Journal of Empirical Finance 68 (2022) 232–245

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Multiple testing of the forward rate unbiasedness hypothesis across currencies[☆]

Hsuan Fu, Richard Luger [*]

Department of Finance, Insurance and Real Estate, Laval University, Quebec City, Quebec, Canada

ARTICLE INFO

JEL classification:
C14
C15
C33
F31

Keywords:
Foreign exchange rates
Predictability
Simultaneous inference
Unbalanced panel
Conditional heteroskedasticity

ABSTRACT

We develop exact distribution-free test procedures for joint inference about the forward rate unbiasedness hypothesis (FRUH) across multiple currencies. The procedures can be applied with either levels or differences specifications. This unified approach proceeds with sign and signed rank tests for each currency and then uses Monte Carlo resampling to control the overall Type I error rate of either: (i) global FRUH tests obtained via combinations of the p-values; or (ii) individual FRUH tests using multiplicity adjusted p-values. Our framework allows for missing data and for the presence of time-varying conditional covariances between currencies. The usefulness of the new procedures is illustrated with a simulation study and with assessments of the FRUH across 13 currencies in an unbalanced panel. Multiplicity adjusted p-values reveal that the joint FRUH rejections are primarily driven by just a few of the more minor currencies.

1. Introduction

The FRUH states that if foreign exchange market participants are risk neutral and have rational expectations, and the market is perfectly competitive, then the forward rate should be an unbiased predictor of the future spot rate. To introduce this hypothesis more formally, let $f_{i,t}$ denote the log forward rate of currency i at time t for delivery one period later and let $s_{i,t+1}$ denote the corresponding log spot rate at time $t+1$. The FRUH asserts that

$$E_t[s_{i,t+1}] = f_{i,t}, \tag{1}$$

where E_t denotes expectation conditional on $\mathcal{I}_t$, the information available to market participants at time t.[1] Once the forward rate has been determined by the market participants, a random shock that was unpredictable at time t will occur before the spot rate is observed at time $t+1$. This means that under (1) the spot rate is given by

$$s_{i,t+1} = f_{i,t} + u_{i,t+1}, \tag{2}$$

where $E_t[u_{i,t+1}] = 0$, i.e., $u_{i,t+1}$ is orthogonal to $\mathcal{I}_t$. If the unbiasedness condition in (1) fails, then it is possible to implement a carry trade strategy that earns a positive average excess return.[2] There is a large body of literature on testing the FRUH. The various

☆ We would like to thank two anonymous reviewers for their comments and suggestions. This work draws on research supported by the Social Sciences and Humanities Research Council of Canada.

* Corresponding author.
E-mail addresses: hsuan.fu@fsa.ulaval.ca (H. Fu), richard.luger@fsa.ulaval.ca (R. Luger).

[1] As Delcoure et al. (2003) explain, taking the natural logarithms of the spot and forward rates ensures that the FRUH relationship can hold true on both sides of the market; i.e., both $E_t[s_{i,t+1}] = f_{i,t}$ and $E_t[1/s_{i,t+1}] = 1/f_{i,t}$ can hold simultaneously. Owing to Jensen's inequality, those relationships cannot both hold true simultaneously with the raw spot and forward rates.

[2] In a currency carry trade, investors sell forward currencies that are at a forward premium (i.e., the forward rate exceeds the spot exchange rate) and buy forward currencies that are at a forward discount. Transactions costs aside, this strategy is equivalent to borrowing low interest rate currencies, lending high interest rate currencies, and not hedging exchange rate risk (Burnside et al., 2008). A second profit opportunity in currency arbitrage is available through violations of triangular parity; see Gradojevic et al. (2020).

https://doi.org/10.1016/j.jempfin.2022.07.005
Received 1 February 2021; Received in revised form 3 March 2022; Accepted 4 July 2022
Available online 5 July 2022

approaches that have been proposed can be classified according to whether they assess the FRUH relationship: (i) in levels or differences; and (ii) with one currency at a time or with multiple currencies jointly.

The early literature (e.g., Frenkel, 1976; Cornell, 1977; Levitch, 1979) focused on the levels specification

$$s_{i,t+1} = a_i + \beta_i f_{i,t} + u_{i,t+1}, \tag{3}$$

which is consistent with the FRUH in (1) when: $a_i = 0$, $\beta_i = 1$, and $E[u_{i,t+1} \mid f_{i,t}] = 0$. If $E[u_{i,t+1} \mid f_{i,t}] = 0$ is assumed, then the implied moment condition $E[u_{i,t+1} f_{i,t}] = 0$ provides the basis for OLS estimation in (3). The OLS estimates of β_i in such regressions are typically close to 1, in line with the FRUH. The levels data $\{s_{i,t}, f_{i,t}\}$, however, are usually found to be non-stationary, which means that traditional inference methods are inappropriate for testing the FRUH; see Meese and Singleton (1982), Meese (1989), and Baillie and Bollerslev (1989).

The other form that has been widely used in the literature (e.g., Bilson, 1981; Fama, 1984; Froot and Frankel, 1989; Liu and Maynard, 2005) is the differences specification

$$s_{i,t+1} - s_{i,t} = a_i + \beta_i (f_{i,t} - s_{i,t}) + u_{i,t+1}, \tag{4}$$

which avoids working with the non-stationary levels data. This specification is consistent with the FRUH when: $a_i = 0$, $\beta_i = 1$, and $E[u_{i,t+1} \mid f_{i,t}, s_{i,t}] = 0$. Levels versus differences regressions have been studied in Phillips and McFarland (1997), Goodhart et al. (1997), Maynard and Phillips (2001), and Maynard (2003). These studies show that differences regressions have some disadvantages in comparison to levels regressions.[3]

Besides the econometric issues with OLS regressions, it is important to recognize that levels and differences specifications are equivalent only when $\beta_i = 1$. Otherwise, as Goodhart et al. (1997) point out, (3) and (4) will have power in different directions away from $\beta_i = 1$. To see this, note that the differences specification in (4) can be rewritten as

$$s_{i,t+1} = a_i + \beta_i f_{i,t} + (1 - \beta_i) s_{i,t} + u_{i,t+1},$$

making clear that $E_t[s_{i,t+1}]$ is a restricted linear combination of $f_{i,t}$ and $s_{i,t}$. Therefore, the levels specification in (3) will detect departures from the FRUH in which $E_t[s_{i,t+1}]$ depends on $\beta_i f_{i,t}$ (for some $\beta_i \neq 1$) and not at all on $s_{i,t}$. On the other hand, the differences specification in (4) will tend to have better discriminatory power when the term $(1 - \beta_i) s_{i,t}$ plays an important role in $E_t[s_{i,t+1}]$ and $f_{i,t} \neq s_{i,t}$. However when a_i is zero and $f_{i,t} \approx s_{i,t}$ (as often found empirically), β_i will be poorly identified in (4) and the differences specification will be lacking in power since then $E_t[s_{i,t+1}] \approx f_{i,t}$, even if $\beta_i \neq 1$.

Another view of the FRUH can be gleaned from models of cointegration between spot and forward rates (e.g., Hakkio and Rush, 1989; Barnhart and Szakmary, 1991; Hai et al., 1997; Zivot, 2000). Indeed considering that spot and forward prices are generally found to be non-stationary each with a unit root, then in the levels specification (3) the FRUH requires that $s_{i,t+1}$ and $f_{i,t}$ be cointegrated with cointegrating vector $(1, -1)$ and that $u_{i,t+1}$ be stationary. If $a_i = 0$ and $E[u_{i,t+1}] = 0$ unconditionally, then the forward rate will not systematically under- or over-predict the future spot rate in the long run. When the short-run restriction $E_t[u_{i,t+1}] = 0$ holds, the forward rate is a conditionally unbiased predictor.

The differences specification (4) could detect short-run deviations from the FRUH in which $(1, -1)$ cointegration holds and $u_{i,t+1}$ is stationary, but $E_t[u_{i,t+1}] \neq 0$. For instance, consider the differences specification with $\beta_i = 1$ and $u_{i,t+1} = \gamma_i (f_{i,t} - s_{i,t}) + e_{i,t+1}$ where $E_t[e_{i,t+1}] = 0$. In this case, (4) garners power since it can be rewritten as

$$s_{i,t+1} = a_i + (1 + \gamma_i) f_{i,t} - \gamma_i s_{i,t} + e_{i,t+1},$$

meaning that $E_t[s_{i,t+1}]$ takes once more the form of a restricted linear combination between $f_{i,t}$ and $s_{i,t}$ when $\gamma_i \neq 0$. Again the parameter of interest, γ_i in this case, will be poorly identified when $f_{i,t}$ and $s_{i,t}$ are close.

In practice, it is often the case that the investigator has data on several currencies $i = 1, \ldots, N$ and wishes to test the FRUH for all of them. The joint FRUH null hypothesis then becomes

$$E_t[s_{i,t+1}] = f_{i,t}, \text{ for } i = 1, \ldots, N. \tag{5}$$

The simultaneous testing of the N individual-currency null hypotheses comprising (5) gives rise to a multiple comparisons problem. Indeed if the multiplicity of tests is not taken into account, then the probability that some of the true null hypotheses are rejected by chance alone may be unduly large; see Hochberg and Tamhane (1987) and Hsu (1996) for textbook treatments of multiple comparisons.

One way to control the overall significance level is to conduct a joint test in the context of a system of equations. Seemingly unrelated regressions (SUR) models have been used by Bilson (1981), Fama (1984), Cornell (1989), Barnhart and Szakmary (1991), and Hodgson et al. (2004) to test the differences formulation of the joint FRUH. With a set of simultaneous levels regressions, SUR techniques have also been used by Bailey et al. (1984), Barnhart and Szakmary (1991), Evans and Lewis (1995), and Hodgson et al. (2004). Cointegrated panel regression techniques have been used by Ho (2002), McMillan (2005), Delcoure et al. (2003), and Westerlund (2007) in order to further assess the levels formulation of the joint FRUH.

An important advantage of such systems approaches is that they take into account the cross-sectional dependence that exists between the equations' error terms. The presence of this dependence is to be expected when testing the joint FRUH given the

[3] For instance, Phillips and McFarland (1997) show that differences regressions lead to a lower rate of estimator convergence under the null hypothesis; and, when $\beta_i \neq 1$, the OLS estimator of β_i converges in probability to zero, even when the forward rate has predictive ability in the levels specification.

integration of world financial markets and the mere fact that most exchange rates are quoted against a common currency (i.e., the US dollar). Moreover, as Geweke and Feige (1979) argue, a test that accounts for the information contained in all N currencies is expected to be more decisive than a test based on each individual currency treated in isolation.

A practical limitation, however, when performing a joint FRUH test in the context of a system of equations is that the analysis is usually limited to time periods over which all the currencies have joint observations (balanced panels). This precludes joint FRUH tests with currencies observed over differing time periods. For example, suppose the euro (which was officially introduced in 1999) is considered along with other currencies whose observation period began prior to 1999. In order to exploit the contemporaneous correlation structure of the error terms, systems approaches would typically require the investigator to drop the observations on those other currencies prior to 1999 and perform a joint FRUH test using only the data observed jointly since the euro's inception. The resulting loss of information is obviously undesirable.

In this paper, we develop exact distribution-free test procedures for simultaneous inference about the FRUH either in levels or in differences with currencies that may be observed over differing time periods (i.e., with unbalanced panels). We achieve this by using test statistics for each currency along with Monte Carlo resampling to control the overall Type I error rate of either: (i) global FRUH tests obtained via combinations of the p-values; or (ii) individual FRUH tests using multiplicity adjusted p-values. The developed procedures are based on the sign and signed rank statistics proposed in Campbell and Dufour (1997) and further extended in Gungor and Luger (2020). These tests make no assumptions whatsoever about the process governing the forward rates, $f_{i,t}$. A related application of the sign test is given in Maynard (2006), who uses the Campbell and Dufour (1997) methodology for individual FRUH assessments.

Our approach is in line with Westfall and Young (1993) who advocate at length the use of resampling-based methods to obtain multiple testing procedures which take into account the dependence structure between test statistics and achieve control of the familywise error rate. Our resampling scheme assumes that the error terms either have a zero median or they are symmetric, both in a conditional multivariate sense. These assumptions allow for missing data and for the presence of multivariate conditional heteroskedasticity of unknown form in the spot and forward rates. The Monte Carlo resampling technique accounts for the possibly time-varying cross-sectional dependence across currencies in the unbalanced panel.

The rest of the paper is organized as follows. Section 2 develops the simulation-based inference procedures for joint FRUH testing. Section 3 presents the results of simulation experiments designed to compare the empirical Type I rejection rates and discriminatory power of the test procedures. Section 4 presents an empirical application with 13 major currencies and Section 5 offers some concluding remarks.

2. Simulation-based inference

For $i = 1, \dots, N$, the bivariate time series $\{(s_{i,t}, f_{i,t})\}$ is observed over time index values $t = \tau_{i,1}, \dots, \tau_{i,T_i}$, where we use the convention $\min_i \tau_{i,1} = 1$, $\max_i \tau_{i,T_i} = T$, and $1 \le \tau_{i,1} < \tau_{i,T_i} \le T$. This setup will be especially useful when dealing with currencies observed over differing time periods. A leading example is the euro, which was launched in January 1999 and subsequently replaced a number of national currencies. Our empirical application examines the British pound (GBP) and the euro (EUR), among other currencies. The GBP series comprises $T_{\text{GBP}} = 453$ observations from October 1983 ($\tau_{\text{GBP},1} = 1$) to June 2021 ($\tau_{\text{GBP},T_{\text{GBP}}} = 453$), while the EUR series has $T_{\text{EUR}} = 270$ observations from January 1999 ($\tau_{\text{EUR},1} = 184$) to June 2021 ($\tau_{\text{EUR},T_{\text{EUR}}} = 453$).

2.1. Sign and signed rank tests

The FRUH in (1) for currency i can be expressed as $H_{0,i} : \beta_i = 1$ in the specification

$$y_{i,t+1} = \beta_i x_{i,t} + u_{i,t+1}, \tag{6}$$

for $i = 1, \dots, N$, and where we will assume two-sided alternatives $\beta_i \neq 1$. This corresponds to a levels specification when $y_{i,t+1} = s_{i,t+1}$ and $x_{i,t} = f_{i,t}$; and to a differences specification when $y_{i,t+1} = s_{i,t+1} - s_{i,t}$ and $x_{i,t} = f_{i,t} - s_{i,t}$. As we already mentioned, the choice of a levels or differences specification depends on the alternative for which power is desired. By not including an intercept term, (6) imposes the risk neutrality assumption.[4] Notice also that (6) makes no assumptions about the process governing the forward rates, $f_{i,t}$. This means that the forward rates may exhibit any degree of persistence and may be subject to unmodelled structural breaks, time-varying parameters, or any other non-linearities.

The simultaneous treatment of the N currencies will be facilitated by defining a generic fill-in variable $z^+_{i,t}$ that equals $z_{i,t}$ when $\tau_{i,1} \le t \le \tau_{i,T_i}$, and 0 otherwise. The added zeros are not used as data; they are merely used to make dimensionality adjustments. With such definitions for $y^+_{i,t}$, $x^+_{i,t}$, and $u^+_{i,t}$, $t = 1, \dots, T$, the column vectors

$$\mathbf{y}^+_i = \begin{bmatrix} y^+_{i,2} \\ \vdots \\ y^+_{i,T} \end{bmatrix}, \quad \mathbf{x}^+_{i,-1} = \begin{bmatrix} x^+_{i,1} \\ \vdots \\ x^+_{i,T-1} \end{bmatrix}, \quad \mathbf{u}^+_i = \begin{bmatrix} u^+_{i,2} \\ \vdots \\ u^+_{i,T} \end{bmatrix},$$

[4] If one does not wish to assume $a_i = 0$, then an asymptotically justified approach is to replace $y_{i,t} - x_{i,t-1}$ in the proposed tests by $y_{i,t} - x_{i,t-1} - \hat{a}_i$, where $\hat{a}_i$ is a consistent estimate of a_i under $H_{0,i}$.

for $i = 1, \ldots, N$, each have $T-1$ rows. This is ensured by having zeros in positions $1 \leq t < \tau_{i,1}$ and $\tau_{i,T_i} < t \leq T$; i.e., when currency i did not yet exist or ceased to exist. The collection of time series can then be arranged to get the horizontal stacked form representation

$$\mathbf{Y}^+ = \mathbf{X}_{-1}^+ \boldsymbol{\beta} + \mathbf{U}^+, \tag{7}$$

where $\mathbf{Y}^+ = [\mathbf{y}_1^+, \ldots, \mathbf{y}_N^+]$, $\mathbf{X}_{-1}^+ = [\mathbf{x}_{1,-1}^+, \ldots, \mathbf{x}_{N,-1}^+]$, and $\mathbf{U}^+ = [\mathbf{u}_1^+, \ldots, \mathbf{u}_N^+]$ are each $(T-1) \times N$ matrices; and $\boldsymbol{\beta} = \text{diag}(\beta_1, \ldots, \beta_N)$ is an $N \times N$ diagonal matrix. Note that row t of $\mathbf{U}^+$ corresponds to $\mathbf{u}_t^+ = (u_{1,t}^+, \ldots, u_{N,t}^+)$, the vector of cross-sectional filled-in error terms.

In the context of representation (7), the joint FRUH in (5) across currencies can be stated as

$$H_0 : \boldsymbol{\beta} = \mathbf{I}_N, \tag{8}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix. The general alternative is $\beta_i \neq 0$ for at least one i. For further reference, define the sign function as $\text{sign}(z) = 1$, if $z > 0$; $\text{sign}(z) = -1$, if $z < 0$; and $\text{sign}(z) = 0$, if $z = 0$. When $\mathbf{A}$ is a vector or a matrix, $\text{sign}(\mathbf{A})$ applies the sign function element-wise. We also let $\mathbf{r} = (r_2, \ldots, r_T)'$ be a vector of independent Rademacher random variables such that $\Pr(r_t = 1) = \Pr(r_t = -1) = 1/2$, for all t, and we use the symbol $\stackrel{d}{=}$ to denote an equality in distribution.[5]

Observe that any $z \in \mathbb{R}$ can be decomposed as $z = \text{sign}(z)|z|$. For one group of tests, we merely assume that the error terms in (7) are continuous with a conditional "multivariate median" (Small, 1990) at the origin such that

$$\text{sign}(u_{1,t}^+, \ldots, u_{N,t}^+) \stackrel{d}{=} \text{sign}\big(r_t(u_{1,t}^+, \ldots, u_{N,t}^+)\big), \tag{9}$$

given $\mathcal{I}_{t-1}$. We also propose tests under the stronger assumption that the error terms in (7) are continuous and symmetrically distributed (Serfling, 2006) in the sense that, conditional on $\mathcal{I}_{t-1}$, we have

$$(u_{1,t}^+, \ldots, u_{N,t}^+) \stackrel{d}{=} r_t(u_{1,t}^+, \ldots, u_{N,t}^+). \tag{10}$$

The assumed continuity of the actual error terms entails that $\Pr(u_{i,t} = 0) = 0$, so that $u_{i,t}^+ = 0$ occurs only because currency i is not observed in period t. It is obvious that (10) implies (9), but not vice versa, since (9) restricts only the joint behaviour of the contemporaneous sign vector $(\text{sign}(u_{1,t}^+), \ldots, \text{sign}(u_{N,t}^+))$ and leaves free the corresponding vector of absolute values $(|u_{1,t}^+|, \ldots, |u_{N,t}^+|)$.

From Randles and Wolfe (1979, Lemma 1.3.28), the assumption in (9) implies that the signs of the error terms are uncorrelated over time. The tests derived under (9) should thus be interpreted as tests of whether the conditional median of $s_{i,t+1}$ equals $f_{i,t}$ (i.e., median unbiasedness). Under the symmetry assumption in (10), the conditional mean and median of $u_{i,t+1}^+$ both equal zero, and the errors are serially uncorrelated; i.e., $E_t[u_{i,t+1}^+] = 0$. In this case, the tests that rest on (10) yield an assessment of $E_t[s_{i,t+1}] = f_{i,t}$ (i.e., mean unbiasedness), assuming of course that the first moments are well defined.

Note that the reflective symmetry assumption in (10) allows for an arbitrary (possibly time-varying) contemporaneous covariance structure among $(u_{1,t}^+, \ldots, u_{N,t}^+)$. In fact, (10) is compatible with a large class of multivariate GARCH (Silvennoinen and Teräsvirta, 2009) and multivariate stochastic volatility models (Chib et al., 2009). Indeed, a typical starting point for these models is to write the conditional cross-sectional covariance matrix of model errors at time t as $\boldsymbol{\Sigma}_t$ and to assume that the errors are governed by

$$\mathbf{u}_t'^+ = (\boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\eta}_t')^+,$$

where the elements of the $N \times 1$ vector $\boldsymbol{\eta}_t' = (\eta_{1,t}, \ldots, \eta_{N,t})'$ correspond to a joint draw from a symmetric distribution (e.g., multivariate normal or multivariate Student-t). In the present context, the fill-in elements of $(\boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\eta}_t')^+$ equal the corresponding realized values of $\boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\eta}_t'$ when $\tau_{i,1} \leq t \leq \tau_{i,T_i}$, and 0 otherwise. Here $\boldsymbol{\Sigma}_t^{1/2}$ is an $N \times N$ "square root" matrix such that $\boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\Sigma}_t^{1/2} = \boldsymbol{\Sigma}_t$. If $\boldsymbol{\Sigma}_t^{1/2}$ and $\boldsymbol{\eta}_t'$ are conditionally independent given past information, then (10) is satisfied.[6]

Consider the following non-parametric analogue of the t-statistic:

$$\mathcal{S}_i = \sum_{t=\tau_{i,2}}^{\tau_{i,T_i}} \frac{1}{2} \left(\text{sign}\big[(y_{i,t} - x_{i,t-1}) x_{i,t-1}\big] + 1 \right). \tag{11}$$

This sign statistic belongs to a broader class of linear signed rank statistics defined by

$$\mathcal{Z}_i = \sum_{t=\tau_{i,2}}^{\tau_{i,T_i}} \frac{1}{2} \left(\text{sign}\big[(y_{i,t} - x_{i,t-1}) x_{i,t-1}\big] + 1 \right) \varphi_i(R_{i,t}), \tag{12}$$

where $R_{i,t}$ is the rank of $|y_{i,t} - x_{i,t-1}|$ when $|y_{i,\tau_{i,2}} - x_{i,\tau_{i,1}}|, \ldots, |y_{i,\tau_{i,T_i}} - x_{i,\tau_{i,T_i-1}}|$ are placed in ascending order. Note that $R_{i,\tau_{i,2}}, \ldots, R_{i,\tau_{i,T_i}}$ is an arrangement of the first $T_i - 1 = \tau_{i,T_i} - \tau_{i,1}$ positive integers: $1, \ldots, T_i - 1$. The set of scores $\varphi_i(t)$, $t = 1, \ldots, T_i - 1$, are such that $0 \leq \varphi_i(1) \leq \ldots \leq \varphi_i(T_i - 1)$ with $\varphi_i(T_i - 1) > 0$. The sign statistic in (11) is obtained from the constant scores $\varphi_i(t) = 1$. Another familiar member of this class is the Wilcoxon signed rank statistic

$$\mathcal{W}_i = \sum_{t=\tau_{i,2}}^{\tau_{i,T_i}} \frac{1}{2} \left(\text{sign}\big[(y_{i,t} - x_{i,t-1}) x_{i,t-1}\big] + 1 \right) R_{i,t}, \tag{13}$$

obtained by setting $\varphi_i(t) = t$, for $t = 1, \ldots, T_i - 1$.

[5] For further discussion about the 'equal in distribution' technique, the reader may consult Randles and Wolfe (1979, §1.3).

[6] In a similar context, Hodgson et al. (2004) develop joint FRUH tests within a SUR system with error terms assumed to be i.i.d. according to a multivariate symmetric distribution. That is far more restrictive than (10) since assuming i.i.d. errors rules out the possibility of time-dependent variances and covariances.

For notational convenience let $\tilde{\mathbf{U}}^{+} = [\tilde{\mathbf{u}}_1^{+}, \ldots, \tilde{\mathbf{u}}_N^{+}] = \tilde{\mathbf{r}} \odot \mathbf{U}_0^{+}$, where $\mathbf{U}_0^{+} = \mathbf{Y}^{+} - \mathbf{X}_{-1}^{+}\boldsymbol{\beta}_0$ with $\boldsymbol{\beta}_0 = \mathbf{I}_N$. Here the vector $\tilde{\mathbf{r}} = (\tilde{r}_2, \ldots, \tilde{r}_T)'$ comprises random draws from the Rademacher distribution and the symbol $\odot$ means that the scalar element $\tilde{r}_t$ of $\tilde{\mathbf{r}}$ multiplies every element on row t of $\mathbf{U}_0^{+}$, yielding rows $(\tilde{u}_{1,t}^{+}, \ldots, \tilde{u}_{N,t}^{+}) = (\tilde{r}_t u_{1,t,0}^{+}, \ldots, \tilde{r}_t u_{N,t,0}^{+})$, for $t = 2, \ldots, T$. This randomization scheme preserves the cross-sectional covariance structure.[7]

From Proposition 1 in Gungor and Luger (2020) it is straightforward to see that if H_0 is true, then, for $i = 1, \ldots, N$, we have

$$\mathcal{S}_i \overset{d}{=} \sum_{t=\tau_{i,2}}^{\tau_{i,T_i}} \frac{1}{2}\left(\text{sign}\left[\tilde{u}_{i,t}^{+} x_{i,t-1}\right] + 1\right) \overset{d}{=} \sum_{t=\tau_{i,2}}^{\tau_{i,T_i}} \mathcal{B}_t, \tag{14}$$

when assumption (9) holds, and, given $|y_{i,\tau_{i,2}} - x_{i,\tau_{i,1}}|, \ldots, |y_{i,\tau_{i,T_i}} - x_{i,\tau_{i,T_i-1}}|$, we furthermore have

$$\mathcal{Z}_i \overset{d}{=} \sum_{t=\tau_{i,2}}^{\tau_{i,T_i}} \frac{1}{2}\left(\text{sign}\left[\tilde{u}_{i,t}^{+} x_{i,t-1}\right] + 1\right) \varphi_i(R_{i,t}) \overset{d}{=} \sum_{t=\tau_{i,2}}^{\tau_{i,T_i}} \mathcal{B}_t \varphi_i(t), \tag{15}$$

when assumption (10) holds. Here $\mathcal{B}_{\tau_{i,2}}, \ldots, \mathcal{B}_{\tau_{i,T_i}}$ are independent Bernoulli variables such that $\Pr(\mathcal{B}_t = 1) = \Pr(\mathcal{B}_t = 0) = 1/2$, $t = \tau_{i,2}, \ldots, \tau_{i,T_i}$. Notice in (14) and (15) that the characterizations in the middle (involving the terms $\tilde{u}_{i,t}^{+}$) capture the dependence across currencies, while the characterizations on the right (involving the Bernoulli variables) do not. Both these characterizations are used next to obtain joint FRUH tests.

From Randles and Wolfe (1979, §10.2), we know that the distribution of the standardized signed rank statistic

$$\mathcal{Z}_i^{*} = \left(\mathcal{Z}_i - E(\mathcal{Z}_i)\right) \big/ \sqrt{\text{Var}(\mathcal{Z}_i)},$$

with

$$E(\mathcal{Z}_i) = \frac{1}{2} \sum_{t=\tau_{i,2}}^{\tau_{i,T_i}} \varphi_i(t), \quad \text{Var}(\mathcal{Z}_i) = \frac{1}{4} \sum_{t=\tau_{i,2}}^{\tau_{i,T_i}} \varphi_i^2(t),$$

converges to a standard normal as the sample size grows. For the sign statistic in (11) we have $E(\mathcal{S}_i) = (T_i - 1)/2$, $\text{Var}(\mathcal{S}_i) = (T_i - 1)/4$; while for the Wilcoxon signed rank statistic in (13) the needed moments are $E(\mathcal{W}_i) = T_i(T_i - 1)/2$, $\text{Var}(\mathcal{W}_i) = T_i(T_i - 1)(2T_i - 1)/24$. If we let $\Phi(\cdot)$ denote the standard normal cumulative distribution function, the associated two-sided marginal p-values can be defined as $p_i^{\mathcal{Z}} = 2\left(1 - \Phi(|\mathcal{Z}_i^{*}|)\right)$, for $i = 1, \ldots, N$.

2.2. Combined p-values

We now consider combining the attained significance levels $p_1^{\mathcal{Z}}, \ldots, p_N^{\mathcal{Z}}$ in order to obtain a global test of the joint FRUH in (8). To do so, we apply test procedures based on two well-known combination rules (originally proposed for independent statistics):

1. Procedures based on the minimum p-value (Tippett, 1931; Wilkinson, 1951):

$$\mathcal{Z}_{\min} = \min\{p_1^{\mathcal{Z}}, \ldots, p_N^{\mathcal{Z}}\}.$$

2. Procedures based on the product of the individual p-values (Fisher, 1932; Pearson, 1933):

$$\mathcal{Z}_{\times} = \prod_{i=1}^{N} p_i^{\mathcal{Z}}.$$

These rules will lead us to reject H_0 when $\mathcal{Z}_{\min}$ (or $\mathcal{Z}_{\times}$) is sufficiently small. Even though the marginal p-values $p_1^{\mathcal{Z}}, \ldots, p_N^{\mathcal{Z}}$ may have a very complex dependence structure, their distribution is easy to simulate under the joint FRUH. For further discussion and other examples of test combination rules, see Folks (1984), Dufour et al. (2015), and Gungor and Luger (2015, 2020).

For a random draw $(\tilde{u}_{1,t}^{+}, \ldots, \tilde{u}_{N,t}^{+})$, let the associated values appearing in (15) be denoted as

$$\tilde{\mathcal{Z}}_i = \sum_{t=\tau_{i,2}}^{\tau_{i,T_i}} \frac{1}{2}\left(\text{sign}\left[\tilde{u}_{i,t}^{+} x_{i,t-1}\right] + 1\right) \varphi_i(R_{i,t}), \quad i = 1, \ldots, N. \tag{16}$$

In turn, these yield the simulated raw (unadjusted) p-values

$$\tilde{p}_i^{\mathcal{Z}} = 2\left(1 - \Phi(|\tilde{\mathcal{Z}}_i^{*}|)\right), \tag{17}$$

where

$$\tilde{\mathcal{Z}}_i^{*} = \left(\tilde{\mathcal{Z}}_i - \frac{1}{2} \sum_{t=\tau_{i,2}}^{\tau_{i,T_i}} \varphi_i(t)\right) \Bigg/ \sqrt{\frac{1}{4} \sum_{t=\tau_{i,2}}^{\tau_{i,T_i}} \varphi_i^2(t)},$$

[7] Observe that the resampling proceeds conditional on the absolute values of the error terms, since only their signs are randomized.

for $i = 1, \ldots, N$. If $\mathcal{Z}_i$ is the sign statistic $\mathcal{S}_i$ in (11), then from Proposition 2 in Gungor and Luger (2020) we have under H_0 that

$$\begin{aligned} \mathcal{S}_{\min} &\overset{d}{=} \tilde{\mathcal{S}}_{\min} = \min\{\tilde{p}_1^{\mathcal{S}}, \ldots, \tilde{p}_N^{\mathcal{S}}\}, \\ \mathcal{S}_{\times} &\overset{d}{=} \tilde{\mathcal{S}}_{\times} = \prod_{i=1}^{N} \tilde{p}_i^{\mathcal{S}}, \end{aligned} \tag{18}$$

when assumption (9) holds. Furthermore if $\mathcal{Z}_i$ involves ranks, then we have under H_0 that

$$\begin{aligned} \mathcal{Z}_{\min} &\overset{d}{=} \tilde{\mathcal{Z}}_{\min} = \min\{\tilde{p}_1^{\mathcal{Z}}, \ldots, \tilde{p}_N^{\mathcal{Z}}\}, \\ \mathcal{Z}_{\times} &\overset{d}{=} \tilde{\mathcal{Z}}_{\times} = \prod_{i=1}^{N} \tilde{p}_i^{\mathcal{Z}}, \end{aligned} \tag{19}$$

given the absolute values $|y_{i,\tau_{i,2}} - x_{i,\tau_{i,1}}|, \ldots, |y_{i,\tau_{i,T_i}} - x_{i,\tau_{i,T_i-1}}|$, $i = 1, \ldots, N$, when assumption (10) holds.

Let $\mathcal{Z}_\bullet$ denote either $\mathcal{Z}_{\min}$ or $\mathcal{Z}_\times$ computed with the actual data. The characterization of their exact distributions in (18) and (19) paves the way for the computation of Monte Carlo (MC) p-values (Dwass, 1957; Barnard, 1963; Birnbaum, 1974), as follows:

1. Choose B so that αB is an integer, where $\alpha \in (0, 1)$ is the desired significance level.[8]
2. Using draws from the Rademacher distribution, generate $B - 1$ simulated statistics $\tilde{\mathcal{Z}}_{\bullet,1}, \ldots, \tilde{\mathcal{Z}}_{\bullet,B-1}$ following (16)–(19).
3. Compute the global MC p-value of $\mathcal{Z}_\bullet$ as

$$\tilde{p}(\mathcal{Z}_\bullet) = \frac{G(\mathcal{Z}_\bullet; \tilde{\mathcal{Z}}_{\bullet,1}, \ldots, \tilde{\mathcal{Z}}_{\bullet,B-1})}{B},$$

where $G(\mathcal{Z}_\bullet; \tilde{\mathcal{Z}}_{\bullet,1}, \ldots, \tilde{\mathcal{Z}}_{\bullet,B-1}) = 1 + \sum_{b=1}^{B-1} \mathbb{I}[\tilde{\mathcal{Z}}_{\bullet,b} < \mathcal{Z}_\bullet]$, with $\mathbb{I}$ denoting the indicator function.

The decision rule is then to reject H_0 in (8) at level α if $\tilde{p}(\mathcal{Z}_\bullet) \le \alpha$; otherwise, retain the joint FRUH. Note that $G(\mathcal{Z}_\bullet; \tilde{\mathcal{Z}}_{\bullet,1}, \ldots, \tilde{\mathcal{Z}}_{\bullet,B-1})$ is the rank achieved by $\mathcal{Z}_\bullet$ when the values $\tilde{\mathcal{Z}}_{\bullet,1}, \ldots, \tilde{\mathcal{Z}}_{\bullet,B-1}, \mathcal{Z}_\bullet$ are sorted in ascending order. The unlikely occurrence of ties can be dealt with using lexicographic (tie-breaking) ranks. It is easy to see that the B random variables $\tilde{\mathcal{Z}}_{\bullet,1}, \ldots, \tilde{\mathcal{Z}}_{\bullet,B-1}, \mathcal{Z}_\bullet$ are *exchangeable*. From Dufour (2006), we therefore have that $\Pr\left(\tilde{p}(\mathcal{Z}_\bullet) \le \alpha \mid H_0 \text{ is true}\right) = \alpha$. We refer the reader to Dufour and Khalaf (2001) and Kiviet (2012) for a general overview of MC test techniques and further references.

2.3. Multiplicity adjusted p-values

Besides the global inference based on the test combination rules, we can also test the FRUH for each currency given a suitably defined overall Type I error rate. We do so by testing $H_{0,1}, \ldots, H_{0,N}$ individually while keeping under control the familywise error rate (FWER), i.e., the probability of falsely rejecting at least one true FRUH. Westfall and Young (1993) propose several resampling-based methods to adjust individual p-values so as to account for the multiplicity effect. These methods yield FWER adjusted p-values for $H_{0,i}$, generically defined by $p_{i,\text{Adj}} = \inf\{\alpha : H_{0,i} \text{ is rejected at FWER} = \alpha\}$. In comparison with a global p-value, the advantage of adjusted p-values is that they pinpoint the currencies standing in violation of the FRUH. Next we describe how to obtain adjusted MC p-values.

2.3.1. Single-step adjustments

The Westfall and Young (1993) single-step (SS) adjusted p-values are defined in the present context by

$$p_{i,\text{SS}}^{\mathcal{Z}} = \Pr\left(\min_{1 \le j \le N} \tilde{p}_j^{\mathcal{Z}} \le p_i^{\mathcal{Z}} \mid H_0 \text{ is true}\right),$$

which is the probability that the minimum p-value in the resampling distribution is smaller than the p-value observed with the actual data, under the joint FRUH. Based on Westfall and Young (1993, Algorithm 2.5), the MC version of this adjusted p-value is computed as follows:

1. Choose B so that αB is an integer, where $\alpha \in (0, 1)$ is the desired FWER.
2. For $b = 1, \ldots, B - 1$, repeat the following steps:
   (a) Using draws from the Rademacher distribution, simulate raw p-values $\tilde{p}_{1,b}^{\mathcal{Z}}, \ldots, \tilde{p}_{N,b}^{\mathcal{Z}}$ according to (16)–(17).
   (b) Find $\tilde{m}_b^{\mathcal{Z}} = \min_{1 \le j \le N} \tilde{p}_{j,b}^{\mathcal{Z}}$.
3. For $i = 1, \ldots, N$, compute the SS adjusted MC p-value as

$$\tilde{p}_{i,\text{SS}}^{\mathcal{Z}} = \frac{G(p_i^{\mathcal{Z}}; \tilde{m}_1^{\mathcal{Z}}, \ldots, \tilde{m}_{B-1}^{\mathcal{Z}})}{B},$$

where now $G(p_i^{\mathcal{Z}}; \tilde{m}_1^{\mathcal{Z}}, \ldots, \tilde{m}_{B-1}^{\mathcal{Z}}) = 1 + \sum_{b=1}^{B-1} \mathbb{I}[\tilde{m}_b^{\mathcal{Z}} < p_i^{\mathcal{Z}}]$.

[8] For example, setting $B = 20$ is sufficient to obtain a test with exact level 0.05. A larger number of replications decreases the test's sensitivity to the underlying randomization and typically leads to power gains. Dufour et al. (2004), however, find that increasing B beyond 100 has only a small effect on power.

We can then reject an individual-currency null hypothesis $H_{0,i}$ at FWER α, whenever $\tilde{p}^{\mathcal{Z}}_{i,\mathrm{SS}} \leq \alpha$. Since the adjusted p-values are computed under the joint FRUH, it should not be surprising that the FWER is controlled in the sense that $\Pr\left(\min_{1\leq i\leq N} \tilde{p}^{\mathcal{Z}}_{i,\mathrm{SS}} \leq \alpha \mid H_0 \text{ is true}\right) = \alpha$. This is obvious from the equivalence

$$\left\{ \min_{1\leq i\leq N} \tilde{p}^{\mathcal{Z}}_{i,\mathrm{SS}} \leq \alpha \right\} \Longleftrightarrow \left\{ \tilde{p}(\mathcal{Z}_{\min}) \leq \alpha \right\}, \tag{20}$$

given the same underlying random draws of $(\tilde{u}^+_{1,t}, \dots, \tilde{u}^+_{N,t})$ in the computation of the p-values.

2.3.2. Step-down adjustments

The SS adjustments tend to be conservative since they are all based on a common critical value, obtained from the distribution of the minimum of all p-values. Power improvements may be achieved with step-down (SD) procedures. Let $\pi_1, \dots, \pi_N$ be the index values that define the p-value ordering $p^{\mathcal{Z}}_{\pi_1} \leq p^{\mathcal{Z}}_{\pi_2} \leq \dots \leq p^{\mathcal{Z}}_{\pi_N}$, and let $H_{0,\pi_1}, \dots, H_{0,\pi_N}$ denote the corresponding null hypotheses. The Westfall and Young (1993, p. 66) SD adjusted p-values are defined by

$$p^{\mathcal{Z}}_{\pi_i,\mathrm{SD}} = \max_{k=1,\dots,i} \left\{ \Pr\left(\min_{\ell=k,\dots,N} \tilde{p}^{\mathcal{Z}}_{\pi_\ell} \leq p^{\mathcal{Z}}_{\pi_k} \mid H_0 \text{ is true} \right) \right\},$$

where the minima are taken over successively more restricted sets of p-values as k increases. This is the key that makes $p^{\mathcal{Z}}_{\pi_i,\mathrm{SD}}$ less conservative than $p^{\mathcal{Z}}_{i,\mathrm{SS}}$. Note also that taking successive maxima ensures the monotonicity of the adjusted p-values; i.e., $p^{\mathcal{Z}}_{\pi_1,\mathrm{SD}} \leq p^{\mathcal{Z}}_{\pi_2,\mathrm{SD}} \leq \dots \leq p^{\mathcal{Z}}_{\pi_N,\mathrm{SD}}$ so a particular hypothesis in the ordering gets rejected only if all hypotheses with smaller adjusted p-values were rejected beforehand. Adapting from Westfall and Young (1993, Algorithm 2.8) and Ge et al. (2003, Box 3), the SD adjusted MC p-values are computed according to the following procedure:

1. With the actual raw p-values $p^{\mathcal{Z}}_1, \dots, p^{\mathcal{Z}}_N$, get the index values $\pi_1, \dots, \pi_N$ that define the ordering $p^{\mathcal{Z}}_{\pi_1} \leq p^{\mathcal{Z}}_{\pi_2} \leq \dots \leq p^{\mathcal{Z}}_{\pi_N}$.[9]
2. Choose B so that αB is an integer, where $\alpha \in (0,1)$ is the desired FWER.
3. For $b = 1, \dots, B-1$, repeat the following steps:
   (a) Using draws from the Rademacher distribution, simulate raw p-values $\tilde{p}^{\mathcal{Z}}_{1,b}, \dots, \tilde{p}^{\mathcal{Z}}_{N,b}$ according to (16)–(17).
   (b) Find the successive minima of the simulated raw p-values as
   $$\begin{aligned} \tilde{q}^{\mathcal{Z}}_{N,b} &= \tilde{p}^{\mathcal{Z}}_{\pi_N,b}, \\ \tilde{q}^{\mathcal{Z}}_{i,b} &= \min\left(\tilde{q}^{\mathcal{Z}}_{i+1,b}, \tilde{p}^{\mathcal{Z}}_{\pi_i,b}\right), \text{ for } i = N-1, \dots, 1. \end{aligned}$$
4. Compute the SD adjusted MC p-value as
   $$\tilde{p}^{\mathcal{Z}}_{\pi_i,\mathrm{SD}} = \frac{G(p^{\mathcal{Z}}_{\pi_i}; \tilde{q}^{\mathcal{Z}}_{i,1}, \dots, \tilde{q}^{\mathcal{Z}}_{i,B-1})}{B},$$
   for $i = 1, \dots, N$, where $G(p^{\mathcal{Z}}_{\pi_i}; \tilde{q}^{\mathcal{Z}}_{i,1}, \dots, \tilde{q}^{\mathcal{Z}}_{i,B-1}) = 1 + \sum_{b=1}^{B-1} \mathbb{I}[\tilde{q}^{\mathcal{Z}}_{i,b} < p^{\mathcal{Z}}_{\pi_i}]$, and with the monotonicity of the p-values enforced by setting
   $$\begin{aligned} \tilde{p}^{\mathcal{Z}}_{\pi_1,\mathrm{SD}} &\leftarrow \tilde{p}^{\mathcal{Z}}_{\pi_1,\mathrm{SD}}, \\ \tilde{p}^{\mathcal{Z}}_{\pi_i,\mathrm{SD}} &\leftarrow \max\left(\tilde{p}^{\mathcal{Z}}_{\pi_{i-1},\mathrm{SD}}, \tilde{p}^{\mathcal{Z}}_{\pi_i,\mathrm{SD}}\right), \text{ for } i = 2, \dots, N. \end{aligned}$$

The decision rule is then to reject hypothesis H_{0,π_i}, $i = 1, \dots, N$, at FWER α when $\tilde{p}^{\mathcal{Z}}_{\pi_i,\mathrm{SD}} \leq \alpha$. For a general proof of FWER control with SD procedures, we refer the reader to Westfall and Young (1993, §2.8). Compared to the SS adjustments, the SD approach can improve power since it results in uniformly smaller (or equal) adjusted p-values, while retaining the same FWER protection.

3. Simulation experiments

In this section, we present the results of simulation experiments comparing the performance of the proposed test procedures. For this purpose, we generate spot and forward rates data according to

$$\begin{aligned} s_{i,t+1} &= \beta_i f_{i,t} + \gamma_i s_{i,t} + u_{i,t+1}, \\ f_{i,t+1} &= \phi_i f_{i,t} + v_{i,t+1}, \end{aligned}$$

for $i = 1, \dots, N$, with $\beta_i = \beta$, $\gamma_i = \gamma$, $\phi_i = \phi$, and error terms $\boldsymbol{\varepsilon}_t = (u_{1,t}, \dots, u_{N,t}, v_{1,t}, \dots, v_{N,t})$ following a multivariate stochastic volatility process (Chib et al., 2009) of the form:

$$\begin{aligned} \boldsymbol{\varepsilon}'_t &= \mathbf{V}_t^{1/2} \boldsymbol{\eta}'_t, \quad \mathbf{V}_t^{1/2} = \mathrm{diag}\left(\exp(h_{1,t}/2), \dots, \exp(h_{2N,t}/2)\right), \\ h_{j,t+1} &= \phi_h h_{j,t} + \sigma_h e_{j,t}, \quad e_{j,t} \sim \mathrm{N}(0,1), \quad j = 1, \dots, 2N, \end{aligned}$$

[9] Note that the index values $\pi_1, \dots, \pi_N$ are fixed throughout the simulation steps.

Table 1
Empirical FWER and power of joint FRUH tests: Normal errors, $T = 100$.

ϕ	ρ	σ_h	Levels					Differences				
			t	$\mathcal{S}_{\min}$	$\mathcal{S}_{\times}$	$\mathcal{W}_{\min}$	$\mathcal{W}_{\times}$	t	$\mathcal{S}_{\min}$	$\mathcal{S}_{\times}$	$\mathcal{W}_{\min}$	$\mathcal{W}_{\times}$
Panel A: $\beta = 1$, $\gamma = 0$												
0.95	0.15	0.1	13.6	5.1	5.3	5.2	5.1	5.4	5.0	4.8	5.0	5.4
0.95	0.15	0.5	9.6	4.9	5.3	5.2	5.3	13.0	4.6	4.9	4.6	4.8
0.95	0.85	0.1	13.4	4.7	4.6	4.9	4.9	5.3	5.0	4.9	4.5	4.6
0.95	0.85	0.5	9.3	5.1	4.9	4.8	4.6	12.9	4.7	5.0	4.6	5.0
1	0.15	0.1	30.1	5.0	5.4	5.0	4.8	4.5	4.6	4.8	5.0	4.9
1	0.15	0.5	21.1	5.6	5.2	5.2	5.3	13.2	4.6	4.9	4.5	4.7
1	0.85	0.1	29.2	4.5	4.7	4.8	4.7	4.6	5.1	5.1	4.8	4.8
1	0.85	0.5	20.4	4.9	4.7	4.7	4.6	12.9	4.9	4.7	4.4	4.4
Panel B: $\beta = 0.95$, $\gamma = 0$												
0.95	0.15	0.1	23.4	8.8	9.5	10.5	11.8	4.2	4.7	4.8	4.6	4.8
0.95	0.15	0.5	47.0	35.7	37.5	37.0	39.1	5.3	6.0	6.3	6.1	6.3
0.95	0.85	0.1	22.4	8.4	9.5	9.9	11.6	4.1	5.0	5.1	4.9	4.8
0.95	0.85	0.5	46.0	36.2	37.1	36.5	37.5	5.4	6.2	6.2	5.7	5.8
1	0.15	0.1	70.5	38.8	43.3	53.2	62.8	45.4	30.6	32.3	41.6	44.2
1	0.15	0.5	61.7	68.3	71.2	70.5	74.3	10.3	41.4	42.3	44.3	45.4
1	0.85	0.1	68.5	38.2	41.8	54.3	60.6	45.1	31.1	31.9	41.5	43.1
1	0.85	0.5	63.4	69.4	71.3	71.4	73.3	10.6	42.0	42.9	44.5	45.5
Panel C: $\beta = 1.2$, $\gamma = -0.2$												
0.95	0.15	0.1	5.4	4.0	3.8	4.0	4.0	10.8	6.7	7.4	7.2	7.9
0.95	0.15	0.5	3.9	3.7	3.8	3.5	3.3	41.4	37.3	39.3	39.3	41.5
0.95	0.85	0.1	5.5	4.4	3.8	3.6	3.6	11.0	6.4	6.6	6.8	7.1
0.95	0.85	0.5	4.0	3.4	3.6	3.4	3.3	41.8	37.3	38.7	39.7	41.5
1	0.15	0.1	4.5	5.0	5.3	4.7	4.7	9.2	6.1	6.1	7.1	7.1
1	0.15	0.5	3.7	5.6	5.5	5.3	5.5	39.0	35.9	37.9	38.4	40.6
1	0.85	0.1	4.6	4.9	4.7	4.7	4.5	7.7	6.6	6.5	7.5	7.0
1	0.85	0.5	3.6	5.0	5.3	4.9	4.8	38.9	36.1	37.1	38.3	39.9

This table reports the empirical probabilities (in percentages) of rejecting at least one individual FRUH given $\alpha = 5\%$ and $N = 2$, where $\{(s_{1,t}, f_{1,t}) : t = 1, \ldots, T\}$ and $\{(s_{2,t}, f_{2,t}) : t = 0.25T, \ldots, 0.75T\}$. Panel A reports the empirical FWER, while Panels B and C show the empirical power. The power results for the t-tests are based on FWER adjusted p-values.

where $\boldsymbol{\eta}_t'$ is drawn either from a normal distribution, $\mathsf{N}(\mathbf{0}, \boldsymbol{\Lambda})$, or from a Student-$t$ distribution, $\mathrm{t}_3(\mathbf{0}, \boldsymbol{\Lambda})$, both of dimension $2N$. We present results for $N = 2$ currencies and set the corresponding variance–covariance matrix as

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1 & \rho & 0.99 & \rho \\ \rho & 1 & \rho & 0.99 \\ 0.99 & \rho & 1 & \rho \\ \rho & 0.99 & \rho & 1 \end{bmatrix}$$

so the *within* currency correlations equal 0.99, while *in between* currencies the correlations are determined by ρ. This structure reflects what we find empirically, i.e., the error correlations between spot and forward rates are nearly one for a given currency, whereas those correlations are lower and more variable between currencies.

We let the longest series $\{(s_{1,t}, f_{1,t})\}$ range over $\tau_{1,1} = 1, \ldots, \tau_{1,T_1} = T$ and we let the shorter bivariate series $\{(s_{2,t}, f_{2,t})\}$ be observed over $\tau_{2,1} = 0.25T, \ldots, \tau_{2,T_2} = 0.75T$. This setup serves to illustrate the fact that our statistical framework allows the series to have different start and end dates. We consider $T = 100$ and 200, and model parameters taking values as: $(\beta, \gamma) = (1, 0)$ [the null hypothesis], $(0.95, 0)$, $(1.2, -0.2)$ [the alternative hypotheses]; $\phi = 0.95$, 1; $\rho = 0.15$, 0.85; and $\phi_h = 0.95$, $\sigma_h = 0.1$, 0.5. In each case, we test the FRUH in levels and in differences. All the tests are performed with $\alpha = 5\%$ and with $B - 1 = 99$ MC resampling draws; and the reported results are based on 5000 replications of each data-generating configuration.

As benchmark we use the standard t-statistics from regressions (3) and (4) obtained using the data observed over time index $t = \tau_{i,1}, \ldots, \tau_{i,T_i}$. In order to control for the multiplicity of t-tests, we apply the Holm (1979) procedure which proceeds as follows: Let the t-statistics' ordered raw p-values be denoted by $p^t_{\pi_1} \leq p^t_{\pi_2} \leq \ldots \leq p^t_{\pi_N}$, and reject H_{0,π_i} if $p^t_{\pi_k} \leq \alpha/(N - k + 1)$, for $k = 1, \ldots, i$. The corresponding Holm adjusted p-values are given by

$$p^t_{\pi_i,\mathrm{Holm}} = \max_{k=1,\ldots,i} \left\{ \min\left((N - k + 1) p^t_{\pi_k}, 1 \right) \right\}, \tag{21}$$

which is less conservative than the standard Bonferroni correction that would multiply the raw p-values by N at each step; see Westfall and Young (1993, p. 64) and Ge et al. (2003) for more details.

To compare the test procedures we compute the empirical probabilities of rejecting at least one individual FRUH. This yields the empirical FWER when the joint FRUH holds, and provides a measure of test power when it does not hold. Tables 1 and 2 show these rejection rates (in percentages) under normally distributed errors when $T = 100$ and $T = 200$, respectively; while Table 3 corresponds to the heavier-tailed t_3 setting when $T = 100$. The power results for the t-tests are based on FWER adjusted p-values.

Table 2
Empirical FWER and power of joint FRUH tests: Normal errors, $T = 200$.

ϕ	ρ	σ_h	Levels					Differences				
			t	$\mathcal{S}_{\min}$	$\mathcal{S}_{\times}$	$\mathcal{W}_{\min}$	$\mathcal{W}_{\times}$	t	$\mathcal{S}_{\min}$	$\mathcal{S}_{\times}$	$\mathcal{W}_{\min}$	$\mathcal{W}_{\times}$
Panel A: $\beta = 1$, $\gamma = 0$												
0.95	0.15	0.1	10.2	5.0	5.1	4.7	4.9	5.6	4.7	4.5	4.7	4.5
0.95	0.15	0.5	7.3	4.7	4.6	4.5	4.6	19.9	4.8	4.4	4.8	4.5
0.95	0.85	0.1	10.3	4.5	4.9	4.7	4.7	5.5	4.6	4.6	5.0	4.8
0.95	0.85	0.5	6.6	4.8	4.8	4.5	4.7	19.3	5.0	4.6	4.7	4.8
1	0.15	0.1	32.2	4.7	4.8	4.7	4.7	5.3	5.1	4.9	5.1	4.8
1	0.15	0.5	19.2	4.8	5.2	5.1	5.0	20.0	4.7	4.7	5.0	5.2
1	0.85	0.1	30.5	5.0	4.7	4.7	4.6	4.8	5.1	5.2	4.7	5.0
1	0.85	0.5	18.9	5.1	4.8	4.7	4.5	19.7	4.8	5.1	4.6	4.9
Panel B: $\beta = 0.95$, $\gamma = 0$												
0.95	0.15	0.1	60.0	20.3	24.0	27.8	35.6	4.6	4.8	4.6	4.9	4.9
0.95	0.15	0.5	69.9	63.4	67.9	63.0	66.9	5.4	7.2	7.5	7.4	7.4
0.95	0.85	0.1	58.1	19.4	23.1	25.8	32.8	4.5	5.0	5.1	4.9	4.8
0.95	0.85	0.5	71.9	62.5	66.0	62.3	65.7	5.5	6.7	6.8	7.0	7.3
1	0.15	0.1	99.6	91.4	95.1	98.8	99.8	85.8	78.5	82.2	88.0	90.7
1	0.15	0.5	90.8	96.8	98.1	97.4	98.5	22.2	80.4	82.0	81.0	82.1
1	0.85	0.1	99.5	91.1	94.8	99.0	99.6	84.7	77.7	80.9	86.9	89.5
1	0.85	0.5	91.3	97.1	98.0	97.5	98.2	22.0	80.2	81.2	80.0	81.3
Panel C: $\beta = 1.2$, $\gamma = -0.2$												
0.95	0.15	0.1	6.3	4.3	4.3	3.8	3.9	15.5	7.7	8.1	8.9	9.3
0.95	0.15	0.5	3.5	3.1	3.0	2.8	2.7	52.9	57.9	61.1	59.7	63.7
0.95	0.85	0.1	6.4	4.2	3.9	3.9	4.3	15.8	7.6	7.7	8.7	9.3
0.95	0.85	0.5	3.6	3.2	3.0	2.6	3.0	53.6	56.8	60.9	58.7	62.9
1	0.15	0.1	4.6	5.1	5.0	4.6	4.5	11.5	7.1	7.7	8.8	8.9
1	0.15	0.5	2.8	5.0	4.9	4.4	4.5	50.3	55.1	59.1	57.9	61.4
1	0.85	0.1	4.5	4.5	4.6	4.6	4.6	11.9	7.7	7.8	8.8	9.0
1	0.85	0.5	3.1	5.0	5.1	4.4	4.3	51.0	55.1	58.8	57.7	61.4

See notes of Table 1.

Note also that no results are explicitly reported in Tables 1–3 for the SS and SD p-value adjustment methods, since, as expected from (20), these have the same probability of rejecting at least one $H_{0,i}$ as their global inference counterparts based on the minimum p-value rule. To further evaluate the power properties of the SS and SD adjustment methods, we consider the *average power rate* (APR) defined as

$$\text{APR} = \frac{E\left(\text{number of } H_{0,i}\text{'s rejected}\right)}{\text{number of false } H_{0,i}\text{'s}}.$$

For normally distributed errors and $T = 200$, Table 4 reports the empirical APRs (in percentages) obtained from SS and SD adjustments to the p-values of the proposed $\mathcal{S}$ and $\mathcal{W}$ statistics. The main takeaways from these simulation experiments are summarized as follows:

1. From Panel A in Tables 1–3, we see that the FWER can be severely oversized with the Holm adjusted t-tests. When performed with the levels specification, the over-rejection problem becomes worse as ϕ increases towards 1, while it is mainly effected by increases in σ_h in the differences specification. Table 2 reveals that these t-tests continue to be oversized even as the sample size doubles to $T = 200$. Comparing Tables 1 and 3 shows that the situation remains essentially the same whether the errors are normally or t_3 distributed. In accordance with the theory, the empirical FWER of the sign and signed rank tests is close to the nominal 5% level, no matter the sample size, degree of tail heaviness, or the values of ϕ, ρ, and σ_h.
2. Panel B in Tables 1–3 makes clear that proceeding in levels delivers good power as the value of β moves away from 1 and $\gamma = 0$, while the differences tests lag behind. From Panel C, however, we see that the levels tests have no power when $E_t[s_{i,t+1}]$ is a linear combination of both $f_{i,t}$ and $s_{i,t}$. As explained in the introduction, this can be interpreted as a short-run deviation from the FRUH and we see that the differences tests have discriminatory power against such alternatives. In Panel B, power rises markedly across the board as ϕ increases towards 1; while in Panel C the differences approach garners power as σ_h increases.
3. The sign and signed rank tests tend to be more powerful than the t-tests as ϕ and σ_h increase (see Panels B and C in Tables 1–3) and as tail-heaviness increases (compare Tables 1 and 3). In fact, the power of the t-tests is seen to decline in Panel B when $\phi = 1$ and σ_h increases from 0.1 to 0.5. Increasing ρ from 0.15 to 0.85 appears to have little effect on test power. As expected, we see power improving as T grows from 100 in Table 1 to 200 in Table 2.
4. In Tables 1–3 we see the $\mathcal{W}$ statistics delivering better power than the $\mathcal{S}$ statistics, and that power gains are achieved by combining p-values using their product instead of just using the minimum p-value. Turning to Table 4, we observe an overall pattern of APRs mimicking the power patterns in Tables 1–3. The $\mathcal{W}$-based APRs tend to be higher than the $\mathcal{S}$-based ones, and the SD adjustments are seen to improve the APR relative to the SS adjustments.

Table 3
Empirical FWER and power of joint FRUH tests: t_3 errors, $T = 100$.

ϕ	ρ	σ_h	Levels					Differences				
			t	$\mathcal{S}_{\min}$	$\mathcal{S}_{\times}$	$\mathcal{W}_{\min}$	$\mathcal{W}_{\times}$	t	$\mathcal{S}_{\min}$	$\mathcal{S}_{\times}$	$\mathcal{W}_{\min}$	$\mathcal{W}_{\times}$
Panel A: $\beta = 1$, $\gamma = 0$												
0.95	0.15	0.1	12.8	5.3	5.4	5.0	5.0	4.9	4.7	5.1	4.7	4.9
0.95	0.15	0.5	8.5	5.0	5.2	5.1	5.1	10.4	4.6	5.0	4.6	5.0
0.95	0.85	0.1	12.2	4.7	4.6	4.6	4.7	5.2	5.0	5.1	4.9	4.9
0.95	0.85	0.5	8.5	4.4	4.4	4.7	4.5	10.0	4.7	4.7	4.8	4.5
1	0.15	0.1	30.2	5.6	5.3	5.1	5.3	4.2	4.8	4.8	4.8	5.2
1	0.15	0.5	20.7	5.2	5.3	4.9	5.2	10.3	5.4	5.2	4.9	4.6
1	0.85	0.1	28.8	4.9	4.6	4.3	4.0	4.6	4.6	4.7	5.1	5.2
1	0.85	0.5	20.5	4.9	4.6	4.6	4.9	10.2	5.1	4.7	4.5	4.7
Panel B: $\beta = 0.95$, $\gamma = 0$												
0.95	0.15	0.1	23.5	19.9	23.1	21.0	25.0	4.4	5.1	4.9	4.7	5.2
0.95	0.15	0.5	48.3	52.6	55.8	51.0	54.3	5.5	6.4	6.2	6.2	6.4
0.95	0.85	0.1	24.0	18.3	22.3	20.6	23.8	4.6	5.0	4.9	5.4	5.3
0.95	0.85	0.5	49.2	52.7	55.3	51.0	52.1	5.6	6.2	6.5	6.3	6.3
1	0.15	0.1	65.9	69.5	73.9	75.4	83.6	48.9	55.7	57.5	61.5	65.0
1	0.15	0.5	63.5	82.3	84.4	81.5	84.0	12.8	57.3	58.0	57.6	59.8
1	0.85	0.1	63.7	67.5	71.8	75.8	80.1	45.1	55.0	57.4	60.5	63.6
1	0.85	0.5	63.8	81.8	83.5	82.0	83.0	12.9	56.7	57.4	57.1	58.4
Panel C: $\beta = 1.2$, $\gamma = -0.2$												
0.95	0.15	0.1	5.5	4.4	4.6	3.9	3.9	10.2	7.7	8.1	7.8	8.7
0.95	0.15	0.5	3.6	3.9	3.8	3.3	2.9	44.5	42.8	46.2	42.9	45.9
0.95	0.85	0.1	5.4	3.8	4.0	3.5	3.7	9.6	7.7	8.5	8.2	8.6
0.95	0.85	0.5	4.1	3.5	3.6	2.9	2.8	44.7	43.1	46.2	42.5	45.9
1	0.15	0.1	4.8	5.3	5.2	4.8	5.1	7.9	7.0	7.5	7.3	7.9
1	0.15	0.5	3.5	5.7	5.2	5.2	5.2	42.1	40.2	42.9	40.7	42.7
1	0.85	0.1	4.9	4.6	4.7	4.4	3.9	6.7	7.2	7.3	7.6	8.3
1	0.85	0.5	3.7	4.8	4.6	5.0	4.5	41.7	40.4	43.1	40.1	43.1

See notes of Table 1.

Table 4
Empirical APR with multiplicity adjusted p-values: Normal errors, $T = 200$.

ϕ	ρ	σ_h	Levels				Differences			
			$\mathcal{S}_{\mathrm{SS}}$	$\mathcal{S}_{\mathrm{SD}}$	$\mathcal{W}_{\mathrm{SS}}$	$\mathcal{W}_{\mathrm{SD}}$	$\mathcal{S}_{\mathrm{SS}}$	$\mathcal{S}_{\mathrm{SD}}$	$\mathcal{W}_{\mathrm{SS}}$	$\mathcal{W}_{\mathrm{SD}}$
Panel A: $\beta = 0.95$, $\gamma = 0$										
0.95	0.15	0.1	10.8	11.5	15.1	16.4	2.4	2.5	2.5	2.6
0.95	0.15	0.5	38.9	42.1	38.5	41.6	3.6	3.7	3.7	3.8
0.95	0.85	0.1	10.3	11.2	14.5	15.7	2.5	2.6	2.5	2.6
0.95	0.85	0.5	38.9	42.0	38.3	41.6	3.4	3.5	3.6	3.7
1	0.15	0.1	63.3	69.0	76.9	83.7	50.4	54.0	60.4	65.1
1	0.15	0.5	77.5	81.3	78.4	82.4	53.0	55.8	53.8	56.7
1	0.85	0.1	62.9	68.2	77.1	84.0	50.7	54.8	61.4	65.7
1	0.85	0.5	77.8	81.5	78.2	82.4	53.5	56.5	53.6	56.5
Panel B: $\beta = 1.2$, $\gamma = -0.2$										
0.95	0.15	0.1	2.2	1.9	2.2	1.9	3.9	4.5	4.1	4.7
0.95	0.15	0.5	1.5	1.4	1.6	1.4	34.8	35.8	37.3	38.5
0.95	0.85	0.1	2.1	2.0	2.2	2.1	3.9	4.4	4.0	4.6
0.95	0.85	0.5	1.6	1.3	1.6	1.3	34.3	35.7	36.8	38.3
1	0.15	0.1	2.6	2.4	2.7	2.4	3.7	4.5	3.8	4.6
1	0.15	0.5	2.5	2.2	2.6	2.3	32.8	34.4	35.3	37.0
1	0.85	0.1	2.3	2.4	2.5	2.5	4.0	4.5	4.1	4.6
1	0.85	0.5	2.5	2.3	2.6	2.4	32.8	34.6	35.0	37.1

This table reports the empirical APRs (in percentages) from SS and SD multiplicity adjustments to the p-values of the $\mathcal{S}$ and $\mathcal{W}$ statistics given $\alpha = 5\%$ and $N = 2$, where $\{(s_{1,t}, f_{1,t}) : t = 1, \ldots, T\}$ and $\{(s_{2,t}, f_{2,t}) : t = 0.25T, \ldots, 0.75T\}$.

4. Empirical application

Our empirical application uses end-of-month series of spot and one-month forward exchange rates for $N = 13$ major currencies, each relative to the US dollar. The included currencies are the Australian dollar (AUD), Canadian dollar (CAD), Swiss franc (CHF), German deutschemark (DEM), Danish krone (DKK), European euro (EUR), French franc (FRF), British pound (GBP), Italian lira (ITL), Japanese yen (JPY), Norwegian krone (NOK), New Zealand dollar (NZD), and Swedish krona (SEK). Table 5 gives a complete list of the included countries along with the start and end dates for each currency's data series. Note that the number of observations

Table 5
Data overview.

Currency code	Country	Start date	End date	No. Obs.
AUD	Australia	Dec 1984	June 2021	439
CAD	Canada	Dec 1984	June 2021	439
CHF	Switzerland	Oct 1983	June 2021	453
DEM	Germany	Oct 1983	Dec 1998	183
DKK	Denmark	Dec 1984	June 2021	439
EUR	Euro area	Jan 1999	June 2021	270
FRF	France	Oct 1983	Dec 1998	183
GBP	United Kingdom	Oct 1983	June 2021	453
ITL	Italy	Mar 1984	Dec 1998	178
JPY	Japan	Oct 1983	June 2021	453
NOK	Norway	Dec 1984	June 2021	439
NZD	New Zealand	Dec 1984	June 2021	439
SEK	Sweden	Dec 1984	June 2021	439

Table 6
Marginal FRUH assessments.

Currency code	Levels			Differences		
	t	$\mathcal{S}_i$	$\mathcal{W}_i$	t	$\mathcal{S}_i$	$\mathcal{W}_i$
AUD	**4.6**	25.1	**3.4**	**1.5**	**0.2**	**0.2**
CAD	11.6	15.1	**2.8**	**4.1**	**1.6**	20.2
CHF	15.2	39.7	32.8	18.7	45.1	54.6
DEM	23.0	18.2	33.4	71.7	37.3	87.4
DKK	**0.0**	21.4	15.3	**2.0**	**0.3**	**0.0**
EUR	12.6	42.7	70.8	**3.7**	**4.4**	**3.5**
FRF	8.8	**0.3**	9.0	27.2	7.5	**2.3**
GBP	**4.9**	70.6	36.0	**0.4**	13.2	**4.0**
ITL	26.2	**1.9**	5.4	74.5	**2.9**	7.7
JPY	**0.6**	30.0	50.5	**2.2**	**3.0**	**3.3**
NOK	**3.3**	15.1	14.9	21.1	**2.1**	**2.1**
NZD	**0.5**	**0.1**	**0.3**	**0.0**	**0.1**	**0.1**
SEK	14.0	50.3	30.9	**4.2**	**0.1**	**0.0**

The reported results are the p-values (in percentages) of the t-test, the sign test and the Wilcoxon signed rank test for marginal FRUH assessments. The entries in bold are instances of statistical significance at the 5% level.

Table 7
Joint FRUH assessments.

Panel A: Multiplicity adjusted p-values

Currency code	Levels					Differences				
	t	$\mathcal{S}_{\mathrm{SS}}$	$\mathcal{S}_{\mathrm{SD}}$	$\mathcal{W}_{\mathrm{SS}}$	$\mathcal{W}_{\mathrm{SD}}$	t	$\mathcal{S}_{\mathrm{SS}}$	$\mathcal{S}_{\mathrm{SD}}$	$\mathcal{W}_{\mathrm{SS}}$	$\mathcal{W}_{\mathrm{SD}}$
AUD	42.2	93.0	80.2	25.4	21.6	16.5	**4.0**	**3.4**	**2.8**	**2.2**
CAD	69.8	77.0	71.7	21.9	20.5	29.9	19.0	13.6	91.0	48.8
CHF	69.8	99.4	84.7	95.7	82.0	93.5	100.0	62.6	100.0	77.4
DEM	69.8	84.6	72.9	96.3	82.0	100.0	99.6	62.6	100.0	87.2
DKK	**0.2**	89.1	75.9	69.7	54.6	20.6	**5.0**	**3.7**	**1.0**	**1.0**
EUR	69.8	99.5	84.7	100.0	82.0	29.9	40.4	19.2	33.7	23.0
FRF	62.1	**3.4**	**3.2**	50.3	38.6	93.5	61.0	27.6	24.4	17.7
GBP	42.2	100.0	84.7	97.6	82.0	5.3	80.7	35.7	37.7	23.0
ITL	69.8	19.3	17.1	35.4	27.8	100.0	31.1	18.0	57.7	26.8
JPY	7.3	97.5	82.5	99.6	82.0	20.6	32.1	18.0	32.8	23.0
NOK	33.7	77.0	71.7	68.6	54.6	93.5	24.0	15.7	21.7	17.4
NZD	6.6	**1.1**	**1.1**	**2.8**	**2.8**	**0.1**	**1.9**	**1.7**	**2.2**	**2.0**
SEK	69.8	99.8	84.7	94.9	82.0	29.9	**1.2**	**1.2**	**0.3**	**0.3**

Panel B: Combined p-values

	Levels				Differences			
	$\mathcal{S}_{\min}$	$\mathcal{S}_{\times}$	$\mathcal{W}_{\min}$	$\mathcal{W}_{\times}$	$\mathcal{S}_{\min}$	$\mathcal{S}_{\times}$	$\mathcal{W}_{\min}$	$\mathcal{W}_{\times}$
All currencies	**1.1**	**2.0**	**2.8**	**4.9**	**1.2**	**0.1**	**0.3**	**0.1**

This table reports the p-values (in percentages) for joint FRUH inferences across currencies. Panel A shows multiplicity adjusted p-values, which includes the Holm adjustments for the t-tests; and SS and SD adjustments for the sign and Wilcoxon signed rank tests. Panel B reports the combined p-values based on the minimum and product rules. Entries in bold are instances of statistical significance at the 5% level.

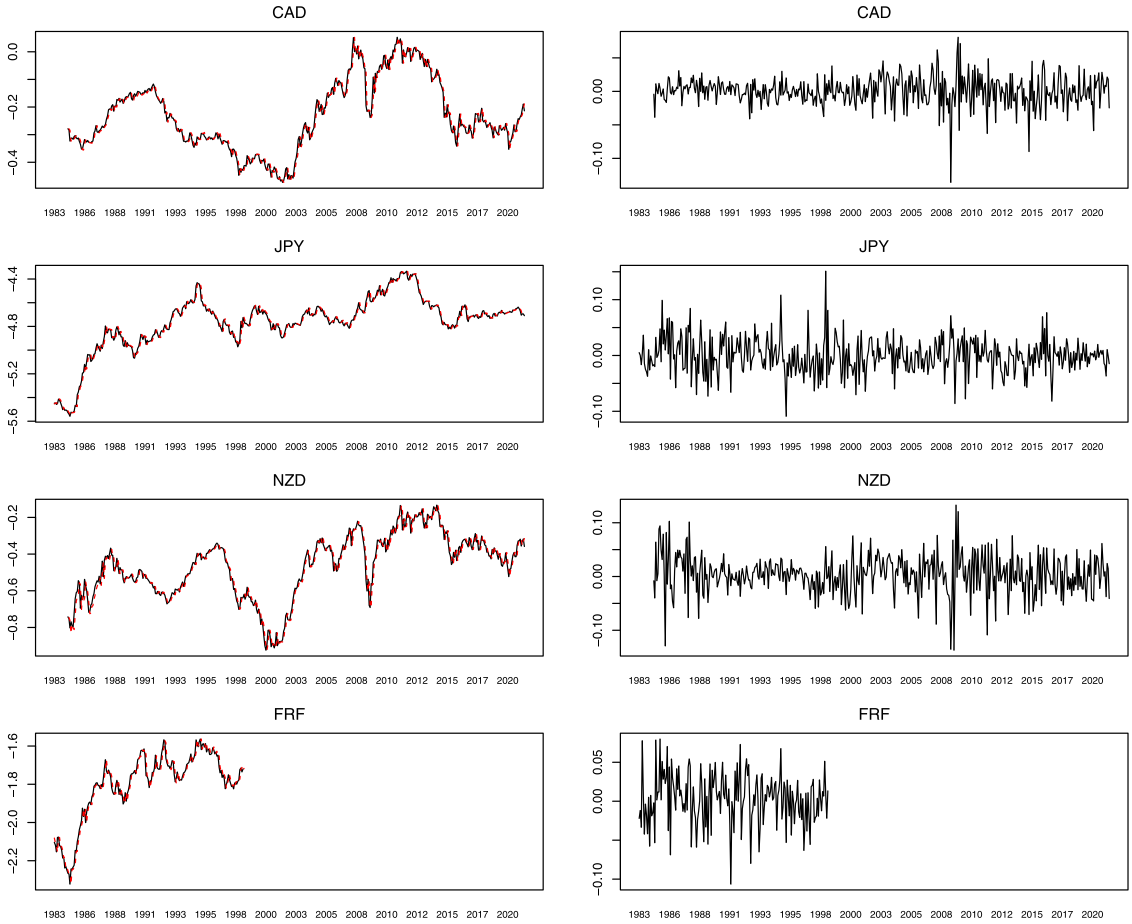

Fig. 1. Monthly exchange rate data for the Canadian dollar, Japanese yen, New Zealand dollar, and French franc (prior to the euro) all relative to the US dollar. For each currency, the plots in the left column show $s_{i,t+1}$ (solid black line) along with $f_{i,t}$ (dashed red line), while the plots in the right column show $s_{i,t+1} - f_{i,t}$.

varies across currencies, owing to differences in the time spans covered by each currency. The number of data points varies from a low of 178 with ITL to a high of 453 with CHF, GBP, and JPY. We obtained these data via the Datastream platform.

Fig. 1 illustrates the salient features of the data with CAD, JPY, NZD, and FRF as examples. For each of these currencies, the plots in the left column show $s_{i,t+1}$, the log spot rate (solid black line), and $f_{i,t}$, the lagged value of the log forward rate (dashed red line). In each case we see the two rates tracking each other very closely, in accordance with the FRUH.[10] The right column of Fig. 1 shows the difference between the log spot rate and the lagged value of the log forward rate; i.e., $s_{i,t+1} - f_{i,t}$, which corresponds to $u_{i,t+1}$ in (6) when the FRUH holds true. We see episodes of increased volatility and prominent outliers in these series, which is consistent with the presence of time-dependent conditional variances. Note also that since the French franc was replaced by the euro, the FRF data are missing from 1999 onwards.

We begin our assessment by testing the FRUH with each currency taken one at a time, thereby ignoring the dependencies between currencies. Table 6 reports the marginal p-values (in percentages) of the standard t-test, the sign test $\mathcal{S}_i$ and the Wilcoxon signed rank test $\mathcal{W}_i$, for a single currency i. We see that the FRUH holds unambiguously in the levels and differences specifications only with CHF and DEM. At the other extreme, the FRUH is convincingly rejected in the case of NZD. In the 10 remaining cases, the decision to reject or not each individual-currency FRUH at the conventional 5% level varies wildly between tests and specifications. There does however tend to be more rejections occurring with the differences specifications. In particular, the $\mathcal{S}_i$ and $\mathcal{W}_i$ tests agree on non-rejections only in the levels specifications for DKK, EUR, GBP, JPY, NOK, and SEK.

A more coherent picture emerges from Table 7, which shows the results of the joint FRUH assessments. For each currency, Panel A reports the SS and SD multiplicity adjusted p-values of the $\mathcal{S}$ and $\mathcal{W}$ statistics; along with the Holm adjusted p-values $p^{t}_{\pi_i,\mathrm{Holm}}$ of the standard t-statistics, given in (21).[11] Panel B reports the combined p-values obtained with the minimum and product rules. Here the MC p-values were computed using $B - 1 = 999$ resampling draws.

[10] The lines are so close together that the reader might need to zoom in to tell them apart.

[11] In addition to Holm's method, we also computed Šidák adjusted p-values (see, e.g., Ge et al., 2003, §3.2) and obtained essentially the same results.

From the combined p-values in Panel B, the joint FRUH appears massively rejected across the board. A closer examination of Panel A reveals the sources of those rejections.[12] We see that NZD continues to be the currency with the most egregious p-values, strongly rejecting the FRUH in both levels and differences. After that, the rejections from the sign and signed rank tests occur with the differences specifications for AUD, DKK, and SEK, as in Table 6. For the 9 other currencies (CAD, CHF, DEM, EUR, FRF, GBP, ITL, JPY, NOK), however, the sign and signed rank tests now show support for the FRUH with p-values well above the traditional 5% FWER cutoff. An exception is FRF for which $\mathcal{S}_{\text{SS}}$ and $\mathcal{S}_{\text{SD}}$ indicate rejections while all the other statistics do not.

As expected, the $\mathcal{S}_{\min}$ ($\mathcal{W}_{\min}$) p-value in Panel B is identical to the minimum of the corresponding $\mathcal{S}_{\text{SS}}$ ($\mathcal{W}_{\text{SS}}$) p-values in Panel A; and all the SD multiplicity adjusted p-values are uniformly smaller or equal to their SS counterparts. Note also the quite general agreement between the $\mathcal{S}$- and $\mathcal{W}$-based inferences. We know that inference based on the standard t-tests can be misleading even after adjusting their p-values for the multiplicity effect, as seen in the simulation experiments. In light of this, the fairly broad agreement in Table 7 between inferences based on the sign and signed rank tests and the t-tests is all the more interesting. The striking contrast between Tables 6 and 7 illustrates the distinction between marginal and joint inference about the FRUH.

5. Concluding remarks

In this paper we have described how distribution-free Monte Carlo resampling can be used to ensure the simultaneous correctness of a set of FRUH inferences. Our approach can be applied with specifications in levels or differences, and proceeds with sign and signed rank tests for each currency. A resampling scheme is used to control the overall Type I error rate of either a global FRUH test obtained via combinations of the marginal p-values, or individual FRUH tests using multiplicity adjusted p-values. This resampling proceeds conditional on the absolute values of the error terms, since only their signs are randomized.

The Lehmann and Stein (1949) impossibility theorem shows that such sign-based tests are the *only* ones that yield valid inference in the presence of non-normalities and heteroskedasticity of unknown form; see also Dufour (2003) for more on this point. Another appealing feature of the proposed test procedures is that they allow for joint FRUH assessments with unbalanced panels comprising currencies observed over differing time periods.

Of course, the test procedures developed in this paper rest on certain auxiliary statistical assumptions and these are maintained along with the FRUH under the null. Indeed, as Meese (1989, p. 157) states:

> Modern empirical work recognizes that any test concerning the behavior of exchange rate risk premiums is necessarily a joint hypothesis test of an equilibrium model of exchange risk and return, an assumption about expectations formation, and a set of auxiliary statistical assumptions under which formal inference proceeds.

This means in the present context that the test procedures provide joint assessments of the FRUH as predicted under rational expectations and risk neutrality, along with the maintained zero median and symmetry assumptions for the error terms.

CRediT authorship contribution statement

Hsuan Fu: Conceptualization, Validation, Data curation, Writing – review & editing. **Richard Luger:** Conceptualization, Methodology, Software, Writing – original draft.

References

Anatolyev, S., Gospodinov, N., Jamali, I., Liu, X., 2017. Foreign exchange predictability and the carry trade: A decomposition approach. J. Empir. Financ. 42, 199–211.
Bailey, R., Baillie, R., McMahon, P., 1984. Interpreting econometric evidence on efficiency in the foreign exchange market. Oxf. Econ. Pap. 36 (1), 67–85.
Baillie, R., Bollerslev, T., 1989. Common stochastic trends in a system of exchange rates. J. Finance 44 (1), 167–181.
Barnard, G., 1963. Comment on 'The spectral analysis of point processes' by M.S. Bartlett. J. R. Stat. Soc. Ser. B Stat. Methodol. 25 (2), 294.
Barnhart, S., Szakmary, A., 1991. Testing the unbiased forward rate hypothesis: Evidence on unit roots, co-integration, and stochastic coefficients. J. Financ. Quant. Anal. 26 (2), 245–267.
Bilson, J., 1981. The "speculative efficiency" hypothesis. J. Bus. 54 (3), 435–451.
Birnbaum, Z., 1974. Computers and unconventional test statistics. In: Proschan, F., Serfling, R. (Eds.), Reliability and Biometry. SIAM, Philadelphia, pp. 441–458.
Burnside, C., Rebelo, S., Eichenbaum, M., 2008. Carry trade: The gains of diversification. J. Eur. Econom. Assoc. 6 (2/3), 581–588.
Campbell, B., Dufour, J.-M., 1997. Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter. Internat. Econom. Rev. 38 (1), 151–173.
Cenedese, G., Sarno, L., Tsiakas, I., 2014. Foreign exchange risk and the predictability of carry trade returns. J. Bank. Financ. 42, 302–313.
Chib, S., Omori, Y., Asai, M., 2009. Multivariate stochastic volatility. In: Andersen, T., Davis, R., Kreiss, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer-Verlag, Berlin, pp. 365–400.
Cornell, B., 1977. Spot rates, forward rates and exchange market efficiency. J. Financ. Econ. 5 (1), 55–65.
Cornell, B., 1989. The impact of data errors on measurement of the foreign exchange risk premium. J. Int. Money Finance 8 (1), 147–157.
Delcoure, N., Barkoulas, J., Baum, C., Chakraborty, A., 2003. The forward rate unbiasedness hypothesis reexamined: Evidence from a new test. Glob. Finance J. 14 (1), 83–93.
Dufour, J.-M., 2003. Identification, weak instruments, and statistical inference in econometrics. Can. J. Econ. 36 (4), 767–808.
Dufour, J.-M., 2006. Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. J. Econometrics 133 (2), 443–477.

[12] Such FRUH rejections are consistent with the predictability of carry trade returns found in a number of studies (e.g., Cenedese et al., 2014; Anatolyev et al., 2017).

Dufour, J.-M., Khalaf, L., 2001. Monte Carlo test methods in econometrics. In: Baltagi, B. (Ed.), A Companion to Theoretical Econometrics. Basil Blackwell, Oxford, UK, pp. 494–510.
Dufour, J.-M., Khalaf, L., Bernard, J.-T., Genest, I., 2004. Simulation-based finite-sample tests for heteroskedasticity and ARCH effects. J. Econometrics 122 (2), 317–347.
Dufour, J.-M., Khalaf, L., Voia, M., 2015. Finite-sample resampling-based combined hypothesis tests, with applications to serial correlation and predictability. Comm. Statist. Simulation Comput. 44 (9), 2329–2347.
Dwass, M., 1957. Modified randomization tests for nonparametric hypotheses. Ann. Math. Stat. 28 (1), 181–187.
Evans, M., Lewis, K., 1995. Do long-term swings in the dollar affect estimates of the risk premia? Rev. Financ. Stud. 8 (3), 709–742.
Fama, E., 1984. Forward and spot exchange rates. J. Monetary Econ. 14 (3), 319–338.
Fisher, R., 1932. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.
Folks, J., 1984. Combination of independent tests. In: Krishnaiah, P., Sen, P. (Eds.), Handbook of Statistics 4: Nonparametric Methods. North-Holland, Amsterdam, pp. 113–121.
Frenkel, J., 1976. A monetary approach to the exchange rate: Doctrinal aspects and empirical evidence. Scand. J. Econ. 78 (2), 200–224.
Froot, K., Frankel, J., 1989. Forward discount bias: Is it an exchange risk premium? Q. J. Econ. 104 (1), 139–161.
Ge, Y., Dudoit, S., Speed, T., 2003. Resampling-based multiple testing for microarray data analysis. Test 12 (1), 1–77.
Geweke, J., Feige, E., 1979. Some joint tests of the efficiency of markets for forward foreign exchange. Rev. Econ. Stat. 61 (3), 334–341.
Goodhart, C., McMahon, P., Ngama, Y., 1997. Why does the spot-forward discount fail to predict changes in future spot rates? Int. J. Finance Econ. 2 (2), 121–129.
Gradojevic, N., Erdemlioglu, D., Gençay, R., 2020. A new wavelet-based ultra-high-frequency analysis of triangular currency arbitrage. Econ. Model. 85, 57–73.
Gungor, S., Luger, R., 2015. Bootstrap tests of mean-variance efficiency with multiple portfolio groupings. L'Actualité Économique 91 (1/2), 35–65, Special issue in English on *Identification, Simulation, and Finite-Sample Inference*.
Gungor, S., Luger, R., 2020. Small-sample tests for stock return predictability with possibly non-stationary regressors and GARCH-type effects. J. Econometrics 218 (2), 750–770.
Hai, W., Mark, N., Wu, Y., 1997. Understanding spot and forward exchange rate regressions. J. Appl. Econometrics 12 (6), 715–734.
Hakkio, C., Rush, M., 1989. Market efficiency and cointegration: An application to the sterling and deutschemark exchange markets. J. Int. Money Finance 8 (1), 75–88.
Ho, T.-W., 2002. The forward rate unbiasedness hypothesis revisited. Appl. Financial Econ. 12 (11), 799–804.
Hochberg, Y., Tamhane, A., 1987. Multiple Comparison Procedures. John Wiley & Sons, New York.
Hodgson, D., Linton, O., Vorkink, K., 2004. Testing forward exchange rate unbiasedness efficiency: A semiparametric approach. J. Appl. Econ. 7 (1), 325–353.
Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6 (2), 65–70.
Hsu, J., 1996. Multiple Comparisons: Theory and Methods. Chapman & Hall/CRC, Boca Raton.
Kiviet, J., 2012. Monte Carlo simulation for econometricians. Found. Trends Econom. 5 (1/2), 1–181.
Lehmann, E., Stein, C., 1949. On the theory of some non-parametric hypotheses. Ann. Math. Stat. 20 (1), 28–45.
Levitch, R., 1979. On the efficiency of markets for foreign exchange. In: Dornbusch, R., Frenkel, J. (Eds.), International Economic Policy: Theory and Evidence. Johns Hopkins University Press, Baltimore, pp. 246–267.
Liu, W., Maynard, A., 2005. Testing forward rate unbiasedness allowing for persistent regressors. J. Empir. Financ. 12 (5), 613–628.
Maynard, A., 2003. Testing for forward-rate unbiasedness: On regression in levels and in returns. Rev. Econ. Stat. 85 (2), 313–327.
Maynard, A., 2006. The forward premium anomaly: statistical artefact or economic puzzle? New evidence from robust tests. Can. J. Econ. 39 (4), 1244–1281.
Maynard, A., Phillips, P., 2001. Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly. J. Appl. Econometrics 16 (6), 671–708.
McMillan, D., 2005. Cointegrating behaviour between spot and forward exchange rates. Appl. Financial Econ. 15 (16), 1135–1144.
Meese, R., 1989. Empirical assessment of foreign currency risk premiums. In: Stone, C. (Ed.), Financial Risk: Theory, Evidence and Implications. Springer, Dordrecht, pp. 157–180.
Meese, R., Singleton, K., 1982. On unit roots and the empirical modeling of exchange rates. J. Finance 37 (4), 1029–1035.
Pearson, K., 1933. On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. Biometrika 25 (3/4), 379–410.
Phillips, P., McFarland, J., 1997. Forward exchange market unbiasedness: The case of the Australian dollar since 1984. J. Int. Money Finance 16 (6), 885–907.
Randles, R., Wolfe, D., 1979. Introduction to the Theory of Nonparametric Statistics. Wiley, New York.
Serfling, R., 2006. Multivariate symmetry and asymmetry. In: Kotz, S., Balakrishnan, N., Read, C., Vidakovic, B. (Eds.), Encyclopedia of Statistical Sciences, second ed. Wiley-Interscience, Hoboken, N.J., pp. 5338–5345.
Silvennoinen, A., Teräsvirta, T., 2009. Multivariate GARCH models. In: Andersen, T., Davis, R., Kreiss, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer-Verlag, Berlin, pp. 201–229.
Small, C., 1990. A survey of multidimensional medians. Internat. Statist. Rev. 58 (3), 263–277.
Tippett, L., 1931. The Methods of Statistics. Williams and Norgate, London.
Westerlund, J., 2007. Estimating cointegrated panels with common factors and the forward rate unbiasedness hypothesis. J. Financ. Econom. 5 (3), 491–522.
Westfall, P., Young, S., 1993. Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. Wiley, New York.
Wilkinson, B., 1951. A statistical consideration in psychological research. Psychol. Bull. 48 (2), 156–158.
Zivot, E., 2000. Cointegration and forward and spot exchange rate regressions. J. Int. Money Finance 19 (6), 785–812.