Journal of Econometrics 171 (2012) 86–97

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

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

Econometric analysis of present value models when the discount factor is near one

Kenneth D. West *

University of Wisconsin, United States

ARTICLE INFO

Article history:
Received 1 July 2011
Received in revised form
9 April 2012
Accepted 13 July 2012
Available online 20 July 2012

JEL classification:
C58
F31
F37
G12
G15
G17

Keywords:
Uncovered interest parity
Asset price
Exchange rate
Forward rate
Forward premium anomaly
Expected return
Random walk
Efficient markets
Martingale model

ABSTRACT

This paper develops asymptotic econometric theory to help understand data generated by a present value model with a discount factor near one. A leading application is to exchange rate models. A key assumption of the asymptotic theory is that the discount factor approaches one as the sample size grows. The finite sample approximation implied by the asymptotic theory is quantitatively congruent with the modest departures from random walk behavior that are typically found and with imprecise estimation of a well-studied regression relating spot and forward exchange rates.

1. Introduction

This paper develops and evaluates, via simulations and analytical calculations, asymptotic econometric theory intended to better explain the behavior of data generated by a present value model with a discount factor near one. A leading area of application is floating exchange rates. The research is motivated by two stylized facts that are well known in the asset pricing literature.

The first stylized fact is that it is difficult to document predictability in changes in asset prices such as floating exchange rates or equity prices, and that the bulk of the modest evidence we have of predictability comes from cross-correlations of asset price changes with other variables rather than univariate autocorrelations of asset price changes. Random walk like behavior in asset prices is *not* necessarily an implication of an efficient markets model such as the classic one exposited by Samuelson (1965). For example, in the exchange rate context, that efficient markets model, which assumes risk neutrality, requires that forward rates be unbiased and efficient predictors of spot rates. Yet forward rates are dominated by lagged spot rates as predictors of exchange rates. The classic reference on predicting exchange rates with forward rates and other variables is Meese and Rogoff (1983); recent updates are Engel et al. (2007) and Molodtsova and Papell (2009). Difficulties in prediction perhaps are less striking for equities than for exchange rates but are still notable; see Campbell and Thompson (2008) and Goyal and Welch (2008) for recent studies.

The second stylized fact pertains to exchange rates but not stock prices. It concerns the pattern of signs in regressions of levels or changes in exchange rates on forward rates on forward premia. (Here and throughout, exchange rates and forward rates are measured in logs.) In particular, the regression of the exchange rate change on the previous period's forward premium (i.e., the difference between the previous period's forward and actual exchange rates) yields a coefficient estimate that is sensitive to sample and to currency, but with tendency to be less than one. To

* Tel.: +1 6082632516; fax: +1 6082622033.
E-mail address: kdwest@wisc.edu.

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.07.002

Table 1
Forward-spot regressions, US dollar.

	Canada	France	Germany	Italy	Japan	UK
(1) $\hat{\beta}_r$	−0.204	0.486	−0.767	1.697	−2.703	−1.433
S.e.	(0.774)	(0.864)	(0.818)	(0.806)	(0.918)	(0.925)
(2) $\hat{\beta}_f$	0.982	0.927	0.976	0.939	0.986	0.904
S.e.	(0.025)	(0.034)	(0.027)	(0.030)	(0.025)	(0.036)
(3) $\hat{\beta}_m$	1.159	1.149	1.054	1.239	1.016	1.144
S.e.	(0.089)	(0.115)	(0.092)	(0.108)	(0.094)	(0.092)
Sample period	79:2–11:1	79:2–98:4	79:2–11:1	79:2–98:4	79:2–11:1	79:2–11:1

Notes:
1. Data are quarterly. Three-month interest rates and end of quarter bilateral US dollar exchange rates 1979–2000 were obtained from Chinn and Meredith (2004), extended to 2001–2011 using LIBOR rates obtained from economagic.com and Bloomberg and end of quarter exchange rates obtained from International Financial Statistics. Forward rates were constructed using covered interest parity.
2. Let y_t be the log of the exchange rate, f_t the log of the three-month-ahead forward rate. Least squares estimates and standard errors of the following regressions are reported: Δy_t = const. + $\hat{\beta}_r(f_{t-1} - y_{t-1})$ + residual; y_t = const. + $\hat{\beta}_f f_{t-1}$ + residual; $y_t - y_{t-2}$ = const. + $\hat{\beta}_m(f_{t-1} - y_{t-2})$ + residual.

illustrate, consider Table 1, which reports three related regressions using quarterly bilateral US dollar exchange rates against the other six G7 countries. The first line of the table reports estimates of the regression of the exchange rate change on the previous quarter's forward premium. Five of the six estimates are below one, and four of the six are negative; the estimates range from −2.703 to 1.697. A value of one would obtain if forward rates were unbiased predictors of spot rates. This value is reflected in the second line of the table, which reports estimates of the regression of the level of the exchange rate on the level of the previous quarter's forward rate.[1] The coefficient estimates are near one. Since the first regression is a transformation of the second regression obtained by subtracting the lag of the exchange rate from the left-hand and right-hand sides, a coefficient of one would still obtain, at least in population, if the forward rate were an unbiased predictor of the spot rate. (The third regression will be discussed below.) One class of explanations for a value less than one in regressions such as those reported in the first line of Table 1 turns on time varying risk premia. Hansen and Hodrick (1983) and Fama (1984) are early papers on this topic; Verdelhan (2010) is a recent contribution. Other explanations for a value less than one turn on expectational or informational biases (e.g., Frankel and Froot, 1987, Bacchetta and van Wincoop, 2006) or on liquidity, transaction or portfolio costs (e.g., Bacchetta and van Wincoop, 2010).

To my knowledge, a systematic econometric theory to rationalize the first stylized fact – that is, to rationalize our findings of modest evidence against the random walk – has not been proposed. Among the possible econometric explanations for the second stylized fact is that the regression produces biased estimates. Indeed, the fact that the regression cannot be reliably used to predict exchange rate changes (the first stylized fact) suggests that the point estimates should not be taken at face value. Maynard and Phillips (2001), building on Baillie and Bollerslev (1994, 2000), develop a theory in which fractional integration of the forward premium leads to bias in the point estimate, even asymptotically. A number of authors have argued that, with samples of size available in practice, the regression produces estimates that are biased downward relative to ones predicted by asymptotic theory. Evans and Lewis (1995) present analytical arguments for bias that turn on small-probability events not being realized in the samples that we observe. They and other authors have found downward biases in simulations.

This paper offers econometric theory to help rationalize both stylized facts. Its starting point is Engel and West (2005). That paper shows that, in present value models with constant discount factors, random walk like behavior is expected in the population as the discount factor approaches unity. The present paper develops econometric theory motivated by that analytical result. I derive an asymptotic approximation that, given a DGP, allows one to determine how close to a random walk an asset price will behave in a finite sample: under this approximation, hypothesis tests of zero predictability of asset price changes have size greater than the nominal size that would apply for a pure random walk, but less than the unit size that would apply under the usual asymptotics. In congruence with the first stylized fact, calibration of exchange rate and stock price data sets implies size barely above nominal size for univariate autocorrelations, and modestly above nominal size for tests of cross-correlation.

This econometric theory also leads to a reinterpretation of the regression that is at the heart of the second stylized fact (a regression that, incidentally, is not discussed in Engel and West, 2005). According to the theory developed here, the regression of the exchange rate change on the previous period's forward premium is inconsistent: even if the forward rate is the expectation of the exchange rate, and even with large sample sizes, estimates of the slope coefficient will not have high probability of being clustered near one. Hence relatively little weight should be put on the estimates of this regression. Finally, the theory suggests that a closely related regression – that of the two-period change in the exchange rate on the difference between the previous period's forward rate and the two-period lagged exchange rate – should produce a coefficient of one. Some of the explanations of the second stylized fact given above imply that this coefficient will be less than one. A tendency of this regression to yield a coefficient near one has not achieved the status of a stylized fact. But this indeed may be the case, as is indicated by the estimates in the third line of Table 1 (which range from 1.054 to 1.239) and by the similar values reported for a different sample by McCallum (1994).

While this paper is motivated by results in the stock price and especially the exchange rate literature, the analytical results are derived for a general present value model. Hence they may be of interest in interpretation of results for other asset prices. For that reason, much of the exposition below refers to "the asset price".

It should be noted that the paper assumes throughout that risk premia are constant (a zero risk premium is a special case), and abstracts as well from possible contributions from informational biases or liquidity or transactions costs. On the roles of such factors for exchange rates, see the references above; for stock returns, see, e.g., Campbell et al. (1997) or Barberis and Thaler (2003). Whatever those roles may be for asset prices and returns, my aim in the present paper is to supply a relatively clean analysis of behavior

[1] Since the exchange and forward rate arguably are $I(1)$ variables, the standard errors should be interpreted with caution. Another reason for caution in interpreting all the point estimates in the table is that, as stated in the notes to the table, forward rates were constructed using covered interest parity, and thus likely depart slightly from measured forward rates (Bhar et al., 2004). I include this table not to present new empirical results, but to focus the econometric discussion.

that is implied in a relatively simple setting. I defer to future work consideration of the effects of such complications.

Section 2 outlines the model. Section 3 develops asymptotic theory for correlations and predictability (the first stylized fact). Section 4 develops asymptotic theory for forward-spot regressions (the second stylized fact). Section 5 presents preliminary Monte Carlo results. Section 6 concludes. The Appendix contains the proofs.

2. Model and assumptions

2.1. The model

The model is

$$y_t = c_y + (1-b)\sum_{j=0}^{\infty} b^j E_t x_{t+j}. \tag{2.1}$$

In (2.1), y_t is the scalar (log) asset price (exchange rate, stock price); c_y is a constant that will not play a role in the analysis; b is a discount factor $0 < b < 1$; x_t is a scalar; E_t is mathematical expectations, assumed equivalent to linear projections. In the monetary model of the exchange rate, x_t is a linear combination of foreign and domestic money supplies, output levels, money demand and purchasing power parity shocks; b depends on the interest semielasticity of money demand; c_y is proportional to a constant (possibly zero) risk premium. In Campbell and Shiller's (1989) log-linearized stock price model, with constant expected returns, x_t is log dividends and c_y and b depend on the average log dividend–price ratio.[2] Observe that, even with constant expected returns, y_t does not, in general, follow a random walk: insofar as Δx_t is predictable, then Δy_t will be predictable as well, with predictable movements in Δy_t offsetting predictable movements in Δx_t to keep the total return (capital gain plus dividend) unpredictable.

I note that the results presented here are applicable to data generated by a present value model in which the factor $(1-b)$ is omitted, i.e., the model $\tilde{y}_t = \text{constant} + \sum_{j=0}^{\infty} b^j E_t x_{t+j}$. This is discussed in Section 4.[3]

2.2. Technical assumptions

I begin by spelling out assumptions about the data generating process. These assumptions allow (but do not require) the following: forecasts of x_t that are made using private data not observable to the econometrician; unobservable fundamentals (x_t is a linear combination of variables, some of which are unobservable); and complex dynamics (e.g., data not following a finite parameter ARMA process). Technical statements of assumptions now follow; readers interested in results but not proofs are advised that they can skip to Section 2.3 with no loss of ability to follow the rest of the text. Formally, assume that there is an $n \times 1$ vector of variables w_t used to forecast x_t. The vector w_t at a minimum includes x_t. It may include additional observable variables as well as variables used by private agents that are not observable to the econometrician. The first difference of this vector is stationary, with mean $E\Delta w_t$, an $n \times 1$ Wold innovation e_t and Wold representation

$$\Delta w_t = E\Delta w_t + \Psi(L)e_t, \quad \Psi(L)e_t = \sum_{j=0}^{\infty}\Psi_j e_{t-j}, \tag{2.2}$$

$$Ee_t e_s' = 0_{n\times n} \quad \text{for } t \neq s,\ Ee_t e_t' \text{ a constant matrix of rank } n;$$
$$\Delta x_t = \alpha'\Delta w_t \quad \text{for a suitable } (n\times 1) \text{ vector } \alpha,\ \alpha'\Psi(1) \neq 0.$$

Then, in (2.1), $E_t x_{t+j}$ is defined as

$$E_t x_{t+j} = x_t + E(\Delta x_{t+1} + \Delta x_{t+2} + \cdots \Delta x_{t+j} | 1, e_t, e_{t-1}, e_{t-2}, \ldots). \tag{2.3}$$

The condition $\alpha'\Psi(1) \neq 0$ says that x_t is $I(1)$, i.e., Δx_t is not an overdifferenced stationary process. I allow $\Psi(1)$ to have deficient rank: there may be cointegration across the variables used to forecast x_t, and there may be some overdifferenced stationary variables in Δw_t.

For a matrix A, let "$|\cdot|$" denote maximum element, $\|\cdot\|$ denote Euclidean norm. I assume either

$$\text{(i) } \sum_{j=0}^{\infty} j\|\Psi_j\| < \infty. \tag{2.4a}$$
(ii) e_t i.i.d. with finite fourth moments; or

$$\begin{aligned}&\text{(i) } \Delta w_t \text{ is fourth-order stationary and follows a stationary}\\ &\quad \text{finite-order ARMA process with}\\ &\quad E(e_t \mid 1, e_t, e_{t-1}, e_{t-2}, \ldots) = 0.\\ &\text{(ii) Let } q_t = (\Delta w_t', e_t')'. \text{ Then } q_t \text{ is strong mixing, with}\\ &\quad E\|q_t\|^{4\nu} < \infty \text{ for some } \nu > 4, \text{ and with mixing coefficients}\\ &\quad \text{of size } \nu/(\nu-1).\end{aligned} \tag{2.4b}$$

Eqs. (2.4a) and (2.4b) each rule out fractionally integrated processes, which have received some attention in empirical work on asset prices (e.g., Aloy et al., 2010 is a relatively recent example). The assumption of fourth-order stationarity is for convenience; moment drift could be accommodated, at the expense of more complicated notation.

2.3. The model, continued

We shall need to split Δy_t into three parts: (1) a deterministic component, denoted μ, (2) an unpredictable component, denoted η_t, and (3) a zero-mean component, denoted $(1-b)z_{t-1}$, that, in general, will be predictable. The deterministic term μ will not play an interesting role in the analysis, but is included because some asset prices (such as stock prices) do drift (upwards) over time; the focus is on the second and third components η_t and z_{t-1}. Specifically, define

$$\mu = E\Delta x_t, \qquad \eta_t = (\Delta y_t - E_{t-1}\Delta y_t),$$
$$z_{t-1} = \sum_{j=0}^{\infty} b^j E_{t-1}(\Delta x_{t+j} - \mu). \tag{2.5}$$

Using the definitions in (2.5), we can first difference and then rearrange (2.1) as

$$\begin{aligned}\Delta y_t &= (\Delta y_t - E_{t-1}\Delta y_t) + E_{t-1}\Delta y_t\\ &= \sum_{j=0}^{\infty} b^j (E_t - E_{t-1})\Delta x_{t+j}\\ &\quad + \left[\mu + (1-b)\sum_{j=0}^{\infty} b^j E_{t-1}(\Delta x_{t+j} - \mu)\right]\\ &\equiv \eta_t + [\mu + (1-b)z_{t-1}].\end{aligned} \tag{2.6}$$

[2] Campbell and Shiller (1989) consider cum-dividend stock prices. If the stock price is, instead, ex-dividend, as in Campbell et al. (1997, p. 264), then the summation begins at $t+1$ instead of t and the first term that is discounted is x_{t+2} rather than x_{t+1}. Modulo straightforward changes in dating, the analysis about to be presented can be adapted to ex-dividend pricing as well.

[3] The analysis does not, however, immediately extend to models in which $y_t = \text{const.} + (1-b)\sum_{j=0}^{\infty} b^j E_t x_{t+j} + \sum_{j=0}^{\infty} b^j E_t \theta_{t+j}$, where $x_t \sim I(1)$ and θ_t is stationary. This is the form of some asset pricing models in which θ_t is a time varying risk premium.

Table 2
Basic variables and parameters.

Variable/parameter	Equation	Description
x_t	(2.1)	$I(1)$ variable in present value determining log asset price y_t
$\Delta y_t, y_t$	(2.1)	Asset return, log asset price
z_t	(2.5)	Stationary fundamental that is proportional to the predictable component of $\Delta y_t - E\Delta y_t$
η_t	(2.5)	Innovation in Δy_t and y_t
$\Delta y_t^*, y_t^*, z_t^*, \eta_t^*$	(2.7), (4.1)	Random walk return, asset price, fundamental, innovation, obtained by setting $b = 1$ in solutions for $\Delta y_t, y_t, z_t, \eta_t$ that result when $b < 1$
f_{t-1}	(4.1)	$E_{t-1}y_t$
b	(2.1)	Discount factor in present value
δ	(2.8)	Parameter determining how close discount factor is to unity
μ	(2.5)	$E\Delta x_t$

Note: These are the descriptions of variables and parameters in the formal analysis in Section 2 through Section 4 of the paper.

Consider the random variables that results by setting $b = 1$ in the definitions of η_t and z_t. Write the results as

$$\Delta y_t^* = \eta_t^* + \mu, \quad \eta_t^* \equiv \sum_{j=0}^{\infty}(E_t - E_{t-1})\Delta x_{t+j},$$
$$z_t^* \equiv \sum_{j=0}^{\infty} E_t(\Delta x_{t+j+1} - \mu). \tag{2.7}$$

Assumption (2.4) ensures that η_t^* and z_t^* are stationary. For convenience of exposition, I refer to Δy_t^* as the *random walk return* and z_t^* as the *random walk fundamental*; the random walk return is the innovation in the random walk fundamental. For convenience, Table 2 lists these and other basic variables.

2.4. A new assumption

Clearly, if b is very near one, then one might need very large samples to distinguish between the martingale difference (2.7) and the potentially serially correlated process given in (2.6). The conventional asymptotic approximation that lets $T \to \infty$ might suggest that estimation and inference will proceed with precision that is impractical given the current sample size. I suggest instead an alternative assumption, which, as we shall see, does lead to an asymptotic approximation in which it is very difficult to distinguish between y_t^* and y_t. In particular, I assume that, as $T \to \infty$,

$$\sqrt{T}(1 - b) = \delta, \quad 0 < \delta < \infty. \tag{2.8}$$

Eq. (2.8) states that, as T grows bigger, b grows closer to 1. One can rewrite (2.8) as $b = 1-(\delta/\sqrt{T})$. The results of course follow if (2.8) is replaced by $\sqrt{T}(1-b) \to \delta$. I assume that x_t and presample values of y_t are unchanged as T grows, with changes in b affecting y_t for $t \geq 1$. The rate $\sqrt{T}$, rather than T^θ for arbitrary $\theta > 0$, was chosen because, as we shall see, this will imply consistency with what I view as the stylized finding that hypothesis tests yield some mild evidence against a random walk model and considerable dispersion in estimates such as those reported in line (1) of Table 1. The proofs in the Appendix in fact provide for a rate of $T^\theta(1-b) = \delta$, for arbitrary $\theta > 0$. I note in the text below how results vary for arbitrary $\theta > 0$, as well as for the usual fixed b asymptotics which in the present context can be described as setting $T^\theta(1 - b) = \delta$ for $\theta = 0$ and $\delta = 1 - b$.

Of course, under (2.8) one should not think of the data as literally being generated with b changing as additional observations appear. In general equilibrium models, changes in b would likely change the stochastic process followed by x_t, and, as just stated, my asymptotics hold this process fixed.[4] Instead, (2.8) should be thought of as a way of delivering an asymptotic approximation that rationalizes finite sample distributions better than does the usual approximation derived with b fixed.

3. Asymptotic results for correlations

Consider first univariate autocovariances and autocorrelations of the observable change in the asset price Δy_t and the unobservable martingale difference Δy_t^*. Under conventional asymptotics, sample autocovariances and autocorrelations of Δy_t converge to their population counterparts; if these counterparts are nonzero, conventional hypothesis tests have unit asymptotic probability of rejecting the null that the given moment is zero. As for the martingale difference Δy_t^*, its autocovariances and autocorrelations are zero. If we were able to observe data on Δy_t^* to, construct sample autocovariances cross-covariances, autocorrelations and cross-correlations, these sample quantities converge in probability to zero; conventional hypothesis tests at a given nominal size have asymptotic rejection probability equal to size.

The first result of this section is that, under condition (2.8), sample autocovariances and autocorrelations constructed from Δy_t converge in probability to zero, as do sample cross-covariances and cross-correlations between Δy_t and lags of forecasting variables such as Δx_t. Thus these moments behave like those of the unobservable variable Δy_t^*. In contrast, hypothesis tests do not behave like those applied to the moments of Δy_t^*. Instead, such tests have asymptotic probability of rejecting that is greater than nominal size but less than unity. The implication is that regressions of Δy_t on forecasting variables will yield some, though perhaps not decisive, evidence against predictability. A calibration from exchange rate and stock price data indeed suggests asymptotic probability of rejection that is barely above nominal size for autocorrelations, but is modestly above nominal size for certain cross-correlations. In my view, this matches the empirical evidence.

Specifically: Let a "-" over a variable denote the usual sample mean, e.g., $\Delta\bar{y} \equiv T^{-1}\sum_{t=1}^{T}\Delta y_t$. Let $\hat{\gamma}_j$ and $\hat{\gamma}_j^*$ denote the usual estimates of the lag j autocovariance of Δy_t and Δy_t^* for $j \geq 0$:

$$\hat{\gamma}_j = T^{-1}\sum_{t=j+1}^{T}(\Delta y_t - \Delta\bar{y})(\Delta y_{t-j} - \Delta\bar{y}),$$
$$\hat{\gamma}_j^* = T^{-1}\sum_{t=j+1}^{T}(\Delta y_t^* - \Delta\bar{y}^*)(\Delta y_{t-j}^* - \Delta\bar{y}^*). \tag{3.1}$$

(Of course, one does not observe Δy_t^*, so one could not in practice construct $\hat{\gamma}_j^*$.) Under the mild conditions stated in (2.4), $\hat{\gamma}_j^* \to_p 0$ for $j \neq 0$. Under the additional condition (2.8), $\hat{\gamma}_j$ behaves similarly. Specifically:

[4] It is straightforward to allow certain forms of minor dependence of the x_t process on b. For example, suppose that α depends on b. (Recall that $\Delta x_t = \alpha'\Delta w_t$.) Index the dependence by writing $\alpha = \alpha(b)$. Then as long as $\alpha(b) \to \alpha(1)$, with $\alpha(1)$ finite and $\alpha(1)'w_t \sim I(1)$, the results about to be presented go through. I rule out such dependence to avoid notational clutter.

Theorem 3.1. *For $j > 0$, $\sqrt{T}(\hat{\gamma}_j - \hat{\gamma}_j^*) \to_p \delta E\eta_{t-j}^* z_{t-1}^*$.*

It follows from Theorem 3.1 that $\hat{\gamma}_j - \hat{\gamma}_j^* \to_p 0$, implying that $\hat{\gamma}_j \to_p 0$ for $j \neq 0$.

The term $E\eta_{t-j}^* z_{t-1}^*$ is the covariance between (1) the j period lag of the random walk return Δy_{t-j}^* ($=\eta_{t-j}^* + \mu$, see (2.7)) and (2) the random walk fundamental z_{t-1}^*. To see why the normalized discrepancy between $\hat{\gamma}_j$ and $\hat{\gamma}_j^*$ is proportional to the covariance $E\eta_{t-j}^* z_{t-1}^*$, suppose that $j = 1$ and $\mu = 0$ for notational simplicity. Let "côv" denote a sample covariance. Then $\hat{\gamma}_1 = \widehat{\mathrm{cov}}[\eta_t + (1-b)z_{t-1}, \eta_{t-1} + (1-b)z_{t-2}] = \widehat{\mathrm{cov}}(\eta_t, \eta_{t-1}) + (1-b)\widehat{\mathrm{cov}}(\eta_t, z_{t-2}) + (1-b)\widehat{\mathrm{cov}}(\eta_{t-1}, z_{t-1}) + (1-b)^2\widehat{\mathrm{cov}}(z_{t-1}, z_{t-2})$. Thus

$$\begin{aligned}\sqrt{T}(\hat{\gamma}_1 - \hat{\gamma}_1^*) &= \sqrt{T}[\widehat{\mathrm{cov}}(\eta_t, \eta_{t-1}) - \widehat{\mathrm{cov}}(\eta_t^*, \eta_{t-1}^*)] \\ &\quad + \sqrt{T}(1-b)\widehat{\mathrm{cov}}(\eta_t, z_{t-2}) \\ &\quad + \sqrt{T}(1-b)\widehat{\mathrm{cov}}(\eta_{t-1}, z_{t-1}) \\ &\quad + \sqrt{T}(1-b)^2\widehat{\mathrm{cov}}(z_{t-1}, z_{t-2}).\end{aligned}$$

Oversimplifying somewhat, the Appendix shows that, as $T \to \infty$, z_t and η_t behave very much like z_t^* and η_t^* – indeed, so much so that to derive the limiting value of the terms on the right-hand side, one can replace z_t with z_t^* and η_t with η_t^*. Hence the first term converges to zero, the second term to $\delta E\eta_t^* z_{t-2}^* = \delta \times 0 = 0$; the final term is $(1-b)$ times a term that converges to $\delta E z_{t-1}^* z_{t-2}^*$ and hence converges to zero because $1 - b \to 0$. The only nonzero quantity comes from the third term, which converges to $\delta E\eta_{t-1}^* z_{t-1}^*$.

Before interpreting Theorem 3.1, let me present the parallel result for correlations between Δy_t and stationary variables in the information set used to forecast x_t. Let u_t be any such variable, with a leading example being $u_t = \Delta x_t$. Let $\hat{\gamma}_{j,yu}$ and $\hat{\gamma}_{j,yu}^*$ denote the usual estimators of the lag j cross-covariance of u_{t-j} with Δy_t and Δy_t^* for $j \geq 0$, $\hat{\gamma}_{j,yu} = T^{-1}\sum_{t=j+1}^{T}(\Delta y_t - \Delta\bar{y})(u_{t-j} - \Delta\bar{u})$, and similarly for $\hat{\gamma}_{j,yu}^*$. Then we have the following.

Theorem 3.2. *For $j > 0$, $\sqrt{T}(\hat{\gamma}_{j,yu} - \hat{\gamma}_{j,yu}^*) \to_p \delta E u_{t-j} z_{t-1}^*$.*

To interpret Theorem 3.1, let $\hat{\rho}_j \equiv \hat{\gamma}_j/\hat{\gamma}_0$ and $\hat{\rho}_j^* \equiv \hat{\gamma}_j^*/\hat{\gamma}_0^*$ denote the sample autocorrelations of Δy_t and Δy_t^*. Suppose that η_t^* is i.i.d., implying that $\sqrt{T}\hat{\rho}_j^* \sim_A N(0, 1)$, $\hat{\rho}_j^* \approx N(0, 1/T)$. Then Theorem 3.1 implies $\hat{\rho}_j \approx N(k/\sqrt{T}, 1/T)$, $k = \delta E\eta_{t-j}^* z_{t-1}^*/\gamma_0^*$. Thus, for any T, when $\sqrt{T}(1-b) = \delta$, the asymptotic approximation implies that $\hat{\rho}_j$ is centered k standard deviations away from zero. Let ζ denote a standard normal variable, and consider tests of a given nominal size, say 0.10, in which case the critical value is 1.645. Then the asymptotic size of such a test is

$$\begin{aligned}&\mathrm{prob}(\zeta < -1.645 - k) + \mathrm{prob}(\zeta > 1.645 - k), \\ &\quad k = \delta E\eta_{t-j}^* z_{t-1}^*/\gamma_0^*. \end{aligned} \tag{3.2}$$

To interpret Theorem 3.2, consider the asymptotic behavior of the usual t-test for $H_0 : \beta = 0$ in the least squares regressions such as $\Delta y_t = \text{const.} + \beta u_{t-j} + \text{residual}$. With ζ a standard normal variable, the asymptotic size of a (say) 0.10 test t-test for $H_0 : \beta = 0$ in this regression is

$$\begin{aligned}&\mathrm{prob}(\zeta < -1.645 - k) + \mathrm{prob}(\zeta > 1.645 - k), \\ &\quad k = \frac{\delta E u_{t-j} z_{t-1}^*}{[\mathrm{var}(u_t)\gamma_0^*]^{\frac{1}{2}}}.\end{aligned} \tag{3.3}$$

Here is a calibration of (3.2) and (3.3) for tests for predictability of changes exchange rates and for stock prices. We do not observe η_{t-j}^*, z_{t-1}^* or γ_0^*, but perhaps we will not be led too far astray by substituting Δy_{t-j}, $y_{t-1} - x_{t-1}$ and $\hat{\gamma}_0$. For exchange rates, I used Engel and West's (2005) quarterly data on bilateral US dollar exchange rates versus the other G7 countries. I set x_t = log money − log output, as in the simplest monetary model of the exchange rate. (See Engel and West, 2005 for discussion of both the model and the data.) For stock prices, I used long-term annual data available from Robert Shiller's web site. I set x_t = log dividends, as in a Campbell and Shiller's (1989) log linearized stock price model with constant expected returns. In accordance with the baseline simulation presented below, I set $\delta = 0.5$ (a value rationalized in the simulations in Section 5).

Univariate autocorrelations implied low probability of finding evidence against the random walk. For $j = 1$, use of (3.2) yielded size between 0.10 and 0.128 for exchange rates versus the six other G7 countries and yielded size of 0.109 for the stock price data set—barely above the 0.10 size that would apply under a pure random walk. For cross-correlations, I set the predictor u in (3.3) to be $u_{t-1} = \Delta x_{t-1}$ in the exchange rate data and u_{t-1} = the log ratio of a ten-year moving average of earnings to price in the stock price data.[5] The implied size ranged from just above 0.10 to 0.34 in the exchange rate data, and was about 0.23 in the stock price data—modestly but still distinctly above the 0.10 that would apply under a pure random walk. That rejection is more common with cross-correlations but that evidence of predictability is still modest is of course consistent with empirical evidence for both exchange rates and equity returns (e.g., Engel et al., 2007 and Campbell et al., 1997).

Six comments on Theorems 3.1 and 3.2.

1. All else equal, higher predictability (i.e., higher size, size closer to unity) is associated with δ being bigger, i.e., with discount factor b being farther from 1. This is consistent with Engel and West (2005).
2. As well, all else equal, higher predictability is associated in (3.3) with a higher absolute value of the correlation between the predictor and the random walk fundamental z_{t-1}^*. This indicates that, in searching for or evaluating predictors of the change in the asset price, the correlation between a given predictor and a proxy for the unobserved variable z_{t-1}^* can provide corroborating or perhaps discouraging evidence of the strength or likely robustness of the predictor's ability to predict.
3. The theorems generalize in straightforward fashion to a multivariate regression of Δy_t on a stationary vector that includes some combination of lags of Δy_t, Δx_t and other elements of the information set used for forecasting. The least squares estimate converges in probability to zero; when normalized by $\sqrt{T}$, the difference between this estimate and the estimate of a hypothetical regression of Δy_t^* on the same variables converges in probability to a constant vector; the usual tests for zero coefficients have asymptotic size greater than nominal size but do not have unit asymptotic probability of rejecting.
4. The theorems generalize to apply to regressions of multiperiod changes $y_t - y_{t-h}$ for $h > 1$ on elements of the information set used for forecasting. For equity returns, it is a stylized finding that there is more evidence against unpredictability in long horizon than in one-period-ahead forecasts (e.g., Campbell et al., 1997). This can be rationalized in the present framework via particular DGPs; see Engel et al. (2011) for an example.
5. For $j = 0$ in Theorem 3.1 and $j \leq 0$ in Theorem 3.2, it is still the case that $\sqrt{T}(\hat{\gamma}_0 - \hat{\gamma}_0^*) \to_p$ nonzero constant and $\sqrt{T}(\hat{\gamma}_{j,yx} - \hat{\gamma}_{j,yx}^*) \to_p$ nonzero constant, with the constant proportional to δ; the factor of proportionality is different than that stated in the two theorems.

[5] Write this predictor as $u_{t-1} = n_{t-1} - y_{t-1}$, where n_{t-1} is the ten-year average of log earnings. Eq. (3.3) was applied as follows. Define $u_{2t-1} \equiv n_{t-1} - x_{t-1}$ and write $u_{t-1} = x_{t-1} - y_{t-1} + u_{2t-1} = \text{const.} + z_{t-1} + u_{2t-1}$; compute the numerator of (3.3) as $\delta \times [\mathrm{variance}(z_t) + \mathrm{covariance}(z_{t-1}, u_{2t-1})]$ and similarly for the denominator.

6. Consider a generalization of assumption (2.8):

$$T^{\theta}(1-b)=\delta,\quad 0<\delta<\infty,\ 0<\theta. \tag{3.4}$$

Then Theorems 3.1 and 3.2 continue to hold with $\sqrt{T}$ replaced by T^{θ}: for $j>0$, $T^{\theta}(\hat{\gamma}_j-\hat{\gamma}_j^*)\to_p \delta E\eta_{t-j}^* z_{t-1}^*$, $T^{\theta}(\hat{\gamma}_{j,yu}-\hat{\gamma}_{j,yu}^*)\to_p \delta Eu_{t-j}z_{t-1}^*$. Thus estimates of correlations with lagged variables tend to zero: for any $\theta>0$ in (3.4), $\hat{\gamma}_j\to_p 0$ ($j\neq 0$), $\hat{\gamma}_{j,yu}\to_p 0$ ($j>0$). In this sense, implications for point estimates are the same for all $\theta>0$. But the hypothesis tests constructed under the null of a random walk asset price behave differently. For $\theta>\frac{1}{2}$, actual size equals nominal size; for $\theta<\frac{1}{2}$, actual size is unity. The logic is outlined in a footnote.[6] I focus on $\theta=\frac{1}{2}$ because, as illustrated in the calibrations in the discussion below Theorem 3.2, it implies that results of hypothesis tests will accord with what I view as the stylized empirical fact that there is some, though not great, evidence against the random walk model.

4. Asymptotic results for forward-spot regressions

Suppose that y_t is the (spot) exchange rate. It is well known that in practice the regression of y_t on the one-period-ahead forward rate yields a coefficient numerically close to unity, while the regression of Δy_t on the forward premium ($\equiv$ difference between the forward rate at $t-1$ and y_{t-1}) does not: indeed, it tends to deliver a coefficient well less than one, though with a large standard error. (See lines (1) and (2) in Table 1.) To analyze these regressions requires some additional notation. First, define f_{t-1} as the one-period-ahead expectation of the asset price:

$$f_{t-1}\equiv E_{t-1}y_t. \tag{4.1}$$

In certain contexts, the variable f_{t-1} is the one-period-ahead forward rate (possibly up to a constant risk premium), so $f_{t-1}-y_{t-1}$ is the forward premium (again apart from a constant risk premium). In any event, whatever the observable counterpart (if any) to $E_{t-1}y_t$, I define f_{t-1} as $E_{t-1}y_t$. Next, let us define the random walk asset price y_t^* whose first difference is the random walk return Δy_t^* defined in (2.7) above. Begin by noting that, with a little algebra, it may be shown that $\Delta y_t=\Delta x_t+b\Delta z_t$.[7] So the level variable consistent with this model and given presample values y_0, x_0 and z_0 for $t\geq 0$ is

$$y_t=c_0+x_t+bz_t,\quad c_0\equiv y_0-x_0-bz_0. \tag{4.2}$$

I define y_t^* by dropping c_0 and setting $b=1$:

$$y_t^*=x_t+z_t^*. \tag{4.3}$$

An arbitrary constant (including c_0) can be added to the right-hand side of (4.3) without changing any of the analysis; I set this constant to zero for notational simplicity.

Under conventional asymptotics, a regression of y_t on f_{t-1} ($\equiv E_{t-1}y_t$), or of Δy_t on $f_{t-1}-y_{t-1}$, will yield a coefficient near unity with high probability, for sufficiently large T. Write these regressions as

$$\Delta y_t=\text{constant}+\beta_r(f_{t-1}-y_{t-1})+\text{residual}, \tag{4.4}$$

$$y_t=\text{constant}+\beta_f f_{t-1}+\text{residual}. \tag{4.5}$$

Let $\hat{\beta}_r$ and $\hat{\beta}_f$ denote the least squares estimators of the slope coefficients.

Theorem 4.1. (a) $\hat{\beta}_r-1\sim_A N(0,V/\delta^2)$, *with* $V>0$ *the asymptotic variance of the estimator of the slope coefficient in a least squares regression of* Δy_t^* *on a constant and* z_{t-1}^*.

(b) *Suppose* $E\Delta x_t\equiv\mu=0$. *Let* $\hat{\beta}_f^*$ *denote the slope coefficient in a regression of* y_t^* *on a constant and* y_{t-1}^*. *Then* $T(\hat{\beta}_f-\hat{\beta}_f^*)\to_p 0$.

Both parts of the theorem generalize for multiperiod forecasts using overlapping data.

Consider first part (b) of the theorem. Since y_t^* is a random walk, $T(\hat{\beta}_f^*-1)$ is nondegenerate asymptotically, with a nonstandard distribution; $\hat{\beta}_f^*\to_p 1$ at a superconsistent rate. Thus part (b) of Theorem 4.1 reassures us that the framework used here continues to deliver a familiar result—the estimator of β_f is superconsistent. To see why, recall from Eqs. (4.2) and (4.3) that $y_t=c_0+x_t+bz_t$, $y_t^*=x_t+z_t^*$. From the point of view of the $I(1)$ variables y_t and y_t^*, the finite variance discrepancy between bz_t and z_t^* is very small for b near 1, so small that part (b) of the theorem is implied.[8]

Part (a) suggests a possible reason that estimates of $\hat{\beta}_r$ are not clustered near unity: the estimator is inconsistent. (Note that the quantity that is asymptotically normal is $\hat{\beta}_r-1$ rather than the usual $\sqrt{T}(\hat{\beta}_r-1)$.) The reason for the inconsistency is that, under condition (2.8), the forward rate f_{t-1} approaches y_{t-1} as $T\to\infty$. Thus the variance of the right-hand-side variable in (4.4) gets smaller as T increases.

To understand why the variance V is that of a regression of Δy_t^* on z_{t-1}^*, observe that $f_{t-1}-y_{t-1}=(1-b)z_{t-1}+\mu$.[9] For large T, a regression of Δy_t on a constant and $z_{t-1}+\mu$ will behave like that of a regression of Δy_t^* on z_{t-1}^* and hence imply a variance V for the regression coefficient; the actual regression is of Δy_t on $(1-b)z_{t-1}+\mu$, with the factor of $(1-b)$ contributing the factor of $1/\delta^2$. The variance V/δ^2 of the limiting normal distribution of $\hat{\beta}_r-1$ is larger when δ is smaller: the "closer" is b to 1 (in the sense of small δ), the noisier is the slope coefficient. Though not stated in the theorem, the conventional t-test of H_0: $\beta_r=1$ is well-behaved asymptotically, despite the inconsistency in estimation of β_r.

The fact that the limiting (nondegenerate) distribution is centered around 1 means that Theorem 4.1(a) does not rationalize a systematic tendency for $\hat{\beta}_r$ to be less than unity, which seems to be the case in practice. One possible explanation is small sample bias. This is suggested by the simulations discussed in the next section.

Other econometric explanations have been offered for the failure to find $\hat{\beta}_r\approx 1$. One such explanation is that the regression is imbalanced: the forward premium $f_{t-1}-y_{t-1}$ is fractionally integrated, it is argued, while Δy_t is stationary. Maynard and Phillips (2001) develop an asymptotic theory that derives the implications for $\hat{\beta}_r$. That theory also implies that a regression of y_t-y_{t-2} on $f_{t-1}-y_{t-2}$ will also produce a slope coefficient that

[6] I illustrate with tests for univariate autocorrelation as discussed below Theorem 3.2. Choose j so that the covariance on the right-hand side of Theorem 3.1 is nonzero, i.e., $\delta E\eta_{t-j}^* z_{t-1}^*\neq 0$. Recall that $\gamma_{j=0}^*$ for $j\neq 0$ and, by a standard argument, assumption (2.4) ensures $\sqrt{T}\hat{\gamma}_j^*=O_p(1)$. (a) When $\theta>\frac{1}{2}$, $T^{\theta}(\hat{\gamma}_j-\hat{\gamma}_j^*)\to_p \delta E\eta_{t-j}^* z_{t-1}^*$ means $\sqrt{T}(\hat{\gamma}_j-\hat{\gamma}_j^*)\to_p 0$. So the limiting distribution of $\sqrt{T}\hat{\gamma}_j$ is that of $\sqrt{T}\hat{\gamma}_j^*$, and similarly for $\sqrt{T}\hat{\rho}_j$ and $\sqrt{T}\hat{\rho}_j^*$. (b) When $\theta<\frac{1}{2}$, $T^{\theta}\hat{\gamma}_j^*\to_p 0$. Since $T^{\theta}(\hat{\gamma}_j-\hat{\gamma}_j^*)\to_p \delta E\eta_{t-j}^* z_{t-1}^*$, we have $T^{\theta}\hat{\gamma}_j\to_p \delta E\eta_{t-j}^* z_{t-1}^*$ and $T^{\theta}\hat{\rho}_j\to_p \delta E\eta_{t-j}^* z_{t-1}^*/\gamma_0^*$. But for a critical value of (say) 1.645, as T grows, we increasingly tend to get $|\sqrt{T}\hat{\rho}_j|>1.645$, because $|\sqrt{T}\hat{\rho}_j|>1.645 \Leftrightarrow T^{\frac{1}{2}-\theta}|T^{\theta}\hat{\rho}_j|>1.645$.

[7] Specifically, rewrite (2.1) as $y_t=c_y+x_t+\sum_{j=1}^{\infty}b^jE_t\Delta x_{t+j}$. First differencing this gives $\Delta y_t=\Delta x_t+\sum_{j=1}^{\infty}b^jE_t\Delta x_{t+j}-\sum_{j=1}^{\infty}b^jE_{t-1}\Delta x_{t+j-1}=\Delta x_t+b\sum_{j=0}^{\infty}b^jE_t(\Delta x_{t+j+1}-\mu)-b\sum_{j=0}^{\infty}b^jE_{t-1}(\Delta x_{t+j}-\mu)\equiv\Delta x_t+b\Delta z_t$.

[8] If $E\Delta x_t\equiv\mu\neq 0$, then $T^{3/2}(\hat{\beta}_f-1)$ is nondegenerate asymptotically and behaves like $T^{3/2}(\hat{\beta}_f^*-1)$.

[9] This follows from the fact that Eq. (2.6) implies that $E_{t-1}y_t-y_{t-1}\equiv(1-b)z_{t-1}+\mu$.

converges in probability to zero.[10] But a less well-known regularity from the exchange rate literature is that a such regression in fact tends to produce a coefficient near unity. (See line (3) in Table 1.) Write the regression as

$$y_t - y_{t-2} = \text{constant} + \beta_m(f_{t-1} - y_{t-2}) + \text{residual}. \tag{4.6}$$

Let $\hat{\beta}_m$ be the least squares estimator of β_m.

Theorem 4.2. $\sqrt{T}(\hat{\beta}_m - 1) \sim_A N(0, V)$, *with* $V > 0$ *the asymptotic variance of the estimator of the slope coefficient in a least squares regression of* $y_t^* - y_{t-2}^*$ *on a constant and* $y_{t-1}^* - y_{t-2}^*$.

A similar result applies when any lag y_{t-j} for $j > 1$ is subtracted from both sides of (4.5). The usual least squares t-statistic for H_0: $\beta_m = 1$ is asymptotically $N(0, 1)$.

Four comments.

1. Theorem 4.2 and a generalization of Theorem 4.1 obtain when one maintains (3.4), i.e., $T^\theta(1 - b) = \delta$ for arbitrary $\theta > 0$. The generalization of Theorem 4.1 is

$$T^{\frac{1}{2}-\theta}(\hat{\beta}_r - 1) \sim_A N(0, V/\delta^2). \tag{4.7}$$

For $\theta < \frac{1}{2}$, one implication of (4.7) is that $\hat{\beta}_r \to_p 1$; for $\theta > \frac{1}{2}$, one implication is that $\hat{\beta}_r - 1$ becomes increasingly disperse as T grows. The $\theta = \frac{1}{2}$ result stated in Theorem 4.1 is congruent with what I view as the stylized fact that the dispersion in estimates of β_r do not seem to either grow or shrink as we get more and more data.

The next two comments build on the results in not only this section but also the previous.

2. It may be helpful to contrast these results with those that obtain under the usual asymptotics with fixed b. In terms of the present section, $\hat{\beta}_r \to_p 1$ and $\sqrt{T}(\hat{\beta}_r - 1)$ is asymptotically normal; Theorem 4.2 continues to apply. In terms of the previous section, except in the special case in which x_t follows a random walk (i.e., $E_t\Delta x_{t+j} = 0$ for all $j > 0$), one or more sample autocovariances and autocorrelations of Δy_t converge in probability to nonzero constants. So, too, do cross-correlations of Δy_t with lags of Δx_t and other stationary variables that help predict future Δx's. Convergence is at the usual $\sqrt{T}$ rate. Hypothesis tests constructed under the null of no correlation are consistent. Hence, with the exception of Theorem 4.2, results are qualitatively different for my $b \to 1$ asymptotics than for the usual fixed b asymptotics.

3. For the most part, the results in both this section and the previous apply when data is generated without the $(1 - b)$ factor scaling the present value in (2.1). Specifically, suppose that the asset price, call it $\tilde{y}_t$, is determined via

$$\tilde{y}_t = \text{constant} + \sum_{j=0}^{\infty} b^j E_t x_{t+j}. \tag{4.8}$$

Continue to maintain assumption (2.8) (i.e., $\sqrt{T}(1 - b) = \delta$) and the assumptions in Section 2.2. Let me first note some results that do *not* still apply. Estimators of autocovariances and cross-covariances of the first difference of the asset price no longer converge in probability to zero. For the first difference of the asset price $\Delta\tilde{y}_t$, let $\Delta\bar{\tilde{y}}$ be the sample mean, and for $j \geq 0$ let $\tilde{\gamma}_j$ be the sample autocovariance: $\Delta\bar{\tilde{y}} = T^{-1}\sum_{t=1}^{T}\Delta\tilde{y}_t$, $\tilde{\gamma}_j \equiv T^{-1}\sum_{t=j+1}^{T}(\Delta\tilde{y}_t - \Delta\bar{\tilde{y}})(\Delta\tilde{y}_{t-j} - \Delta\bar{\tilde{y}})$. Then it is no longer true that $\tilde{\gamma}_j \to_p 0$. This is because, under (4.8) and the assumptions in Section 2, as $b \to 1$, the variance of $\Delta\tilde{y}_t - E_{t-1}\Delta\tilde{y}_t$ grows without bound. Similarly, and in contrast to Theorem 3.2, sample cross-covariances do not converge in probability to zero.

To obtain limiting results under (4.8), one needs to normalize such moments. Fortunately, natural normalizations, which do not require unusual construction, will suffice. Consider, for example, sample autocorrelations. Let $\tilde{\rho}_j \equiv \tilde{\gamma}_j/\tilde{\gamma}_0$. Define in the usual way sample autocovariances and autocorrelations $\hat{\gamma}_j$ and $\hat{\rho}_j$ from the artificial and unobserved random variable $y_t \equiv (1 - b)\tilde{y}_t$. Clearly, $\tilde{\rho}_j = \hat{\rho}_j$, because a factor of $(1 - b)$ cancels out of the numerator and denominator of $\hat{\rho}_j$. And the asymptotic behavior of $\hat{\rho}_j$ was characterized in the previous section. Hence, under (4.8), the technical assumptions in Section 2.2 and assumption (2.8),

(a) $\tilde{\rho}_j \to_p 0$;

(b) hypothesis tests on the null of zero autocorrelation of $\Delta\tilde{y}_t$ of nominal size (say) 0.10 have size greater than 0.10 and less than 1.0; and

(c) there is continued applicability of results listed in comment 6 of the previous section, describing what happens when assumption (2.8) is generalized by (3.4) (i.e., $T^\theta(1 - b) = \delta$ for some $\theta > 0$).

The same cancellation of a factor of $(1 - b)$ occurs in Theorems 4.1 and 4.2 and in t- or F-statistics of a regression of $\Delta\tilde{y}_t$ on lagged stationary elements of the information set used to forecast x_t. Hence those results continue to apply to model (4.8).

4. We see from the previous comments that Theorem 4.2 holds as stated above, without reference to θ or δ, under each of the setups described in those previous comments. Perhaps this suggests an advantage to regression (4.6) relative to the much more commonly estimated regression (4.4).

5. Simulation evidence

A small simulation was executed to evaluate the finite sample accuracy of the proposed asymptotic approximation. The DGP for Δx_t was as follows:

$$\Delta x_t = \Delta x_{1t} + u_{3t-1}, \tag{5.1a}$$

$$\Delta x_{1t} = \varphi\Delta x_{1t-1} + u_{2t} + u_{1t-1} + u_{1t-2}, \tag{5.1b}$$

$$e_t \equiv (u_{1t}, u_{2t}, u_{3t})' \sim \text{i.i.d. } N(0, \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)). \tag{5.1c}$$

It may be shown that the implied process for z_t (and thus for $f_t - y_t \equiv (1 - b)z_t$) follows an ARMA(1,1). The use of three shocks serves two purposes: first, there must be two or more shocks to preclude a singularity in the interest parity regressions discussed in the previous section, and, second, the introduction of a moving average components allows the first one or two autocorrelations of Δx_t and $f_t - y_t$ to be well below φ, even if φ is near one so that the subsequent autocorrelations decline slowly. (See Baillie and Bollerslev, 1994 and Maynard and Phillips, 2001 on slow decay of autocorrelations of $f_t - y_t$.)

For the baseline DGP, called DGP A in one of the tables, the parameters were $\varphi = 0.95$, $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 1, 50)$, $\sqrt{T}(1 - b) = 0.5$. Sample sizes of $T = 100$, 400, and 1600 were generated; implied values of b are 0.95, 0.975, and 0.9875 for the three values of T. These values are consistent with calibrations of standard stock price and exchange rate models to annual and higher frequencies; see Campbell et al. (1997) and Engel and West (2005). Table 3 has some implied moments. The DGP for Δx_t does not vary with T. But moments of Δy and $f - y$ do vary with T, since the implied value of b varies with T. Table 3 demonstrates one sense in which a random walk better approximates Δy_t as T grows with $\sqrt{T}(1-b)$ held fixed: the first-order autocorrelation of Δy_t falls from 0.08 to 0.03 to 0.01; the cross-correlation between Δy_t and $f_{t-1} - y_{t-1}$ falls from 0.16 to 0.08 to 0.04. All these values are more or less in line with point estimates from a typical exchange rate dataset. The

10 Maynard and Phillips (2001) show that, when $f_{t-1} - y_{t-1}$ is fractionally integrated, $\hat{\beta}_r \to_p 0$. The intuition is that, since the dependent variable is stationary and the regressor fractionally integrated, a stationary regression residual is delivered only with a zero slope coefficient. When the regressor is changed from $f_{t-1} - y_{t-1}$ to $f_{t-1} - y_{t-2} = f_{t-1} - y_{t-1} + y_{t-1} - y_{t-2}$, the regressor is still fractionally integrated, and the zero slope coefficient is still implied.

Table 3
Population moments, baseline DGP.

Correlations in columns (2)–(6)						
T	$\Delta x_t, \Delta x_{t-1}$	$f_t - y_t, f_{t-1} - y_{t-1}$	$\Delta y_t, \Delta y_{t-1}$	$\Delta y_t, f_{t-1} - y_{t-1}$	$\Delta y_t, \Delta x_{t-1}$	$\sigma_{\Delta y}/\sigma_{f-y}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
100	0.49	0.95	0.08	0.16	0.11	6.2
400	0.49	0.95	0.03	0.08	0.05	12.4
1600	0.49	0.95	0.01	0.04	0.03	24.8

Notes:
1. The model is $y_t = (1-b)\sum_{j=0}^{\infty} b^j E_t x_{t+j}$, with $f_{t-1} = E_{t-1}y_t$.
2. The moments are calculated for the DGP given in Eq. (5.1), with parameters $\varphi = 0.95$, $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 1, 50)$, $\sqrt{T}(1-b) = 0.5$.

Table 4
Simulation results, baseline DGP.

A. Quantiles of point estimates

(1)	(2)			(3)			(4)			(5)			(6)		
T		$\hat{\rho}_1$			$\hat{\rho}_{1,yx}$			$\hat{\beta}_r$			$\hat{\beta}_f$			$\hat{\beta}_m$	
	0.05	0.50	0.95	0.05	0.50	0.95	0.05	0.50	0.95	0.05	0.50	0.95	0.05	0.50	0.95
100	−0.12	0.04	0.20	−0.15	0.03	0.21	−1.60	0.36	1.40	0.90	0.97	1.00	0.85	1.01	1.17
400	−0.06	0.03	0.11	−0.08	0.03	0.13	−0.68	0.71	1.60	0.97	0.99	1.00	0.94	1.02	1.10
1600	−0.03	0.01	0.05	−0.04	0.02	0.07	−0.35	0.84	1.79	0.99	1.00	1.00	0.97	1.01	1.05
∞	0.00	0.00	0.00	0.00	0.00	0.00	−0.03	1.00	2.03	1.00	1.00	1.00	1.00	1.00	1.00

B. Actual and asymptotic rejection rates of two-tailed nominal 0.10 tests

Hypothesis					
(1)	(2)	(3)	(4)	(5)	(6)
T	$\rho_1 = 0$	$\rho_{1,yx} = 0$	$\beta_r = 1$	$\beta_f = 1$	$\beta_m = 1$
100	0.126	0.162	0.200	0.107	0.123
400	0.144	0.231	0.127	0.112	0.136
1600	0.153	0.259	0.115	0.100	0.139
∞	0.137	0.270	0.100	0.100	0.100

Notes:
1. The model and DGP are described in notes to Table 3.
2. The "∞" lines give asymptotic values, computed analytically. All other results are based on 5000 repetitions.
3. ρ_1 is the first-order autocorrelation of Δy_t; $\rho_{1,yx}$ is the correlation between Δy_t and Δx_{t-1}; β_r, β_f and β_m are slopes in linear projections of $y_t - y_{t-1}$ on $f_{t-1} - y_{t-1}$, y_t on f_{t-1}, and $y_t - y_{t-2}$ on $f_{t-1} - y_{t-2}$.
4. Panel A gives the 0.05, median and 0.95 quantiles for the indicated statistics. For example, the figure of "−0.12" in the 0.05 column underneath column (2), panel A, means that, in 250 of the 5000 repetitions, the estimated first-order autocorrelation of Δy_t was less than or equal to −0.12.
5. For t-statistics computed as described in the next note, columns (2)–(4) and (6) in panel B give the fraction of samples in which the t-statistic was greater than 1.645 in absolute value; column (5) gives the fraction of t-statistics less than −2.86 or greater than −0.07.
6. In column (2) in panel B, t-statistics were computed by dividing the point estimate by the square root of sample size. In columns (3)–(6), t-statistics for H_0 : slope coefficient = 1 were computed according to the usual least squares formula. According to the theory in this paper, the values in columns (2) and (3) will remain above 0.10 for large T (see (3.2) and (3.3)), while the values in columns (4)–(6) will approach 0.10.

main effect of increasing T is to increase the standard deviation of Δy_t relative to $f_t - y_t$.

Table 4 has simulation results. These and all other simulation results are based on 5000 repetitions. Consider correlations first. Upon comparing columns 2 and 3 in Table 4, panel A, with columns 4 and 5 in Table 3, we see that the median values of the point estimates of the first-order autocorrelation of Δy_t and the correlation between Δy_t and Δx_{t-1} are close to the population values. Columns 2 and 3 in Table 4, panel B, show that the power of tests for zero correlation is low. The implied asymptotic probabilities of rejection of a zero first-order autocorrelation coefficient and of zero correlation between Δy_t and Δx_{t-1}, computed in accordance with (3.2) and (3.3) and reported in the line labelled "∞", are 0.137 and 0.270, respectively, quite close to the actual rejection rates of 0.12–0.15 and 0.16–0.26.

Column (4) in Table 4 considers the regression of Δy_t on $f_{t-1} - y_{t-1}$. According to the asymptotic theory proposed here, this regression is inconsistent. Indeed, while the estimates of the slope coefficient β_r narrow somewhat as T increases, they are still quite disperse, even for $T = 1600$. The asymptotic width of the central 90% of the distribution is also wide, and includes a greater than 5% mass less than zero (see "∞" line in panel A). For the finite samples considered in the simulation, the median bias (downwards) in the estimate of this slope coefficient is apparent, with median $\hat{\beta}_r = 0.36$ for $T = 100$, for example. To introduce an analytical complement to the simulation evidence, consider the following approximation to the mean bias in the estimator of $\hat{\beta}_r$. Suppose counterfactually that right-hand-side variable $f_t - y_t$ followed an AR(1) with parameter ρ and i.i.d. normal innovation u_t, $f_t - y_t = \rho(f_{t-1} - r_{t-1}) + u_t$. Then the results of Stambaugh (1999) apply, and $E\hat{\beta}_r - 1 \approx -[\mathrm{cov}(u_t, \eta_t)/\mathrm{var}(u_t)](1 + 3\rho)/T$. In the DGPs in the simulations, $f_t - y_t$ does not follow an AR(1). But let ρ be the first-order autocorrelation of $f_t - y_t$, define $u_t \equiv f_t - y_t - \rho(f_{t-1} - r_{t-1})$, and compute an approximate analytical bias as $-[\mathrm{cov}(u_t, \eta_t)/\mathrm{var}(u_t)](1 + 3\rho)/T$ on the thought that the AR(1) formula might be close enough to give insight even though u_t so defined is serially correlated according to the DGP actually used. Here are some relevant quantities:

T	Median $\hat{\beta}_r$	Median bias	$\%\hat{\beta}_r < 0$	Mean $\hat{\beta}_r$	Mean bias	Approx. bias (analytical)	
100	0.36	−0.64	36%	0.18	−0.82	−0.69	(5.2)
400	0.71	−0.29	19%	0.62	−0.38	−0.36	
1600	0.84	−0.16	12%	0.80	−0.20	−0.19.	

Table 5
Simulation results, alternative DGPs, $T = 100$.

A. Population moments

Correlations

(1)	(2)	(3)	(4)	(5)	(6)	(7)
DGP	$\Delta x_t, \Delta x_{t-1}$	$f_t - y_t, f_{t-1} - y_{t-1}$	$\Delta y_t, \Delta y_{t-1}$	$\Delta y_t, f_{t-1} - y_{t-1}$	$\Delta y_t, \Delta x_{t-1}$	$\sigma_{\Delta y}/\sigma_{f-y}$
A	0.49	0.95	0.08	0.16	0.11	6.2
B	0.40	0.77	0.13	0.18	0.09	5.7
C	0.32	0.94	0.02	0.06	0.03	15.8

B. Quantiles of point estimates

(1)	(2)			(3)			(4)			(5)			(6)		
DGP	$\hat{\rho}_1$			$\hat{\rho}_{1,yx}$			$\hat{\beta}_r$			$\hat{\beta}_f$			$\hat{\beta}_m$		
	0.05	0.50	0.95	0.05	0.50	0.95	0.05	0.50	0.95	0.05	0.50	0.95	0.05	0.50	0.95
A	−0.12	0.04	0.20	−0.15	0.03	0.21	−1.60	0.36	1.40	0.90	0.97	1.00	0.85	1.01	1.17
B	−0.08	0.10	0.27	−0.15	0.06	0.23	−0.44	0.76	1.62	0.90	0.97	1.00	0.88	1.03	1.17
C	−0.15	0.01	0.16	−0.17	−0.01	0.15	−5.75	−0.68	2.04	0.88	0.96	1.00	0.83	1.00	1.16

C. Actual rejection rates of two-tailed nominal 0.10 tests

Hypothesis

(1)	(2)	(3)	(4)	(5)	(6)
DGP	$\rho_1 = 0$	$\rho_{1,yx} = 0$	β	$\beta_f = 1$	$\beta_m = 1$
A	0.126	0.162	0.200	0.107	0.123
B	0.273	0.188	0.128	0.103	0.154
C	0.087	0.107	0.201	0.101	0.108

Notes:
1. The results for DGP *A*, which sets $\varphi = 0.95$, $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 1, 50)$, $\sqrt{T}(1-b) = 0.5$, are repeated from earlier tables. DGP *B* sets $\varphi = 0.8$, $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 1, 15)$, $\sqrt{T}(1-b) = 1.0$. DGP *C* sets $\varphi = 0.95$, $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 1, 100)$, $\sqrt{T}(1-b) = 0.2$.
2. See notes to earlier tables.

The values for the median $\hat{\beta}_r$ are repeated from panel A in Table 4, and the bias is computed by subtraction from 1. The column $\%\hat{\beta}_r < 0$ gives the percentage of the 5000 samples in which the estimate of β_r was negative (not reported in Table 4). One can see in (5.2) that the mean bias is more marked than the median bias, because of the left tail of the distribution. Finally, the values for analytical bias match tolerably well with the mean bias found in the simulation. Thus small sample bias in conjunction with inconsistency of the estimator helps explain the finding that $\hat{\beta}_r$ tends to be below one.

Columns (5) and (6) in Table 4 consider the regression of y_t on f_{t-1} and $y_t - y_{t-2}$ on $f_{t-1} - y_{t-2}$. In accordance with the asymptotic theory, both regressions are centered around 1. Use of Dickey–Fuller critical values (column (5), panel (B)) or standard normal critical values (column (6), panel (B)) results in reasonably well-sized tests.

Table 5 considers simulation results after modest permutations in the DGP, for $T = 100$. Parameter values are given in the notes to the table. DGP *A* is the DGP used in Tables 1 and 2. Figures for DGP *A* are repeated from those tables, for convenience. DGP *B* puts slightly more predictability into Δy_t, DGP *C* puts slightly less. The basic patterns seen in Tables 1 and 2 for DGP *A* are also found for DGP *B* and *C*. (1) Poor power in tests of serial correlation. (2) Wide dispersion of estimates of $\hat{\beta}_r$, with downward bias. (3) Good performance of estimates and hypothesis tests on β_f and β_m.

6. Conclusions

Asset prices such as stock prices or floating exchange rates behave like a random walk in that their changes are hard to predict. The limited predictability that we find tends to come not from lagged changes of the asset price but from lags of fundamental variables such as interest rates, money supplies, outputs, earnings and dividends. Also, the behavior of regressions of exchange rate on forward rates or forward premia has a distinctive pattern that is difficult to explain. A regression of the exchange rate on the lagged forward premium delivers a coefficient that generally is less than one and often negative; a seemingly similar regression of the two-period change in exchange rates on the difference between the forward rate and the lagged exchange rate typically delivers a coefficient very near one.

An asymptotic approximation that allows the discount factor b to approach one delineates how close to a random walk an asset price will behave in a finite sample: hypothesis tests of zero predictability have size that varies in a well-defined way with the DGP; this size is less than unity but greater than the nominal size that would apply for changes in a pure random walk. The theory rationalizes patterns seen in practice in tests of predictability. The theory also rationalizes the behavior of forward-spot regressions. As well, simulations generally support the asymptotic approximation.

One priority for future work is to better explain the small sample bias observed in the regression of the exchange rate change on the lagged forward premium. A second priority is to include time varying risk premia, informational biases and portfolio adjustment costs: for both theoretical and empirical reasons, full understanding of the behavior of exchange rates and other asset prices requires allowing substantive roles for one or more of these factors.

Acknowledgments

I thank Nonarit Bisonyabut, Roberto Duncan, Terry Iverson and Ping Yu for research assistance, two anonymous referees, seminar audiences at Duke, the Triangle Econometrics Conference, the Far Eastern Meetings of the Econometric Society and the University of Wisconsin for helpful comments, and the National Science Foundation for financial support.

Appendix

In the proofs, c denotes a generic constant, not necessarily the same from proof to proof. For an arbitrary matrix A, $|A|$ denotes $\max_{i,j}|A_{ij}|$. All limits are taken as $T \to \infty$. When the bounds on a summation are not shown, the summation runs over t from $t = 1$ to $t = T$. A "¯" over a variable denotes a sample mean, e.g., $\bar{z} = T^{-1}\sum z_{t-1}$. For given b, $0 < b \le 1$, and for e_t defined in (2.2), write

$$z_t = \sum_{j=0}^{\infty} a_{jb} e_{t-j}, \qquad z_t^* = \sum_{j=0}^{\infty} a_j^* e_{t-j}, \tag{A.1}$$

where a_{jb} and a_j^* are $1 \times n$, the "b" subscript in a_{jb} emphasizes dependence of the moving average weights on b, and $a_j^* \equiv a_{j1}$. For Ψ_j and α defined in (2.2), define the $1 \times n$ vectors ψ_j and $\psi(b)$ as

$$\psi_j = \alpha'\Psi_j,$$
$$\psi(b) = \alpha'\Psi(b) \equiv \alpha' \sum_{j=0}^{\infty} b^j \Psi_j \equiv \sum_{j=0}^{\infty} b^j \psi_j. \tag{A.2}$$

Throughout the Appendix, I maintain assumption (3.4), with assumption (2.8) a special case.

Lemma A.1. (a) *For a_{jb}, a_j^*, ψ_j and $\psi(b)$ defined in (A.1) and (A.2), $a_{jb} = \sum_{i=0}^{\infty} b^i \psi_{j+1+i}$, $a_j^* = \sum_{i=0}^{\infty} \psi_{j+1+i}$.* (b) *$\eta_t = \psi(b)e_t$.*

Proof. (a) For Δw_t defined in (2.2) and for $0 < b \le 1$, let $q_t = \sum_{j=0}^{\infty} b^j E_t(\Delta w_{t+j} - E\Delta w_t) = \text{(say)} \sum_{j=0}^{\infty} A_{jb} e_{t-j}$. Then $A_{jb} = \sum_{i=0}^{\infty} b^i \Psi_{i+j}$. Since $z_t = \alpha' E_t q_{t+1}$, we have $a_{jb} \equiv \alpha' A_{j+1,b}$, which is the desired result.

(b) Follows from $(E_t - E_{t-1})\Delta w_{t+j} = \Psi_j e_t$, $\Delta x_t = \alpha'\Delta w_t$, and the fact that $\eta_t = \sum_{j=0}^{\infty} b^j (E_t - E_{t-1})\Delta x_{t+j}$. □

Lemma A.2. (a) $\sum_{j=0}^{\infty} j|\Psi_j| < \infty$. (b) *For θ defined in (3.4), $T^\theta[\Psi(b) - \Psi(1)] \to -\delta \sum_{j=1}^{\infty} j\Psi_j$.* (c) *$\sum_{j=0}^{\infty} |a_{jb}| < c < \infty$ for a constant c that does not depend on b.* (d) *$\sum_{j=0}^{\infty} |Ez_t z_{t-j}| < c < \infty$ for a constant c that does not depend on b.* (e) *$\sum_{j=0}^{\infty} (a_{jb} - a_j^*)(a_{jb} - a_j^*)' \to 0$.* (f) *$E(c + z_t - z_t^*)^2 \to c^2$.*

Proof. (a) Assumed directly in (2.4a); under (2.4b), follows from the fact that Δw_t follows a finite parameter ARMA process.

(b) We have $T^\theta[\Psi(b) - \Psi(1)] = T^\theta[\sum_{j=1}^{\infty}(b^j - 1)\Psi_j] = T^\theta[\sum_{j=1}^{\infty}(b - 1)(1 + b + \cdots + b^{j-1})\Psi_j] = -\delta \sum_{j=1}^{\infty}[(1 + b + \cdots + b^{j-1})\Psi_j]$. So the result follows if $\sum_{j=1}^{\infty} j\Psi_j - \sum_{j=1}^{\infty}[(1 + b + \cdots + b^{j-1})\Psi_j] \to 0$ as $b \to 1$. We have $|\sum_{j=1}^{\infty} j\Psi_j - \sum_{j=1}^{\infty}[(1 + b + \cdots + b^{j-1})\Psi_j]| = |(1 - b)\sum_{j=2}^{\infty} \Psi_j + (1 - b^2)\sum_{j=3}^{\infty} \Psi_j + \cdots| \le (1 - b)\sum_{j=1}^{\infty} j|\Psi_j| \to 0$.

(c) Define A_{jb} as in the proof of Lemma A.1, $A_{jb} = \sum_{i=0}^{\infty} b^i \Psi_{i+j}$. Then $|A_{jb}| \le \sum_{i=0}^{\infty} |\Psi_{i+j}| \Rightarrow \sum_{j=0}^{\infty} |A_{jb}| \le \sum_{j=0}^{\infty}(j + 1)|\Psi_j|$. Since $a_{jb} = \alpha' A_{j+1,b}$, $\sum_{j=0}^{\infty} |a_{jb}| \le n|\alpha| \sum_{j=0}^{\infty}(j + 1)|\Psi_j|$, which is finite by part (a).

(d) Follows from part (c).

(e) Write $\sum_{j=0}^{\infty}(a_{jb} - a_j^*)(a_{jb} - a_j^*)' = \sum\sum_{i,k=0}^{\infty} d_{ik,b}\psi_i\psi_k'$ for scalars $d_{ik,b}$. With some straightforward algebra, it can be shown that $|d_{ik,b}| \le (1 - b)ik$. Then $|\sum_{j=0}^{\infty}(a_{jb} - a_j^*)(a_{jb} - a_j^*)'| \le n^2(1 - b)\sum\sum_{i,k=0}^{\infty} ik|\psi_i|\,|\psi_k| = n^2(1 - b)(\sum_{i=0}^{\infty} i|\psi_i|)^2 \to 0$.

(f) It suffices to show $E(z_t - z_t^*)^2 \to 0$. But $E(z_t - z_t^*)^2 = \sum_{j=0}^{\infty}(a_{jb} - a_j^*)Ee_te_t'(a_{jb} - a_j^*)' \le n^4|Ee_te_t'|\,|\sum_{j=0}^{\infty}(a_{jb} - a_j^*)(a_{jb} - a_j^*)'| \to 0$ by part (e) of this lemma. □

Lemma A.3. *For $\ell = 1, 2, 3, 4$, let the scalar $q_{\ell t}$ denote either z_t or an element of e_t. Let p denote a fourfold product of z and of elements of e, at arbitrary dates, $t, t - m, t - j$ and $t - k$: $p = q_{1t}q_{2t-m}q_{3t-j}q_{4t-k}$. Then $|Ep| < c < \infty$ for a constant c that does not depend on b.*

Proof. I illustrate when $q_{\ell s} = z_s$; the argument when one or more of the q's is an element of e is similar though simpler. Assume e is a scalar for notational simplicity. We have

$$\begin{aligned}
|Ez_tz_{t-m}z_{t-j}z_{t-k}| &= \left| E\sum\sum\sum\sum_{f,g,h,i=0}^{\infty} \right. \\
&\qquad \left. \times\, a_{fb}e_{t-f}a_{gb}e_{t-m-g}a_{hb}e_{t-j-h}a_{ib}e_{t-k-i} \right| \\
&\le \sum\sum\sum\sum_{f,g,h,i=0}^{\infty} |a_{fb}a_{gb}a_{hb}a_{ib}|E \\
&\qquad \times |e_{t-f}e_{t-m-g}e_{t-j-h}e_{t-k-i}| \\
&\le Ee_t^4 \sum\sum\sum\sum_{f,g,h,i=0}^{\infty} |a_{fb}a_{gb}a_{hb}a_{ib}| \\
&= Ee_t^4 \left(\sum_{f=0}^{\infty} |a_{fb}|\right)^4 \le c < \infty,
\end{aligned}$$

where the last two inequalities follow from Lemma A.2(c). □

Lemma A.4. (a) *Let $q_t^* = (e_t', \Delta w_t', z_t^*)'$. Then, for any j, $T^{-1}\sum_{t=j+1}^{T} q_t^* q_{t-j}^{*\prime} \to_p Eq_t^* q_{t-j}^{*\prime}$.* (b) *For θ defined in (3.4), $T^{-\frac{1}{2}-\theta}\sum_{t=1}^{T} z_{t-1}e_t \to_p 0$.* (c) *For any j, $T^{-1}\sum_{t=j+1}^{T} z_t e_{t-j}' \to_p Ez_t^* e_{t-j}'$.* (d) *For any j, $T^{-1}\sum_{t=j+1}^{T} z_t z_{t-j} \to_p Ez_t^* z_{t-j}^*$.*

Proof. (a) For moments involving e_t and Δw_t but not z_t^*, the result follows immediately under (2.4a) from Hannan (1970, p. 210) and under (2.4b) from White (1984, p. 47). But the same references apply as well to moments involving z_t^*. Under (2.4a), Lemma A.2(c) implies that z_t^*'s moving average weights are absolutely summable, which makes Hannan (1970, p. 210) applicable; under (2.4b), since Δw_t follows a finite parameter ARMA process, z_t is a function of a finite number of lags of Δw_t and e_t (Hansen and Sargent, 1981), implying that z_t also obeys the mixing and moment conditions of (2.4b).

(b) It is easy to show that $T^{-\frac{1}{2}-\theta}\sum_{t=1}^{T} z_{t-1}^* e_t$ converges in mean square and hence in probability to zero. Hence it suffices to show $T^{-\frac{1}{2}-\theta}\sum_{t=1}^{T}(z_{t-1} - z_{t-1}^*)e_t \to_p 0$. We have $T^{-\frac{1}{2}-\theta}\sum_{t=1}^{T}(z_{t-1} - z_{t-1}^*)e_t = T^{-\frac{1}{2}-\theta}\sum_{t=1}^{T}\{\sum_{j=0}^{\infty}[(a_{jb} - a_j^*)e_{t-j}]e_t\}$. The T terms in braces are pairwise uncorrelated. Assume e_t is a scalar for notational simplicity. Then the expectation of the square is $T^{-2\theta}[T^{-1}\sum_{t=1}^{T} E\{\sum_{j=0}^{\infty}[(a_{jb} - a_j^*)^2 e_{t-j}^2]e_t^2\}] = T^{-2\theta}E\sum_{j=0}^{\infty}[(a_{jb} - a_j^*)^2 e_{t-j}^2]e_t^2 \le T^{-2\theta}Ee_t^4\sum_{j=0}^{\infty}(a_{jb} - a_j^*)^2 \to 0$ by $\theta > 0$ and Lemma A.2(e).

(c) I will illustrate for given $j \ge 0$. Assume e_t is a scalar for simplicity. We wish to show $T^{-1}\sum_{t=j+1}^{T}(z_t - z_t^*)e_{t-j} \to_p 0$. We have

$$\begin{aligned}
T^{-1}\sum_{t=j+1}^{T}(z_t - z_t^*)e_{t-j} &= T^{-1}\sum_{t=j+1}^{T}\left[\sum_{k=0}^{\infty}(a_{kb} - a_k^*)e_{t-k}\right]e_{t-j} \\
&= \sum_{k=0}^{j}(a_{kb} - a_k^*)\left[T^{-1}\sum_{t=j+1}^{T} e_{t-k}e_{t-j}\right] \\
&\quad + T^{-1}\sum_{t=j+1}^{T}\left[e_{t-j}\sum_{k=j+1}^{\infty}(a_{kb} - a_k^*)e_{t-k}\right].
\end{aligned}$$

The first term converges in probability to zero because, for given k, $0 \le k \le j$, (i) $T^{-1}\sum_{t=j+1}^{T} e_{t-k}e_{t-j} \to_p Ee_{t-j}e_{t-k}$ ($=0$ if $j \ne k$, $= Ee_t^2$ if $j = k$), and (ii) $a_{kb} - a_k^* \to 0$. As for the second term, the $(T-j)$ terms $e_{t-j}\sum_{k=j+1}^{\infty}(a_{kb} - a_k^*)e_{t-k}$ are pairwise uncorrelated and the logic used in part (b) of this lemma delivers the result.

(d) I will illustrate for $j = 0$. We have $T^{-1}\sum_{t=1}^{T} z_t^2 = -T^{-1}\sum_{t=1}^{T}(z_t - z_t^*)^2 + 2T^{-1}\sum_{t=1}^{T} z_t(z_t - z_t^*) + T^{-1}\sum_{t=1}^{T} z_t^{*2}$, and the desired result follows if the first two terms converge in probability to zero. We have $T^{-1}\sum_{t=1}^{T}(z_t - z_t^*)^2 \to_p 0$ by Lemma A.2(f). Also $|T^{-1}\sum_{t=1}^{T} z_t(z_t - z_t^*)| \le (T^{-1}\sum_{t=1}^{T} z_t^2)^{\frac{1}{2}}[T^{-1}\sum_{t=1}^{T}(z_t - z_t^*)^2]^{\frac{1}{2}}$, and $T^{-1}\sum_{t=1}^{T} z_t(z_t - z_t^*) \to_p 0$ will follow if $T^{-1}\sum_{t=1}^{T} z_t^2 = O_p(1)$. But this does indeed follow, since $E[T^{-1}\sum_{t=1}^{T} z_t^2]^2 < c < \infty$ by Lemma A.3. □

Lemma A.5. (a) $T^{-\frac{1}{2}}\sum \eta_t^* \sim_A N(0, E\eta_t^{*2})$, $T^{-\frac{1}{2}}\sum z_{t-1}^*\eta_t^* \sim_A N(0, E(z_{t-1}^*\eta_t^*)^2)$. (b) $T^{-\frac{1}{2}}\sum z_{t-1}\eta_t - T^{-\frac{1}{2}}\sum z_{t-1}^*\eta_t^* \to_p 0$.

Proof. (a) Follows from White (1984, p. 124).

(b) Assume e_t is a scalar for notational simplicity. We have $T^{-\frac{1}{2}}\sum z_{t-1}\eta_t - T^{-\frac{1}{2}}\sum z_{t-1}^*\eta_t^* = T^{-\frac{1}{2}}\sum z_{t-1}(\eta_t - \eta_t^*) + T^{-\frac{1}{2}}\sum(z_{t-1} - z_{t-1}^*)\eta_t^*$. The first term is $T^{\theta}[\psi(b) - \psi(1)][T^{-\frac{1}{2}-\theta}\sum z_{t-1}e_t] \to_p 0$, since $T^{\theta}[\psi(b) - \psi(1)] \to c < \infty$ by Lemma A.2(b) and $T^{-\frac{1}{2}-\theta}\sum z_{t-1}e_t \to_p 0$ by Lemma A.4(b). The expectation of the square of the second term is $T^{-1}\sum E(z_{t-1} - z_{t-1}^*)^2\eta_t^{*2} = E(z_{t-1} - z_{t-1}^*)^2\eta_t^{*2} = |\sum\sum_{j,k=0}^{\infty} E(a_{jb} - a_j^*)(a_{kb} - a_k^*)e_{t-j-1}e_{t-k-1}\eta_t^{*2}| \le (Ee_t^4)^{\frac{1}{2}}(E\eta_t^{*4})^{\frac{1}{2}}\sum\sum_{j,k=0}^{\infty}|(a_{jb} - a_j^*)(a_{kb} - a_k^*)| \le (Ee_t^4)^{\frac{1}{2}}(E\eta_t^{*4})^{\frac{1}{2}}(\sum_{j=0}^{\infty}|a_{jb} - a_j^*|)^2 \to 0$ by Lemma A.2(e). □

Lemma A.6. *Let* $q_t^* = (e_t', \Delta w_t', z_t^*, \eta_t^*)'$, $q_t = (e_t', \Delta w_t', z_t, \eta_t)'$. *Then, for any* j, $T^{-1}\sum_{t=j+1}^{T} q_t q_{t-j}' \to_p Eq_t^* q_{t-j}^{*\prime}$.

Proof. Follows from Lemma A.4 and algebra similar to that used the proof of that lemma. □

Proof of Theorem 3.1 and comment 6 in Section 3. For algebraic simplicity, assume that it is known that $\mu = 0$ and the estimators are uncentered rather than centered, e.g., $\hat{\gamma}_j = T^{-1}\sum_{t=j+1}^{T}\Delta y_t\Delta y_{t-j}$; terms relating to the sample means of Δy_t and Δy_t^* can be handled by an argument such as that below. For θ defined in (3.4), we have

$$
\begin{aligned}
T^{\theta}(\hat{\gamma}_j - \hat{\gamma}_j^*) = T^{\theta}&\left\{T^{-1}\sum_{t=j+1}^{T}[\eta_t\eta_{t-j} - \eta_t^*\eta_{t-j}^*]\right\}\\
&+ T^{\theta}(1-b)\left[T^{-1}\sum_{t=j+1}^{T}\eta_t z_{t-1-j}\right]\\
&+ T^{\theta}(1-b)\left[T^{-1}\sum_{t=j+1}^{T}z_{t-1}\eta_{t-j}\right]\\
&+ T^{\theta}(1-b)^2\left[T^{-1}\sum_{t=j+1}^{T}z_{t-1}z_{t-1-j}\right]. \qquad \text{(A.3)}
\end{aligned}
$$

Let $\psi_i(b)$ and e_{it} denote typical elements of the n-dimensional vectors $\psi(b)$ and e_t. The first term on the right-hand side of (A.3) is $T^{\theta}\{T^{-1}\sum_{t=j+1}^{T}[\psi(b)e_t\psi(b)e_{t-j} - \psi(1)e_t\psi(1)e_{t-j}]\} = \sum\sum_{i,k=1}^{n}\{T^{\theta}[\psi_i(b)\psi_k(b) - \psi_i(1)\psi_k(1)]T^{-1}\sum_{t=j+1}^{T} e_{it}e_{kt-j}\}$. But, for each i and k, $T^{\theta}[\psi_i(b)\psi_k(b) - \psi_i(1)\psi_k(1)]$ converges to a finite constant by Lemma A.2(b), and $T^{-1}\sum_{t=j+1}^{T} e_{it}e_{kt-j} \to_p 0$ by Lemma A.6. Hence this first term converges in probability to zero. The second term converges in probability to zero, because $T^{\theta}(1-b) = \delta$, and, by Lemma A.6, $T^{-1}\sum_{t=j+1}^{T}\eta_t z_{t-1-j} \to_p E_t\eta_t^* z_{t-1-j}^* = 0$. By Lemma A.6, the third term converges in probability to $\delta E z_{t-1}^*\eta_{t-j}^*$. The fourth term converges in probability to zero, because $T^{\theta}(1-b)^2 \to 0$ and $T^{-1}\sum_{t=j+1}^{T} z_{t-1}z_{t-1-j} \to_p Ez_t^* z_{t-j}^*$ by Lemma A.6. □

Proof of Theorem 3.2. Similar to the proof of Theorem 3.1. □

Lemma A.7. *Suppose* $\mu = 0$. *Then* $T^{-2}\sum(y_{t-1}^*)^2 = O_p(1)$ *and* $T^{-1}\sum y_{t-1}^*\eta_t^* = O_p(1)$.

Proof. Phillips (1987). □

Proof of Theorem 4.1 and result (4.7). (a) Since $f_{t-1} - y_{t-1} = (1-b)z_{t-1}$, we have $T^{\frac{1}{2}-\theta}[T^{\theta}(1-b)](\hat{\beta}_r - 1) = [T^{-1}\sum(z_{t-1} - \bar{z})^2]^{-1} \times T^{-\frac{1}{2}}\sum(z_{t-1} - \bar{z})\eta_t$. From Lemma A.6, $T^{-1}\sum(z_{t-1} - \bar{z})^2 - T^{-1}\sum {z^*_{t-1}}^2 \to_p 0$, and, from Lemma A.5, $T^{-\frac{1}{2}}\sum(z_{t-1} - \bar{z})\eta_t - T^{-\frac{1}{2}}\sum z_{t-1}^*\eta_t^* \to_p 0$, implying that $T^{-\frac{1}{2}-\theta}[T^{\theta}(1-b)](\hat{\beta}_r - 1) = (T^{-1}\sum {z_{t-1}^*}^2)^{-1}(T^{-\frac{1}{2}}\sum z_{t-1}^*\eta_t^*) + o_p(1)$. Since $T^{\theta}(1-b) = \delta$ and, by Lemmas A.4 and A.5, $(T^{-1}\sum {z_{t-1}^*}^2)^{-1}(T^{-\frac{1}{2}}\sum z_{t-1}^*\eta_t^*) \sim_A N(0, V)$, $V = E(z_{t-1}^*\eta_t^*)^2/(E{z_{t-1}^*}^2)^2$, the result follows.

(b) Let $\theta \equiv (\alpha\beta_f)'$ denote the constant and slope, X the $T \times 2$ matrix whose t'th row is $(1 f_{t-1})$, N the $T \times 1$ vector whose t'th element is η_t, D a 2×2 diagonal matrix $\mathrm{diag}(T^{\frac{1}{2}}, T)$, $A = DX'XD$, $B = DX'N$. Let θ^*, X^*, N^*, A^* and B^* denote the corresponding quantities for the regression with ${y^*}_{t-1}$. Then $D(\hat{\theta} - \hat{\theta}^*) = A^{-1}B - A^{*-1}B^* = (A^{-1} - A^{*-1})B + A^{*-1}(B - B^*)$. Given that a standard argument (e.g., Hamilton, 1994, pp. 490–492 for the i.i.d. case) establishes that A^* is invertible for large T, it suffices to show that $A - A^* \to_p 0$ and $B - B^* \to_p 0$. I will illustrate by presenting the argument for (i) the (2, 2) element of $A - A^*$, which is $T^{-2}\sum f_{t-1}^2 - T^{-2}\sum {y_{t-1}^*}^2$, and (ii) the second element of $B - B^*$, which is $T^{-1}\sum f_{t-1}\eta_t - T^{-1}\sum {y^*}_{t-1}\eta_t^*$. It will be useful to note first that (2.1) may be rewritten as $y_t = c_y + x_{t-1} + \sum_{j=0}^{\infty} b^j E_t\Delta x_{t+j} \Rightarrow E_{t-1}y_t = c_y + x_{t-1} + \sum_{j=0}^{\infty} b^j E_{t-1}\Delta x_{t+j}$. Since $\sum_{j=0}^{\infty} b^j E_{t-1}\Delta x_{t+j} \equiv z_{t-1}$ when $\mu = 0$, we have, using Eqs. (4.1) and (4.3),

$$
f_{t-1} = c_y + x_{t-1} + z_{t-1} \Rightarrow f_{t-1} - y_{t-1}^* = c_y + z_{t-1} - z_{t-1}^*. \qquad \text{(A.4)}
$$

(i) $T^{-2}\sum f_{t-1}^2 - T^{-2}\sum {y_{t-1}^*}^2 = T^{-2}\sum(f_{t-1} - y_{t-1}^*)^2 + 2T^{-2}\sum[y_{t-1}^*(f_{t-1} - {y^*}_{t-1})]$. The desired result follows from (A.4), Lemma A.2(f) and $T^{-2}\sum[y_{t-1}^*(f_{t-1} - y_{t-1}^*)] \le [T^{-2}\sum(y_{t-1}^*)^2]^{\frac{1}{2}}[T^{-2}\sum(f_{t-1} - y_{t-1}^*)^2]^{\frac{1}{2}} \to_p 0$ by Lemmas A.7 and A.2(f).

(ii) $T^{-1}\sum f_{t-1}\eta_t - T^{-1}\sum y_{t-1}^*\eta_t^* = T^{-1}\sum(c_y + z_{t-1} - z_{t-1}^*)\eta_t + T^{-1}\sum y_{t-1}^*(\eta_t - \eta_t^*)$. The first term is $o_p(1)$ from Lemma A.2(f) and A.6, the second is $o_p(1)$ by Lemma A.7 and $\psi(b) - \psi(1) \to 0$. □

Proof of Theorem 4.2. Straightforward. □

References

Aloy, Marcel, Boutahar, Mohamed, Gente, Karine, Péguin-Feissolle, Anne, 2010. Fractional integration and cointegration in stock prices and exchange rates. Economics Bulletin 30 (1), 115–129. AccessEcon.

Bacchetta, Philippe, van Wincoop, Eric, 2006. Can information heterogeneity help explain the exchange rate determination puzzle? American Economic Review 96, 552–576.

Bacchetta, Philippe, van Wincoop, Eric, 2010. Infrequent portfolio decisions: a solution to the forward discount puzzle. American Economic Review 100, 837–869.

Baillie, Richard T., Bollerslev, Tim, 1994. The long memory of the forward premium. Journal of International Money and Finance 13 (5), 565–571.

Baillie, Richard T., Bollerslev, Tim, 2000. The forward premium anomaly is not as bad as you think. Journal of International Money and Finance 19 (4), 471–488.
Barberis, Nicholas, Thaler, Richard, 2003. A survey of behavioral finance. In: Constantinides, G., Stulz, R., Harris, M. (Eds.), Handbook of the Economics of Finance. North Holland, Amsterdam, pp. 1053–1124.
Bhar, Ramprasad, Kim, Suk-Joong, Pham, Toan M., 2004. Exchange rate volatility and its impact on the transaction costs of covered interest rate parity. Japan and the World Economy 16 (4), 503–525.
Campbell, John Y., Lo, Andrew W., MacKinlay, A. Craig, 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton.
Campbell, John Y., Shiller, Robert J., 1989. The dividend price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1 (3), 195–228.
Campbell, John Y., Thompson, Samuel, 2008. Predicting the equity premium out of sample: can anything beat the historical average? Review of Financial Studies 21 (4), 1509–1531.
Chinn, Menzie D., Meredith, Guy, 2004. Monetary policy and long-horizon uncovered interest parity. IMF Staff Papers 51 (3), 409–430.
Engel, Charles, Mark, Nelson C., West, Kenneth D., 2007. Exchange rate models are not as bad as you think. In: Acemoglu, D., Rogoff, K., Woodford, M. (Eds.), NBER Macroeconomics Annual 2007. University of Chicago Press, Chicago, pp. 381–443.
Engel, Charles, Wang, Jian, Wu, Jason, 2011. Long horizon forecasts of exchange rates when the discount factor is close to unity. Manuscript. University of Wisconsin.
Engel, Charles, West, Kenneth D., 2005. Exchange rates and fundamentals. Journal of Political Economy 113 (2), 485–517.
Evans, Martin D., Lewis, Karen K., 1995. Do long term swings in the dollar affect estimates of the risk premia? Review of Financial Studies 8, 709–742.
Fama, Eugene F., 1984. Forward and spot exchange rates. Journal of Monetary Economics 14 (3), 319–338.
Frankel, Jeffrey A., Froot, Kenneth A., 1987. Using survey data to test standard propositions regarding exchange rate expectations. American Economic Review 77, 133–153.
Goyal, Amit, Welch, Ivo, 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21 (4), 1455–1508.
Hamilton, James D., 1994. Time Series Analysis. Princeton University Press, Princeton.
Hannan, E.J., 1970. Multiple Time Series. John Wiley and Sons, New York.
Hansen, Lars Peter, Hodrick, Robert J., 1983. Risk averse speculation in the forward exchange market: an econometric analysis of linear models. In: Frankel, J. (Ed.), Exchange Rates and International Economics. University of Chicago Press, Chicago, pp. 113–142.
Hansen, Lars Peter, Sargent, Thomas J., 1981. A note on Weiner–Kolmogorov prediction formulas for rational expectations models. Economics Letters 8, 255–260.
Maynard, Alex, Phillips, Peter C.B., 2001. Rethinking an old puzzle: econometric evidence on the forward discount anomaly. Journal of Applied Econometrics 16, 671–708.
McCallum, Bennett T., 1994. A reconsideration of the uncovered interest parity relationship. Journal of Monetary Economics 33, 105–132.
Meese, Richard A., Rogoff, Kenneth, 1983. Empirical exchange rate models of the seventies: do they fit out of sample? Journal of International Economics 14, 3–24.
Molodtsova, Tanya, Papell, David H., 2009. Out-of-sample exchange rate predictability with Taylor rule fundamentals. Journal of International Economics 77, 167–180.
Phillips, Peter C.B., 1987. Time series regression with a unit root. Econometrica 55 (2), 277–301.
Samuelson, Paul A., 1965. Proof that properly anticipated prices fluctuate randomly. Industrial Management Review 6, 41–49.
Stambaugh, Robert F., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.
Verdelhan, Adrian, 2010. A habit based explanation of the exchange rate risk premium. Journal of Finance 65, 123–145.
White, Halbert, 1984. Asymptotic Theory for Econometricians. Academic Press, New York.