root and, in addition, there is a $(1,-1)$ cointegration relationship between both variables. In this paper, we specify instead a continuous-time model for the infinitesimal increment of the exchange rate and the forward premium. In particular, we follow Phillips (1991) and Chambers (2003) and directly build a continuous-time model in which the $(1,-1)$ cointegration relationship is satisfied:[4]

$$p_\tau(t) = u_1^{(\tau)}(t), \tag{2}$$

$$ds(t) = u_2^{(\tau)}(t)dt + \sigma_s \zeta_s(dt), \tag{3}$$

where $\mathbf{u}^{(\tau)}(t) = \left[u_1^{(\tau)}(t), u_2^{(\tau)}(t)\right]'$ is a covariance-stationary continuous-time process,[5] and $\zeta_s(dt)$ is a continuous-time white-noise with mean $E\left[\zeta_s(dt)\right] = 0$ and variance $E\left[\zeta_s(dt)^2\right] = dt$.[6]

In this context, we can express condition (1) as:

$$E_t\left[s(t+\tau) - s(t)\right] = E_t\left[\int_0^\tau ds(t+h)\right] = p_\tau(t), \tag{4}$$

which by the law of iterated projections (see Hansen and Sargent, 1991a) we can in turn write as:

$$E_t^*\left[s(t+\tau) - s(t)\right] = p_\tau(t), \tag{5}$$

where $E_t^*(\cdot)$ denotes projections onto the linear span of any variables known at time t that includes $p_\tau(t)$. Thus, UIP imposes a set of conditions on the temporal evolution of the τ-period forward premia and the exchange rate. As a limiting example, let the forward

[4]See also Comte (1999) for a discussion on the relationship between discrete and continuous-time cointegration, and an error correction model in continuous-time that we could alternatively use to test the UIP hypothesis in a continuous-time framework.

[5]Note that if we drop the $\zeta(dt)$ term from (3), then we obtain Phillips (1991)'s continuous-time cointegrated system in triangular form representation. In that case, (3) could be expressed as $Ds(t) = u_2(t)$ where $D \equiv d/dt$ is the mean square differential operator. This implies that the sample paths for the spot exchange rate $s(t)$ would be differentiable and, therefore, that the infinitesimal change in $s(t)$ would be smooth. However, the assumption of differentiable exchange rate paths does not seem to be supported by data.

[6]In particular, a continuous-time white-noise process $\boldsymbol{\zeta} = (\zeta_1, \ldots \zeta_n)'$ is a vector of random measures, defined on all subsets of the line $-\infty < t < \infty$ with finite Lebesgue measure, such that $E\left[\boldsymbol{\zeta}(dt)\right] = 0$, $E\left[\boldsymbol{\zeta}(dt)\boldsymbol{\zeta}(dt)'\right] = \boldsymbol{\Sigma}dt$, where $\boldsymbol{\Sigma}$ *is* a positive definite matrix and $E\left[\zeta_i(\Delta)\zeta_j(\Delta')\right] = 0$ for any disjoint sets Δ and Δ' on the line $-\infty < t < \infty$ (see Assumption 1 in Bergstrom, 1983).

contract period τ go to zero, as in the model of Mark and Kyu Moh (2007). Then, the restriction $E_t^*[ds(t)] = p_0(t)$ will be satisfied if and only if $u_1^{(0)}(t) = u_2^{(0)}(t)\ \forall t$, which forces the movements of the forward premia and the exchange rate drift to be exactly the same. The case of $\tau = 0$, though, is not empirically relevant because instantaneous forward contracts do not exist. For the general case of $\tau > 0$, the following proposition summarizes the conditions which guarantee that UIP holds at that horizon:

Proposition 1 *Assume that the temporal evolution of the τ-period forward premium and the spot exchange rate is given by (2) and (3), where $\mathbf{u}^{(\tau)}(t) = \left[u_1^{(\tau)}(t), u_2^{(\tau)}(t)\right]'$ is a covariance stationary continuous-time process whose Wold decomposition is given by:*

$$\mathbf{u}^{(\tau)}(t) = \int_0^\infty \boldsymbol{\phi}^{(\tau)}(h)\boldsymbol{\zeta}_u^{(\tau)}(t - dh), \tag{6}$$

where $\boldsymbol{\zeta}_u^{(\tau)}(t)$ is a two-dimensional white noise process with mean zero and instantaneous covariance matrix given by $E\left[\boldsymbol{\zeta}_u^{(\tau)}(dt)\boldsymbol{\zeta}_u^{(\tau)}(dt)'\right] = \boldsymbol{\Sigma}_u^{(\tau)}dt$, and $\boldsymbol{\phi}^{(\tau)}(h)$ is a 2×2 matrix of square integrable functions such that tr $\left[\int_0^\infty \boldsymbol{\phi}^{(\tau)}(h)\boldsymbol{\Sigma}_u^{(\tau)}\boldsymbol{\phi}^{(\tau)}(h)'dh\right] < \infty$. Then, the uncovered interest parity condition in terms of linear projections (5) holds if and only if:

$$\phi_{11}^{(\tau)}(h) = \int_0^\tau \phi_{21}^{(\tau)}(h+r)dr \qquad \forall h, \tag{7}$$

$$\phi_{12}^{(\tau)}(h) = \int_0^\tau \phi_{22}^{(\tau)}(h+r)dr \qquad \forall h, \tag{8}$$

where $\phi_{ij}^{(\tau)}(h)$ is the ij-element of $\boldsymbol{\phi}^{(\tau)}(h)$.

Proof. First note that the LHS of the UIP condition in continuous time (5) can be written as:

$$E_t^*\left[\int_0^\tau ds(t+r)\right] = E_t^*\left[\int_0^\tau u_2^{(\tau)}(t+r)dr + \int_0^\tau \zeta_s(t+dr)\right] = E_t^*\left[\int_0^\tau u_2^{(\tau)}(t+r)dr\right],$$

while the Wold decomposition (6) implies that:

$$u_2^{(\tau)}(t+r) = \int_{-r}^\infty \phi_{21}^{(\tau)}(h+r)\zeta_{u1}^{(\tau)}(t-dh) + \int_{-r}^\infty \phi_{22}^{(\tau)}(h+r)\zeta_{u2}^{(\tau)}(t-dh).$$

Thus to obtain the required projection conditioned on information available at time t we simply need to apply an annihilation operator that zeros out $\phi_{21}^{(\tau)}(h+r)$ and $\phi_{22}^{(\tau)}(h+r)$ for $t \in [-r, 0]$ (see Hansen and Sargent, 1991b), which simply reflects the fact that future increments of $\boldsymbol{\zeta}_u^{(\tau)}(t)$ are linearly unpredictable while past changes are known. In this way, we obtain

$$E_t^*\left[u_2^{(\tau)}(t+r)\right] = \int_0^\infty \phi_{21}^{(\tau)}(h+r)\zeta_{u1}^{(\tau)}(t-dh) + \int_0^\infty \phi_{22}^{(\tau)}(h+r)\zeta_{u2}^{(\tau)}(t-dh),$$

which in turn yields:

$$E_t^*\left[s(t+\tau)-s(t)\right]=\int_0^\tau\int_0^\infty \phi_{21}^{(\tau)}(h+r)\zeta_{u1}^{(\tau)}(t-dh)dr+\int_0^\tau\int_0^\infty \phi_{22}^{(\tau)}(h+r)\zeta_{u2}^{(\tau)}(t-dh)dr. \tag{9}$$

On the other hand, (6) also implies that:

$$p_\tau(t)=\int_0^\infty \phi_{11}^{(\tau)}(h)\zeta_{u1}^{(\tau)}(t-dh)+\int_0^\infty \phi_{12}^{(\tau)}(h)\zeta_{u2}^{(\tau)}(t-dh). \tag{10}$$

Given that the integrals in (9) are defined in the wide sense with respect to time, we can first change the order of integration, and then equate the right hand sides of (9) and (10). On this basis, it is straightforward to see that UIP is equivalent to the conditions (7) and (8). ■

This proposition is the continuous-time analogue to the results in the appendix of Hansen and Hodrick (1980), who derived the restrictions that UIP implies on the Wold decomposition of discrete-time processes. However, a direct test of (7) and (8) is difficult in practice because it requires the estimation of the bivariate (continuous-time) Wold decomposition in (6). To avoid such a difficulty, we follow the literature on UIP testing in discrete-time (see e.g. Baillie et al. 1984 and Hakkio 1981) and translate those restrictions into testable hypothesis on the continuous-time analogue of a state-space model. Given that we concentrate on a single forward contract at a time, hereinafter we will drop the superscript τ on $\mathbf{u}(t)$, $\boldsymbol{\phi}(h)$ and $\boldsymbol{\zeta}_u(t)$ to simplify the notation.

2.2 A continuous-time state-space approach

The following proposition provides the continuous-time analogue to the rational expectations cross-equation restrictions in Campbell and Shiller (1987) (cf. 35) by translating the UIP restrictions (5) into testable hypotheses on the continuous-time analogue of a state-space model:

Proposition 2 *Assume that the temporal evolution of the τ-period forward premium and the spot exchange rate is given by (2) and (3), where $\mathbf{u}(t)=[u_1(t),u_2(t)]'$ are the first two elements of a $n\times 1$ vector $\mathbf{x}(t)$ that follows a multivariate Orstein-Uhlenbeck (OU) process characterized by the following system of linear stochastic differential equations with constant coefficients:*

$$d\mathbf{x}(t)=\mathbf{A}\mathbf{x}(t)dt+\mathbf{R}^{1/2}\boldsymbol{\zeta}_x(dt), \tag{11}$$

where $\boldsymbol{\zeta}_x(t)$ is a continuous-time white-noise process with mean $E\left[\boldsymbol{\zeta}_x(dt)\right]=0$ and covariance matrix $E\left[\boldsymbol{\zeta}_x(dt)\boldsymbol{\zeta}_x(dt)'\right]=\mathbf{I}dt$, $\mathbf{I}$ being the identity matrix; and all the eigenvalues of

$\mathbf{A}$ *are negative to guarantee the stationarity of the process. Then, the uncovered interest parity condition in terms of linear projections (5) holds if and only if:*

$$\mathbf{e}_2'\mathbf{A}^{-1}(e^{\mathbf{A}\tau} - \mathbf{I}) = \mathbf{e}_1', \tag{12}$$

where $\mathbf{e}_j$ *is a* $n \times 1$ *vector with a one in the* j^{th} *position and zeroes in the others.*

Proof. By combining equations (3) and (11), we can conveniently write our continuous-time model as the following augmented OU process:

$$\begin{aligned} d\begin{bmatrix} \mathbf{x}(t) \\ s(t) \end{bmatrix} &= \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{e}_2' & 0 \end{pmatrix}\begin{bmatrix} \mathbf{x}(t) \\ s(t) \end{bmatrix} dt + \begin{pmatrix} \mathbf{R}^{1/2} & 0 \\ \boldsymbol{\sigma}_{xs}' & \sigma_s \end{pmatrix}\begin{bmatrix} \boldsymbol{\zeta}_x(dt) \\ \zeta_s(dt) \end{bmatrix}, \\ d\boldsymbol{\xi}(t) &= \mathbf{B}\boldsymbol{\xi}(t)dt + \mathbf{S}^{1/2}\boldsymbol{\zeta}(dt), \end{aligned} \tag{13}$$

where $E\left[\boldsymbol{\zeta}_x(dt)\zeta_s(dt)\right] = 0$ without loss of generality.

Under some regularity conditions (see e.g. Bergstrom, 1984), the OU process (13) generates discrete observations that regardless of the discretization interval h will exactly satisfy the following VAR(1) model:

$$\boldsymbol{\xi}_t = \mathbf{F}^{(h)}\boldsymbol{\xi}_{t-h} + \boldsymbol{\eta}_t^{(h)}, \tag{14}$$

where $\mathbf{F}^{(h)} = \exp(\mathbf{B}h) = \mathbf{I} + \sum_{j=1}^{\infty}(\mathbf{B}h)^j/j!$, and the error term $\boldsymbol{\eta}_t^{(h)} = \int_{t-h}^{t} e^{\mathbf{B}(t-r)}\mathbf{S}^{1/2}\boldsymbol{\zeta}(dr)$ satisfies $E\left[\boldsymbol{\eta}_t^{(h)}\right] = \mathbf{0}$, $E\left[\boldsymbol{\eta}_t^{(h)}\boldsymbol{\eta}_t^{(h)\prime}\right] = \boldsymbol{\Omega}^{(h)} = \int_0^h e^{\mathbf{B}r}\mathbf{S}e^{\mathbf{B}'r}dr$, and $E\left[\boldsymbol{\eta}_t^{(h)}\boldsymbol{\eta}_{t-s}^{(h)\prime}\right] = \mathbf{0}$ for $s \geq h$. We can then exploit this VAR structure to generate the corresponding forecasts of $\boldsymbol{\xi}_{t+h}$ given the information at time t as:

$$E_t^*\boldsymbol{\xi}_{t+h} = \mathbf{F}^{(h)}\boldsymbol{\xi}_t. \tag{15}$$

Hence, by setting the discretization frequency, h, equal to the maturity of the contract, τ, and exploiting that $\exp(\mathbf{B}h) = \sum_{k=0}^{\infty}\frac{1}{k!}(\mathbf{B}h)^k$, we finally arrive at:

$$E_t^*\begin{pmatrix} \mathbf{x}_{t+\tau} \\ s_{t+\tau} \end{pmatrix} = \begin{bmatrix} e^{\mathbf{A}\boldsymbol{\tau}} & \mathbf{0} \\ \mathbf{e}_2'\mathbf{A}^{-1}(e^{\mathbf{A}\tau} - \mathbf{I}) & 1 \end{bmatrix}\begin{pmatrix} \mathbf{x}_t \\ s_t \end{pmatrix}.$$

Given that the forward premium $p_{t,\tau}$ is the first element of $\mathbf{x}_t$ (and therefore of $\mathbf{u}_t$), it is straightforward to prove that the UIP condition in terms of linear projections $E_t^*\left(s_{t+\tau} - s_t\right) = p_{t,\tau}$, is equivalent to equation (12). ■

Many models of empirical interest can be cast in terms of (11) in Proposition 2. For example, the continuous-time VAR(p) model:

$$d\left[D^{p-1}\mathbf{u}(t)\right] = \left[\boldsymbol{\Phi}_0\mathbf{u}(t) + \boldsymbol{\Phi}_1 D\mathbf{u}(t) + \ldots + \boldsymbol{\Phi}_{p-1}D^{p-1}\mathbf{u}(t)\right]dt + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\zeta}(dt),$$

which allows for rich dynamics, can be written in companion form as:

$$d\begin{bmatrix} \mathbf{u}(t) \\ D\mathbf{u}(t) \\ \vdots \\ D^{p-2}\mathbf{u}(t) \\ D^{p-1}\mathbf{u}(t) \end{bmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} \\ \mathbf{\Phi}_0 & \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \dots & \mathbf{\Phi}_{p-1} \end{pmatrix} \begin{bmatrix} \mathbf{u}(t) \\ D\mathbf{u}(t) \\ \vdots \\ D^{p-2}\mathbf{u}(t) \\ D^{p-1}\mathbf{u}(t) \end{bmatrix} dt + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{\Sigma}^{1/2} \end{bmatrix} \boldsymbol{\zeta}(dt),$$

where $D \equiv d/dt$ is the mean square differential operator (see Bergstrom, 1984, for details). Similarly, continuous-time VARMA models can be cast in terms of (11) using the companion form VAR representation provided in Chambers and Thornton (2009).

Alternatively, we could consider the continuous-time analogue to the discrete-time VAR approach in Bekaert and Hodrick (1992), and augment $\mathbf{x}(t)$ with dividend yields or other forecasting instruments that may be useful in predicting exchange rates.

Further, we can also focus on models where some of the elements in $\mathbf{x}(t)$ are unobserved. For example, the following model where $\mathbf{u}(t)$ is jointly determined with the unobservable variable $\psi(t)$:

$$d\begin{bmatrix} u_1(t) \\ u_2(t) \\ \psi(t) \end{bmatrix} = \begin{pmatrix} \varphi_{11} & 0 & \pi \\ \varphi_{21} & \varphi_{22} & 0 \\ -\pi & 0 & \varphi_{11} \end{pmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \\ \psi(t) \end{bmatrix} dt + \begin{pmatrix} \gamma_{11} & 0 & 0 \\ \gamma_{21} & \gamma_{22} & 0 \\ 0 & 0 & \gamma_{11} \end{pmatrix} \begin{bmatrix} \zeta_{u1}(dt) \\ \zeta_{u2}(dt) \\ \zeta_{\psi}(dt) \end{bmatrix},$$

delivers an AR(1) process in discrete time with negative autocorrelation for $u_1(t)$ (see Harvey, 1989).

Notice that our general framework also nests the case of Gaussian continuous-time processes where $\boldsymbol{\zeta}(dt) = d\mathbf{W}(t)$ and $\mathbf{W}(t)$ is a standard Wiener process, as in Renault et al. (1998). An important advantage of such models is that conditional expectations are easy to compute as they coincide with linear projections. Yet, our modeling framework also allows for volatility clustering and/or the presence of jumps in exchange rates because, as noted by Berstrom (1984), the discrete-time representation of the model in continuous-time in equation (14) remains valid even when the continuous-time white-noise process, $\boldsymbol{\zeta}(t)$, is not Gaussian. For example, equation (13) may correspond to the vector Lévy-driven OU process in Barndorff-Nielsen and Shephard (2001), in which $\boldsymbol{\zeta}(dt)$ are instantaneous time-homogeneous independent increments that include Gaussian, compensated Poisson, Gamma and inverse Gaussian processes among others, as well as linear combinations of these.

We could also assume that $\boldsymbol{\xi}(t)$ follows the multivariate affine diffusion:

$$d\boldsymbol{\xi}(t) = \mathbf{B}\boldsymbol{\xi}(t)dt + \mathbf{S}^{1/2}\mathbf{C}^{1/2}(t)d\mathbf{W}(t), \tag{16}$$

where the matrix $\mathbf{C}(t)$ is diagonal with elements $c_{ii}(t) = \alpha_i + \boldsymbol{\beta}_i'\boldsymbol{\xi}(t)$. This would allow us to consider not only square root type processeses, but also stochastic volatility models in which volatility follows a mean reverting OU process itself. Note that in this case $vec\left[var_t\left(\boldsymbol{\xi}_{t+h}\right)\right] = \mathbf{v}_0 + \mathbf{v}_1\boldsymbol{\xi}_t$ (see Fisher and Gilles, 1996, Meddahi and Renault, 1996, or Duffee, 2002). Similarly, we could allow for conditional volatility in exchange rates and forward premia if we assume that $\boldsymbol{\xi}(t)$ follows a multivariate version of the continuous-time square-root stochastic autoregressive volatility (SR-ARV) process of Meddahi and Renault (2004) with a linear drift.

As in the Gaussian case, discrete observations from vector Lévy-driven OU processes, multivariate affine diffusions and multivariate SR-ARV models will also satisfy equation (15) in terms of conditional expectations despite the presence of jumps or conditional heteroskedastity. Consequently, one can readily obtain the relevant UIP restrictions (12) for these models in terms of conditional expectations also using the results in Proposition 2 by simply replacing $E_t^*(\cdot)$ by $E_t(\cdot)$.

To illustrate our methods, it is pedagogically convenient to study in detail the case in which $\mathbf{u}(t)$ follows a continuous-time VAR(1) process.

Example 1. Suppose that the temporal evolution of the τ-period forward premium and the spot exchange rate is given by:

$$p_\tau(t) = u_1(t),$$
$$ds(t) = u_2(t)dt + \boldsymbol{\alpha}'\boldsymbol{\zeta}_u(dt),$$

with $\mathbf{u}(t)$ following a continuous-time VAR(1) model.

$$d\begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} dt + \begin{pmatrix} \sigma_{11} & 0 \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{bmatrix} \zeta_{u1}(dt) \\ \zeta_{u2}(dt) \end{bmatrix}, \tag{17}$$
$$d\mathbf{u}(t) = \boldsymbol{\Phi}\mathbf{u}(t)dt + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\zeta}_u(dt),$$

where $\boldsymbol{\Phi}$ has two negative eigenvalues to guarantee the stationarity of the process, and $E[\boldsymbol{\zeta}_u(dt)\boldsymbol{\zeta}_u(dt)'] = \mathbf{I}dt$. Note also that we have assumed that $\zeta_s(dt)$ is an exact linear combination of the fundamental shocks driving $\mathbf{u}(t)$.

If we choose $\mathbf{x}(t) = \mathbf{u}(t)$, $\mathbf{A} = \boldsymbol{\Phi}$, and $\mathbf{R} = \boldsymbol{\Sigma}$, this model coincides with the one in equation (11). Thus, we can specialize the conditions in equation (12) to obtain that UIP will hold if and only if:

$$\mathbf{e}_2'\boldsymbol{\Phi}^{-1}(e^{\boldsymbol{\Phi}\tau} - \mathbf{I}) = \mathbf{e}_1'. \tag{18}$$

Not surprisingly, we can arrive to the same condition by exploiting the results in our Proposition 1. To this end, note model (17) implies that the matrix $\boldsymbol{\phi}(h)$ in the Wold decomposition in equation (6) is given by $\boldsymbol{\phi}(h) = e^{\boldsymbol{\Phi}h}$. On this basis, we can jointly express conditions (7) and (8) as

$$\mathbf{e}_1'\boldsymbol{\phi}(h) = \mathbf{e}_2'\int_0^{\tau}\boldsymbol{\phi}(h+r)dr. \tag{19}$$

Substituting $\boldsymbol{\phi}(h)$ by $e^{\boldsymbol{\Phi}h}$ into (19) and solving the integral delivers the restrictions derived in equation (18) by exploiting the discrete-time representation of a continuous-time VAR model.

Example 2. Suppose that the temporal evolution of the τ-period forward premium and the spot exchange rate is given by:

$$d\begin{bmatrix} p_\tau(t) \\ s(t) \end{bmatrix} = \begin{pmatrix} \varphi_{11} & 0 \\ \varphi_{21} & 0 \end{pmatrix}\begin{bmatrix} p_\tau(t) \\ s(t) \end{bmatrix}dt + \begin{pmatrix} \gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \end{pmatrix}\begin{bmatrix} \zeta_{v1}(dt) \\ \zeta_{v2}(dt) \end{bmatrix}, \tag{20}$$

with $\varphi_{11} < 0$ and $E[\boldsymbol{\zeta}(dt)\boldsymbol{\zeta}(dt)'] = \mathbf{I}dt$. A direct application of (14) gives us the following restricted discrete-time VAR(1):

$$\begin{pmatrix} p_{t,\tau} \\ \Delta_h s_t \end{pmatrix} = \begin{bmatrix} e^{\varphi_{11}h} & 0 \\ \frac{\varphi_{21}}{\varphi_{11}}(e^{\varphi_{11}h}-1) & 0 \end{bmatrix}\begin{pmatrix} p_{t-h,\tau} \\ \Delta_h s_{t-h} \end{pmatrix} + \begin{pmatrix} \eta_{1t}^{(h)} \\ \eta_{2t}^{(h)} \end{pmatrix}.$$

Hence, the forward premia at the h interval is a stationary AR(1) process with autocorrelation coefficient $e^{\varphi_{11}h}$, while the spot exchange rate has a unit root. Moreover, there is no feedback from the exchange rate, $s(t)$, to the forward premium, $p_\tau(t)$.

By setting the discretization period h equal to the contract period τ, we obtain that the least squares projection coefficient of $\Delta_\tau s_{t+\tau}$ on $p_{t,\tau}$ is equal to $\varphi_{21}(e^{\varphi_{11}h}-1)/\varphi_{11}$. Thus, the UIP condition, $E_t^*(\Delta_\tau s_{t+\tau}) = p_{t,\tau}$, holds if and only if:

$$\varphi_{21} = \frac{\varphi_{11}}{e^{\varphi_{11}\tau}-1}, \tag{21}$$

This model is a special case of Example 1 because the condition $E_t^*[ds(t)] = u_2(t) = \varphi_{21}p_\tau(t)$ implies that:

$$du_2(t) = \varphi_{21}dp_\tau(t) = \varphi_{21}\varphi_{11}p_\tau(t)dt + \varphi_{21}\gamma_{11}\zeta_1(dt),$$

and thus we can write (20) as

$$d\begin{bmatrix} p_\tau(t) \\ u_2(t) \\ s(t) \end{bmatrix} = \begin{pmatrix} \varphi_{11} & 0 & 0 \\ \varphi_{21}\varphi_{11} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{bmatrix} p_\tau(t) \\ u_2(t) \\ s(t) \end{bmatrix}dt + \begin{pmatrix} \gamma_{11} & 0 \\ \varphi_{21}\gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \end{pmatrix}\begin{bmatrix} \zeta_1(dt) \\ \zeta_2(dt) \end{bmatrix}, \tag{22}$$

which has the form of the model in equation (17). Moreover, given that in this representation the corresponding matrix $\boldsymbol{\Phi}$ has a zero eigenvalue, we can exploit the discretization in equation (15) to show that:

$$E_t^*(s_{t+\tau} - s_t) = \left[\frac{\varphi_{21}}{\varphi_{11}}(e^{\varphi_{11}h}-1) - \varphi_{21}\right]p_{t,\tau} + u_{2t}.$$

Substituting $u_{2t} = \varphi_{21} p_{t,\tau}$ we obtain again that the least squares projection coefficient of $\Delta_\tau s_{t+\tau}$ on $p_{t,\tau}$ is equal to $\varphi_{21}(e^{\varphi_{11}h} - 1)/\varphi_{11}$. Therefore UIP holds when $\varphi_{21} = \varphi_{11}/(e^{\varphi_{11}\tau} - 1)$, which is the same condition derived in (21). Alternatively, we could exploit the Wold decomposition of (20) to arrive to the same expression.

Note, however, that equation (22) is not the only representation of the model in Example 2 in terms of the model in Example 1. In particular, we can also use the fact that $u_2(t) = \varphi_{21} p_\tau(t)$ to write

$$du_2(t) = \varphi_{11} u_2(t)dt + \varphi_{21}\gamma_{11}\zeta_1(dt),$$

which delivers the following alternative expression:

$$d\begin{bmatrix} p_\tau(t) \\ u_2(t) \\ s(t) \end{bmatrix} = \begin{pmatrix} \varphi_{11} & 0 & 0 \\ 0 & \varphi_{11} & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{bmatrix} p_\tau(t) \\ u_2(t) \\ s(t) \end{bmatrix} dt + \begin{pmatrix} \gamma_{11} & 0 \\ \varphi_{21}\gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{bmatrix} \zeta_1(dt) \\ \zeta_2(dt) \end{bmatrix}. \tag{23}$$

Not surprisingly, this representation yields the same UIP condition (21).[7]

2.3 Estimation

We estimate the structural parameters of the continuous-time model (13) by Gaussian pseudo maximum likelihood (PML) estimation using Kalman filtering techniques.[8] To do so, we set h to the sampling frequency, which for simplicity we normalize to 1, so equation (14) becomes

$$\boldsymbol{\xi}_t = \mathbf{F}^{(1)}\boldsymbol{\xi}_{t-1} + \boldsymbol{\eta}_t^{(1)}. \tag{24}$$

In addition, if we further assume for estimation purposes that $\boldsymbol{\zeta}(t)$ in equation (13) is a continuous-time white-noise process with a Gaussian distribution (i.e. a Wiener process), then the error term in equation (24), $\boldsymbol{\eta}_t^{(1)}$, will be an *i.i.d.* sequence of Gaussian random vectors with mean $E\left[\boldsymbol{\eta}_t^{(1)}\right] = \mathbf{0}$, and covariance matrix $E\left[\boldsymbol{\eta}_t^{(1)}\boldsymbol{\eta}_t^{(1)}\right] = \boldsymbol{\Omega}^{(1)} = \int_0^1 e^{\mathbf{B}r}\mathbf{S}e^{\mathbf{B}'r}dr$.

Equation (24) can then be understood as the discrete-time transition equation of the

[7] Given that the model in Example 2 can be nested within the model in Example 1 in several ways, some of the parameters appearing in (17) will not be identified when the true model is given by (20). To avoid this problem, we would recommend estimating the model in Example 2 directly from (20).

[8] As we discuss in Appendix A, we could alternatively eliminate any unobservable variable from the system by substitution. Following such an approach in the case of the model in Example 1, we would end up with a bivariate VARMA(2,1) system in the vector $(p_{t,\tau}, \Delta_h s_t)'$, which could be then estimated by (pseudo) maximum likelihood. Bergstrom and Chambers (1990) estimate a model of durable goods by elimination of the unobservable stock of durable goods.

following model in state-space form:

$$\mathbf{y}_t = \mathbf{d} + \mathbf{H}\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t,$$

$$\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \mathbf{u}_t,$$

$$\left(\begin{array}{c} \boldsymbol{\varepsilon}_t \\ \mathbf{u}_t \end{array}\right) \Bigg| \left(\begin{array}{c} \mathbf{y}_{t-1} \\ \boldsymbol{\alpha}_{t-1} \end{array}\right), \left(\begin{array}{c} \mathbf{y}_{t-2} \\ \boldsymbol{\alpha}_{t-2} \end{array}\right), \ldots \sim N\left[\left(\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array}\right), \left(\begin{array}{cc} \mathbf{Z} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{array}\right)\right],$$

where $\mathbf{y}_t = (p_{t,\tau}, \Delta s_t)'$ when there are no extra predictors in the model, and the matrices $\mathbf{d}$, $\mathbf{H}$, $\mathbf{T}$, $\mathbf{Z}$ and $\mathbf{Q}$ are different depending on the model estimated. In the case of example 1, these matrices are

$$\mathbf{d} = \left(\begin{array}{c} 0 \\ 0 \end{array}\right), \qquad \mathbf{H} = \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 1 \end{array}\right), \qquad \mathbf{Z} = \left(\begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array}\right),$$

$$\mathbf{T} = \left(\begin{array}{cc} e^{\boldsymbol{\Phi}} & 0 \\ \mathbf{e}_2'\boldsymbol{\Phi}^{-1}(e^{\boldsymbol{\Phi}} - \mathbf{I}) & 0 \end{array}\right), \qquad \mathbf{Q} = \boldsymbol{\Omega}^{(1)} = \int_0^1 e^{\mathbf{B}r}\mathbf{S}e^{\mathbf{B}'r}dr,$$

where $\mathbf{B}$ and $\mathbf{S}^{1/2}$ are defined in equation (13) with $\mathbf{A} = \boldsymbol{\Phi}$, $\mathbf{R} = \boldsymbol{\Sigma}$, $\boldsymbol{\sigma}_{xs} = \boldsymbol{\alpha}$, $\sigma_s = 0$.

We evaluate the exact Gaussian pseudo log-likelihood via the usual prediction error decomposition:

$$\ln L(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_t,$$

with

$$l_t = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{O}_t| - \frac{1}{2}\mathbf{v}_t'\mathbf{O}_t^{-1}\mathbf{v}_t, \tag{25}$$

where $\boldsymbol{\theta}$ is the vector of parameters of the continuous-time model, $\mathbf{v}_t$ is the vector of one-step-ahead prediction errors produced by the Kalman filter, and $\mathbf{O}_t$ their conditional variance.

The usual Kalman filter recursions are given by

$$\left.\begin{array}{r} \boldsymbol{\alpha}_{t|t-1} = \mathbf{T}\boldsymbol{\alpha}_{t-1|t-1} \\ \mathbf{P}_{t|t-1} = \mathbf{T}\mathbf{P}_{t-1|t-1}\mathbf{T}' + \mathbf{Q} \\ \mathbf{v}_t = \mathbf{y}_t - \mathbf{d} - \mathbf{H}\boldsymbol{\alpha}_{t|t-1} \\ \mathbf{O}_t = \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}' + \mathbf{Z} \\ \boldsymbol{\alpha}_{t|t} = \boldsymbol{\alpha}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{H}'\mathbf{O}_t^{-1}\mathbf{v}_t \\ \mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}'\mathbf{O}_t^{-1}\mathbf{H}\mathbf{P}_{t|t-1} \end{array}\right\} \tag{26}$$

where $\boldsymbol{\alpha}_{t|t-1} = E_{t-1}(\boldsymbol{\alpha}_t)$ and $\mathbf{P}_{t|t-1} = E\left[(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1})(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1})'\right]$ are the expectation and covariance matrix of $\boldsymbol{\alpha}_t$ conditional on information up to time $t-1$, while $\boldsymbol{\alpha}_{t|t} = E_t(\boldsymbol{\alpha}_t)$ and $\mathbf{P}_{t|t} = E\left[(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t})(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t})'\right]$ are the expectation and covariance matrix of $\boldsymbol{\alpha}_t$ conditional on information up to time t (see Harvey, 1989). Given that we are assuming that the state variables are covariance stationary, we can initialize the filter using $\boldsymbol{\alpha}_0 = E(\boldsymbol{\alpha}_t) = \mathbf{0}$ and $vec(\mathbf{P}_0) = (\mathbf{I} - \mathbf{T} \otimes \mathbf{T})^{-1}\, vec(\mathbf{Q})$.

Finally, we can exploit again the prediction error decomposition in (25) to obtain first and second derivatives of the log likelihood function (see Harvey, 1989), which we need to estimate the variance of the score and the expected value of the Hessian that appears in the asymptotic distribution of Gaussian PML estimators of $\boldsymbol{\theta}$. In particular, the score vector takes the following form:

$$\frac{\partial l_t(\boldsymbol{\theta})}{\partial \theta_i} = s_t(\boldsymbol{\theta}) = -\frac{1}{2} tr\left[\left(\mathbf{O}_t^{-1}\frac{\partial \mathbf{O}_t}{\partial \theta_i}\right)\left(\mathbf{I} - \mathbf{O}_t^{-1}\mathbf{v}_t\mathbf{v}_t'\right)\right] - \frac{\partial \mathbf{v}_t'}{\partial \theta_i}\mathbf{O}_t^{-1}\mathbf{v}_t, \tag{27}$$

while the ij-th element of the conditionally expected Hessian matrix satisfies:

$$-E_{t-1}\left(\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}\right) = -E_{t-1}\left(\frac{\partial s_{it}}{\partial \theta_j}\right) = \frac{1}{2} tr\left(\mathbf{O}_t^{-1}\frac{\partial \mathbf{O}_t}{\partial \theta_i}\mathbf{O}_t^{-1}\frac{\partial \mathbf{O}_t}{\partial \theta_j}\right) + \frac{\partial \mathbf{v}_t'}{\partial \theta_i}\mathbf{O}_t^{-1}\frac{\partial \mathbf{v}_t}{\partial \theta_j}. \tag{28}$$

In turn, these two expressions require the first derivatives of $\mathbf{O}_t$ and $\mathbf{v}_t$, which we can evaluate analytically by an extra set of recursions that run in parallel with the Kalman filter. As Harvey (1989, pp. 140-3) shows, the extra recursions are obtained by differentiating the Kalman filter prediction and updating equations (26).[9] In particular, we make use of these formulae in a scoring algorithm to maximize the exact log-likelihood function with analytical expressions for the score vector and information matrix.[10]

We also use those expressions to obtain heteroskedasticity-robust standard errors and Wald tests. In particular, under standard regularity conditions, it can be shown that $\widehat{\boldsymbol{\theta}}$, the PML estimator of the model parameters, has the following asymptotic distribution:

$$\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N\left[\mathbf{0}, (\mathbf{D}_{\boldsymbol{\theta}}'\mathbf{S}_{\boldsymbol{\theta}}^{-1}\mathbf{D}_{\boldsymbol{\theta}})^{-1}\right],$$

where $\mathbf{D}_{\boldsymbol{\theta}} = E\left[\partial \mathbf{s}_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'\right] = E\left[\partial^2 l_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'\right]$, $\mathbf{S}_{\boldsymbol{\theta}} = E\left[\mathbf{s}_t(\boldsymbol{\theta})\mathbf{s}_t(\boldsymbol{\theta})'\right]$ and the relevant elements of these matrices can be obtained from (27) and (28). On this basis, we can test the null hypothesis $H_0 : \mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$ using the following Wald test:

$$T \cdot \mathbf{r}(\widehat{\boldsymbol{\theta}})'\left[\frac{\partial \mathbf{r}(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}(\widehat{\mathbf{D}}_{\boldsymbol{\theta}}'\widehat{\mathbf{S}}_{\boldsymbol{\theta}}^{-1}\widehat{\mathbf{D}}_{\boldsymbol{\theta}})^{-1}\frac{\partial \mathbf{r}(\widehat{\boldsymbol{\theta}})'}{\partial \boldsymbol{\theta}}\right]\mathbf{r}(\widehat{\boldsymbol{\theta}}), \tag{29}$$

where $\widehat{\mathbf{D}}_{\boldsymbol{\theta}}$ and $\widehat{\mathbf{S}}_{\boldsymbol{\theta}}$ are consistent estimates of $\mathbf{D}_{\boldsymbol{\theta}}$ and $\mathbf{S}_{\boldsymbol{\theta}}$, respectively. In the context of model (17), for example, we have that the UIP restrictions to test are

$$\mathbf{r}(\boldsymbol{\theta}) = \mathbf{e}_2'\boldsymbol{\Phi}^{-1}(e^{\boldsymbol{\Phi}\tau} - \mathbf{I}) - \mathbf{e}_1'.$$

[9] In our continuous-time models the analytical derivatives of the Kalman filter equations with respect to the structural parameters require the derivatives of the exponential of a matrix, which we obtain using the results in Chen and Zadrozny (2001)

[10] Details on how to obtain initial values for the optimization algorithm are provided in Appendix B.

When the continuous-time process, $\boldsymbol{\zeta}(t)$, is not Gaussian, the previous procedure still yields Gaussian PML estimators of $\boldsymbol{\theta}$, which are consistent albeit less efficient. In addition, the Kalman recursions yield minimum mean-squared error predictors based on the fitted model (see Brockwell, 2001). Alternatively, one could estimate by maximum likelihood the parameters of fully specified continuous-time data generating processes that can capture some important high frequency features of exchange rates, such as non-normality or volatility clustering. As we mentioned before, the UIP restrictions in Proposition 2 remain valid in those contexts too. Nonetheless, given that the efficient estimation of multivariate Lévy-driven OU processes or conditionally heteroskedastic affine diffusions is not a trivial task, we only consider Gaussian PML estimators in this paper.

2.4 Monte Carlo simulations of UIP tests

In this section, we carry out an extensive Monte Carlo study to assess the ability of our proposed methodology to test UIP. In addition, we also compare our proposed continuous-time-based test to the two main approaches to test UIP in the existing literature: OLS- and VAR-based tests.

2.4.1 Design

We initially simulate 10,000 samples of 30 years of weekly data ($T = 1,560$) from the continuous-time model (17) under the assumption that the continuous-time white-noise process is Gaussian (i.e. a Wiener process). We fix the contract period τ to 52 (one year). To make them more realistic, we include unconditional means for the observed variables. Therefore, the model that we simulate is given by:

$$\begin{pmatrix} \widetilde{p}_{t,\tau} \\ \Delta\tilde{s}_t \end{pmatrix} = \begin{pmatrix} \mu_p \\ \mu_{\Delta s} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \\ \Delta s_t \end{pmatrix}, \tag{30}$$

$$\begin{pmatrix} \mathbf{u}_t \\ \Delta s_t \end{pmatrix} = \begin{bmatrix} e^{\boldsymbol{\Phi}} & 0 \\ e_2'\boldsymbol{\Phi}^{-1}(e^{\boldsymbol{\Phi}} - I) & 0 \end{bmatrix} \begin{pmatrix} \mathbf{u}_{t-1} \\ \Delta s_{t-1} \end{pmatrix} + \boldsymbol{\eta}_t^{(1)}, \tag{31}$$

where $\widetilde{p}_{t,\tau} = \mu_p + p_{t,\tau}$, $\Delta\tilde{s}_t = \mu_{\Delta s} + \Delta s_t$, $\boldsymbol{\eta}_t^{(1)} = \int_{t-1}^{t} e^{\mathbf{B}(t-r)}\mathbf{S}^{1/2}d\mathbf{W}(r)$, $\mathbf{B}$ and $\mathbf{S}^{1/2}$ are defined in equation (13) with $\mathbf{A} = \boldsymbol{\Phi}$, $\mathbf{R} = \boldsymbol{\Sigma}$, $\boldsymbol{\sigma}_{xs} = \boldsymbol{\alpha}$, $\sigma_s = 0$, $\sigma_{11} = .3$, $\sigma_{21} = -.2$, $\sigma_{22} = .1$, $\alpha_1 = -.1$, $\alpha_2 = 1.5$, $\mu_p = 2$, and $\mu_{\Delta s} = 0$. In order to impose the null hypothesis of UIP, we first decompose $\boldsymbol{\Phi}$ as $\mathbf{PDP}^{-1}$, where $\mathbf{D}$ is a diagonal matrix with elements $d_1 = -.025$ and $d_2 = -.25$, and then choose $\mathbf{P}$ so that $\boldsymbol{\Phi}$ satisfies (18). Such parameter values are chosen in order to match the empirical characteristics of our dataset.

In particular, our choice of d_1 and d_2 imply that the forward premium is stationary but very persistent.

2.4.2 Traditional UIP tests

Ordinary least squares. The first approach is an OLS-based test motivated by the following regression equation:

$$s_{t+\tau} - s_t = \alpha + \beta(f_{t,\tau} - s_t) + w_{t+\tau} \tag{32}$$

where the "Unbiasedness Proposition" implies that $\alpha = 0$ and $\beta = 1$, while we just need $\beta = 1$ to satisfy the "Modified Unbiasedness Proposition". In addition, rational expectations imply that $w_{t+\tau}$ is serially correlated when the sampling interval is shorter than τ because we will have overlapping observations. In particular, Hansen and Hodrick (1980) show how to use overlapping data in order to increase the sample size, which should result in gains in the asymptotic power of UIP tests, using Hansen's (1982) GMM. Yet, sample estimates of heteroskedasticity and autocorrelation consistent (HAC) covariance matrices are very sensitive to the election of bandwidth and kernel, which often results in inferences that are severely distorted (see den Haan and Levin, 1996, and Ligeralde, 1997). To illustrate this point, we compute several OLS-based UIP tests in which asymptotically valid standard errors are estimated using the following different methods:

1. Newey-West (1987) approach (NW), which is the most popular method to construct asymptotic standard errors when testing UIP in a regression setup (see e.g. Bansal and Dahlquist, 2000, and Flood and Rose, 2002). As in the recent literature, we use the optimal data-driven bandwidth selection rule in Andrews (1991).

2. Eichenbaum, Hansen and Singleton (1988) approach (EHS), which exploits that, under the null hypothesis, the error term in the OLS estimation of (32) follows a moving-average (MA) process of finite known order but with unknown coefficients to construct the asymptotic covariance matrix. We follow Eichenbaum, Hansen and Singleton (1988) in using Durbin (1960)'s method to estimate the MA structure.

3. Den Haan and Levin (1996) approach with a VAR order automatically selected using either the Akaike Information Criteria (VARHAC-AIC) or the Bayesian Information Criteria (VARHAC-BIC). Den Haan and Levin (1996) data-driven approach assumes that the moment conditions implicit in the normal equations of (32) have a finite VAR representation, which they exploit to construct their estimated covariance matrix.

4. Non-overlapping observations (NO), which we compute by sampling exchange and interest rates every $\tau = 52$ periods. This approach entails a considerable waste of sample information.

Vector autoregressions in discrete time. We also compare our continuous-time approach with those results that we would have obtained using a VAR-based test. This second approach estimates a joint covariance stationary process for the first difference of the spot exchange rate Δs_t and the forward premia $p_{t,\tau}$ by Gaussian PML. Unlike in an OLS-based test, the difference operator on the spot exchange rate is taken over the sampling interval in order to avoid overlapping residuals. Consequently, the UIP condition (in terms of linear projections) becomes:

$$E_t^* \left(s_{t+\tau} - s_t\right) = E_t^* \left(\sum_{i=1}^{\tau} \Delta s_{t+i}\right) = p_{t,\tau}. \tag{33}$$

In this context, Baillie et al. (1984) and Hakkio (1981) show how to obtain testable restrictions on the companion matrix of a VAR. The rationale for looking at vector autoregressions is that we can always approximate any strictly invertible and covariance stationary discrete-time process by a VAR model with a sufficient number of lags. Moreover, the VAR assumption allows us to use the Campbell and Shiller (1987) methodology for testing present value models. Specifically, we can use the VAR model to produce optimal forecasts of the increment of the spot exchange rate in (33), from which we can obtain the appropriate UIP conditions. As an illustration, assume that $\mathbf{y}_t = (p_{t,\tau}, \Delta s_t)'$ follows the VAR(1) model

$$\mathbf{y}_t = \mathbf{B}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{34}$$

where $\boldsymbol{\varepsilon}_t$ is a two-dimensional vector of white noise disturbances with contemporaneous covariance matrix $E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right) = \boldsymbol{\Upsilon}$. Then, the optimal linear forecast of $\mathbf{y}_{t+i}$ $(i = 1, ..., \tau)$ based on the information set defined by $\mathbf{y}_t$ and its lagged values is given by $E_t^*\mathbf{y}_{t+i} = \mathbf{B}^i\mathbf{y}_t$. Consequently, the projection of Δs_{t+i} will be given by $\mathbf{e}_2'\mathbf{B}^i\mathbf{y}_t$, where $\mathbf{e}_j$ is a vector with a one in the j^{th} position and zeroes in the others. Therefore, the testable restrictions on the VAR parameters that UIP implies for a τ-period forward contract are:[11]

$$\mathbf{e}_2'\mathbf{B}(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B}^\tau) = \mathbf{e}_1', \tag{35}$$

[11] Note that the left hand side (LHS) of (33) can be expressed as:

$$E_t^* \left(\sum_{i=1}^{\tau} \Delta s_{t+i}\right) = \mathbf{e}_2' \left(\sum_{i=1}^{\tau} \mathbf{B}^i\mathbf{y}_t\right) = \mathbf{e}_2'\mathbf{B}(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B}^\tau)\mathbf{y}_t,$$

while the right hand side (RHS) is $p_{t,\tau} = \mathbf{e}_1'\mathbf{y}_t$.

which are the discrete-time analogue to (12).

Although we can always consider (34) as the first order companion form of a higher order VAR, if our estimated model does not provide a good representation of the joint Wold decomposition of Δs_t and $p_{t,\tau}$ because we have selected an insufficient number of lags, say, then we may end up rejecting the UIP hypothesis when in fact it is true. To illustrate this point, we compute VAR-based tests for two lag choices: $p = 1$ and 4.

2.4.3 A fair comparison of UIP tests.

In total, we compare the performance of a test based on the continuous-time specification in equations (30) and (31) to seven other different UIP tests (five OLS- and two VAR-based tests). Nonetheless, one has to be careful in comparing all these different tests of the UIP hypothesis because each of them has a different alternative hypothesis in mind. As a confirmation, note that OLS-based tests have one degree of freedom, VAR(p)-based tests have $2p$ degrees of freedom, while tests based on the continuous-time model (17) have two degrees of freedom. In order to make a fair comparison across models, we follow Hodrick (1992) and Bekaert (1995) and obtain an implied beta from the VAR and the continuous-time approach that is analogous to the regression slope tested in the simple regression approach. Given that the regression coefficient is simply the ratio of the covariance between the expected future rate of depreciation and the forward premium to the variance of the forward premium, the implied slope in the VAR(1) in equation (34) is:

$$\beta^{VAR(1)} = \frac{\mathbf{e}_2'\mathbf{B}(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}-\mathbf{B}^{\tau})\mathbf{\Psi}\mathbf{e}_1}{\mathbf{e}_1'\mathbf{\Psi}\mathbf{e}_1}, \tag{36}$$

where $\mathbf{\Psi}$ is the unconditional covariance matrix of $\mathbf{y}_t = (p_{t,\tau}, \Delta s_t)'$, which can be obtained from the equation $vec(\mathbf{\Psi}) = (\mathbf{I}-\mathbf{B}\otimes\mathbf{B})^{-1}vec(\mathbf{\Upsilon})$. On the other hand, the implied slope for the continuous time model (17) is given by:

$$\beta^{OU} = \frac{\mathbf{e}_2'\mathbf{\Phi}^{-1}(e^{\mathbf{\Phi}\tau}-\mathbf{I})\mathbf{\Lambda}\mathbf{e}_1}{\mathbf{e}_1'\mathbf{\Lambda}\mathbf{e}_1}, \tag{37}$$

where $vec(\mathbf{\Lambda}) = -\left(\mathbf{\Phi}\otimes\mathbf{I}+\mathbf{I}\otimes\mathbf{\Phi}\right)^{-1}vec(\mathbf{\Sigma})$ is the unconditional variance of $\mathbf{u}_t$. Therefore, we will concentrate on the null hypotheses $H_0 : \beta^{VAR(p)} = 1$ for $p = 1$ and 4, as well as $H_0 : \beta^{OU} = 1$. For this reason, we modify the Wald test statistic in (29) using $r(\boldsymbol{\theta}) = \left[\mathbf{e}_2'\mathbf{\Phi}^{-1}(e^{\mathbf{\Phi}\tau}-\mathbf{I})\mathbf{\Lambda}\mathbf{e}_1\right]/(\mathbf{e}_1'\mathbf{\Lambda}\mathbf{e}_1) - 1$ as the restriction on the parameters to be tested:

$$T \cdot \frac{\left[r(\widehat{\boldsymbol{\theta}})\right]^2}{\left[\frac{\partial r(\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}'}(\widehat{\mathbf{D}}_{\boldsymbol{\theta}}{}'\widehat{\mathbf{S}}_{\boldsymbol{\theta}}^{-1}\widehat{\mathbf{D}}_{\boldsymbol{\theta}})^{-1}\frac{\partial r(\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta})}\right]}.$$

2.4.4 Results: Size

Figure 1 summarises the finite sample size properties of each of the aforementioned UIP tests by means of Davidson and MacKinnon's (1998) p-value discrepancy plots, which show the difference between actual and nominal test sizes for every possible nominal size. As expected, given the large degree of overlap, the tests based on OLS regressions with standard errors that rely on the usual GMM asymptotic results suffer considerable size distortions. For example, the test that uses the Newey-West estimator of the long-run covariance matrix of the OLS moment conditions massively over-rejects the UIP hypothesis. In contrast, the actual size of tests based on the EHS and VARHAC-BIC methods are well below their nominal sizes. The size distortions for the EHS method probably reflect the difficulties in estimating a MA(51) structure using Durbin's method, while those in the VARHAC-BIC approach might be caused by the apparent tendency of the BIC lag selection procedure to choose an insufficient number of lags. Although the best OLS-based method for overlapping observations is the VARHAC-AIC approach, it still over-rejects in finite samples. Similarly to the results in Richardson and Stock (1989) and Valkanov (2003), these results suggest that when the degree of overlap becomes non-trivial relative to the sample size, standard GMM asymptotics no longer provides a good approximation to the finite sample distribution of the tests. Finally, tests based on non-overlapping regressions also over-reject the UIP hypothesis due to the small number of observations included in the estimation ($T = 30$).

As for VAR-based tests, we find that we approximate better the autocorrelation structure of $\mathbf{y}_t = (p_{t,\tau}, \Delta s_t)'$ as we increase the order of the VAR from $p = 1$ to $p = 4$. A simple explanation for this phenomenon can be obtained from an inspection of the population values of the implicit beta obtained by estimating a misspecified VAR(p) when the true model is in fact given by the continuous-time process (17). Without loss of generality, assume that $p = 1$ (otherwise, simply write a higher order VAR as an augmented VAR(1)). The companion matrix of a VAR(1) model is defined by the relationship $\mathbf{B} \equiv E\left(\mathbf{y}_t\mathbf{y}'_{t-1}\right)\left[E\left(\mathbf{y}_t\mathbf{y}'_t\right)\right]^{-1}$, while the variance-covariance matrix of the residuals is given by $\boldsymbol{\Upsilon} \equiv E\left(\mathbf{y}_t\mathbf{y}'_{t-1}\right) - E\left(\mathbf{y}_t\mathbf{y}'_{t-1}\right)\left[E\left(\mathbf{y}_t\mathbf{y}'_t\right)\right]^{-1} E\left(\mathbf{y}_t\mathbf{y}'_{t-1}\right)'$. If we then plug in the expressions for $E\left(\mathbf{y}_t\mathbf{y}'_t\right)$ and $E\left(\mathbf{y}_t\mathbf{y}'_{t-1}\right)$ implied by the continuous-time model (17), we will obtain analytical expressions for the population value of $\mathbf{B}(\boldsymbol{\theta})$ and $\boldsymbol{\Upsilon}(\boldsymbol{\theta})$ as a function of the parameters of the continuous-time model, $\boldsymbol{\theta}$. Then, we can use equation (36) to compute the population value of $\beta^{VAR(1)}(\boldsymbol{\theta})$, which we can understand as the implicit beta obtained by postulating a VAR(1) model when in fact the true model is the continuous time process (17). In this way, we obtain $\beta^{VAR(1)}(\boldsymbol{\theta}) = 0.6952$ and $\beta^{VAR(4)}(\boldsymbol{\theta}) = 0.9802$ for the value of $\boldsymbol{\theta}$ in our experimental design. These values indicate that the implicit beta approaches 1 as we increase p, which explains why the test based on a VAR(1) process largely over-rejects in finite samples, while the actual and nominal sizes of the test based on the VAR(4) process are quite close for standard nominal levels.[12] Our results also confirm that testing UIP in such a full-information setup should be considered as a joint test of the UIP hypothesis and the dynamic specification of the model. Consequently, the application of specification tests is especially relevant in this context. We will return to this issue in Section 3.

Finally, note that the test based on our correctly specified continuous-time model provides very reliable inferences.

2.4.5 Results: Power

We run a second Monte Carlo experiment with another 10,000 replications to assess the finite-sample power of the same seven tests. In this case, the design is essentially identical to the previous one, including the eigenvalues of $\mathbf{\Phi}$. The only difference is that we now set $\phi_{11} = -.025$, $\phi_{12} = 1$, $\phi_{21} = 0$, and $\phi_{22} = -.25$ so that UIP is violated because $\mathbf{e}_2'\mathbf{\Phi}^{-1}(e^{\mathbf{\Phi}\tau} - \mathbf{I}) = 4\mathbf{e}_2' \neq \mathbf{e}_1'$. Figure 2 summarises the finite sample power properties for each of the UIP tests by means of Davidson and MacKinnon's (1998) size-power curves, which show power for every possible actual size. The most obvious result from this figure is that the test based on our continuous-time approach has the highest power for any given size, followed by the test based on the VAR(4) model, the OLS-based tests that use overlapping observations, the one that uses non-overlapping data, and finally the VAR(1) test. Intuitively, our continuous-time approach and, to a less extent the VAR(4), have high power because they exploit the correct dynamic properties of the data (see Hallwood and MacDonald, 1994).[13] To interpret our results, it is useful to resort again to the population values of the implicit beta for the VAR models. For this design, we have that $\beta^{VAR(1)}(\boldsymbol{\theta}) = 0.4795$ and $\beta^{VAR(4)}(\boldsymbol{\theta}) = 0.1966$, while the population value of the implicit beta for the continuous-time model (17) calculated according to equation (37) is $\beta^{OU}(\boldsymbol{\theta}) = 0.0879$. Note that under the alternative hypothesis, the smaller the order of the VAR(p) model, the closer the value of $\beta^{VAR(p)}$ is to one, which explains the relative

[12] Given that we can always approximate the autocorrelation structure of any strictly invertible covariance stationary process observed in discrete-time by a VAR model with a sufficient number of lags, we would expect the population value of the implied beta to approach 1 as the number of lags increases.

[13] Note that if the true distribution is Gaussian then our continuous-time approach delivers the maximum likelihood estimator, which is efficient, and gives rise to optimal tests.

ranking of the two VAR-based tests.

In summary, our Monte Carlo results suggest that: (i) in situations where traditional tests of the UIP hypothesis have size distortions, a test based on our continuous-time approach has the right size, and (ii) in situations where existing tests have the right size, our proposed test is more powerful.

2.4.6 A simpler Monte Carlo design

We now briefly describe the results of a second Monte Carlo experiment based instead on the model of example 2, which is such that a discrete time VAR(1) process captures the autocorrelation structure of $\mathbf{y}_t = (p_{t,\tau}, \Delta s_t)'$. In particular, we simulate 10,000 samples of 30 years of weekly data ($T = 1,560$) from the continuous-time model (20) under the assumption that the continuous-time white-noise process is Gaussian. As in the previous exercise, we fix the contract period τ to 52, and include unconditional means for the observed variables. Thus, the model we simulate to assess the finite-sample size of the UIP tests is given by:

$$\begin{pmatrix} \widetilde{p}_{t,\tau} \\ \Delta\tilde{s}_t \end{pmatrix} = \begin{pmatrix} \mu_p \\ \mu_{\Delta s} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p_{t,\tau} \\ \Delta s_t \end{pmatrix},$$

$$\begin{pmatrix} p_{t,\tau} \\ \Delta s_t \end{pmatrix} = \begin{bmatrix} e^{\varphi_{11}} & 0 \\ \frac{\varphi_{21}}{\varphi_{11}}\left(e^{\varphi_{11}} - 1\right) & 0 \end{bmatrix} \begin{pmatrix} p_{t-1,\tau} \\ \Delta s_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t}^{(1)} \\ \eta_{2t}^{(1)} \end{pmatrix},$$

where $\varphi_{11} = -.025$, $\varphi_{21} = \varphi_{11}/\left(e^{\gamma_{11}\tau} - 1\right)$ to guarantee that UIP holds, $\gamma_{11} = .3$, $\gamma_{21} = -.1$, $\gamma_{22} = 1.5$, $\mu_p = 2$ and $\mu_{\Delta s} = 0$.

Once again, we find that tests based on OLS regressions and GMM asymptotics suffer substantial size distortions. In particular, both the test that uses Newey-West standard errors and the one based on non-overlapping regressions present actual sizes well above their nominal ones. Tests based on the EHS and VARHAC methods (both with AIC and BIC lag selection criteria) are somewhat better behaved, yet they still over-reject in finite samples. Not surprisingly, in this case VAR-based tests provide inferences which are as reliable as those obtained by estimating the continuous-time model given in equation (20).

As for the finite-sample power of UIP tests, we simulate another 10,000 replications of the above model with $\varphi_{21} = 0$ so that exchange rates follow a random-walk process. Our results indicate that both the test based on our-continuous-time approach and the one based on a VAR(1) model have the highest power for any given size because, effectively, the continuous time model in (20) implies a VAR(1) structure in discrete time. In terms of power, these two tests are ranked above the tests based on the VAR(4) model and the

OLS-based tests. Our results indicate that, for the parameter configuration and sample size in this Monte Carlo design, the loss of power from estimating an unrestricted VAR(1) is negligible, while the loss of power from estimating a VAR(4) is fairly small.

3 Specification tests that combine different sampling frequencies

3.1 Description

As illustrated in the previous section, misspecification of the joint autocorrelation structure of exchange rates and interest rate differentials can lead to systematic rejections of UIP when, in fact, it holds. For example, we have shown that if we choose an insufficient number of lags in a VAR model, then UIP tests based on this model will tend to over-reject. To some extent, our continuous-time approach also suffers from the same problem, and the power gains that we see in Figure 2 come at a cost: if the joint autocorrelation structure implied by our continuous-time model is not valid, then our proposed UIP test may also become misleading. Therefore, the calculation of dynamic specification tests becomes particularly relevant in our context. In this sense, one possiblity would be to test for lack of serial correlation of the residuals using an LM test, as in Renault et al. (1998).

In this paper, on the other hand, we introduce an alternative specification test that exploits the fact that the structure of a continuous time model is the same regardless of the discretization frequency, h. Under the null hypothesis that our continuous time specification is valid, Gaussian pseudo maximum likelihood parameter estimators are consistent irrespective of the sampling frequency. In contrast, if the dynamic specification is incorrect, then estimators based on different sampling frequencies will have different probability limits.

In order to gain some extra intuition on the specification approach, assume for simplicity that the sampling frequency is weekly and that we want to compare weekly and biweekly estimates of the parameter vector $\boldsymbol{\theta}$.[14] Our test is based on the following algorithm.

1. Estimate $\boldsymbol{\theta}$ using the whole sample by Gaussian PML. Let $E[\mathbf{s}_t^{(1)}(\boldsymbol{\theta})] = \mathbf{0}$ denote the moment conditions that define the pseudo maximum likelihood estimator (PMLE)

[14] In practice, we can modify steps 2 and 3 of our proposed algorithm to obtain estimators for any aggregated frequency of choice.

and call the solution to this equation $\hat{\boldsymbol{\theta}}^{(1)}$.

2a. Artificially generate a new biweekly data set from the original data by discarding all even observations. Then, write the moment conditions that define the PMLE of $\boldsymbol{\theta}$ which only uses odd observations as $E[\mathbf{s}_t^{(2o)}(\boldsymbol{\theta})] = \mathbf{0}$. By convention, $\mathbf{s}_t^{(2o)}(\boldsymbol{\theta}) = \mathbf{0}$ when t is an even number.

2b. Similarly, discard all odd observations and write the moment conditions that define the PMLE which only uses even observations as $E[\mathbf{s}_t^{(2e)}(\boldsymbol{\theta})] = \mathbf{0}$. Again, $\mathbf{s}_t^{(2e)}(\boldsymbol{\theta}) = \mathbf{0}$ when t is an odd number.

3. Obtain a new estimate of the vector parameter, $\hat{\boldsymbol{\theta}}^{(2)}$ from the sum of the moment conditions that define the PMLE for both odd and even observations: $E[\mathbf{s}_t^{(2o)}(\boldsymbol{\theta}) + \mathbf{s}_t^{(2e)}(\boldsymbol{\theta})] = E[\mathbf{s}_t^{(2)}(\boldsymbol{\theta})] = \mathbf{0}$.

4. Finally, compare both estimators $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ using the following test statistic:

$$\left(\hat{\boldsymbol{\theta}}^{(1)} - \hat{\boldsymbol{\theta}}^{(2)}\right)' \left[Var\left(\hat{\boldsymbol{\theta}}^{(1)} - \hat{\boldsymbol{\theta}}^{(2)}\right)\right]^{-1} \left(\hat{\boldsymbol{\theta}}^{(1)} - \hat{\boldsymbol{\theta}}^{(2)}\right) \tag{38}$$

If this test indicates that both estimators are statistically close, accept the hypothesis of correct specification. Reject it otherwise.

As is well-known, if the estimation criterion is the true log-likelihood of the data then the estimator that uses data at the highest frequency is efficient, the variance of the difference of the estimators is the difference of the respective variances, and our test statistic can be viewed as an application of Hausman's (1978) specification test idea. More generally, we explain in the next section how to obtain the relevant expression for $Var(\hat{\boldsymbol{\theta}}^{(1)} - \hat{\boldsymbol{\theta}}^{(2)})$ from the joint asymptotic distribution of the pseudo scores of the model at both frequencies, $\mathbf{s}_t^{(1)}(\boldsymbol{\theta})$ and $\mathbf{s}_t^{(2)}(\boldsymbol{\theta})$. This approach guarantees that the variance of the difference of the estimators is positive definite and that our specification test remains valid in pseudo log-likelihood contexts.

Our specification test is related to Ryu (1994), who considered a continuous-time proportional hazard duration model for grouped data with time-invariant categorical regressors. To test the proportionality assumption on the hazard, Ryu (1994) uses a specification test that compares the estimates obtained from two time intervals with those obtained with a single aggregated time interval. However, a direct translation of his approach to our context would involve the use of non-overlapping data at the lower frequency, and thereby an information loss. In contrast, our approach efficiently exploits all the information available by combining the moment conditions that define the lower frequency estimators for both odd and even observations.

3.2 Implementation

In this section, we explain how to test the validity of the joint autocorrelation structure implied by the continuous-time model in Example 2. In particular, imagine that we want to compare the parameter estimates of the model obtained with weekly and biweekly data. As we saw in section 2.2, observations of this model sampled at the weekly frequency ($h = 1$) will satisfy a VAR(1) process which can be conveniently re-written in state-space form as:

$$\begin{pmatrix} p_{t,\tau} \\ \Delta s_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \end{pmatrix}, \tag{39}$$

$$\begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \end{pmatrix} = \begin{bmatrix} e^{\varphi_{11}} & 0 \\ \frac{\varphi_{21}}{\varphi_{11}}\left(e^{\varphi_{11}} - 1\right) & 0 \end{bmatrix} \begin{pmatrix} \alpha_{1t-1} \\ \alpha_{2t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t}^{(1)} \\ \eta_{2t}^{(1)} \end{pmatrix}. \tag{40}$$

where $\boldsymbol{\eta}_t = [\eta_{1t}^{(1)}, \eta_{2t}^{(1)}]' = \int_{t-1}^{t} e^{\boldsymbol{\Phi}(t-r)}\boldsymbol{\Gamma}^{1/2}\boldsymbol{\zeta}(dr)$. By assuming that the continuous-time white-noise process is Gaussian, we can readily obtain PML estimators of the vector of parameters of the continuous-time model in (20), $\boldsymbol{\theta}$, as the solution to the sample versions of the following moment conditions:

$$E\left[\frac{\partial l^{(1)}(\mathbf{y}_t^{(1)};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = E\left[\mathbf{s}_t^{(1)}(\boldsymbol{\theta})\right] = 0, \tag{41}$$

where $\mathbf{y}_t^{(1)} = [p_{t,\tau}, \Delta s_t]'$, $l^{(1)}(\mathbf{y}_t^{(1)};\boldsymbol{\theta})$ is the (pseudo) log-likelihood contribution of $\mathbf{y}_t^{(1)}$, and the superscript (1) indicates that $\mathbf{y}_t^{(1)}$ is an observation obtained at the highest available frequency (weekly). Analogously, we denote by $\boldsymbol{\theta}^{(1)}$ the value of $\boldsymbol{\theta}$ that solves (41).

Alternatively, we could generate a biweekly data set by discarding all even observations. By treating discarded observations as missing values and using the approach in Mariano and Murasawa (2003) to write the likelihood of this sample, we construct a new series

$$\mathbf{y}_t^{(2o+)} = D_t \cdot \mathbf{y}_t^{(2)} + (1 - D_t)\mathbf{z}_t^e,$$

where D_t is a dummy variable that takes the value of 1 when $\mathbf{y}_t^{(2)} = [p_{t,\tau}, \Delta_2 s_t]'$ is observed because t is an odd number, while $\mathbf{z}_t^e$ is a bivariate random vector drawn from an independent arbitrary distribution that does not depend on $\boldsymbol{\theta}$. Let us define $\mathbf{y}_T^{(2o)} = (\mathbf{y}_1^{(2)}, \mathbf{y}_3^{(2)}, ..., \mathbf{y}_T^{(2)})'$ and $\mathbf{y}_T^{(2o+)} = (\mathbf{y}_1^{(2o+)}, \mathbf{y}_2^{(2o+)}, ..., \mathbf{y}_T^{(2o+)})'$. Given that the $\mathbf{z}_t^e$'s are independent of $\mathbf{y}_T^{(2o)}$ by construction, we can write the joint probability distribution of $\mathbf{y}_T^{(2o+)}$ as

$$f(\mathbf{y}_T^{(2o+)};\boldsymbol{\theta}) = f(\mathbf{y}_T^{(2o)};\boldsymbol{\theta}) \cdot \prod_{t:D_t=0} f(\mathbf{z}_t^e).$$

Thus, the log likelihood function of $\boldsymbol{\theta}$ given $\mathbf{y}_T^{(2o)}$ and the corresponding log likelihood given $\mathbf{y}_T^{(2o+)}$ are identical up to scale, so they will be maximized by the same value. The main advantage of working with the augmented data series $\mathbf{y}_T^{(2o+)}$ is that it no longer contains missing observations and, therefore, it is easy to derive a state space model for $\mathbf{y}_t^{(2o)}$. In particular, the measurement equation is:

$$\begin{pmatrix} p_{t,\tau} \\ \Delta_2 s_t \end{pmatrix} = \begin{pmatrix} D_t & 0 & 0 \\ 0 & D_t & D_t \end{pmatrix} \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \alpha_{2t-1} \end{pmatrix} + (1-D_t)\mathbf{z}_t^e, \tag{42}$$

while the transition equation will be:

$$\begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \alpha_{2t-1} \end{pmatrix} = \begin{pmatrix} e^{\varphi_{11}} & 0 & 0 & 0 \\ \frac{\varphi_{21}}{\varphi_{11}}\left(e^{\varphi_{11}}-1\right) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha_{1t-1} \\ \alpha_{2t-1} \\ \alpha_{2t-2} \end{pmatrix} + \begin{pmatrix} \eta_{1t}^{(1)} \\ \eta_{2t}^{(1)} \\ 0 \end{pmatrix}, \tag{43}$$

which can be understood as an augmented version of (40).

Once again, we can use the Kalman filter to compute the (exact) log-likelihood function of this state space model. Similarly, Gaussian PML estimates of $\boldsymbol{\theta}$ based on the odd observations will satisfy the sample analogues to the moment conditions:

$$E\left[\frac{\partial l^{(2o)}(\mathbf{y}_t^{(2o+)};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = E\left[\mathbf{s}_t^{(2o)}(\boldsymbol{\theta})\right] = \mathbf{0}. \tag{44}$$

As explained before, we overcome the waste of information resulting from discarding one half of the sample by combining the moment conditions (44) with those that one would obtain by discarding all odd observations instead, and then estimating $\boldsymbol{\theta}$ as the solution to the sample analogue of the following set of moment conditions:

$$E\left[\mathbf{s}_t^{(2o)}(\boldsymbol{\theta}) + \mathbf{s}_t^{(2e)}(\boldsymbol{\theta})\right] = E\left[\mathbf{s}_t^{(2)}(\boldsymbol{\theta})\right] = \mathbf{0}, \tag{45}$$

where $\mathbf{s}_t^{(2e)}(\boldsymbol{\theta})$ are the influence functions that define the estimator that only uses even observations. We denote the resulting estimator as $\hat{\boldsymbol{\theta}}^{(2)}$.[15]

[15] To estimate the model at the lower frequency, we minimize the following quadratic form:

$$\left[\frac{1}{T}\sum_{t=1}^{T} \mathbf{s}_t^{(2)}(\boldsymbol{\theta}^{(2)})\right]' \mathbf{W}^{-1} \left[\frac{1}{T}\sum_{t=1}^{T} \mathbf{s}_t^{(2)}(\boldsymbol{\theta}^{(2)})\right].$$

In this context, our testing methodology simply assesses whether the probability limits of $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ coincide. Specifically, define $\boldsymbol{\psi} = \left(\boldsymbol{\theta}^{(1)\prime}, \boldsymbol{\theta}^{(2)\prime}\right)'$, and think of $\hat{\boldsymbol{\psi}}$ as solving the sample versions of the following set of moment conditions:

$$\left\{ \begin{array}{c} E[\mathbf{s}_t^{(1)}(\boldsymbol{\theta}^{(1)})] \\ E[\mathbf{s}_t^{(2)}(\boldsymbol{\theta}^{(2)})] \end{array} \right\} = E\left[\mathbf{s}_t(\boldsymbol{\psi})\right] = \mathbf{0}. \tag{46}$$

Then, we can use standard GMM asymptotic theory to show that:

$$\sqrt{T}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \xrightarrow{d} N\left[\mathbf{0}, (\mathbf{D}_{\boldsymbol{\psi}}'\mathbf{S}_{\boldsymbol{\psi}}^{-1}\mathbf{D}_{\boldsymbol{\psi}})^{-1}\right], \tag{47}$$

where $\mathbf{D}_{\boldsymbol{\psi}} = E\left[\partial \mathbf{s}_t(\boldsymbol{\psi})/\partial \boldsymbol{\psi}'\right]$ and $\mathbf{S}_{\boldsymbol{\psi}} = \sum_{j=-\infty}^{\infty} E\left[\mathbf{s}_t(\boldsymbol{\psi})\mathbf{s}_{t-j}(\boldsymbol{\psi})'\right]$, whether or not $\mathbf{y}_t$ really follows a Gaussian process.

On this basis, we can test the restriction $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(2)}$ by using the Wald test:

$$T \cdot \hat{\boldsymbol{\psi}}'\mathbf{R}'\left[\mathbf{R}(\widehat{\mathbf{D}}_{\boldsymbol{\psi}}'\widehat{\mathbf{S}}_{\boldsymbol{\psi}}^{-1}\widehat{\mathbf{D}}_{\boldsymbol{\psi}})^{-1}\mathbf{R}'\right]^{-1}\mathbf{R}\hat{\boldsymbol{\psi}},$$

where $\mathbf{R} = (\mathbf{I}, -\mathbf{I})$, and $\widehat{\mathbf{D}}_{\boldsymbol{\psi}}$ and $\widehat{\mathbf{S}}_{\boldsymbol{\psi}}$ are consistent estimates of $\mathbf{D}_{\boldsymbol{\psi}}$ and $\mathbf{S}_{\boldsymbol{\psi}}$, respectively. Nevertheless, it is important to remember that the comparison of estimators at different frequencies induces an overlapping problem that in general makes $E\left[\mathbf{s}_t(\boldsymbol{\psi})\mathbf{s}_{t-j}(\boldsymbol{\psi})'\right] \neq 0$ for $j \leq \delta - 1$, where δ is the ratio of the sampling frequencies (=2 in this example). Thus, we have to take into account this MA structure in computing a consistent estimate of $\mathbf{S}_{\boldsymbol{\psi}}$.

Given that we mostly care about the sampling interval in as much as a change in h leads to different conclusions on the validity of the UIP, we simply test if the implied betas remain the same when we vary the sampling frequency instead of testing whether the full parameter vectors $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ coincide. In the context of model (20) in particular, we would test if $\beta^{(1)} = \beta^{(2)}$, with $\beta^{(\tau)} = \phi_{21}\left(e^{\phi_{11}\tau} - 1\right)/\phi_{11}$, using the following Wald statistic

$$T \cdot f(\hat{\boldsymbol{\psi}})\left[\frac{\partial f(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\psi}'}(\widehat{\mathbf{D}}_{\boldsymbol{\psi}}'\widehat{\mathbf{S}}_{\boldsymbol{\psi}}^{-1}\widehat{\mathbf{D}}_{\boldsymbol{\psi}})^{-1}\frac{\partial f(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\psi}}\right]^{-1} f(\hat{\boldsymbol{\psi}}),$$

where $f(\boldsymbol{\psi}) = \varphi_{21}^{(1)}\left(e^{\varphi_{11}^{(1)}\tau} - 1\right)/\varphi_{11}^{(1)} - \varphi_{21}^{(2)}\left(e^{\varphi_{11}^{(2)}\tau} - 1\right)/\varphi_{11}^{(2)}$. By focusing on this particular characteristic of the model we avoid the use of a large number of degrees of freedom, which is likely to improve the finite sample properties of our test.

Similarly, we test the specification of the continuous time model (17) in Example 1 by checking if $\mathbf{r}^{(1)} = \mathbf{r}^{(2)}$, where $\mathbf{r}^{(j)} = \mathbf{e}_2'\left(\boldsymbol{\Phi}^{(j)}\right)^{-1}(e^{\boldsymbol{\Phi}^{(j)}\tau} - \mathbf{I}) - \mathbf{e}_1'$ are the restrictions that UIP implies on this model.

Since equation (45) exactly identifies $\boldsymbol{\theta}^{(2)}$, the above quadratic form will take the value of zero at the optimum for any choice of the weighting matrix $\mathbf{W}$ for T sufficiently large. Still, to improve the covergence properties of our numerical optimisation algorithm, we use the estimated values of $\boldsymbol{\theta}^{(1)}$ as starting values, and choose $\mathbf{W}$ to be the Newey-West estimate of the long-run covariance matrix of the moment conditions $\mathbf{s}_t^{(2)}(\boldsymbol{\theta})$ evaluated at $\hat{\boldsymbol{\theta}}^{(1)}$.

3.3 Monte Carlo simulations of specification tests

3.3.1 Design

In this section, we investigate the performance of the specification test discussed above by means of two additional Monte Carlo studies. In order to assess its finite-sample size properties, we generate 10,000 simulations of 30 years of weekly data ($T = 1,560$) from the continuous-time model (20) in example 2, where once again we fix the contract period to be equal to $\tau = 52$. Similar to what we did in Section 2.4, we add unconditional means to $\mathbf{y}_t = [p_{t,\tau}, \Delta s_t]'$, so that the model we simulate is:

$$\begin{pmatrix} \widetilde{p}_{t,\tau} \\ \Delta\tilde{s}_t \end{pmatrix} = \begin{pmatrix} \mu_p \\ \mu_{\Delta s} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p_{t,\tau} \\ \Delta s_t \end{pmatrix},$$

$$\begin{pmatrix} p_{t,\tau} \\ \Delta s_t \end{pmatrix} = \begin{bmatrix} e^{\varphi_{11}} & 0 \\ \frac{\varphi_{21}}{\varphi_{11}} \left(e^{\varphi_{11}} - 1\right) & 0 \end{bmatrix} \begin{pmatrix} p_{t-1,\tau} \\ \Delta s_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t}^{(1)} \\ \eta_{2t}^{(1)} \end{pmatrix},$$

where $\varphi_{11} = -.025$, $\varphi_{21} = \varphi_{11}/\left(e^{\gamma_{11}\tau} - 1\right)$, $\gamma_{11} = .3$, $\gamma_{21} = -.1$, $\gamma_{22} = 1.5$, $\mu_p = 2$ and $\mu_{\Delta s} = 0$. Importantly, note that we maintain the UIP restriction (21).

Since we compare the value of β that we obtain using the weekly sample, $\beta^{(1)}$, with the one that we would obtain had we sampled the data every two weeks, $\beta^{(2)}$, and these are functions of only φ_{11} and φ_{21}, we exclude the scores corresponding to the remaining parameters at the lowest frequency ($h = 2$).[16]

The comparison between these estimators creates an overlapping problem that introduces an MA(1) structure in the error term, which is nevertheless much simpler than the MA(51) structure in section 2.4. For that reason, we consider again the Newey-West (1987) approach (NW) with the optimal data-driven bandwidth selection rule in Andrews (1991), the Eichenbaum, Hansen and Singleton (1988) approach (EHS), as well as the Den Haan and Levin (1996)'s VARHAC approaches with VAR order selected using either the Akaike Information Criteria (VARHAC-AIC) or the Bayesian Information Criteria (VARHAC-BIC). In this sense, the only change with respect to section 2.4 is that in the EHS approach we explicitly impose that the (pseudo) scores of the model at the highest frequency are serially uncorrelated.[17]

[16] In addition, by focusing on the scores corresponding to φ_{11} and φ_{21}, we are also able to avoid some singularities of the long-run covariance matrix of the scores, $\mathbf{S}_{\boldsymbol{\psi}} = \sum_{j=-\infty}^{\infty} E\left[\mathbf{s}_t(\boldsymbol{\psi})\mathbf{s}_{t-j}(\boldsymbol{\psi})'\right]$. Specifically, the influence functions of the moment conditions that define $\boldsymbol{\mu}^{(2)}$ are a dynamic linear combination of the influence functions defining $\hat{\boldsymbol{\mu}}^{(1)}$. Similarly, under certain conditions, the influence functions of $\boldsymbol{\Gamma}^{(2)}$ are also a dynamic linear combination of influence functions at the highest frequency. Further details on this issue are available upon request.

[17] Yet, we should bear in mind that such an assumption will only be valid under a correct specification of both the conditional mean and conditional variance of the process.

Figure 3 summarises the finite sample size properties of our proposed specification test for each of the HAC covariance estimation methods. As can be seen, all four approaches to computing the covariance matrix of the estimates tend to under-reject the null hypothesis of correct specification, the VARHAC-AIC method being the one that under-rejects the least.

We also generate another 10,000 simulations of 30 years of weekly data from the continuous-time model (17) to assess the finite-sample power of the specification test that takes as its null hypothesis that the correct model is given by (20). Specifically, we simulate again from the model in equations (30) and (31) for $\tau = 52$, except that this time we choose $d_2 = -1.00$ because all four versions of our specification test reject with probability 1 when $d_2 = -7.5$. Once again, we maintain the UIP restrictions in (21).

Figure 4 summarises the finite sample power properties of each of the HAC covariance estimation methods. The first thing to note is that our specification test has non-trivial power against dynamic misspecification of the continuous-time process. We can also see that the NW approach has the highest power, followed by the ones based on the EHS approach, the VARHAC-BIC approach and finally the VARHAC-AIC one.

4 Can we rescue UIP?

In this section, we apply our continuous-time approach to test the UIP hypothesis on the U.S. dollar bilateral exchange rates against the British pound, the German DM-Euro and the Canadian dollar. We use two different data sets for each country. We use the appropriate Eurocurrency interest rates at maturities of one, three, six months, and one year at a weekly frequency over the period January 1977 to December 2005. This allows us to focus on the traditional UIP at short horizons. We also follow Chinn and Meredith (2004) and Bekaert et al. (2007) and study UIP at long horizons. To this end, we use zero-coupon bond yields at maturities one, two and five years at the weekly frequency over the period June 1992 to December 2005. Data on exchange and Eurocurrency interest rates are from Datastream while data on zero-coupon bond yields have been obtained from the corresponding central banks (with the exception of the German data which was obtained from the Bank of England). Finally, note that our choice of sample and countries is restricted by data availability.[18]

[18] Although our data set does not incorporate either the transactions costs inherent in bid-ask spreads or the delivery structure of the market, Bekaert and Hodrick (1993) show that these factors have a negligible effect on the empirical results.

4.1 UIP at short horizons

Panel a of Table 1 reports the estimated coefficients of the continuous-time model (20) in Example 2, as well as the estimate of the implied beta and the Gaussian log-likelihood of the sample, $\ln L(\theta)$. This reveals several interesting facts. First, the estimated φ_{11} is close to zero, which confirms that the forward premium is rather persistent (see e.g. Baillie and Bollerslev, 2000, and Maynard and Phillips, 2001). For example, the monthly autocorrelation coefficient of the one-month forward premium is approximately .95 for the U.K., .97 for Germany, and .90 for Canada. Second, the forward premium is much less volatile than the rate of depreciation, which is consistent with previous studies (e.g. Bekaert and Hodrick, 2001). Third, the correlation between the innovation to the forward premium and the innovation to the rate of depreciation is negative for the U.K. and Germany, and positive for Canada. Finally, the implied beta is always negative and significantly different from one. Therefore, UIP is rejected for all currency pairs and maturities.

As argued before, however, it is important to check the validity of the continuous time model that we estimate. For that reason, in Panel a of Table 2 we report the results of our proposed specification test applied to the estimates of β that we obtain using weekly data with the one that we would obtain had we sampled the data every two weeks. Notice that the difference $\hat{\beta}^{(1)} - \hat{\beta}^{(2)}$ tends to be small and, in fact, is only significantly different from zero for the one-, and three-month Canadian dollar contracts.

For this reason, Table 3 reports the estimated coefficients of the more flexible continuous time model (17) for the cases in which model (20) is rejected. Still, the forward premium continues to be very persistent and less volatile than the rate of depreciation, and the implicit beta remains negative and significantly different from one. This time, though, we cannot reject the dynamic specification of model (17). In particular, the p-values of the specification test lie between .65 (AIC) and .99 (EHS), and .45 (AIC) and .86 (EHS) for the one-, and three-month Canadian dollar contracts, respectively.

Therefore, we are unable to rescue the UIP hypothesis at short-horizons even though we appropriately account for temporal aggregation.

Finally, we also implement the traditional UIP tests described in Section 2.4. Specifically, we compute OLS-based UIP tests for both non-overlapping and overlapping data, in which case the standard errors are obtained using the Newey-West (1987) with the optimal data-driven bandwidth selection rule in Andrews (1991), Eichenbaum, Hansen and Singleton (1988), and Den Haan and Levin (1996)'s VARHAC approaches with VAR order selection computed using either the Akaike Information Criteria (VARHAC-AIC) or the Bayesian Information Criteria (VARHAC-BIC). Similarly, we also compute VAR-based tests for lags $p = 1$ and 4. Not surprisingly, the results reported in Panel a of Table 4 indicate that the estimate of the slope coefficient β is negative. As expected from the Monte Carlo experiment reported in Section 2.4, the results of the OLS-based UIP tests with overlapping observations are somewhat sensitive to the covariance matrix estimator employed (see also Ligeralde, 1997). For example, if we use the EHS or VARHAC-BIC methods to test UIP at the one-year horizon with U.K. data, we find that we cannot reject that $H_0 : \beta = 1$, and the same is true if we use non-overlapping observations. Similar results are obtained if we use the VARHAC-BIC approach to test the UIP hypothesis with German data at the three-month horizon, or at the one-year horizon with the EHS, VARHAC-AIC or VARHAC-BIC approaches. In contrast, tests based on the NW covariance estimator always reject UIP, and the same is true of VAR-based tests.

4.2 UIP at long horizons

Chinn and Meredith (2004) argue that, in contrast to studies which have used short-horizon data (up to one year), it is not possible to reject the UIP hypothesis once one uses interest rates on longer-maturity bonds. Using OLS-based tests with quarterly data, they find that the coefficient on the interest rate differential is positive and close to the UIP value of unity. However, Bekaert et al. (2007) argue that it is unlikely that short-term deviations from the UIP and long-term deviations from the expectations hypothesis of the term structure would exactly offset each other so as to make UIP hold at long horizons. Using a VAR approach and monthly data, Bekaert et al (2007) find that the UIP hypothesis tends to be rejected at both short and long-horizons.

We try to shed some light on this empirical debate by re-examining long-horizon UIP using our continuous-time approach. To do so, we focus on zero-coupon bond yields at maturities one, two and five years at the weekly frequency over the period June 1992 to December 2005.[19] Note that our methodology is especially useful to handle the large degree of overlap (relative to the sample size) that characterizes long-horizon UIP hypothesis tests. As a result, we can use weekly data in contrast to Chinn and Meredith (2004), who use quarterly data presumably to avoid an excessive degree of overlap in their regression tests. Therefore, we expect to achieve power gains over their approach. Moreover, our

[19] When we explicitly compared the results obtained with one-year zero-coupon bond yields with those obtained using one-year Eurocurrency interest rates over the common sample period, we found that the results were qualitatively and quantitatively similar.

test should be free from the potential temporal aggregation biases that might affect the VAR approach in Bekaert et al. (2007).

Panel b of Table 1 reports the estimated coefficients of the continuous-time model (17) in Example 2. Notice again that the estimated φ_{11} is close to zero, and that the forward premium is much less volatile than the rate of depreciation. Both results are consistent with those reported in the previous subsection. We also find that the implied betas for the U.K. at the one and two-year horizons are negative, while the implied beta at the five-year horizon is positive. However, these betas are imprecisely estimated, which implies that we cannot reject that they are different from one. When we look at Germany, we find that the estimated betas are negative for all forecast horizons. However, we can only reject that the one and two-year betas are different from one. The implied beta at the five-year horizon is again very imprecisely estimated so we cannot reject that it is equal to one despite being very negative. Finally, the implied betas for Canada are all negative and statistically different from one.

Once again, we check the validity of the continuous-time model that we estimate by comparing the estimates of β that we obtain using weekly data with the ones that we would have obtained had we sampled the data every two weeks. As reported in Panel b of Table 2, we do not find that the difference between the estimators is significantly different from zero for any of the countries and maturities under consideration.

Given that the number of non-overlapping periods is only 14 for $\tau = 52$, 7 for $\tau = 104$, and 3(!) for $\tau = 260$, we only compute UIP tests with overlapping data in which the standard errors are obtained using the NW, and VARHAC approaches with VAR order selection computed using either the Akaike Information Criteria (VARHAC-AIC) or the Bayesian Information Criteria (VARHAC-BIC).[20] Last, we compute VAR-based tests for lags $p = 1$ and 4. We report these results in Panel b of Table 4. As in the case of short-horizon UIP, we find that the results of the OLS-based UIP tests are somewhat sensitive to the covariance matrix estimator employed. More interesting, the OLS estimate of β at the five-year horizon is positive and larger than one for all countries under consideration, which is consistent with the results in Chinn and Meredith (2007). However, once we test the UIP using a discrete-time VAR approach, we find that the UIP slope is negative, except for the U.K.

Overall, our results are closer to those in Bekaert et al. (2007) in that we find little evidence in favour of UIP at long horizons in our weekly data set. This is in contrast to

[20] When $\tau = 104$ or 260 weeks, the large degree of overlap makes impossible to compute Eichenbaum, Hansen and Singleton (1988) standard errors.

Chinn and Meredith (2004), who cannot reject the validity of the UIP hypothesis at long horizons on the basis of quarterly data.

5 Final Remarks

In this paper we focus on the impact of temporal aggregation on the statistical properties of traditional tests of UIP, where by temporal aggregation we mean the fact that exchange rates evolve on a much finer time-scale than the frequency of observations typically employed by empirical researchers. While in many areas of economics collecting data is very expensive, nowadays the sampling frequency of exchange rates and interest rates is to a large extent chosen by the researcher.

Two main problems arise when we consider the impact of the choice of sampling frequency on traditional UIP tests. In the regression approach, if the period of the forward contract is longer than the sampling interval, the resulting overlapping observations will produce serially correlated regression errors. This fact in turn leads to unreliable finite sample inferences to the extent that, if the degree of overlap becomes non-trivial relative to the sample size, standard GMM asymptotic theory no longer applies. In the VAR approach, in contrast, the problem is that if high frequency observations of the forward premia and the rate of depreciation satisfy a VAR model, then low frequency observations of the same variables will typically satisfy a more complex VARMA model. But since UIP tests in a multivariate framework are joint tests of the UIP hypothesis and the specification of the joint stochastic process for forward premia and exchange rates, dynamic misspecifications will often result in misleading UIP tests.

Motivated by these two problems, we assume that there is an underlying joint process for exchange rates and interest rate differentials that evolves in continuous time. We then estimate the parameters of the underlying continuous process on the basis of discretely sampled data, and test the implied UIP restrictions. Our approach has the advantage that we can accommodate situations with a large ratio of observations per contract period, with the corresponding gains in terms of asymptotic power. At the same time, though, the model that we estimate is the same irrespective of the sampling frequency. Our Monte Carlo results suggest that: (i) in situations where traditional tests of the UIP hypothesis have size distortions, a test based on our continuous-time approach has the right size, and (ii) in situations where existing tests have the right size, our proposed test is more powerful.

However, if the joint autocorrelation structure implied by our continuous-time model is not valid, then our proposed UIP test may also become misleading. For this reason, we introduce a specification test that exploits the fact that the structure of a continuous-time model is the same regardless of the discretization frequency. Specifically, we estimate the model using the whole sample first, then using lower frequency observations only, and decide if those two estimators are "statistically close".

Finally, we apply our continuous-time approach to test UIP at both short and long-horizons on the U.S. dollar bilateral exchange rates against the British pound, the German DM-Euro and the Canadian dollar using weekly data. We use Eurocurrency interest rates of maturities one, three, six-months and one-year to test UIP at short horizons, while we use zero-coupon bond yields of maturities one, two and five years to test it at long horizons. Note that our methodology is especially useful to handle the large degree of overlap (relative to the sample size) that characterizes the UIP hypothesis at long horizons. While Chinn and Meredith (2004) use quarterly data in their regression tests, we use weekly data. Importantly, we also use our proposed specification test to check the validity of the continuous-time processes that we estimate. In this sense, the empirical results obtained with our specification test do not suggest a clear pattern for the relationship between contract length and model specification.

The results that we obtain with correctly specified models continue to reject the UIP hypothesis at short-horizons even after taking care of temporal aggregation problems. Our findings also indicate little support for the UIP at long-horizons. This is in contrast to Chinn and Meredith (2004), who cannot reject the validity of the UIP hypothesis at long-horizons on the basis of quarterly data.

Our Monte Carlo experiments have also confirmed that the UIP regression tests are sensitive to the covariance matrix estimator employed, and that although some automatic lag selection procedures provide more reliable inferences, they are far from perfect. Thus, there is still scope for improvement in this respect. In particular, a fruitful avenue for further research would be to consider bootstrap procedures to reduce size-distortions. However, given that the regressor is not strictly exogenous, a feasible bootstrap procedure may require an auxiliary ad-hoc specification of the data generating process, which would be subject to the same criticisms as the discrete-time VAR approach. In contrast, a parametric bootstrap procedure would be a rather natural choice for our dynamic specification test.

One open question is whether a well-specified continuous-time model such as ours is more apt to handle the persistence of the forward premium than the standard regression-based approach, as our Monte Carlo results seem to suggest. Again, we leave this issue for further research.

Another area that deserves further investigation is the development of alternative continuous time models for exchange rates and interest rate differentials that can account for the rejections of the UIP hypothesis that our empirical results have confirmed. Some progress along these lines can be found, for example, in Diez de los Rios (2009) who proposes a two-country model to explain exchange rates and the term structure of forward premia with two factors.

Similarly, our continuous-time approach can also be used to derive new tests involving long-horizon regressions, such as tests for the validity of the expectation hypothesis of the term structure of interest rates or the long-horizon predictability of excess stock returns or exchange rates. For example, if we were interested in testing the hypothesis of no-predictability of exchange rates, we could do so by replacing (12) by the alternative condition:

$$\mathbf{e}_2'\mathbf{A}^{-1}(e^{\mathbf{A}\tau} - \mathbf{I}) = \mathbf{0}.$$

Finally, it is worth mentioning that our specification test can also be applied to check the dynamic specification of discrete time models such as (34), which have clear implications for the behaviour of exchange rates and interest rate differentials observed at lower frequencies. In fact, our test can in principle be applied to any continuous-time or discrete-time process. This constitutes another interesting avenue for further research.

References

Andrews, D.W.K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817–58.

Baillie, R.T. and T. Bollerslev (2000): "The forward premium anomaly is not as bad as you think," *Journal of International Money and Finance* 19, 471-88.

Baillie, R.T., R.E. Lippens and P.C. McMahon (1984): "Interpreting econometric evidence on efficiency in the foreign exchange market," *Oxford Economic Papers* 36, 67-85.

Bansal R. and M. Dahlquist (2000): "The forward premium puzzle: different tales from developed and emerging economies," *Journal of International Economics*, 51, 115-44.

Bekaert, G. (1995): "The time variation of expected returns and volatility in foreign-exchange markets," *Journal of Business and Economic Statistics* 13, 397-408.

Bekaert, G. and R.J. Hodrick (1992): "Characterizing predictable components in excess returns on equity and foreign exchange markets," *Journal of Finance*, 47, 467-509.

Bekaert, G. and R.J. Hodrick (1993): "On biases in the measurement of foreign exchange risk premiums," *Journal of International Money and Finance,* 12, 115-38.

Bekaert, G. and R.J. Hodrick (2001): "Expectations hypotheses tests," *Journal of Finance* 56, 1357-93.

Bekaert, G., M. Wei and Y. Xing (2007): "Uncovered interest rate parity and the term structure," *Journal of International Money and Finance*, 26, 1038-69.

Bergstrom, A.R. (1983): "Gaussian estimation of structural parameters in higher order continuous time dynamic models," *Econometrica*, 51, 117-152.

Bergstrom, A.R. (1984): "Continuous time stochastic models and issues of aggregation over time," in *Handbook of Econometric*s, vol 2, edited by Z. Griliches and M. D. Intriligator, North-Holland, Amsterdam.

Bergstrom, A.R. and M.J. Chambers (1990): "Gaussian estimation of a continuous time model of demand for consumer durable goods with applications to demand in the United Kingdom, 1973-84," in *Continuous Time Econometric Modeling*, edited by A.R. Bergstrom, Oxford University Press, Oxford.

Brandt M.W. and P. Santa-Clara (2002): "Simulated Likelihood Estimation of Diffussions with an Application to Exchange Rate Dynamics in Incomplete Markets," *Journal of Financial Economics*, 63, 161-210.

Brennan, M.J., and Y. Xia (2006): "International Capital Markets and Foreign Exchange Risk," *Review of Financial Studies*, 19, 753-795.

Brockwell, P.J. (2001): "Continuous-time ARMA processes," in *Handbook of Statistics*, edited by D. N. Shanbhag and C. R. Rao, Elsevier Science, The Netherlands.

Burnside, C., M. Eichenbaum, I. Kleshchelski and S. Rebelo (2006): "The returns to currency speculation," National Bureau of Economic Research Working Paper 12489.

Campbell, J. and R.J. Shiller (1987): "Cointegration and tests of present value models," *Journal of Political Economy* 95, 1062-88.

Chambers, M.J. (2003): "The asymptotic efficiency of cointegration estimators under temporal aggregation," *Econometric Theory* 19, 49-77.

Chambers, M.J. and M.A. Thornton (2008): "Discrete Time Representation of Continuous Time ARMA Processes," Mimeo University of Essex.

Chen, B. and P.A. Zadrozny (2001): "Analytic derivatives of the matrix exponential for estimation of linear continuous-time models," *Journal of Economic Dynamics and Control* 25, 1867-79.

Chinn, M.D. and G. Meredith (2004): "Monetary policy and long-horizon uncovered interest parity," *IMF Staff Papers*, 51, 409-430.

Comte, F. (1999): "Discrete and continuous time cointegration," *Journal of Econometrics*, 88, 207-226.

Davidson R. and J.G. MacKinnon (1998): "Graphical methods for investigating the size and power of hypothesis tests," *The Manchester School* 66, 1-26.

De Long J.B., A. Schleifer and L.H. Summers. (1990): "Noise trader risk in financial markets," *Journal of Political Economy* 98, 703-38.

Den Haan, W.J. and A.T. Levin (1996): "A practitioner's guide to robust covariance estimation," National Bureau of Economic Research Technical Working Paper 197.

Diez de los Rios, A. (2009): "Can affine term structure models help us predict exchange rates?," *Journal of Money, Credit and Banking*, 41 (4), pp. 755-766.

Duffee, G.R. (2002): "Term Premia and Interest Rate Forecasts in Affine Models." *Journal of Finance*, 57, 405-443.

Durbin, J. (1960) "The fitting of time series models," *Review of the International Statistical Institute* 28, 233-44.

Eichenbaum, M., L.P. Hansen and K.J. Singleton (1988): "A time series analysis of representative consumer models of consumption and leisure choice under uncertainty," *Quarterly Journal of Economics* 103, 51-78.

Fisher, M. and C. Gilles (1996): "Estimating exponentially-affine models of the term structure," Federal Reserve Board Mimeo.

Flood, R.P and A.K. Rose (2002): "Uncovered interest parity in crisis," *IMF Staff Papers* 49:2, 252-266.

Frenkel, J. A. (1977): "The forward exchange rate, expectations, and the demand for money: the German hyperinflation," *American Economic Review* 67, 653-70.

Hallwood C.P. and R. MacDonald (1994): *International Money and Finance*, 2nd ed. Blackwell, Oxford.

Hakkio C.S. (1981): "Expectations and the forward exchange rate," *International Economic Review* 22, 663-678.

Hansen, L.P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica* 50, 1029-54.

Hansen, L.P. and R.J. Hodrick (1980): "Forward exchange rates as optimal predictors of future spot rates: an econometric analysis," *Journal of Political Economy* 88, 829-853.

Hansen, L.P. and R.J. Hodrick (1983): "Risk averse speculation in the forward foreign exchange market: an econometric analysis of linear models," in *Exchange Rates and International Macroeconomics*, ed. by J.A. Frenkel. Chicago: University of Chicago Press for National Bureau of Economic Research.

Hansen, L.P. and T.J. Sargent (1991): "Lecture notes on least-squares prediction theory," in *Rational Expectations Econometrics*, edited by L.P. Hansen and T. J. Sargent. Westview Press, 1991.

Hansen, L.P. and T.J. Sargent (1991): "Prediction formulas for continuous-time linear rational expectations models," in *Rational Expectations Econometrics*, edited by L.P. Hansen and T. J. Sargent. Westview Press, 1991.

Hodrick, R.J. (1992): "Dividend yields and expected stock returns: alternative procedures for inference and measurement," *Review of Financial Studies* 5, 357-86.

Harvey, A. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge.

Kyu Moh, Y. (2006): "Continuous-time model of uncovered interest parity with regulated jump-diffusion interest differential," *Applied Economics*, 38, 2523-2533.

Lewis, K. (1989): "Changing beliefs and systematic rational forecast errors with evidence from foreign exchange," *American Economic Review* 79, 621-36.

Ligeralde, A.J. (1997): "Covariance matrix estimators and tests of market efficiency," *Journal of International Money and Finance* 16, 323-43.

Marcellino, M. (1999): "Some consequences of temporal aggregation in empirical analysis," *Journal of Business and Economic Statistics* 17, 129-36.

Mariano R. and Y. Murasawa (2003): "A new coincident index of business cycles based on monthly and quarterly series," *Journal of Applied Econometrics* 18, 427-443.

Mark, N.C. and Y. Kyu Moh (2007): "Official interventions and the forward premium anomaly," *Journal of Empirical Finance* 14, 499-522.

Mark, N.C. and Y. Wu (1998): "Rethinking deviations from uncovered interest rate parity: the role of covariance risk and noise," *Economic Journal* 108, 1686-706.

Maynard, A. and P.C.B. Phillips (2001): "Rethinking an old empirical puzzle: econometric evidence on the forward discount anomaly," *Journal of Applied Econometrics* 16, 671-708.

Meddahi, N. and E. Renault (1996): "Aggregations and Marginalization of GARCH and Stochastic Volatility Models," Université de Montréal Mimeo.

Meddahi, N. and E. Renault (2004): "Temporal aggregation of volatility models," *Journal of Econometrics*, 119, 355-379.

McCrorie, J. R. and M.J. Chambers (2006): "Granger causality and the sampling of economic processes," *Journal of Econometrics* 132, 311-36.

Newey, W.K. and K.D. West (1987): "A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica* 55, 703-8.

Phillips, P.C.B. (1991): "Error correction and long run equilibrium in continuous time," *Econometrica* 59, 967-80.

Renault, E., K. Sekkat and A. Szafarz (1998): "Testing for spurious causality in exchange rates," *Journal of Empirical Finance*, 5, 47-66

Richardson, M. and J.M. Stock (1989): "Drawing inferences from statistics based on multiyear asset returns," *Journal of Financial Economics* 25, 323-48.

Ryu, K. (1994): "Group duration analysis of the proportional hazard model: minimum chi-squared estimators and specification Tests,", *Journal of the American Statistical Association*, 89, 1386-97.

Valkanov, R. (2003): "Long-horizon regressions: theoretical results and applications," *Journal of Financial Economics* 68, 201-232.

Appendix

A Discrete-time VARMA(2,1) representation of the continuous-time model in Example 1

In this appendix we derive an alternative discrete-time representation of the continuous-time model in example 1 by eliminating the unobservable variable $u_2(t)$ from the system by substitution (as in Bergstrom and Chambers, 1990). As noted in the main text of the paper, we can express the model in equation (17) as the following augmented OU model:

$$d\begin{bmatrix} \mathbf{u}(t) \\ s(t) \end{bmatrix} = \begin{pmatrix} \mathbf{\Phi} & 0 \\ \mathbf{e}_2' & 0 \end{pmatrix} \begin{bmatrix} \mathbf{u}(t) \\ s(t) \end{bmatrix} dt + \begin{pmatrix} \mathbf{\Sigma}^{1/2} & 0 \\ \boldsymbol{\alpha}' & 0 \end{pmatrix} \begin{bmatrix} \boldsymbol{\zeta}_u(dt) \\ \zeta_s(dt) \end{bmatrix}, \tag{48}$$
$$d\boldsymbol{\xi}(t) = \mathbf{B}\boldsymbol{\xi}(t)dt + \mathbf{S}^{1/2}\boldsymbol{\zeta}(dt),$$

which generates discrete time observations that satisfy the following VAR(1) model:

$$\boldsymbol{\xi}_t = \mathbf{F}\boldsymbol{\xi}_{t-1} + \boldsymbol{\eta}_t, \tag{49}$$

where $\mathbf{F} = \exp(\mathbf{B})$, the error term satisfies $E[\boldsymbol{\eta}_t] = 0$, $E[\boldsymbol{\eta}_t\boldsymbol{\eta}_t'] = \mathbf{\Omega} = \int_0^1 e^{\mathbf{B}r}\mathbf{S}e^{\mathbf{B}'r}dr$, and $E\left[\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-s}'\right] = \mathbf{0}$ for $s \geq 1$; and where, without loss of generality, we have set $h = 1$ and dropped superscripts (with respect to the notation used in the main text) for clarity of exposition.

In particular, note that both the structure of $\mathbf{B}$ in (48), and the discrete-time VAR(1) representation in (49) imply that:

$$p_{t,\tau} = f_{11}p_{t-1} + f_{12}u_{2t-1} + \eta_{1t}, \tag{50}$$
$$u_{2t} = f_{21}p_{t-1} + f_{22}u_{2t-1} + \eta_{2t}, \tag{51}$$
$$\Delta s_t = f_{31}p_{t-1} + f_{32}u_{2t-1} + \eta_{3t}, \tag{52}$$

with where f_{ij} is the ij-th element of $\mathbf{F}$.

Lagging (52) and solving for u_{2t-2}, we obtain:

$$u_{2t-2} = \frac{1}{f_{32}}\Delta s_{t-1} - \frac{f_{31}}{f_{32}}p_{t-2,\tau} - \frac{1}{f_{32}}\eta_{3t-1}. \tag{53}$$

Finally, substituting lagged (51) and (53) into (50) and (52), we obtain the following VARMA(2,1) system for $\mathbf{y}_t = (p_{t,\tau}, \Delta s_t)'$:

$$\mathbf{y}_t = \mathbf{\Pi}_1\mathbf{y}_{t-1} + \mathbf{\Pi}_2\mathbf{y}_{t-2} + \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1\boldsymbol{\varepsilon}_{t-1},$$

where

$$\mathbf{\Pi}_1 = \begin{pmatrix} f_{11} & f_{12}f_{22}f_{32}^{-1} \\ f_{31} & f_{22} \end{pmatrix}, \qquad \mathbf{\Pi}_2 = \begin{pmatrix} f_{12}f_{21} - f_{12}f_{22}f_{31}f_{32}^{-1} & 0 \\ f_{32}f_{21} - f_{22}f_{31} & 0 \end{pmatrix},$$

and the error term $\boldsymbol{\varepsilon}_t$ satisfies $E[\boldsymbol{\varepsilon}_t] = 0$, $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] = \boldsymbol{\Sigma}$, and $E\left[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t-s}'\right] = \mathbf{0}$ for $s \geq 1$; and where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Theta}_1$ satisfy the equations:

$$\begin{aligned} \boldsymbol{\Sigma} + \boldsymbol{\Theta}_1 \boldsymbol{\Sigma} \boldsymbol{\Theta}_1' &= \mathbf{C}_0 \boldsymbol{\Omega} \mathbf{C}_0' + \mathbf{C}_1 \boldsymbol{\Omega} \mathbf{C}_1', \\ \boldsymbol{\Theta}_1 \boldsymbol{\Sigma} &= \mathbf{C}_1 \boldsymbol{\Omega} \mathbf{C}_0', \end{aligned}$$

with

$$\mathbf{C}_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{C}_1 = \begin{pmatrix} 0 & f_{12} & -f_{12} f_{22} f_{32}^{-1} \\ 0 & f_{32} & -f_{22} \end{pmatrix}.$$

B Initial values for the optimization algorithm

Example 1. We obtain initial values for the scoring algorithm by exploiting the Euler discretization of the model in equation (17), which is given by:

$$\begin{pmatrix} p_{t,\tau} \\ u_{2t} \\ \Delta s_t \end{pmatrix} = \begin{pmatrix} 1+\phi_{11} & \phi_{12} & 0 \\ \phi_{21} & 1+\phi_{22} & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} p_{t-1,\tau} \\ u_{2t-1} \\ \Delta s_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t}^{euler} \\ \eta_{2t}^{euler} \\ \eta_{3t}^{euler} \end{pmatrix}.$$

We proceed as follows:

1. We first compute the sample average of the forward premium and the rate of depreciation to estimate μ_p and $\mu_{\Delta s}$.

2. Then, we estimate the VAR(p) model

   $$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{A}_2 \mathbf{y}_{t-2} + \ldots \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{e}_t,$$

   for $\mathbf{y}_t = (p_{t,\tau}, \Delta s_t)'$ where $p_{t,\tau}$ and Δs_t are the demeaned forward premium and rate of depreciation, respectively.

3. Given that $E_{t-1}\Delta s_t$ is exactly equal to u_{2t} in this discretization scheme, we use the VAR coefficient estimators to construct estimates $\hat{u}_{2t}$ of the conditional mean of Δs_t using the fact that

   $$E_{t-1}\Delta s_t = \mathbf{e}_2' \left(\mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{A}_2 \mathbf{y}_{t-2} + \ldots \mathbf{A}_p \mathbf{y}_{t-p} \right).$$

   As a by-product, we also obtain $\hat{e}_{2t}$ as an estimate of η_{3t}^{euler}.

4. Next, we estimate the VAR(1) model $\widehat{\mathbf{u}}_t = \mathbf{F}\widehat{\mathbf{u}}_{t-1} + \mathbf{v}_t$ for $\widehat{\mathbf{u}}_t = (p_{t,\tau}, \hat{u}_{2t})'$. From here, we can obtain an estimate of $\boldsymbol{\Phi}$ as $\hat{\boldsymbol{\Phi}} = \hat{\mathbf{F}} - \mathbf{I}$. In addition, we also obtain $\widehat{\mathbf{v}}_t$ as an estimate of $(\eta_{1t}^{euler}, \eta_{2t}^{euler})'$.

5. Finally, we obtain estimates of $\boldsymbol{\Sigma}^{1/2}$ and $\boldsymbol{\alpha}$ in the following way. We first estimate $\boldsymbol{\Omega}$, which is the covariance matrix of $(\eta_{1t}^{euler}, \eta_{3t}^{euler})$, with the sample covariance of $\widehat{\mathbf{z}}_t = (\widehat{\eta}_{1t}^{euler}, \widehat{\eta}_{3t}^{euler})'$. Next, we use $\widehat{l}_{11}$, $\widehat{l}_{21}$ and $\widehat{l}_{22}$ as estimates of σ_{11}, α_1 and α_2, respectively, where $\mathbf{LL}'$ is the Cholesky decomposition of $\boldsymbol{\Omega}$. Finally, we estimate σ_{21} and σ_{22} as the coefficients in the regression of $\widehat{\eta}_{2t}^{euler}$ on $\widehat{\mathbf{z}}_t^*$, where $\mathbf{z}_t^* = \mathbf{L}^{-1}\mathbf{z}_t$.

Example 2. To obtain initial values for the scoring algorithm, we exploit the fact that the discrete-time representation in equation (20) is a VAR(1) model with coefficient restrictions. We proceed as follows:

1. We first compute the sample average of the forward premium and the rate of depreciation to estimate μ_p and $\mu_{\Delta s}$.

2. Then, we estimate the VAR(1) model

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{e}_t \qquad E\left[\mathbf{e}_t\mathbf{e}_t'\right] = \mathbf{\Omega}.$$

   for $\mathbf{y}_t = (p_{t,\tau}, \Delta s_t)'$ where $p_{t,\tau}$ and Δs_t are the demeaned forward premium and rate of depreciation, respectively; and subject to the restrictions that $a_{21} = a_{22} = 0$.

3. Finally, we recover the structural parameters in $\mathbf{\Phi}$ and $\mathbf{\Gamma}$ from the restricted reduced form parameters in the VAR(1) using the fact that:

$$\begin{aligned}
\varphi_{11} &= \log(a_{11}), \\
\varphi_{21} &= \frac{a_{21}\varphi_{11}}{e^{\varphi_{11}} - 1}, \\
g_{11} &= \frac{2\varphi_{11}\omega_{11}}{e^{2\varphi_{11}} - 1}, \\
g_{21} &= \frac{\varphi_{11}}{e^{\varphi_{11}} - 1}\left[\omega_{21} - \frac{\sqrt{g_{11}}\varphi_{21}}{2\varphi_{11}^2}\left(e^{\varphi_{11}} - 1\right)^2\right], \\
g_{22} &= \omega_{22} - \frac{\sqrt{g_{11}}\varphi_{21}^2}{2\varphi_{11}^3}\left(2\varphi_{11} + e^{2\varphi_{11}} - 4e^{\varphi_{11}} + 3\right) - \frac{2\sqrt{g_{11}}\varphi_{21}}{\varphi_{11}^3}\left(e^{\varphi_{11}} - \varphi_{11} - 1\right),
\end{aligned}$$

   where ω_{ij} is the ij^{th} element of $\mathbf{\Omega}$ and g_{ij} is the ij^{th} element of $\mathbf{\Gamma}$.

Table 1
Estimates of the Continuous-Time Model in Example 2
Panel a: Uncovered Interest Parity at Short Horizons

Contract	φ_{11}	φ_{21}	γ_{11}	γ_{21}	γ_{22}	μ_p	$\mu_{\Delta s}$	β	$\ln L(\boldsymbol{\theta})$
U.K.									
$\tau = 4$ weeks	-0.015	-0.537	0.041	-0.110	1.367	-0.192	0.014	-2.084	76.11
	(0.010)	(0.199)	(0.002)	(0.038)	(0.042)	(0.086)	(0.061)	(0.780)	
$\tau = 13$ weeks	-0.010	-0.206	0.095	-0.049	1.371	-0.582	0.019	-2.519	-1203.59
	(0.008)	(0.070)	(0.004)	(0.041)	(0.042)	(0.336)	(0.079)	(0.886)	
$\tau = 26$ weeks	-0.009	-0.112	0.170	-0.054	1.370	-1.082	0.020	-2.605	-2087.63
	(0.007)	(0.037)	(0.006)	(0.054)	(0.041)	(0.648)	(0.083)	(0.934)	
$\tau = 52$ weeks	-0.008	-0.063	0.297	-0.094	1.368	-1.960	0.021	-2.652	-2924.44
	(0.007)	(0.021)	(0.011)	(0.056)	(0.041)	(1.236)	(0.089)	(1.063)	
Germany									
$\tau = 4$ weeks	-0.008	-0.222	0.032	-0.168	1.466	0.130	0.025	-0.873	326.32
	(0.005)	(0.175)	(0.002)	(0.053)	(0.033)	(0.085)	(0.044)	(0.688)	
$\tau = 13$ weeks	-0.006	-0.075	0.075	-0.175	1.466	0.384	0.026	-0.934	-955.95
	(0.004)	(0.060)	(0.004)	(0.051)	(0.033)	(0.276)	(0.045)	(0.760)	
$\tau = 26$ weeks	-0.005	-0.039	0.141	-0.154	1.468	0.783	0.026	-0.957	-1903.93
	(0.003)	(0.031)	(0.008)	(0.048)	(0.033)	(0.536)	(0.045)	(0.766)	
$\tau = 52$ weeks	-0.005	-0.022	0.239	-0.184	1.465	1.595	0.026	-1.004	-2701.73
	(0.002)	(0.017)	(0.013)	(0.049)	(0.032)	(0.994)	(0.046)	(0.800)	
Canada									
$\tau = 4$ weeks	-0.026	-0.362	0.032	0.048	0.724	-0.065	-0.010	-1.375	1401.08
	(0.009)	(0.139)	(0.002)	(0.022)	(0.019)	(0.031)	(0.021)	(0.529)	
$\tau = 13$ weeks	-0.018	-0.128	0.076	0.075	0.722	-0.197	-0.010	-1.480	112.22
	(0.006)	(0.047)	(0.004)	(0.025)	(0.019)	(0.107)	(0.022)	(0.547)	
$\tau = 26$ weeks	-0.015	-0.065	0.131	0.090	0.721	-0.375	-0.010	-1.384	-708.88
	(0.005)	(0.024)	(0.005)	(0.022)	(0.019)	(0.216)	(0.022)	(0.531)	
$\tau = 52$ weeks	-0.014	-0.037	0.221	0.087	0.721	-0.701	-0.011	-1.349	-1502.97
	(0.004)	(0.013)	(0.008)	(0.021)	(0.019)	(0.395)	(0.022)	(0.502)	

Note: Robust standard errors in parenthesis. Sample Period: January 1977 to December 2005; 1,513 weekly observations.

Table 1
Estimates of the Continuous-Time Model in Example 2
Panel b: Uncovered Interest Parity at Long Horizons

Contract	φ_{11}	φ_{21}	γ_{11}	γ_{21}	γ_{22}	μ_p	$\mu_{\Delta s}$	β	$\ln L(\boldsymbol{\theta})$
U.K.									
$\tau = 52$ weeks	-0.003	-0.003	0.132	-0.437	1.152	-1.667	0.018	-0.132	-645.90
	(0.005)	(0.064)	(0.009)	(0.101)	(0.043)	(1.870)	(0.063)	(3.092)	
$\tau = 104$ weeks	-0.008	-0.002	0.341	-0.331	1.187	-3.037	0.006	-0.176	-1338.65
	(0.009)	(0.036)	(0.021)	(0.081)	(0.052)	(2.046)	(0.054)	(2.548)	
$\tau = 260$ weeks	-0.012	0.005	0.510	-0.373	1.175	-3.404	0.001	0.437	-1614.91
	(0.008)	(0.023)	(0.021)	(0.074)	(0.053)	(1.722)	(0.046)	(1.793)	
Germany									
$\tau = 52$ weeks	-0.001	-0.053	0.129	-0.191	1.391	-0.820	0.065	-2.686	-767.65
	(0.001)	(0.035)	(0.007)	(0.066)	(0.042)	(2.900)	(0.177)	(1.770)	
$\tau = 104$ weeks	-0.003	-0.035	0.326	-0.143	1.395	-1.033	0.045	-3.137	-1424.13
	(0.004)	(0.020)	(0.019)	(0.056)	(0.042)	(4.357)	(0.173)	(1.985)	
$\tau = 260$ weeks	-0.007	-0.033	0.479	-0.203	1.387	1.104	0.019	-3.893	-1690.00
	(0.006)	(0.018)	(0.016)	(0.063)	(0.042)	(2.871)	(0.117)	(3.382)	
Canada									
$\tau = 52$ weeks	-0.009	-0.039	0.151	0.183	0.826	-0.427	0.000	-1.623	-504.97
	(0.006)	(0.025)	(0.012)	(0.032)	(0.027)	(0.622)	(0.036)	(1.029)	
$\tau = 104$ weeks	-0.013	-0.019	0.357	0.087	0.842	-1.430	0.002	-1.073	-1126.63
	(0.007)	(0.014)	(0.021)	(0.033)	(0.027)	(0.990)	(0.035)	(0.799)	
$\tau = 260$ weeks	-0.010	-0.014	0.425	0.148	0.833	-1.798	-0.001	-1.216	-1242.49
	(0.006)	(0.010)	(0.017)	(0.038)	(0.027)	(1.321)	(0.034)	(0.990)	

Note: Robust standard errors in parenthesis. Sample Period: June 1992 to December 2005; 692 weekly observations.

Table 2
Specification Tests
Panel a: Uncovered Interest Parity at Short Horizons

	$\beta^{(1)} - \beta^{(2)}$	NW	EHS	AIC	BIC
U.K.					
$\tau = 4$ weeks	0.036	[0.836]	[0.688]	[0.497]	[0.727]
$\tau = 13$ weeks	0.010	[0.955]	[0.931]	[0.946]	[0.829]
$\tau = 26$ weeks	-0.025	[0.882]	[0.798]	[0.478]	[0.825]
$\tau = 52$ weeks	-0.061	[0.782]	[0.588]	[0.275]	[0.729]
Germany					
$\tau = 4$ weeks	0.026	[0.840]	[0.675]	[0.548]	[0.989]
$\tau = 13$ weeks	0.025	[0.853]	[0.639]	[0.453]	[0.435]
$\tau = 26$ weeks	0.021	[0.873]	[0.716]	[0.406]	[0.735]
$\tau = 52$ weeks	0.021	[0.878]	[0.704]	[0.335]	[0.743]
Canada					
$\tau = 4$ weeks	-0.084	[0.526]	[0.369]	[0.044]	[0.405]
$\tau = 13$ weeks	-0.110	[0.382]	[0.137]	[0.011]	[0.193]
$\tau = 26$ weeks	-0.075	[0.500]	[0.251]	[0.173]	[0.379]
$\tau = 52$ weeks	-0.084	[0.397]	[0.131]	[0.241]	[0.151]

Note: p-values of the null hypothesis $H_0 : \beta^{(1)} - \beta^{(2)} = 0$ are presented in square brackets. Sample Period: January 1977 to December 2005; 1,513 weekly observations.

Panel b: Uncovered Interest Parity at Long Horizons

	$\beta^{(1)} - \beta^{(2)}$	NW	EHS	AIC	BIC
U.K.					
$\tau = 52$ weeks	-0.034	[0.933]	[0.944]	[0.675]	[0.734]
$\tau = 104$ weeks	0.195	[0.642]	[0.701]	[0.176]	[0.218]
$\tau = 260$ weeks	-0.059	[0.825]	[0.860]	[0.481]	[0.722]
Germany					
$\tau = 52$ weeks	-0.043	[0.886]	[0.929]	[0.490]	[0.394]
$\tau = 104$ weeks	0.086	[0.802]	[0.862]	[0.640]	[0.239]
$\tau = 260$ weeks	-0.027	[0.939]	[0.958]	[0.773]	[0.896]
Canada					
$\tau = 52$ weeks	0.075	[0.890]	[0.922]	[0.617]	[0.534]
$\tau = 104$ weeks	0.258	[0.549]	[0.677]	[0.133]	[0.062]
$\tau = 260$ weeks	0.018	[0.967]	[0.979]	[0.970]	[0.865]

Note: p-values of the null hypothesis $H_0 : \beta^{(1)} - \beta^{(2)} = 0$ are presented in square brackets. Sample Period: June 1992 to December 2005; 692 weekly observations.

Table 3
Estimates of the Continuous-Time Model in Example 1
Uncovered Interest Parity at Short Horizons

Contract	ϕ_{11}	ϕ_{21}	ϕ_{12}	ϕ_{22}	σ_{11}	σ_{21}	σ_{22}	α_1	α_2	μ_p	$\mu_{\Delta s}$	β	$\ln L(\boldsymbol{\theta})$
Canada													
τ = 4 weeks	-0.034	-0.029	-0.021	-0.289	0.032	-0.057	-0.031	0.064	0.736	-0.065	-0.009	-1.094	1407.84
	(0.012)	(0.063)	(0.024)	(0.211)	(0.002)	(0.027)	(0.025)	(0.025)	(0.023)	(0.029)	(0.017)	(0.465)	
τ = 13 weeks	-0.024	-0.012	-0.050	-0.280	0.075	-0.059	-0.032	0.094	0.734	-0.198	-0.009	-0.837	120.74
	(0.008)	(0.019)	(0.053)	(0.176)	(0.004)	(0.025)	(0.024)	(0.027)	(0.023)	(0.099)	(0.017)	(0.497)	

Note: Robust standard errors in parenthesis. Sample Period: January 1977 to December 2005; 1,513 weekly observations.

Table 4
Comparison of Uncovered Interest Parity Tests: Implicit Betas
Panel a: Uncovered Interest Parity at Short Horizons

	NW	EHS	AIC	BIC	NO	VAR(1)	VAR(4)	OU(2)	OU(1)
U.K.									
$\tau = 4$ weeks	-2.071	-2.071	-2.071	-2.071	-2.435	-2.002	-2.107	-2.084	
	(0.917)	(0.714)	(0.965)	(1.031)	(0.915)	(0.790)	(0.836)	(0.780)	
	[0.001]	[0.000]	[0.001]	[0.003]	[0.000]	[0.000]	[0.000]	[0.000]	
$\tau = 13$ weeks	-2.155	-2.155	-2.155	-2.155	-1.859	-2.401	-2.172	-2.519	
	(1.064)	(1.336)	(0.962)	(1.575)	(1.028)	(0.873)	(0.934)	(0.886)	
	[0.003]	[0.018]	[0.001]	[0.045]	[0.005]	[0.000]	[0.001]	[0.000]	
$\tau = 26$ weeks	-2.051	-2.051	-2.051	-2.051	-2.047	-2.407	-2.042	-2.605	
	(1.127)	(1.420)	(1.229)	(1.446)	(1.228)	(0.884)	(0.938)	(0.934)	
	[0.007]	[0.032]	[0.013]	[0.035]	[0.005]	[0.001]	[0.001]	[0.000]	
$\tau = 52$ weeks	-1.507	-1.507	-1.507	-1.507	-1.587	-2.257	-1.989	-2.652	
	(1.090)	(1.448)	(0.937)	(1.595)	(1.809)	(0.894)	(0.919)	(1.063)	
	[0.022]	[0.083]	[0.007]	[0.116]	[0.153]	[0.000]	[0.001]	[0.001]	
Germany									
$\tau = 4$ weeks	-0.828	-0.828	-0.828	-0.828	-0.873	-0.881	-0.835	-0.873	
	(0.665)	(0.828)	(0.698)	(0.778)	(0.861)	(0.703)	(0.736)	(0.689)	
	[0.006]	[0.027]	[0.009]	[0.019]	[0.030]	[0.008]	[0.013]	[0.007]	
$\tau = 13$ weeks	-0.785	-0.785	-0.785	-0.785	-0.658	-0.946	-0.699	-0.934	
	(0.671)	(0.896)	(0.626)	(1.109)	(0.897)	(0.780)	(0.827)	(0.760)	
	[0.008]	[0.046]	[0.004]	[0.108]	[0.028]	[0.013]	[0.040]	[0.011]	
$\tau = 26$ weeks	-0.911	-0.911	-0.911	-0.911	-0.900	-0.971	-0.697	-0.957	
	(0.710)	(0.888)	(0.873)	(0.964)	(0.865)	(0.790)	(0.835)	(0.766)	
	[0.007]	[0.031]	[0.029]	[0.047]	[0.028]	[0.013]	[0.042]	[0.011]	
$\tau = 52$ weeks	-0.756	-0.756	-0.756	-0.756	-0.488	-1.029	-0.823	-1.004	
	(0.702)	(1.492)	(0.942)	(1.039)	(0.787)	(0.834)	(0.880)	(0.800)	
	[0.012]	[0.239]	[0.062]	[0.091]	[0.059]	[0.015]	[0.038]	[0.012]	
Canada									
$\tau = 4$ weeks	-1.081	-1.081	-1.081	-1.081	-1.358	-1.364	-1.119	-1.375	-1.094
	(0.352)	(0.416)	(0.320)	(0.323)	(0.484)	(0.499)	(0.508)	(0.529)	(0.465)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
$\tau = 13$ weeks	-0.842	-0.842	-0.842	-0.842	-0.683	-1.460	-0.940	-1.480	-0.837
	(0.417)	(0.402)	(0.395)	(0.489)	(0.534)	(0.514)	(0.546)	(0.547)	(0.497)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.002]	[0.000]	[0.000]	[0.000]	[0.000]
$\tau = 26$ weeks	-0.746	-0.746	-0.746	-0.746	-0.516	-1.358	-0.983	-1.384	
	(0.473)	(0.474)	(0.582)	(0.490)	(0.511)	(0.496)	(0.527)	(0.531)	
	[0.000]	[0.000]	[0.003]	[0.000]	[0.003]	[0.000]	[0.000]	[0.000]	
$\tau = 52$ weeks	-0.976	-0.976	-0.976	-0.976	-0.919	-1.317	-1.072	-1.349	
	(0.628)	(1.004)	(0.632)	(0.615)	(0.750)	(0.463)	(0.518)	(0.502)	
	[0.000]	[0.049]	[0.002]	[0.001]	[0.011]	[0.000]	[0.000]	[0.000]	

Note: Robust standard errors in parenthesis. p-values for the null hypothesis $H_0 : \beta = 1$ are provided in square brackets. Sample Period: January 1977 to December 2005; 1,513 weekly observations.

Table 4
Comparison of Uncovered Interest Parity Tests: Implicit Betas
Panel b: Uncovered Interest Parity at Long Horizons

	NW	AIC	BIC	VAR(1)	VAR(4)	OU(2)
U.K.						
$\tau = 52$ weeks	0.147	0.147	0.147	1.039	2.424	-0.132
	(2.393)	(1.451)	(1.868)	(2.695)	(2.114)	(3.092)
	[0.722]	[0.648]	[0.788]	[0.988]	[0.501]	[0.714]
$\tau = 104$ weeks	-0.827	-0.827	-0.827	0.290	1.266	-0.176
	(1.622)	(1.219)	(1.912)	(1.850)	(1.620)	(2.548)
	[0.260]	[0.134]	[0.339]	[0.701]	[0.870]	[0.644]
$\tau = 260$ weeks	1.611	1.611	1.611	0.516	1.009	0.437
	(0.564)	(0.889)	(0.584)	(1.411)	(0.773)	(1.793)
	[0.279]	[0.492]	[0.296]	[0.731]	[0.991]	[0.754]
Germany						
$\tau = 52$ weeks	-2.375	-2.375	-2.375	-2.028	-1.701	-2.686
	(1.877)	(1.310)	(3.004)	(1.663)	(1.785)	(1.770)
	[0.072]	[0.010]	[0.261]	[0.069]	[0.130]	[0.037]
$\tau = 104$ weeks	-3.066	-3.066	-3.066	-2.206	-1.898	-3.137
	(1.233)	(0.972)	(1.594)	(1.548)	(1.797)	(1.985)
	[0.001]	[0.000]	[0.011]	[0.038]	[0.107]	[0.037]
$\tau = 260$ weeks	1.566	1.566	1.566	-2.559	-1.505	-3.893
	(1.165)	(3.246)	(3.057)	(2.209)	(1.926)	(3.382)
	[0.627]	[0.862]	[0.853]	[0.107]	[0.193]	[0.148]
Canada						
$\tau = 52$ weeks	-1.703	-1.703	-1.703	-1.583	-1.385	-1.623
	(1.570)	(2.211)	(1.888)	(0.952)	(0.987)	(1.029)
	[0.085]	[0.222]	[0.152]	[0.007]	[0.016]	[0.011]
$\tau = 104$ weeks	-1.672	-1.672	-1.672	-1.018	-1.293	-1.073
	(1.382)	(1.818)	(1.696)	(0.715)	(0.946)	(0.799)
	[0.052]	[0.142]	[0.115]	[0.005]	[0.015]	[0.009]
$\tau = 260$ weeks	1.733	1.733	1.733	-1.218	-1.137	-1.216
	(1.075)	(6.517)	(6.517)	(1.025)	(0.971)	(0.990)
	[0.495]	[0.910]	[0.910]	[0.030]	[0.028]	[0.025]

Note: Robust standard errors in parenthesis. p-values for the null hypothesis $H_0 : \beta = 1$ are provided in square brackets. Sample Period: June 1992 to December 2005; 692 weekly observations.

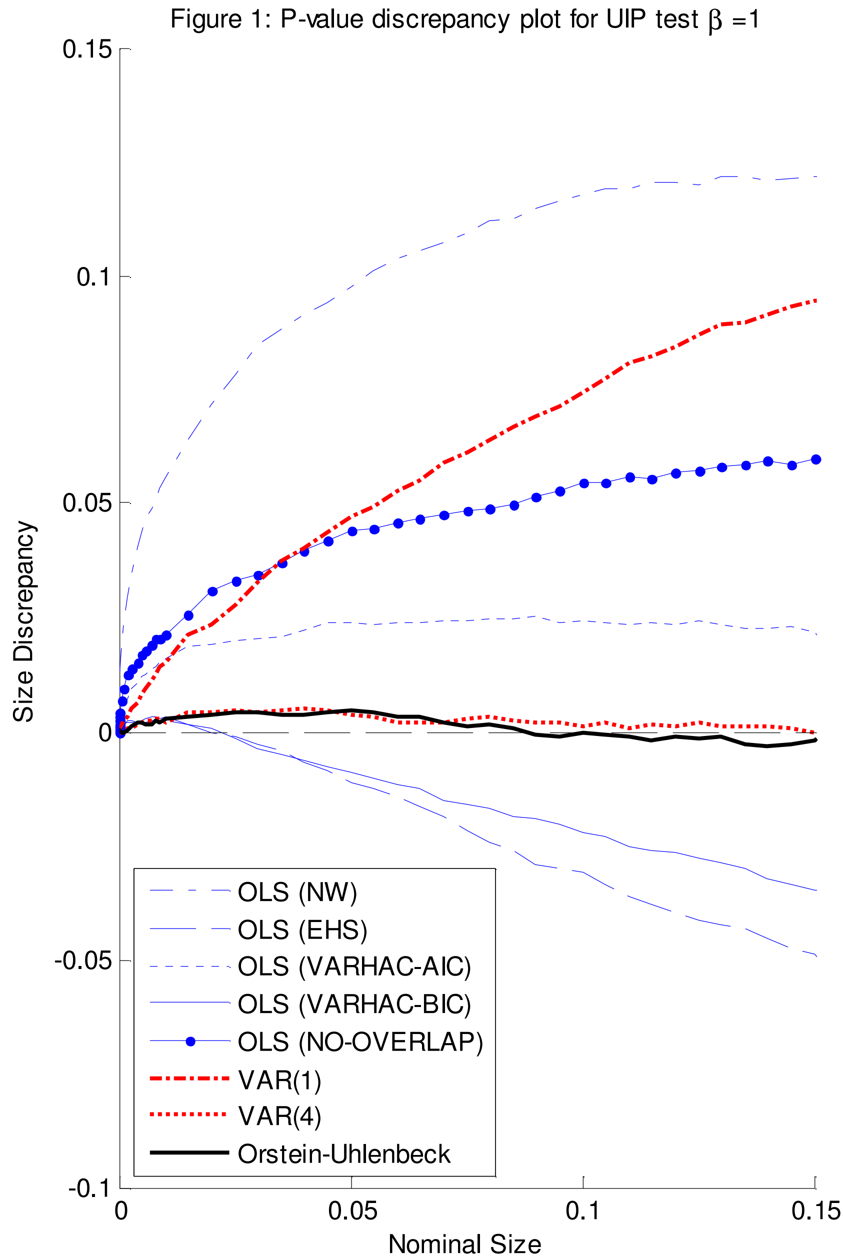
Figure 1: P-value discrepancy plot for UIP test β =1
0.15
0.1
0.05
0
-0.05
-0.1
Size Discrepancy
0
0.05
0.1
0.15
Nominal Size
OLS (NW)
OLS (EHS)
OLS (VARHAC-AIC)
OLS (VARHAC-BIC)
OLS (NO-OVERLAP)
VAR(1)
VAR(4)
Orstein-Uhlenbeck

Figure 2: Size-adjusted power for UIP test β =1

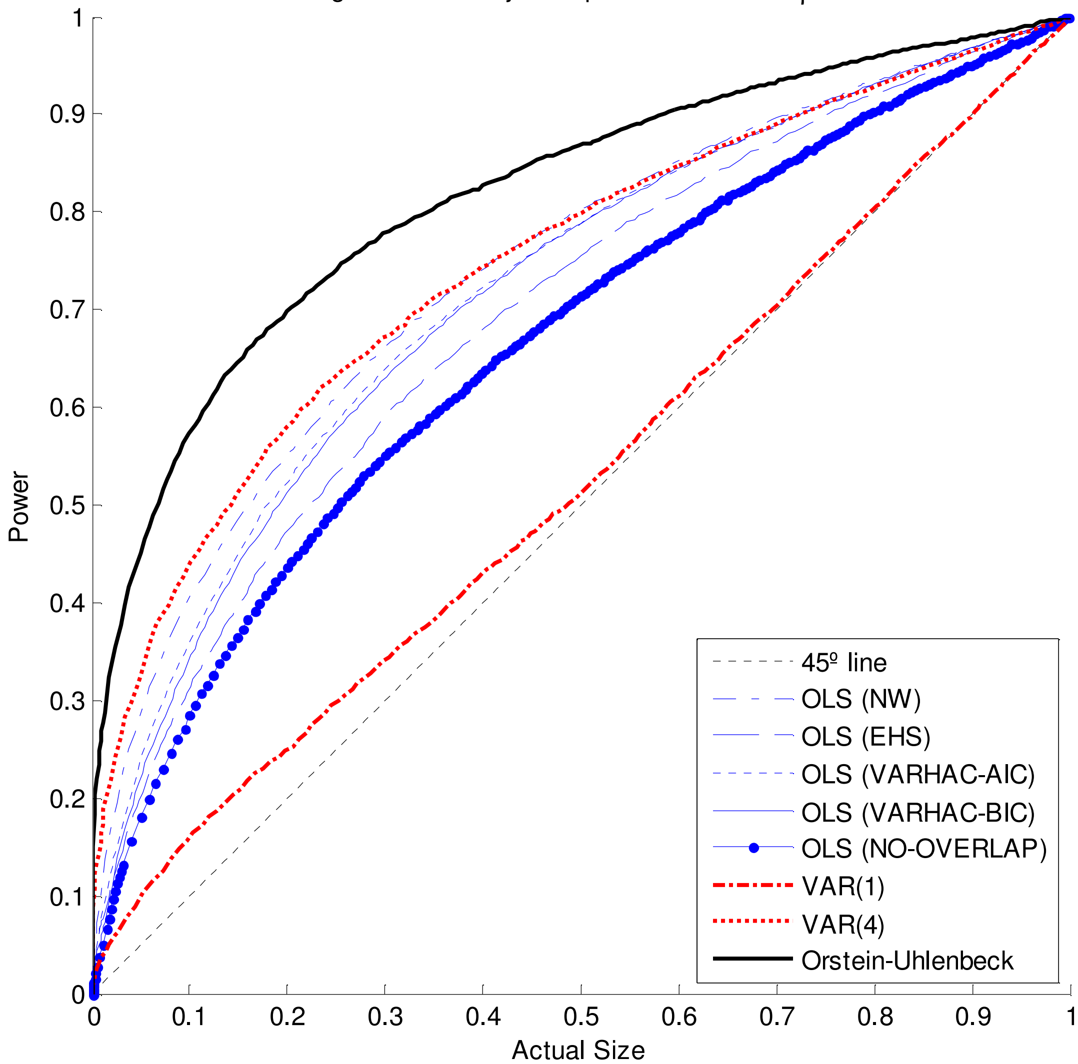

Figure 3: P-value discrepancy plot for specification test $\beta^{(1)}=\beta^{(2)}$

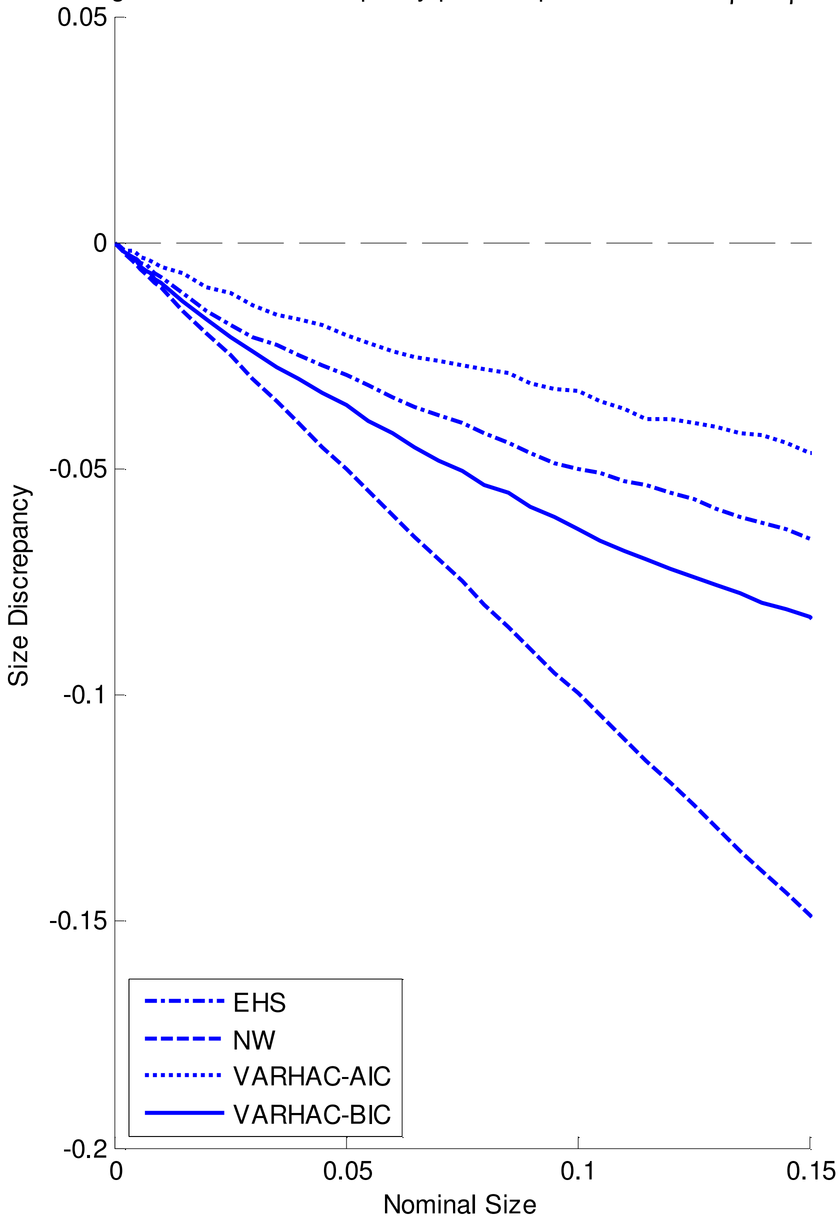

Figure 4: Size-adjusted power for specification test $\beta^{(1)}=\beta^{(2)}$