# Testing Uncovered Interest Parity: A Continuous-Time Approach

by Antonio Diez de los Rios and Enrique Sentana

# Testing Uncovered Interest Parity:
# A Continuous-Time Approach

**by**

**Antonio Diez de los Rios[1] and Enrique Sentana[2]**

[1]Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
diez@bankofcanada.ca

[2]CEMFI
sentana@cemfi.es

# Acknowledgements

# Abstract

Nowadays researchers can choose the sampling frequency of exchange rates and interest rates. If the number of observations per contract period is large relative to the sample size, standard GMM asymptotic theory provides unreliable inferences in UIP regression tests. We specify a bivariate continuous-time model for exchange rates and forward premia robust to temporal aggregation, unlike the discrete time models in the literature. We obtain the UIP restrictions on the continuous-time model parameters, which we estimate efficiently, and propose a novel specification test that compares estimators at different frequencies. Our empirical results based on correctly specified models reject UIP.

*JEL classification: F31, G15*
*Bank classification: Exchange rates; Econometric and statistical methods*

# Résumé

De nos jours, les chercheurs peuvent choisir la fréquence d'échantillonnage des taux de change et des taux d'intérêt. Si la période couverte par le contrat compte un nombre d'observations élevé par rapport à la taille de l'échantillon, le recours à une approximation asymptotique pour tester l'hypothèse de parité des taux d'intérêt non couverte à l'aide de la méthode des moments généralisés peut conduire à des conclusions fallacieuses. Le modèle en temps continu que définissent les auteurs pour l'évolution du taux de change et du report n'est pas sensible à l'agrégation temporelle, contrairement à ceux en temps discret que l'on trouve dans la littérature. Les auteurs obtiennent des estimations efficaces des paramètres du modèle en temps continu, en testant les restrictions associées à la parité des taux d'intérêt non couverte, et proposent un test de spécification novateur qui permet de comparer les estimateurs à différentes fréquences. Les résultats empiriques tirés de l'estimation de modèles bien spécifiés conduisent au rejet de la parité des taux non couverte.

*Classification JEL : F31, G15*
*Classification de la Banque : Taux de change; Modèles économétriques et statistiques*

# 1  Introduction

During the last twenty-five years the majority of studies have rejected the hypothesis of uncovered interest parity (UIP), 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 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).

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 UIP has been lessened in recent studies. In particular, Flood and Rose (2002) find that UIP 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 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 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

---

[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).

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 onto traditional UIP 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 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 UIP regression tests (see e.g. Richardson and Stock, 1989). For example, imagine that we are interested in testing UIP using weekly data on 3-month forward contracts, as in Hansen and Hodrick (1980). Since the degree of overlapping is only 12 periods, we may expect the usual asymptotic results to be reliable if the sample size is reasonably large. But if we decide to use daily data instead, then we will have an overlapping degree of 60 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 overlapping, 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 test the constraints that UIP 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 overlapping 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 UIP in a multivariate framework is a joint test of the UIP hypothesis and the dynamic specification of the model, and like in many other contexts, having a misspecified model will often result in misleading UIP tests.

Motivated by these two problems, we use a continuous-time approach to derive a new test of uncovered interest rate parity. In particular, we assume that there is an underlying continuous-time joint process for exchange rates and interest rate differentials, which can be observed at discrete points of 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.

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 generation 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.

We begin our analysis by deriving the conditions that uncovered interest parity imposes on the Wold decomposition of continuous-time processes.[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 UIP hypothesis have size distortions, the test based on our continuous-time

---

[2]Throughout this paper, we equate linear projections to conditional expectations.

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 novel Hausman 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 the UIP hypothesis on the U.S. dollar bilateral exchange rates against the British pound, the German DM-Euro and the Canadian dollar using weekly data over the period from January 1977 to December 2005. 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 uncovered interest parity hypothesis even after taking care of temporal aggregation problems.

The paper is organized as follows. Section 2 details our dynamic framework, the testable restrictions that uncovered interest parity imposes on continuous-time models, 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. Proofs and auxiliary results are gathered in appendices.

## 2   A continuous-time framework

### 2.1   Conditions for UIP

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 $s_t$ 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 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 $\tau$-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. In order to economise on the use of constants, we will also understand $p_{t,\tau}$ and $\Delta s_t$ 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 $\Delta s_t$ 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 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 borrow from Phillips (1991) and Chambers (2003) to state the following continuous-time model in which the $(1, -1)$ cointegration relationship is also satisfied:

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

$$ds(t) = u_2(t)dt + dW_s(t), \tag{3}$$

where $\mathbf{u}(t) = [u_1(t), u_2(t)]'$ is a covariance stationary, continuous-time residual,[4] and where we have dropped the dependence of $u_1(t)$ on $\tau$ because we are concentrating on a single forward contract.

---

[3]Most often, uncovered interest parity 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 $\tau$-period interest rates on a deposit denominated in domestic and foreign currency, respectively.

[4]Note that if we drop the $dW_s(t)$ term from (3), then we obtain Phillips (1991)'s continuous-time cointegrated system in triangular form representation. In that case, (3) can 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 $s(t)$ are differentiable, and therefore that the infinitesimal change in $s(t)$ is smooth. However, the assumption of differentiable exchange rate paths does not seem to be supported by data.

In this context, UIP is expressed 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 imposes a set of conditions on the temporal evolution of the forward premia and the exchange rate. As an extreme example, let the forward contract period $\tau$ go to zero (see Mark and Moh, 2006). Then, the restriction $E_t \left[ ds(t) \right] = p_0(t)$ will be satisfied if and only if $u_1(t) = u_2(t) \ \forall t$, 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 $\tau \neq 0$, the following proposition summarizes the conditions which guarantee that UIP holds:

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

$$\mathbf{u}(t) = \int_0^\infty \boldsymbol{\phi}(h) d\mathbf{W}_u(t-h), \tag{5}$$

$\mathbf{W}_u(t)$ *is a 2-dimensional Wiener process with instantaneous covariance matrix given by* $E \left[ d\mathbf{W}_u(t) d\mathbf{W}_u(t)' \right] = \boldsymbol{\Sigma}_u dt$, *and* $\boldsymbol{\phi}(h)$ *is a* $2 \times 2$ *matrix of square integrable functions such that* $tr \left[ \int_0^\infty \boldsymbol{\phi}(h) \boldsymbol{\Sigma}_u \boldsymbol{\phi}(h)' dh \right] < \infty$. *Then, the Uncovered Interest Parity condition (4) holds if and only if:*

$$\phi_{11}(h) = \int_0^\tau \phi_{21}(h+r) dr \qquad \forall h, \tag{6}$$

$$\phi_{12}(h) = \int_0^\tau \phi_{22}(h+r) dr \qquad \forall h, \tag{7}$$

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

This proposition is the continuous-time analogue to the results provided in the appendix of Hansen and Hodrick (1980), who derived the restrictions that UIP implies on the Wold decomposition of discrete-time processes. In the next subsection, we will illustrate it with two empirically realistic examples.

## 2.2   Examples

A multivariate Orstein-Uhlenbeck (O-U) model is a continuous-time process characterized by the system of linear stochastic differential equations with constant coefficients:

$$d\boldsymbol{\xi}(t) = \mathbf{A}\boldsymbol{\xi}(t)dt + \mathbf{S}^{1/2}d\mathbf{W}(t). \tag{8}$$

Under some regularity conditions (see e.g. Bergstrom 1984), (8) generates discrete observations that will exactly satisfy the following VAR model:

$$\boldsymbol{\xi}_t = \mathbf{F}\boldsymbol{\xi}_{t-h} + \boldsymbol{\eta}_t, \tag{9}$$

where $h$ is the discretization interval, $\mathbf{F} = \exp(\mathbf{A}h)$, $\boldsymbol{\eta}_t|\boldsymbol{\xi}_{t-h}, \boldsymbol{\xi}_{t-2h}, \ldots \sim N(\mathbf{0}, \boldsymbol{\Omega})$ and $\boldsymbol{\Omega} = \int_0^h e^{\mathbf{A}r}\mathbf{S}e^{\mathbf{A}'r}dr$. This result illustrates the main advantages of O-U models: they have an exact discrete-time representation whose structure is the same regardless the discretization frequency.

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

$$d\begin{bmatrix} p_\tau(t) \\ s(t) \end{bmatrix} = \begin{pmatrix} \varphi_{11} & 0 \\ \varphi_{21} & 0 \end{pmatrix}\begin{bmatrix} p_\tau(t) \\ s(t) \end{bmatrix}dt + \begin{pmatrix} \gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \end{pmatrix}\begin{bmatrix} dW_1(t) \\ dW_2(t) \end{bmatrix}, \tag{10}$$

$$d\mathbf{v}(t) = \boldsymbol{\Phi}\mathbf{v}(t)dt + \boldsymbol{\Gamma}^{1/2}d\mathbf{W}(t),$$

with $\varphi_{11} < 0$ and $E[d\mathbf{W}(t)d\mathbf{W}(t)'] = \mathbf{I}dt$, where $\mathbf{I}$ is the identity matrix. A direct application of (9) gives us that the discrete-time representation of the continuous time model (10) will be given by the following restricted VAR(1):

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

$$\mathbf{v}_t = e^{\boldsymbol{\Phi}}\mathbf{v}_{t-h} + \boldsymbol{\eta}_t,$$

where $\boldsymbol{\eta}_t = \int_{t-h}^t e^{\boldsymbol{\Phi}(t-r)}\boldsymbol{\Gamma}^{1/2}d\mathbf{W}(r)$. Hence, it follows that 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, $p_\tau(t)$. Rewriting this model in terms of $\mathbf{x}_t = (p_{t,\tau}, \Delta_h s_t)'$ we obtain:

$$\begin{pmatrix} p_{t,\tau} \\ \Delta_h s_t \end{pmatrix} = \begin{pmatrix} e^{\varphi_{11}h} & 0 \\ \frac{\varphi_{21}}{\varphi_{11}}(e^{\varphi_{11}h} - 1) & 0 \end{pmatrix}\begin{pmatrix} p_{t-h,\tau} \\ \Delta_h s_{t-h} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}, \tag{12}$$

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-h} + \boldsymbol{\eta}_t,$$

While we could use Proposition 1 to obtain the UIP conditions for this example, it is easier to set the discretization period $h$ equal to the contract period $\tau$. By doing so, we obtain

that the least squares projection coefficient of $\Delta_\tau s_{t+\tau}$ on $p_{t,\tau}$ is equal to $\varphi_{21}\left(e^{\varphi_{11}\tau}-1\right)/\varphi_{11}$. Thus, $E_t\left(\Delta_\tau s_{t+\tau}\right)=p_{t,\tau}$ is guaranteed if and only if:

$$\varphi_{21} = \frac{\varphi_{11}}{e^{\varphi_{11}\tau}-1}. \tag{13}$$

Finally, estimation of the parameters of the continuous-time model (10) can be performed by maximum likelihood on the basis of (12) by exploiting the fact that:

$$\boldsymbol{\eta}_t|\boldsymbol{\xi}_{t-h},\boldsymbol{\xi}_{t-2h},\ldots \sim N(\mathbf{0},\boldsymbol{\Omega})$$

with $\boldsymbol{\Omega} = \int_0^h e^{\boldsymbol{\Phi}r}\boldsymbol{\Gamma}e^{\boldsymbol{\Phi}'r}dr$.

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

$$p_\tau(t) = u_1(t),$$
$$ds(t) = u_2(t)dt + \boldsymbol{\alpha}'d\mathbf{W}(t),$$

with

$$d\begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}\begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}dt + \begin{pmatrix} \sigma_{11} & 0 \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\begin{bmatrix} dW_1(t) \\ dW_2(t) \end{bmatrix}, \tag{14}$$
$$d\mathbf{u}(t) = \boldsymbol{\Phi}\mathbf{u}(t)dt + \boldsymbol{\Sigma}^{1/2}d\mathbf{W}(t),$$

where $\boldsymbol{\Phi}$ has two negative eigenvalues to guarantee the stationarity of the process, and $E[d\mathbf{W}(t)d\mathbf{W}(t)'] = \mathbf{I}dt$.

To obtain the discrete-time representation of this model, it is convenient to write it first as the following augmented O-U process:

$$d\begin{bmatrix} p_\tau(t) \\ u_2(t) \\ s(t) \end{bmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} & 0 \\ \phi_{21} & \phi_{22} & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{bmatrix} p_\tau(t) \\ u_2(t) \\ s(t) \end{bmatrix}dt + \begin{pmatrix} \sigma_{11} & 0 \\ \sigma_{21} & \sigma_{22} \\ \alpha_1 & \alpha_2 \end{pmatrix}\begin{bmatrix} dW_1(t) \\ dW_2(t) \end{bmatrix}, \tag{15}$$
$$d\boldsymbol{\xi}(t) = \mathbf{A}\boldsymbol{\xi}(t)dt + \mathbf{S}^{1/2}d\mathbf{W}(t),$$

and then apply equation (9), together with the fact that $e^{\mathbf{A}h} = \sum_{k=0}^{\infty}\frac{1}{k!}(\mathbf{A}h)^k$. In this way, we obtain that:

$$\begin{pmatrix} \mathbf{u}_t \\ s_t \end{pmatrix} = \begin{pmatrix} e^{\boldsymbol{\Phi}h} & 0 \\ \mathbf{e}_2'\boldsymbol{\Phi}^{-1}(e^{\boldsymbol{\Phi}h}-\mathbf{I}) & 1 \end{pmatrix}\begin{pmatrix} \mathbf{u}_{t-h} \\ s_{t-h} \end{pmatrix} + \boldsymbol{\eta}_t,$$

where $\boldsymbol{\eta}_t = \int_{t-h}^t e^{\mathbf{A}(t-r)}\boldsymbol{\Sigma}^{1/2}d\mathbf{W}(r)$. Therefore, discretely sampled observations of model (14) will satisfy the following state-space model:

$$\begin{pmatrix} p_{t,\tau} \\ \Delta_h s_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \\ \Delta_h s_t \end{pmatrix}, \tag{16}$$

$$\begin{pmatrix} \mathbf{u}_t \\ \Delta_h s_t \end{pmatrix} = \begin{pmatrix} e^{\boldsymbol{\Phi} h} & 0 \\ \mathbf{e}_2'\boldsymbol{\Phi}^{-1}(e^{\boldsymbol{\Phi} h} - \mathbf{I}) & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_{t-h} \\ \Delta_h s_{t-h} \end{pmatrix} + \boldsymbol{\eta}_t, \tag{17}$$

Once again, setting $h = \tau$ gives us the projections of $s_{t+\tau} - s_t$ onto $(\mathbf{u}_t', s_t - s_{t-\tau})'$. In this case, it is straightforward to prove that the UIP condition $E_t(\Delta_\tau s_{t+\tau}) = p_{t,\tau}$ is equivalent to:

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

where $\mathbf{e}_j$ is a vector of the same dimension as $\mathbf{u}(t)$ with a one in the $j^{th}$ position, and zeroes in the others.

Estimation can also be done by maximum likelihood using the fact that:

$$\boldsymbol{\eta}_t | \boldsymbol{\xi}_{t-h}, \boldsymbol{\xi}_{t-2h}, \ldots \sim N(\mathbf{0}, \boldsymbol{\Omega})$$

with $\boldsymbol{\Omega} = \int_0^h e^{\mathbf{A}r}\mathbf{S}e^{\mathbf{A}'r}dr$. However, note that $u_{2t}$ is an unobservable factor. For that reason, we resort to the Kalman filter to evaluate the exact Gaussian likelihood function of this model.

Finally, it is worth pointing out that the model in Example 1 can be nested within the model in Example 2 in several different ways. In particular, since the condition $E_t[ds(t)] = u_2(t) = \varphi_{21}p_\tau(t)$ implies that:

$$du_2(t) = \varphi_{21}dp_\tau(t) = \varphi_{21}\varphi_{11}p_\tau(t)dt + \varphi_{21}\gamma_{11}dW_1(t),$$

we can write:

$$d\begin{bmatrix} p_\tau(t) \\ u_2(t) \\ s(t) \end{bmatrix} = \begin{pmatrix} \varphi_{11} & 0 & 0 \\ \varphi_{21}\varphi_{11} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{bmatrix} p_\tau(t) \\ u_2(t) \\ s(t) \end{bmatrix} dt + \begin{pmatrix} \gamma_{11} & 0 \\ \varphi_{21}\gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{bmatrix} dW_1(t) \\ dW_2(t) \end{bmatrix}.$$

However, we can also use the fact that $u_2(t) = \varphi_{21}p_\tau(t)$ to prove that

$$du_2(t) = \varphi_{11}u_2(t)dt + \varphi_{21}\gamma_{11}dW_1(t),$$

which delivers the following alternative expression:

$$d\begin{bmatrix} p_\tau(t) \\ u_2(t) \\ s(t) \end{bmatrix} = \begin{pmatrix} \varphi_{11} & 0 & 0 \\ 0 & \varphi_{11} & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{bmatrix} p_\tau(t) \\ u_2(t) \\ s(t) \end{bmatrix} dt + \begin{pmatrix} \gamma_{11} & 0 \\ \varphi_{21}\gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{bmatrix} dW_1(t) \\ dW_2(t) \end{bmatrix}.$$

Therefore, if the true model is given by (10), then some of the parameters appearing in (14) will not be identified.

## 2.3 Comparison to OLS and discrete-time VAR

Let us quickly review the two main approaches to test UIP in the existing literature:

**Ordinary Least Squares.** We can combine the equation $E_t\left(s_{t+\tau} - s_t\right) = f_{t,\tau} - s_t$ with the assumption that expectations are rational to obtain:

$$s_{t+\tau} - s_t = f_{t,\tau} - s_t + w_{t+\tau},$$

where $w_{t+\tau}$ is a rational expectations error that is mean independent from any variable in the information set at time $t$. This equation has motivated the regression equation

$$s_{t+\tau} - s_t = \alpha + \beta(f_{t,\tau} - s_t) + w_{t+\tau} \tag{19}$$

as the usual starting point to test the UIP theory. In the context of (19), the "Unbiasedness Proposition" implies that $\alpha = 0$ and $\beta = 1$, while we just need $\beta = 1$ to satisfy the "Modified Unbiasedness Proposition". In addition, the rational expectations assumption implies that $w_{t+\tau}$ is not autocorrelated as long the sampling interval is equal or larger than $\tau$. For this reason, several authors (see e.g. Frenkel, 1977 among others) sample exchange and interest rates every $\tau$ periods to produce a data set with non-overlapping residuals, even though doing so entails a considerable waste of sample information. In contrast, 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) Generalized Method of Moments, they obtain asymptotic standard errors that take into account the serial correlation of the regression errors when the prediction horizon is higher than the sampling interval of the data. They also explain how to compute standard errors robust to conditional heteroskedasticity.

Nonetheless, it is well known that the 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). Moreover, if the degree of overlapping becomes non-trivial relative to the sample size, standard GMM asymptotic theory no longer

applies (see Richardson and Stock, 1989). We will revisit these issues in the Monte Carlo experiments of Section 2.4.

**Vector Autoregressions (VAR).** This second approach estimates a joint covariance stationary process for the first difference of the spot exchange rate $\Delta s_t$ and the forward premia $p_{t,\tau}$ by Gaussian pseudo maximum likelihood (PML). In this case, the difference operation on the spot exchange rate is taken over the sampling interval in order to avoid overlapping residuals. Consequently, the UIP condition becomes:

$$E_t\left(s_{t+\tau} - s_t\right) = E_t\left(\sum_{i=1}^{\tau} \Delta s_{t+i}\right) = p_{t,\tau}. \tag{20}$$

The constraints that this condition imposes on the joint process for $\Delta s_t$ and $p_{t,\tau}$ can be found by using the Wold decomposition of the joint process to obtain the projection of each $\Delta s_{t+i}$ ($i = 1, \ldots, \tau$) onto the information set defined by $\{\Delta s_t, p_{t,\tau}, \Delta s_{t-1}, p_{t-1,\tau}, \ldots\}$ (see Hansen and Hodrick, 1980). In this sense, our Proposition 1 can be seen as a limiting case of this methodology.

In this context, Baillie, Lippens and McMahon (1984) and Hakkio (1981) show how to translate the restrictions on the Wold decomposition in Hansen and Hodrick (1980) into testable hypotheses on 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 (20), from which we can obtain the appropriate UIP conditions. As an illustration, assume that $\mathbf{x}_t = (p_{t,\tau}, \Delta s_t)'$ follows the VAR(1) model

$$\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{21}$$

where $\boldsymbol{\varepsilon}_t$ is a 2-dimensional vector of white noise disturbances with contemporaneous covariance matrix $E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right) = \boldsymbol{\Upsilon}$. Then, the optimal forecast of $\mathbf{x}_{t+i}$ ($i = 1, \ldots, \tau$) based on the information set defined by $\mathbf{x}_t$ and its lagged values is given by $E_t\mathbf{x}_{t+i} = \mathbf{B}^i\mathbf{x}_t$. Consequently, the projection of $\Delta s_{t+i}$ will be given by $\mathbf{e}_2'\mathbf{B}^i\mathbf{x}_t$, where $\mathbf{e}_j$ is a vector with a one in the $j^{th}$ position and zeroes in the others. Since the LHS of (20) can be expressed as:

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

11

while the right hand side (RHS) is:

$$p_{t,\tau} = \mathbf{e}_1' \mathbf{x}_t,$$

it follows that the testable restrictions on the VAR parameters that UIP implies for a $\tau$-period forward contract are:

$$\mathbf{e}_2' \mathbf{B} (\mathbf{I} - \mathbf{B})^{-1} (\mathbf{I} - \mathbf{B}^\tau) = \mathbf{e}_1'. \tag{22}$$

Although we can always consider (21) 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 $\Delta s_t$ 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. Therefore, testing UIP in a full information set-up 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. Again, we will show examples of the effects of such misspecifications in our Monte Carlo experiments.

## 2.4 Monte Carlo Simulations

In this section, we carry out an extensive Monte Carlo study to assess the ability of our proposed methodology to test UIP. Further, we also compare our approach to the traditional UIP tests described in the previous section.

We initially simulate 10,000 samples of 30 years of weekly data ($T = 1,560$) from the continuous-time model (14) in which we fix the contract period $\tau$ 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:

$$\begin{pmatrix} \widetilde{p}_{t,\tau} \\ \Delta \tilde{s}_t \end{pmatrix} = \begin{pmatrix} \mu_p \\ \mu_{\Delta s} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \\ \Delta s_t \end{pmatrix}, \tag{23}$$

$$\begin{pmatrix} \mathbf{u}_t \\ \Delta s_t \end{pmatrix} = \begin{pmatrix} e^{\mathbf{\Phi}} & 0 \\ \mathbf{e}_2' \mathbf{\Phi}^{-1} (e^{\mathbf{\Phi}} - I) & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_{t-1} \\ \Delta_1 s_{t-1} \end{pmatrix} + \boldsymbol{\eta}_t, \tag{24}$$

where $\widetilde{p}_{t,\tau} = \mu_p + p_{t,\tau}$, $\Delta \tilde{s}_t = \mu_{\Delta s} + \Delta s_t$, $\boldsymbol{\eta}_t = \int_{t-h}^{t} e^{\mathbf{A}(t-r)} \boldsymbol{\Sigma}^{1/2} d\mathbf{W}(r)$ with $\mathbf{A}$, $\boldsymbol{\Sigma}^{1/2}$ defined in equation (15), $\sigma_{11} = .2$, $\sigma_{21} = -.3$, $\sigma_{22} = -.1$, $\alpha_1 = -.2$, $\alpha_2 = -1.5$, $\mu_p = 2$, and $\mu_{\Delta s} = 0$. In order to impose the null hypothesis of UIP, we first decompose $\mathbf{\Phi}$ as $\mathbf{PDP}^{-1}$,

12

where $\mathbf{D}$ is a diagonal matrix with elements $d_1 = -.025$ and $d_2 = -.25$, and then choose $\mathbf{P}$ so that $\mathbf{\Phi}$ satisfies (18).

Given the highly non-linear nature of the model, we obtain good initial values for the optimization algorithm by exploiting its Euler discretization as explained in appendix B. Then, we employ a scoring algorithm to maximize the exact log-likelihood function, with analytical expressions for the score vector and information matrix obtained by differentiating the Kalman filter prediction and updating equations as in Harvey (1989, pp 140-3). We also use those analytical expressions to obtain heteroskedasticity-robust standard errors and Wald tests. Additional computational details can be found in Appendix C.

In addition, we compare the finite-sample performance of our proposed continuous-time approach to another six different UIP tests. In particular, for the same simulated data we compute OLS-based UIP tests in which asymptotically valid standard errors are estimated using the following four different methods:

1. Newey-West (1987) approach (NW) with a fixed bandwidth, 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). We use a bandwidth given by the rule of thumb $T^{1/3}$, where $T$ is the sample size (see 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 (19) follows a moving-average (MA) process of finite known order but with unknown coefficients to construct the asymptotic covariance matrix.[5] We follow Eichenbaum, Hansen and Singleton (1988) in using Durbin (1960)'s method to estimate the MA structure.

3. Den Haan and Levin (1996)'s 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)'s data-dependent approach assumes that the moment conditions implicit in the normal equations of (19) have a finite VAR representation, which they exploit to construct their estimated covariance matrix.

Furthermore, we also compute VAR-based tests with $p = 1$ and 4 lags.

---

[5]Hence, we cannot use Hodrick's (1992) standard errors.

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, simply 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 (14) 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 (21) will be:

$$\beta^{VAR(1)} = \frac{\mathbf{e}_2' \mathbf{B}(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B}^\tau)\mathbf{\Psi}\mathbf{e}_1}{\mathbf{e}_1' \mathbf{\Psi} \mathbf{e}_1}, \tag{25}$$

where $\mathbf{\Psi}$ is the unconditional covariance matrix of $\mathbf{x}_t = (p_{t,\tau}, \Delta s_t)'$, 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 (14) is given by

$$\beta^{OU} = \frac{\mathbf{e}_2' \mathbf{\Phi}^{-1}(e^{\mathbf{\Phi}\tau} - \mathbf{I})\mathbf{\Lambda}\mathbf{e}_1}{\mathbf{e}_1' \mathbf{\Lambda} \mathbf{e}_1}, \tag{26}$$

where $vec(\mathbf{\Lambda}) = -(\mathbf{\Phi} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{\Phi})^{-1} vec(\mathbf{\Sigma})$ 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 $H_0 : \beta^{OU} = 1$.

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 overlapping, 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. This result is not entirely surprising since the autocorrelation structure of $w_{t+\tau}$ in (19) is not well-captured by a bandwidth of $T^{1/3} \simeq 11$ when the correct order of the MA structure is $\tau - 1 = 51$. 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 is the VARHAC approach with AIC order selection, it still over-rejects in finite samples.

As for VAR-based tests, we find that we approximate better the autocorrelation structure of $\mathbf{x}_t = (p_{t,\tau}, \Delta s_t)'$ as we increase the order 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 (14). 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 $\mathbf{B} \equiv E\left(\mathbf{x}_t\mathbf{x}'_{t-1}\right)\left[E\left(\mathbf{x}_t\mathbf{x}'_t\right)\right]^{-1}$, while the variance-covariance matrix of the residuals is given by $\mathbf{\Upsilon} \equiv E\left(\mathbf{x}_t\mathbf{x}'_{t-1}\right) - E\left(\mathbf{x}_t\mathbf{x}'_{t-1}\right)\left[E\left(\mathbf{x}_t\mathbf{x}'_t\right)\right]^{-1}E\left(\mathbf{x}_t\mathbf{x}'_{t-1}\right)'$. If we then plug in the expressions for $E\left(\mathbf{x}_t\mathbf{x}'_t\right)$ and $E\left(\mathbf{x}_t\mathbf{x}'_{t-1}\right)$ implied by the continuous-time model (14),[6] we will obtain analytical expressions for the population value of $\mathbf{B}(\boldsymbol{\theta})$ and $\mathbf{\Upsilon}(\boldsymbol{\theta})$ as a function of the parameters of the continuous-time model, $\boldsymbol{\theta}$. Then, we can use equation (25) 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 (14). 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 confirm 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.

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

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

---

[6]In this case $E\left(\mathbf{x}_t\mathbf{x}'_t\right) = \mathbf{H}\mathbf{\Psi}\mathbf{H}'$ and $E\left(\mathbf{x}_t\mathbf{x}'_{t-1}\right) = \mathbf{H}\mathbf{F}\mathbf{\Psi}\mathbf{H}'$ where $\mathbf{H}$ is a matrix that has ones in the (1,1) and (2,3) positions and zeros in the others; $\mathbf{F} = e^{\mathbf{A}}$, $\mathbf{\Omega} = \int_0^h e^{\mathbf{A}r}\mathbf{S}e^{\mathbf{A}'r}dr$ with $\mathbf{A}$ and $\mathbf{S}$ defined in equation (15); and $\mathbf{\Psi}$ is the unconditional variance of $\boldsymbol{\xi}_t$, which is computed from $vec(\mathbf{\Psi}) = (\mathbf{I} - \mathbf{F} \otimes \mathbf{F})^{-1}vec(\mathbf{\Omega})$.

each of the UIP tests by means of Davidson and MacKinnon's (1998) size-power curves, which shows 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, and finally the VAR(1) one. 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).[7] 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 (14) calculated according to equation (26) 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 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.

# 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.

---

[7]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.

For this reason, we introduce a novel Hausman test that exploits the fact that the structure of a continuous time model is the same regardless of the discretization frequency, $h$. As a result, we can first estimate the model using the whole sample (e.g. weekly data), then using data on odd observations only, say (e.g. bi-weekly data), and finally decide if those two estimators are "statistically close". Under the null hypothesis that our continuous time specification is valid, Gaussian pseudo maximum likelihood parameter estimators are consistent regardless of the sampling frequency, but those that use the whole sample will be efficient. In contrast, if the dynamic specification is incorrect, then estimators based on different sampling frequencies will have different probability limits.

In order to illustrate the implementation of our proposed specification test, consider again the continuous-time model (10) in Example 1, and imagine that we want to compare the estimators of the model obtained using weekly and bi-weekly data.[8] As we saw in Section 2.2, observations of this model sampled at the weekly frequency will satisfy the following VAR(1) process:

$$\begin{pmatrix} p_{t,\tau} \\ \Delta s_t \end{pmatrix} = \begin{pmatrix} e^{\varphi_{11}} & 0 \\ \frac{\varphi_{21}}{\varphi_{11}} \left(e^{\varphi_{11}} - 1\right) & 0 \end{pmatrix} \begin{pmatrix} p_{t-1,\tau} \\ \Delta s_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t}^{(1)} \\ \eta_{2t}^{(1)} \end{pmatrix}, \tag{27}$$

while bi-weekly observations will satisfy the alternative VAR(1) process:

$$\begin{pmatrix} p_{t,\tau} \\ \Delta_2 s_t \end{pmatrix} = \begin{pmatrix} e^{2\varphi_{11}} & 0 \\ \frac{\varphi_{21}}{\varphi_{11}} \left(e^{2\varphi_{11}} - 1\right) & 0 \end{pmatrix} \begin{pmatrix} p_{t-2,\tau} \\ \Delta_2 s_{t-2} \end{pmatrix} + \begin{pmatrix} \eta_{1t}^{(2)} \\ \eta_{2t}^{(2)} \end{pmatrix}, \tag{28}$$

where

$$\begin{pmatrix} \eta_{1t}^{(2)} \\ \eta_{2t}^{(2)} \end{pmatrix} = \begin{pmatrix} \eta_{1t}^{(1)} \\ \eta_{2t}^{(1)} \end{pmatrix} + \begin{pmatrix} e^{\varphi_{11}} & 0 \\ \frac{\varphi_{21}}{\varphi_{11}} \left(e^{\varphi_{11}} - 1\right) & 1 \end{pmatrix} \begin{pmatrix} \eta_{1t-1}^{(1)} \\ \eta_{2t-1}^{(1)} \end{pmatrix} \tag{29}$$

Thus, if we denote by $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ the estimators that we obtain from (27) and (28), respectively, then our testing methodology simply assesses whether their probability limits coincide. To see how, define $\boldsymbol{\psi} = \left(\boldsymbol{\theta}^{(1)\prime}, \boldsymbol{\theta}^{(2)\prime}\right)^{\prime}$, and think of $\hat{\boldsymbol{\psi}}$ as solving the sample versions of the following set of moment conditions:

$$\left\{ \begin{array}{c} E[\mathbf{s}^{(1)}(\boldsymbol{\theta}^{(1)})] \\ E[\mathbf{s}^{(2)}(\boldsymbol{\theta}^{(2)})] \end{array} \right\} = E\left[\mathbf{s}_t(\boldsymbol{\psi})\right] = \mathbf{0}, \tag{30}$$

where the influence functions $\mathbf{s}^{(1)}(\boldsymbol{\theta}^{(1)})$ and $\mathbf{s}^{(2)}(\boldsymbol{\theta}^{(2)})$ are the Gaussian pseudo-scores of models (27) and (28), respectively. Then, we can use standard GMM asymptotic theory to

---

[8]In practice, we can compare any two different frequencies of choice. For instance, in the Monte Carlo simulations of the specification tests and the empirical application, we compare weekly and monthly estimates instead.

show that:

$$\sqrt{T}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{D}'^{-1}\mathbf{V}\mathbf{D}^{-1}\right),$$

where $\mathbf{D} = E\left[\partial \mathbf{s}_t(\boldsymbol{\psi})/\partial \boldsymbol{\psi}'\right]$ and $\mathbf{V} = \sum_{j=-\infty}^{\infty} E\left[\mathbf{s}_t(\boldsymbol{\psi})\mathbf{s}_{t-j}(\boldsymbol{\psi})'\right]$. On this basis, we can test the restriction $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(2)}$ using the Wald test:

$$T \cdot \hat{\boldsymbol{\psi}}' \mathbf{R}'(\mathbf{R}\hat{\mathbf{D}}'^{-1}\hat{\mathbf{V}}\hat{\mathbf{D}}^{-1}\mathbf{R}')^{-1}\mathbf{R}\hat{\boldsymbol{\psi}},$$

where $\mathbf{R} = (\mathbf{I}, -\mathbf{I})$, and $\hat{\mathbf{D}}$ and $\hat{\mathbf{V}}$ are consistent estimates of $\mathbf{D}$ and $\mathbf{V}$, respectively. Note, however, that the comparison of estimators at different frequencies also 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 $\delta$ is the ratio of the sampling frequencies ($=2$ in this example). For instance, (29) implies that $E\left(\boldsymbol{\eta}_t^{(2)}\boldsymbol{\eta}_{t-1}^{(1)\prime}\right) \neq \mathbf{0}$. Nevertheless, this overlapping problem is far less severe than the problem that plagues the OLS-based UIP tests, unless we decide to compare weekly and yearly estimates. Further details on this specification test can be found in Appendix D.

In any case, it is worth remembering 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. For this reason, rather than testing whether the full parameter vectors $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ coincide, we simply test if the implied betas remain the same when we vary the sampling frequency. In the context of model (10) in particular, we would test if $\beta^{(1)} = \beta^{(2)}$, with $\beta = \phi_{21}\left(e^{\phi_{11}\tau} - 1\right)/\phi_{11}$, using the following Wald statistic

$$T \cdot f(\hat{\boldsymbol{\psi}}) \left(\frac{\partial f(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\psi}'}\hat{\mathbf{D}}'^{-1}\hat{\mathbf{V}}\hat{\mathbf{D}}^{-1}\frac{\partial f(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\psi}}\right)^{-1} f(\hat{\boldsymbol{\psi}}),$$

where $f(\boldsymbol{\psi}) = \varphi_{21}^{(1)}\left(e^{\varphi_{11}^{(1)}\tau} - 1\right)/\varphi_{11}^{(1)} - \varphi_{21}^{(2)}\left(e^{\varphi_{11}^{(2)}\tau} - 1\right)/\varphi_{11}^{(2)}$. In addition, note that 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 can test the specification of the continuous time model (14) in example 2 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.

## 3.2  Monte Carlo Simulations

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 (10) in Example 1, 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 $\mathbf{x}_t = [p_{t,\tau}, \Delta s_t]'$, so that the model that we simulate is:

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

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

where $\varphi_{11} = -.025$, $\varphi_{21} = \varphi_{11}/\left( e^{\gamma_{11}\tau} - 1 \right)$, $\gamma_{11} = .2$, $\gamma_{21} = -.2$, $\gamma_{22} = -1.5$, $\mu_p = 2$ and $\mu_{\Delta s} = 0$. Importantly, note that we maintain the UIP restriction (13). Details on the estimation of this model at two different frequencies can also be found in Appendix D.

Since empirical researchers often decide between working with weekly or monthly data to test UIP, we compare the value of $\beta$ that we obtain using the weekly sample, $\beta^{(1)}$, with the one that we would obtain had we sampled the data once a month. The comparison between weekly and monthly estimators creates an overlapping problem that introduces an MA(3) structure in the moment conditions (30), which is much simpler than the MA(51) structure in section 2.4. For that reason, we consider again the Newey-West (1987) approach (NW), 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 scores of the model at the highest frequency are uncorrelated.

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, the EHS approach tends to over-reject slightly in finite samples, while the VARHAC-BIC approach tends to under-reject by almost the same magnitude. In contrast, both the VARHAC-AIC and NW approaches are rather more conservative.

We also generate another 10,000 simulations of 30 years of weekly data from the continuous-time model (14) to assess the finite-sample power of the specification test that takes as its null hypothesis that the correct model is given by (10). Specifically, we simulate again from the model in equations (23) and (24) 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 $d_2 = -.25$. Once again, note that we maintain the UIP restrictions in (18).

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 version of the test based on the NW approach has the highest power, followed by the ones based on the EHS approach, the VARHAC-AIC approach and finally the VARHAC-BIC one. However, we should remember that the reported results are size-adjusted. In practice, the NW approach has poor size and, therefore, one would not want to use it to test the dynamic specification of the model. If we focus on the EHS and the VARHAC-BIC approaches, which are the ones with the most reliable sizes, the EHS method seems to be the one with the highest power.

# 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 using weekly data over the period from January 1977 to December 2005. As for $\tau$, we use the appropriate Eurocurrency interest rates at maturities of one, three, six, and twelve months.

Table 1 reports the estimated coefficients of the continuous-time model (10) in Example 1, as well as the estimate of the implied beta. This reveals several interesting facts. First, the estimated $\varphi_{11}$ 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 0.95 for the U.K., 0.97 for Germany, and 0.90 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 Table 2 we report the results of our proposed specification test applied to the estimators of $\beta$ that we obtain using weekly data with

the one that we would obtain had we sampled the data monthly. Note that $\hat{\beta}^{(4)}$ tends to be less negative than $\hat{\beta}^{(1)}$, with the exception of the one-month forward contracts for the pound sterling and the Canadian dollar, for which $\hat{\beta}^{(1)} - \hat{\beta}^{(4)}$ equals .033 and .029, respectively. However, the difference is only significantly different from zero for the one-year pound sterling contract, and the one-month, three-month and six-month Canadian dollar contracts.

For this reason, Table 3 presents the estimated coefficients of the more flexible continuous-time model (14) for the cases in which model (10) 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 (14). In particular, the p-values of the specification test lie between 0.98 (BIC) and 0.99 (AIC) for the one-year pound sterling contract. Similarly, the p-values lie between 0.56 (EHS) and 0.81 (NW), 0.66 (EHS) and 0.99 (AIC), and 0.60 (EHS) and 0.96 (BIC) for the one-month, three-month and six-month Canadian dollar contracts, respectively.

Therefore, we are unable to rescue the UIP hypothesis 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 in which the standard errors are obtained using the Newey-West (1987), 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 Table 4 indicate that the estimate of the slope coefficient $\beta$ is negative. As expected from the Monte Carlo experiment reported in Section 2.4, the results of the OLS-based UIP tests are somewhat sensitive to the covariance matrix estimator employed (see also Ligeralde, 1997). For example, if we use the EHS or VARHAC-BIC method to test UIP at the one-year horizon with U.K. data, we find that we cannot reject that $H_0 : \beta = 1$. The same occurs 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.

# 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 overlapping 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 novel Hausman test that exploits the fact that the structure of a continuous

time model is the same regardless 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 the UIP hypothesis on the U.S. dollar bilateral exchange rates against the British pound, the German DM-Euro and the Canadian dollar using weekly data over the period from January 1977 to December 2005. 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 uncovered interest parity hypothesis even after taking care of temporal aggregation problems.

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 (2006).

Finally, it is worth mentioning that our specification test can also be applied to check the dynamic specification of discrete time models such as (21), 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.

Bergstrom, A. (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.

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

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

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.

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.

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. (2006): "Can affine term structure models help us predict exchange rates?," Bank of Canada Working Paper 2006-27.

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.

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): "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.

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.K. Moh (2006): "Official interventions and the forward premium anomaly," forthcoming, *Journal of Empirical Finance.*

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.

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.

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

# Appendix

## A  Proof of Proposition 1

Given that (3) implies that $u_2(t)$ is the drift of $s(t)$, we can express the LHS of the UIP condition in continuous time (4) as:

$$E_t\left[\int_0^\tau ds(t+r)\right] = E_t\left[\int_0^\tau u_2(t+r)dr + \int_0^\tau dW_s(t)\right] = E_t\left[\int_0^\tau u_2(t+r)dr\right].$$

But since the Wold decomposition (5) implies that

$$u_2(t+r) = \int_{-r}^\infty \phi_{21}(h+r)dW_{u1}(t-h) + \int_{-r}^\infty \phi_{22}(h+r)dW_{u2}(t-h),$$

then to obtain the required expectation conditioned on information available at time $t$ we simply need to apply an annihilation operator that zeros out $\phi_{21}(h+r)$ and $\phi_{22}(h+r)$ for $t \in [-r, 0]$ (see Hansen and Sargent, 1991). In this way we obtain

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

which in turn yields:

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

$$(31)$$

On the other hand, (5) also implies that:

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

$$(32)$$

Given that the integrals in (31) 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 (31) and (32). On this basis, it is straightforward to see that UIP is equivalent to the conditions (6) and (7).  $\square$

## B  Initial values for the optimization algorithm

We obtain initial values for the scoring algorithm in Section 2.4 by exploiting the Euler discretization of the model in equation (15), 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 $\mu_p$ and $\mu_{\Delta s}$.

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

$$\mathbf{x}_t = \mathbf{A}_1\mathbf{x}_{t-1} + \mathbf{A}_2\mathbf{x}_{t-2} + \dots \mathbf{A}_p\mathbf{x}_{t-p} + \mathbf{e}_t$$

   for $\mathbf{x}_t = (p_{t,\tau}, \Delta s_t)'$ where $p_{t,\tau}$ and $\Delta s_t$ 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 $\Delta s_t$ using the fact that

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

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

4. Next, we estimate the VAR(1) model $\widehat{\mathbf{u}}_t = \mathbf{F}\widehat{\mathbf{u}}_{t-1} + \mathbf{v}_t$ for $\hat{\mathbf{u}}_t = (p_{t,\tau}, \hat{u}_{2t})'$. From here, we can obtain an estimate of $\mathbf{\Phi}$ as $\hat{\mathbf{\Phi}} = \hat{\mathbf{F}} - \mathbf{I}$. 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 $\mathbf{\Sigma}^{1/2}$ and $\boldsymbol{\alpha}$ in the following way. We first estimate $\mathbf{\Omega}$, 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 $\sigma_{11}$, $\alpha_1$ and $\alpha_2$, respectively, where $\mathbf{L}\mathbf{L}'$ is the Cholesky decomposition of $\mathbf{\Omega}$. Finally, we estimate $\sigma_{21}$ and $\sigma_{22}$ as the coefficients in the regression of $\widehat{\eta}_{2t}^{euler}$ on $\widehat{\mathbf{z}}_t^*$, where $\mathbf{z}_t^* = \mathbf{L}^{-1}\mathbf{z}_t$.

## C   Derivatives of the log-likelihood function

We can express the discrete time versions of the continuous time models in Examples 1 and 2 as:

$$\mathbf{y}_t = \mathbf{d} + \mathbf{Z}\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t,$$
$$\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \mathbf{u}_t,$$
$$\begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{u}_t \end{pmatrix} \Bigg| \begin{pmatrix} \mathbf{y}_{t-1} \\ \boldsymbol{\alpha}_{t-1} \end{pmatrix}, \begin{pmatrix} \mathbf{y}_{t-2} \\ \boldsymbol{\alpha}_{t-2} \end{pmatrix}, \dots \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix}\right].$$

Given this state-space formulation, we can evaluate the exact Gaussian likelihood via the usual prediction error decomposition:

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

with

$$l_t = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{F}_t| - \frac{1}{2}\mathbf{v}_t'\mathbf{F}_t^{-1}\mathbf{v}_t, \tag{33}$$

where $\boldsymbol{\psi}$ 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 $\mathbf{F}_t$ their conditional variance. The Kalman filter recursions are given by

$$\left.\begin{array}{c} \boldsymbol{\alpha}_{t|t-1} = \mathbf{T}\boldsymbol{\alpha}_{t-1|t-1} \\ \mathbf{P}_{t|t-1} = \mathbf{T}\mathbf{P}_{t-1|t-1}\mathbf{T}' + \mathbf{Q} \\ \mathbf{v}_t = \mathbf{y}_t - \mathbf{d} - \mathbf{Z}\boldsymbol{\alpha}_{t|t-1} \\ \mathbf{F}_t = \mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}' + \mathbf{R} \\ \boldsymbol{\alpha}_{t|t} = \boldsymbol{\alpha}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{Z}'\mathbf{F}_t^{-1}\mathbf{v}_t \\ \mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{Z}'\mathbf{F}_t^{-1}\mathbf{Z}\mathbf{P}_{t|t-1} \end{array}\right\} \tag{34}$$

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 stationarity, we initialize the filter using $\boldsymbol{\alpha}_0 = E(\boldsymbol{\alpha}_t) = \mathbf{0}$ and $vec(\mathbf{P}_0) = (\mathbf{I} - \mathbf{T} \otimes \mathbf{T})^{-1} vec(\mathbf{Q})$.

The prediction error decomposition in (33) can also be used 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 appear in the asymptotic distribution of the Gaussian pseudo-ML estimator of $\boldsymbol{\psi}$. In particular, the score vector takes the following form:

$$\frac{\partial l_t(\boldsymbol{\psi})}{\partial \psi_i} = s_t(\boldsymbol{\psi}) = -\frac{1}{2}tr\left[\left(\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \psi_i}\right)\left(\mathbf{I} - \mathbf{F}_t^{-1}\mathbf{v}_t\mathbf{v}_t'\right)\right] - \frac{\partial \mathbf{v}_t'}{\partial \psi_i}\mathbf{F}_t^{-1}\mathbf{v}_t,$$

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

$$-E_{t-1}\left(\frac{\partial^2 l_t}{\partial \psi_i \partial \psi_j}\right) = \frac{1}{2}tr\left(\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \psi_i}\mathbf{F}_t^{-1}\frac{\partial \mathbf{F}_t}{\partial \psi_{ij}}\right) + \frac{\partial \mathbf{v}_t'}{\partial \psi_i}\mathbf{F}_t^{-1}\frac{\partial \mathbf{v}_t}{\partial \psi_j}.$$

In turn, these two expressions require the first derivatives of $\mathbf{F}_t$ 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 (34). 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).

# D   Additional details on the specification test

As in the main text, we concentrate on the continuous-time model (10) in Example 1, and assume that we want to compare the estimators obtained with weekly and bi-weekly

data. For our purposes it is convenient to write the discretisation (27) in state-space form as:

$$\begin{pmatrix} p_{t,\tau} \\ \Delta s_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \end{pmatrix}, \tag{35}$$

$$\begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \end{pmatrix} = \begin{pmatrix} e^{\varphi_{11}} & 0 \\ \frac{\varphi_{21}}{\varphi_{11}} \left( e^{\varphi_{11}} - 1 \right) & 0 \end{pmatrix} \begin{pmatrix} \alpha_{1t-1} \\ \alpha_{2t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t}^{(1)} \\ \eta_{2t}^{(1)} \end{pmatrix}. \tag{36}$$

On this basis, we can use the Kalman filter algorithm described in Appendix C to obtain Gaussian pseudo ML estimators of the parameters of this model as the solution of the sample versions of the following moment conditions:

$$E\left[ \frac{\partial l^{(1)}(y_t^{(1)}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right] = E\left[ \mathbf{s}^{(1)}(y_t^{(1)}; \boldsymbol{\theta}) \right] = 0, \tag{37}$$

where $y_t^{(1)} = [p_{t,\tau}, \Delta s_t]'$, $l^{(1)}(y_t^{(1)}; \boldsymbol{\theta})$ is the log-likelihood contribution of $y_t^{(1)}$, $\boldsymbol{\theta}$ is the vector of parameters of the continuous-time model (10), and the superscript (1) indicates that $y_t^{(1)}$ is an observation obtained at the highest available frequency. Analogously, we denote by $\boldsymbol{\theta}^{(1)}$ the value of $\boldsymbol{\theta}$ that solves (37).

Let us initially assume that we discard all even observations. Although we could use (28) to compute the log likelihood function of such a bi-weekly sample, for our purposes it is more convenient to use the approach in Mariano and Murasawa (2003), who treat discarded observations as missing values. Specifically, we can construct a new series

$$y_t^{(2)odd+} = D_t \cdot y_t^{(2)} + (1 - D_t)z_t^{even},$$

where $D_t$ is a dummy variable that takes the value of 1 when $y_t^{(2)} = (p_{t,\tau}, \Delta_2 s_t)'$ is observed because $t$ is an odd number, while $z_t^{even}$ is a bivariate random vector drawn from an independent arbitrary distribution that does not depend on $\boldsymbol{\theta}$. Let us define $\mathbf{y}_T^{(2)odd} = (y_1^{(2)}, y_3^{(2)}, ..., y_T^{(2)})'$ and $\mathbf{y}_T^{(2)odd+} = (y_1^{(2)odd+}, y_2^{(2)odd+}, ..., y_T^{(2)odd+})'$. Given that the $z_t^{even}$'s are independent of $\mathbf{y}_T^{(2)odd}$ by construction, we can write the joint probability distribution of $\mathbf{y}_T^{(2)odd+}$ as

$$f(\mathbf{y}_T^{(2)odd+}; \boldsymbol{\theta}) = f(\mathbf{y}_T^{(2)odd}; \boldsymbol{\theta}) \cdot \prod_{t:D_t=0} f(z_t^{even}).$$

Thus, the likelihood function of $\boldsymbol{\theta}$ given $\mathbf{y}_T^{(2)odd}$ and the corresponding log likelihood given $\mathbf{y}_T^{(2)odd+}$ are identical up to scale, so they will be maximized by the same value. The main advantage of working with the augmented data series $\mathbf{y}_T^{(2)odd+}$ is that it no longer contains missing observations.

It is then easy to derive a state space model for $y_t^{(2)odd+}$. In particular, the measurement equation is:

$$\begin{pmatrix} p_{t,\tau} \\ \Delta_2 s_t \end{pmatrix} = \begin{pmatrix} D_t & 0 & 0 \\ 0 & D_t & D_t \end{pmatrix} \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \alpha_{2t-1} \end{pmatrix} + (1 - D_t)z_t^{even}, \tag{38}$$

while the transition equation will be:

$$
\begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \alpha_{2t-1} \end{pmatrix} = \begin{pmatrix} e^{\varphi_{11}} & 0 & 0 & 0 \\ \frac{\varphi_{21}}{\varphi_{11}}\left(e^{\varphi_{11}}-1\right) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha_{1t-1} \\ \alpha_{2t-1} \\ \alpha_{2t-2} \end{pmatrix} + \begin{pmatrix} \eta_{1t}^{(1)} \\ \eta_{2t}^{(1)} \\ 0 \end{pmatrix}, \qquad (39)
$$

which can be understood as an augmented version of (36).

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

$$
E\left[\frac{\partial l^{(2)odd}(y_t^{(2)odd+};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = E\left[\mathbf{s}^{(2)odd}(y_t^{(2)odd+};\boldsymbol{\theta})\right] = \mathbf{0}. \qquad (40)
$$

Alternatively, we could discard the odd observations instead. A direct application of the methods explained in the previous paragraphs leads to the alternative set of moment conditions:

$$
E\left[\frac{\partial l^{(2)even}(y_t^{(2)even+};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = E\left[\mathbf{s}^{(2)even}(y_t^{(2)even+};\boldsymbol{\theta})\right] = \mathbf{0}. \qquad (41)
$$

Given that the choice between even and odd observations is arbitrary, and that discarding observations entails a loss of information, we combine (40) and (41) to obtain the following set of moment conditions:

$$
E\left[\mathbf{s}^{(2)odd}(y_t^{(2)odd+};\boldsymbol{\theta}) + \mathbf{s}^{(2)even}(y_t^{(2)even+};\boldsymbol{\theta})\right] = E\left[\mathbf{s}^{(2)}(y_t^{(2)};\boldsymbol{\theta})\right] = \mathbf{0}. \qquad (42)
$$

However, note that the moment conditions that define the unconditional mean parameters $\mu_p$ and $\mu_{\Delta s}$ are identical at both frequencies.[9] Therefore, we drop the moment conditions that define the unconditional means at the second frequency from the set of moment conditions in (42) in order to avoid singularities. Finally, if we denote by $\boldsymbol{\theta}^{(2)}$ the value of $\boldsymbol{\theta}$ that solves (42) in the population, our testing methodology is equivalent to checking whether $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(2)}$.

In practice, we use the following procedure to estimate model (10) at the highest frequency:

1. We estimate $\mu_p^{(1)}$ and $\mu_{\Delta s}^{(1)}$ with the full sample means of the forward premium and the expected rate of depreciation.

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

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

---

[9] Note that we have that $E(p_t) = 0$ implies that $E[p_t D_t + p_t(1-D_t)] = E(p_t) = 0$, while $E(\Delta s_t) = 0$ implies that $E[\Delta_2 s_t D_t + \Delta_2 s_t(1-D_t)] = E(\Delta s_t + \Delta s_{t-1}) = 0$, where, again, $D_t$ is a dummy variable that takes the value of 1 when $t$ is an odd number.

for $\mathbf{x}_t = (p_{t,\tau}, \Delta s_t)'$ where $p_{t,\tau}$ and $\Delta s_t$ are the demeaned forward premium and rate of depreciation, respectively; and subject to the restrictions that $a_{21} = a_{22} = 0$. In particular, note that (12) implies that $\mathbf{A} = e^{\mathbf{\Phi}}$ and $\mathbf{\Omega} = \int_0^1 e^{\mathbf{\Phi}r}\mathbf{\Gamma}e^{\mathbf{\Phi}'r}dr$. Thus, 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:

$$\varphi_{11} = \log(a_{11})$$

$$\varphi_{21} = \frac{a_{21}\varphi_{11}}{e^{\varphi_{11}} - 1}$$

$$g_{11} = \frac{2\varphi_{11}\omega_{11}}{e^{2\varphi_{11}} - 1}$$

$$g_{21} = \frac{\varphi_{11}}{e^{\varphi_{11}} - 1}\left[\omega_{21} - \frac{\sqrt{g_{11}}\varphi_{21}}{2\varphi_{11}^2}\left(e^{\varphi_{11}} - 1\right)^2\right]$$

$$g_{22} = \omega_{22} - \frac{\sqrt{g_{11}}\varphi_{21}^2}{2\varphi_{11}^3}\left(2\varphi_{11} + e^{2\varphi_{11}} - 4e^{\varphi_{11}} + 3\right) - \frac{2\sqrt{g_{11}}\varphi_{21}}{\varphi_{11}^3}\left(e^{\varphi_{11}} - \varphi_{11} - 1\right)$$

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

3. However, the estimates obtained in the previous step only maximize the log-likelihood function of the data conditional on the first observation. For that reason, we feed them as initial values to the scoring algorithm that we use for the numerical maximization of the Kalman filter based exact log-likelihood of the model.

In contrast, to estimate the model at the lower frequency, we minimize the following quadratic form:

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

where $\mathbf{s}_t^{(2)}(\boldsymbol{\theta}^{(2)}) = \mathbf{s}^{(2)}(y_t^{(2)}; \boldsymbol{\theta}^{(2)})$ is defined in equation (42). For the reasons explained before, though, we exclude the scores corresponding to $\boldsymbol{\mu}^{(2)}$, and set the mean parameters to $\hat{\boldsymbol{\mu}}^{(1)}$. Since equation (42) 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 $\mathbf{W}$. Still, we use the estimated values of $\boldsymbol{\theta}^{(1)}$ as starting values for this numerical optimization procedure, 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 $\hat{\boldsymbol{\theta}}^{(1)}$.

<div align="center">

**Table 1**
**Estimates of the continuous-time model in Example 1**

</div>

| Contract | $\varphi_{11}$ | $\varphi_{21}$ | $\gamma_{11}$ | $\gamma_{21}$ | $\gamma_{22}$ | $\mu_p$ | $\mu_{\Delta s}$ | $\beta$ |
|---|---|---|---|---|---|---|---|---|
| **UK** | | | | | | | | |
| 4 weeks | -0.015 | -0.537 | 0.041 | -0.110 | 1.367 | -0.192 | 0.014 | -2.084 |
| | (0.010) | (0.199) | (0.002) | (0.038) | (0.042) | (0.086) | (0.061) | (0.780) |
| 13 weeks | -0.010 | -0.206 | 0.095 | -0.049 | 1.371 | -0.582 | 0.019 | -2.519 |
| | (0.008) | (0.070) | (0.004) | (0.041) | (0.042) | (0.336) | (0.079) | (0.886) |
| 26 weeks | -0.009 | -0.112 | 0.170 | -0.054 | 1.370 | -1.082 | 0.020 | -2.605 |
| | (0.007) | (0.037) | (0.006) | (0.054) | (0.041) | (0.648) | (0.083) | (0.934) |
| 52 weeks | -0.008 | -0.063 | 0.297 | -0.094 | 1.368 | -1.960 | 0.021 | -2.652 |
| | (0.007) | (0.021) | (0.011) | (0.056) | (0.041) | (1.236) | (0.089) | (1.063) |
| **Germany** | | | | | | | | |
| 4 weeks | -0.006 | -0.239 | 0.031 | -0.206 | 1.461 | 0.269 | 0.034 | -0.946 |
| | (0.005) | (0.148) | (0.002) | (0.049) | (0.033) | (0.093) | (0.046) | (0.587) |
| 13 weeks | -0.006 | -0.075 | 0.075 | -0.175 | 1.466 | 0.384 | 0.026 | -0.934 |
| | (0.004) | (0.060) | (0.004) | (0.051) | (0.033) | (0.276) | (0.045) | (0.760) |
| 26 weeks | -0.005 | -0.039 | 0.141 | -0.154 | 1.468 | 0.783 | 0.026 | -0.957 |
| | (0.003) | (0.031) | (0.008) | (0.048) | (0.033) | (0.536) | (0.045) | (0.766) |
| 52 weeks | -0.005 | -0.022 | 0.239 | -0.184 | 1.465 | 1.595 | 0.026 | -1.004 |
| | (0.002) | (0.017) | (0.013) | (0.049) | (0.032) | (0.994) | (0.046) | (0.800) |
| **Canada** | | | | | | | | |
| 4 weeks | -0.026 | -0.362 | 0.032 | 0.048 | 0.724 | -0.065 | -0.010 | -1.375 |
| | (0.009) | (0.139) | (0.002) | (0.022) | (0.019) | (0.031) | (0.021) | (0.529) |
| 13 weeks | -0.018 | -0.128 | 0.076 | 0.075 | 0.722 | -0.197 | -0.010 | -1.480 |
| | (0.006) | (0.047) | (0.004) | (0.025) | (0.019) | (0.107) | (0.022) | (0.547) |
| 26 weeks | -0.015 | -0.065 | 0.131 | 0.090 | 0.721 | -0.375 | -0.010 | -1.384 |
| | (0.005) | (0.024) | (0.005) | (0.022) | (0.019) | (0.216) | (0.022) | (0.531) |
| 52 weeks | -0.014 | -0.037 | 0.221 | 0.087 | 0.721 | -0.701 | -0.011 | -1.349 |
| | (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.

<div align="center">

**Table 2**
**Specification tests**

</div>

|  | $\tau = 4$ weeks | $\tau = 13$ weeks | $\tau = 26$ weeks | $\tau = 52$ weeks |
|---|---|---|---|---|
| **UK** | | | | |
| $\beta^{(1)} - \beta^{(4)}$ | 0.033 | -0.192 | -0.299 | -0.354 |
| NW | [0.918] | [0.596] | [0.436] | [0.442] |
| EHS | [0.774] | [0.165] | [0.079] | [0.007] |
| AIC | [0.878] | [0.230] | [0.061] | [0.019] |
| BIC | [0.816] | [0.244] | [0.088] | [0.114] |
| **Germany** | | | | |
| $\beta^{(1)} - \beta^{(4)}$ | 0.029 | -0.088 | -0.096 | -0.060 |
| NW | [0.901] | [0.778] | [0.757] | [0.852] |
| EHS | [0.739] | [0.378] | [0.365] | [0.522] |
| AIC | [0.643] | [0.869] | [0.465] | [0.620] |
| BIC | [0.878] | [0.419] | [0.333] | [0.514] |
| **Canada** | | | | |
| $\beta^{(1)} - \beta^{(4)}$ | -0.303 | -0.297 | -0.186 | -0.119 |
| NW | [0.144] | [0.173] | [0.378] | [0.555] |
| EHS | [0.034] | [0.006] | [0.021] | [0.084] |
| AIC | [0.031] | [0.533] | [0.494] | [0.213] |
| BIC | [0.002] | [0.030] | [0.111] | [0.161] |

**Note:** p-values of the null hypothesis $H_0 : \beta^{(1)} - \beta^{(4)} = 0$ are presented in square brackets. Sample Period: January 1977 to December 2005; 1,513 weekly observations.

Table 3
**Table 3**
**Estimates of the continuous-time model in Example 2**

| Contract | $\phi_{11}$ | $\phi_{21}$ | $\phi_{12}$ | $\phi_{22}$ | $\sigma_{11}$ | $\sigma_{21}$ | $\sigma_{22}$ | $\alpha_1$ | $\alpha_2$ | $\mu_p$ | $\mu_{\Delta s}$ | $\beta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **UK** | | | | | | | | | | | | |
| 52 | -0.055 | -0.014 | -0.728 | -0.260 | 0.283 | -0.055 | -0.002 | -0.076 | 1.368 | -1.862 | 0.015 | -2.077 |
| | (0.047) | (0.014) | (0.698) | (0.231) | (0.013) | (0.038) | (0.011) | (0.059) | (0.040) | (0.839) | (0.063) | (0.998) |
| **Canada** | | | | | | | | | | | | |
| 4 | -0.034 | -0.029 | -0.021 | -0.289 | 0.032 | -0.057 | -0.031 | 0.064 | 0.736 | -0.065 | -0.009 | -1.094 |
| | (0.012) | (0.063) | (0.024) | (0.211) | (0.002) | (0.027) | (0.025) | (0.025) | (0.023) | (0.029) | (0.017) | (0.465) |
| 13 | -0.024 | -0.012 | -0.050 | -0.280 | 0.075 | -0.059 | -0.032 | 0.094 | 0.734 | -0.198 | -0.009 | -0.837 |
| | (0.008) | (0.019) | (0.053) | (0.176) | (0.004) | (0.025) | (0.024) | (0.027) | (0.023) | (0.099) | (0.017) | (0.497) |
| 26 | -0.007 | -0.001 | 0.122 | -0.119 | 0.132 | -0.037 | 0.008 | 0.097 | 0.715 | -0.381 | -0.008 | -0.669 |
| | (0.007) | (0.005) | (0.091) | (0.080) | (0.005) | (0.013) | (0.011) | (0.023) | (0.020) | (0.282) | (0.022) | (0.642) |

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

<div align="center">

**Table 4**
**Comparison of UIP tests: implicit betas**

</div>

| | NW | EHS | AIC | BIC | VAR(1) | VAR(4) | OU(1) | OU(2) |
|---|---|---|---|---|---|---|---|---|
| **UK** | | | | | | | | |
| $\tau = 4$ weeks | -2.071 | -2.071 | -2.071 | -2.071 | -2.002 | -2.107 | -2.084 | |
| | (0.858) | (0.714) | (0.965) | (1.031) | (0.790) | (0.836) | (0.780) | |
| | [0.000] | [0.000] | [0.001] | [0.003] | [0.000] | [0.000] | [0.000] | |
| $\tau = 13$ weeks | -2.155 | -2.155 | -2.155 | -2.155 | -2.401 | -2.172 | -2.519 | |
| | (0.904) | (1.336) | (0.962) | (1.575) | (0.873) | (0.934) | (0.886) | |
| | [0.000] | [0.018] | [0.001] | [0.045] | [0.000] | [0.001] | [0.000] | |
| $\tau = 26$ weeks | -2.051 | -2.051 | -2.051 | -2.051 | -2.407 | -2.042 | -2.605 | |
| | (0.760) | (1.420) | (1.229) | (1.446) | (0.884) | (0.938) | (0.934) | |
| | [0.000] | [0.032] | [0.013] | [0.035] | [0.001] | [0.001] | [0.000] | |
| $\tau = 52$ weeks | -1.507 | -1.507 | -1.507 | -1.507 | -2.257 | -1.989 | -2.652 | -2.077 |
| | (0.578) | (1.448) | (0.937) | (1.595) | (0.894) | (0.919) | (1.063) | (0.998) |
| | [0.000] | [0.083] | [0.007] | [0.116] | [0.000] | [0.001] | [0.001] | [0.002] |
| **Germany** | | | | | | | | |
| $\tau = 4$ weeks | -0.988 | -0.988 | -0.988 | -0.988 | -0.950 | -0.996 | -0.946 | |
| | (0.572) | (0.316) | (0.597) | (0.637) | (0.602) | (0.627) | (0.587) | |
| | [0.001] | [0.000] | [0.001] | [0.002] | [0.001] | [0.001] | [0.001] | |
| $\tau = 13$ weeks | -0.785 | -0.785 | -0.785 | -0.785 | -0.946 | -0.699 | -0.934 | |
| | (0.623) | (0.896) | (0.626) | (1.109) | (0.780) | (0.827) | (0.760) | |
| | [0.004] | [0.046] | [0.004] | [0.108] | [0.013] | [0.040] | [0.011] | |
| $\tau = 26$ weeks | -0.911 | -0.911 | -0.911 | -0.911 | -0.971 | -0.697 | -0.957 | |
| | (0.488) | (0.888) | (0.873) | (0.964) | (0.790) | (0.835) | (0.766) | |
| | [0.000] | [0.031] | [0.029] | [0.047] | [0.013] | [0.042] | [0.011] | |
| $\tau = 52$ weeks | -0.756 | -0.756 | -0.756 | -0.756 | -1.029 | -0.823 | -1.004 | |
| | (0.383) | (1.492) | (0.942) | (1.039) | (0.834) | (0.880) | (0.800) | |
| | [0.000] | [0.239] | [0.062] | [0.091] | [0.015] | [0.038] | [0.012] | |
| **Canada** | | | | | | | | |
| $\tau = 4$ weeks | -1.081 | -1.081 | -1.081 | -1.081 | -1.364 | -1.119 | -1.375 | -1.094 |
| | (0.354) | (0.416) | (0.320) | (0.323) | (0.499) | (0.508) | (0.529) | (0.465) |
| | [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 | -1.460 | -0.940 | -1.480 | -0.837 |
| | (0.346) | (0.402) | (0.395) | (0.489) | (0.514) | (0.546) | (0.547) | (0.497) |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| $\tau = 26$ weeks | -0.746 | -0.746 | -0.746 | -0.746 | -1.358 | -0.983 | -1.384 | -0.669 |
| | (0.287) | (0.474) | (0.582) | (0.490) | (0.496) | (0.527) | (0.531) | (0.642) |
| | [0.000] | [0.000] | [0.003] | [0.000] | [0.000] | [0.000] | [0.000] | [0.009] |
| $\tau = 52$ weeks | -0.976 | -0.976 | -0.976 | -0.976 | -1.317 | -1.072 | -1.349 | |
| | (0.282) | (1.004) | (0.632) | (0.615) | (0.463) | (0.518) | (0.502) | |
| | [0.000] | [0.049] | [0.002] | [0.001] | [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.

Figure 1: P-value discrepancy plot for UIP test $\beta = 1$

Figure 2: Size-adjusted power for UIP test $\beta = 1$

Legend:
- ---- 45° line
- — · — OLS (NW)
- — — OLS (EHS)
- ---- OLS (VARHAC-AIC)
- —— OLS (VARHAC-BIC)
- — · — VAR(1)
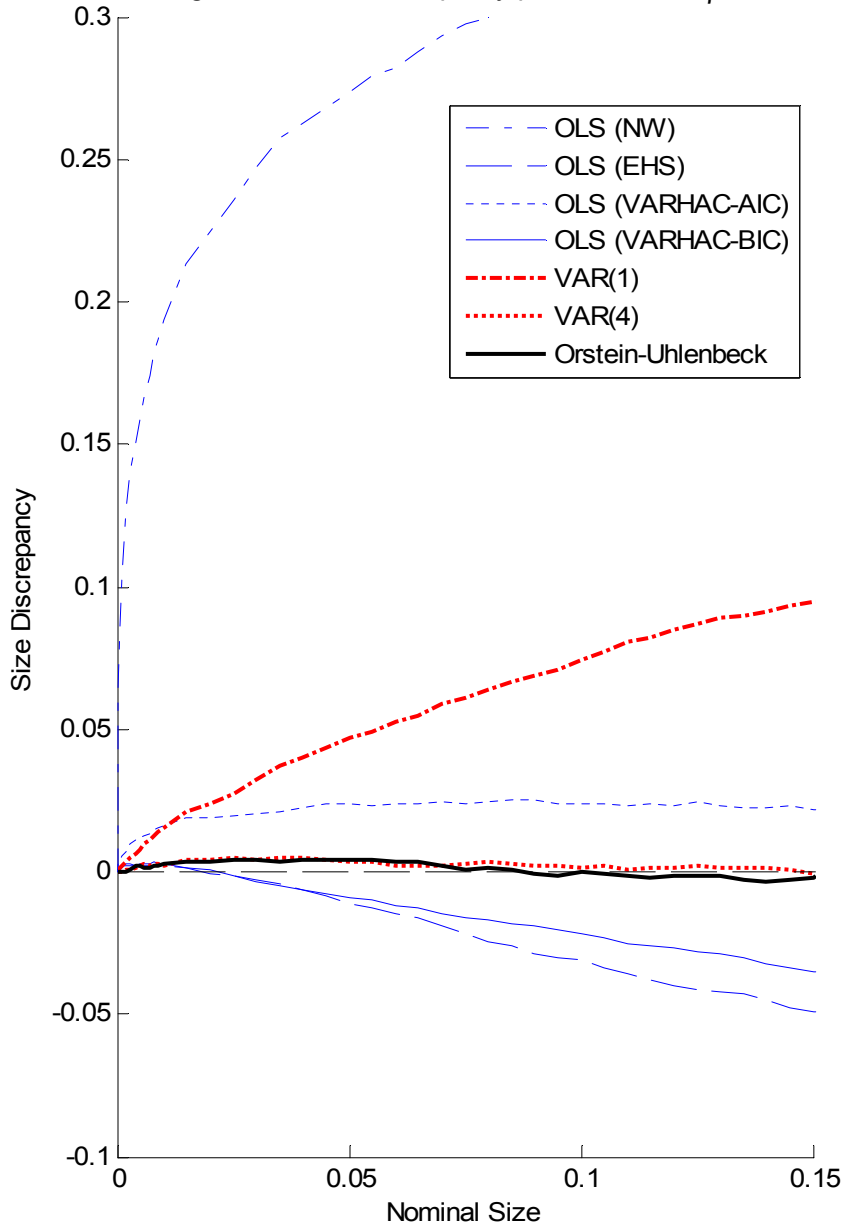- ······ VAR(4)
- —— Orstein-Uhlenbeck

Figure 3: P-value discrepancy plot for specification test $\beta^{(1)}=\beta^{(4)}$
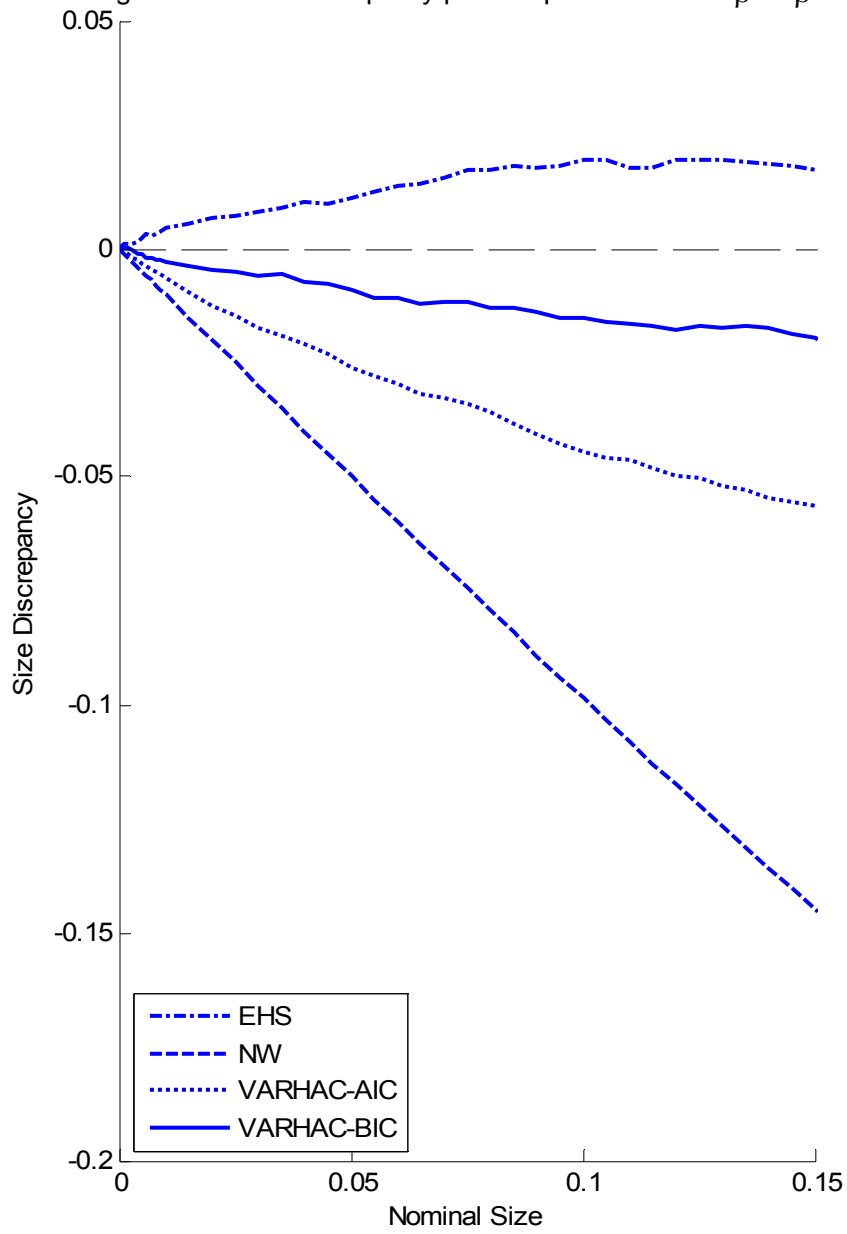
Figure 4: Size-adjusted power for specification test $\beta^{(1)}=\beta^{(4)}$