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 behavior in asset prices is not necessarily an implication of an efficient markets model such as the classic one expositions 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
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: $\Delta y_t = \text{const.} + \hat{\beta}_1 (f_{t-1} - y_{t-1}) + \text{residual}$; $y_t = \text{const.} + \hat{\beta}_m f_{t-1} + \text{residual}$; $y_t - y_{t-2} = \text{const.} + \hat{\beta}_n (f_{t-1} - y_{t-2}) + \text{residual}$.

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 (Bharat et al., 2004). I include this table not to present new empirical results, but to focus the econometric discussion.

---

### Table 1
Forward-spot regressions, US dollar.

<table>
<thead>
<tr>
<th>Country</th>
<th>$\hat{\beta}_1$</th>
<th>S.E.</th>
<th>$\hat{\beta}_m$</th>
<th>S.E.</th>
<th>$\hat{\beta}_n$</th>
<th>S.E.</th>
<th>Sample period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>-0.204</td>
<td>0.774</td>
<td>1.597</td>
<td>0.089</td>
<td>1.159</td>
<td>0.089</td>
<td>79:2–11:1</td>
</tr>
<tr>
<td>France</td>
<td>0.486</td>
<td>0.864</td>
<td>1.149</td>
<td>0.092</td>
<td>1.115</td>
<td>0.115</td>
<td>79:2–11:1</td>
</tr>
<tr>
<td>Germany</td>
<td>-0.767</td>
<td>0.818</td>
<td>1.054</td>
<td>0.027</td>
<td>0.108</td>
<td>0.108</td>
<td>79:2–11:1</td>
</tr>
<tr>
<td>Italy</td>
<td>1.697</td>
<td>0.806</td>
<td>1.239</td>
<td>0.030</td>
<td>1.016</td>
<td>0.094</td>
<td>79:2–11:1</td>
</tr>
<tr>
<td>Japan</td>
<td>-2.703</td>
<td>0.918</td>
<td>1.016</td>
<td>0.025</td>
<td>1.144</td>
<td>0.092</td>
<td>79:2–11:1</td>
</tr>
<tr>
<td>UK</td>
<td>-1.433</td>
<td>0.925</td>
<td>1.144</td>
<td>0.036</td>
<td></td>
<td></td>
<td>79:2–11:1</td>
</tr>
</tbody>
</table>

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: $\Delta y_t = \text{const.} + \hat{\beta}_1 (f_{t-1} - y_{t-1}) + \text{residual}$; $y_t = \text{const.} + \hat{\beta}_m f_{t-1} + \text{residual}$; $y_t - y_{t-2} = \text{const.} + \hat{\beta}_n (f_{t-1} - y_{t-2}) + \text{residual}$.
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

\[ y_t = c_y + (1 - b) \sum_{j=0}^{\infty} b^j E_x t_{t+j}. \] (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; \(E_x\) 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 \(\Delta y_t\) is predictable, then \(\Delta y_t\) will be predictable as well, with predictable movements in \(\Delta y_t\) offsetting predictable movements in \(\Delta x_t\) 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 \(y_t = \text{constant} + \sum_{j=0}^{\infty} b^j E_x t_{t+j}\). 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

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

\[ E e_t' e_t' = 0_{n \times n} \quad \text{for} \; t \neq s, \quad E e_t' e_t' \; \text{a constant matrix of rank} \; n; \]

\[ \Delta x_t = \alpha' \Delta w_t \quad \text{for a suitable} \; (n \times 1) \; \text{vector}\; \alpha, \; \alpha' \Psi (1) \; \neq 0. \]

Then, in (2.1), \(E_x t_{t+j}\) is defined as

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

The condition \(\alpha' \Psi (1) \neq 0\) says that \(x_t\) is \(l(1)\), i.e., \(\Delta x_t\) is not an overdiff erenced 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 overdiff erenced stationary variables in \(\Delta w_t\).

For a matrix \(A\), let \(\text{“} \cdot \text{“} \) denote maximum element, \(\| \cdot \|\) denote Euclidean norm. I assume either

\[ (i) \sum_{j=0}^{\infty} \| e_t \| < \infty. \] (2.4a)

(ii) \(e_t\) i.i.d. with finite fourth moments; or

(i) \(\Delta u_t\) is fourth-order stationary and follows a stationary finite-order ARMA process with

\[ E (e_t | 1, e_t, e_{t-1}, e_{t-2}, \ldots) = 0. \] (2.4b)

(ii) \(\lim_{t \to \infty} E || q_t ||_{4^k} < \infty \quad \text{for some} \; k > 4, \; \text{and with mixing coefficients of size} \; \psi (\psi - 1). \]

Equations (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 \(\Delta y_t\) into three parts: (1) a deterministic component, denoted \(\mu\), (2) an unpredictable component, denoted \(\eta_t\), and (3) a zero-mean component, denoted \((1 - b)z_{t-1}\), that, in general, will be predictable. The deterministic term \(\mu\) 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 \(\eta_t\) and \(z_{t-1}\). Specifically, define

\[ \mu = E \Delta x_t, \quad \eta_t = (\Delta y_t - E_{t-1} \Delta y_t), \]

\[ z_{t-1} = \sum_{j=0}^{\infty} b^j E_{t-1} (\Delta x_{t+j} - \mu). \] (2.5)

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

\[ \Delta y_t = (\Delta y_t - E_{t-1} \Delta y_t) + E_{t-1} \Delta y_t \]

\[ = \sum_{j=0}^{\infty} b^j (E_t - E_{t-1}) \Delta x_{t+j} + \mu + (1 - b) \sum_{j=0}^{\infty} b^j E_{t-1} (\Delta x_{t+j} - \mu) \]

\[ = \eta_t + [\mu + (1 - b)z_{t-1}]. \] (2.6)
Consider the random variables that results by setting $b = 1$ in the definitions of $\eta_t$ and $z_t$. Write the results as

$$
\Delta^*_t = \eta^*_t + \mu, \quad \eta^*_t \equiv \sum_{j=0}^{\infty} (E_t - E_{t-\gamma}) \Delta x_{t+j},
$$

(2.7)

$$
z_*^t \equiv \sum_{j=0}^{\infty} E_t (\Delta x_{t+j+1} - \mu).
$$

Assumption (2.4) ensures that $\eta^*_t$ and $z_*^t$ are stationary. For convenience of exposition, I refer to $\Delta^*_t$ 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 $T \to \infty$,

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

(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 $\sqrt{T} (1-b) \to \delta$. A common problem is the estimation and inference on $\delta$ and changes in $y_t$ affecting $y_t$ for $t \geq 1$. The rate $\sqrt{T}$, rather than $T^\alpha$ for arbitrary $0 < \alpha$, 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 $T^\alpha (1-b) = \delta$, for arbitrary $\alpha > 0$. 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 $T^\alpha (1-b) = \delta$ 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. 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 $\Delta y_t$ and the unobservable martingale difference $\Delta^*_t$. Under conventional asymptotics, sample autocovariances and autocorrelations of $\Delta y_t$ 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 $\Delta^*_t$, its autocovariances and autocorrelations are zero. If we were able to observe data on $\Delta^*_t$ to construct sample autocovariances and autocorrelations, we see that sample quantities converge in probability to the 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 $\Delta y_t$ converge in probability to zero, as do sample cross-covariances and cross-correlations between $\Delta y_t$ and lags of forecasting variables such as $\Delta x_t$. Thus these moments behave like those of the unobservable variable $\Delta^*_t$. In contrast, hypothesis tests do not behave like those applied to the moments of $\Delta^*_t$, instead, such tests have asymptotic probability of rejecting that is greater than nominal size but less than unity. The implication is that regressions of $\Delta y_t$ 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., $\hat{\Delta} y \equiv T^{-1} \sum_{t=1}^{T} \Delta y_t$. Let $\hat{\gamma}_j$ and $\hat{\gamma}^*_{jq}$ denote the usual estimates of the lag $j$ autocovariance of $\Delta y_t$ and $\Delta^*_t$ for $j \geq 0$:

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

(3.1)

$$
\hat{\gamma}^*_{jq} = T^{-1} \sum_{t=j+1}^{T} (\Delta^*_t - \hat{\Delta}^* y) (\Delta^*_t - \hat{\Delta}^* y).
$$

(3.1)

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

---

Table 2

<table>
<thead>
<tr>
<th>Variable/parameter</th>
<th>Equation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_t$</td>
<td>(2.1)</td>
<td>l(1) variable in present value determining log asset price $y_t$</td>
</tr>
<tr>
<td>$\Delta y_t, y_t$</td>
<td>(2.1)</td>
<td>Asset return, log asset price</td>
</tr>
<tr>
<td>$z_t$</td>
<td>(2.5)</td>
<td>Stationary fundamental that is proportional to the predictable component of $\Delta y_t$</td>
</tr>
<tr>
<td>$\eta_t$</td>
<td>(2.5)</td>
<td>Innovation in $\Delta y_t$ and $y_t$</td>
</tr>
<tr>
<td>$\Delta^<em>_t, \eta^</em>_t, \eta^*_t$</td>
<td>(2.7),(4.1)</td>
<td>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 $b &lt; 1$</td>
</tr>
<tr>
<td>$f_{t-1}$</td>
<td>(4.1)</td>
<td>$E_{t-1} y_t$</td>
</tr>
<tr>
<td>$b$</td>
<td>(2.1)</td>
<td>Discount factor in present value</td>
</tr>
<tr>
<td>$\delta$</td>
<td>(2.8)</td>
<td>Parameter determining how close discount factor is to unity</td>
</tr>
<tr>
<td>$\mu$</td>
<td>(2.5)</td>
<td>$E x_t$</td>
</tr>
</tbody>
</table>

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

---

4 It is straightforward to allow certain forms of minor dependence of the $x_t$ process on $b$. For example, suppose that $a$ depends on $b$. (Recall that $\Delta x_t = a' \Delta w_t$.)
Theorem 3.1. For \( j > 0 \), \( \sqrt{T}(\hat{\gamma}_j - \gamma_j^*) \to_p \delta \text{En}_j^* z_{t-1}^* \).

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

The term \( \text{En}_j^* z_{t-1}^* \) is the covariance between \( 1 \) the period lag of the random walk return \( \Delta y_{t-j} \) and \( \gamma_j^* \), see (2.7) and \( 2 \) the random walk fundamental \( z_{t-1}^* \). To see why the normalized discrepancy between \( \hat{\gamma}_j \) and \( \gamma_j^* \) is proportional to the covariance \( \text{En}_j^* z_{t-1}^* \), suppose that \( j = 1 \) and \( \mu = 0 \) for notational simplicity. Let “\( \text{cov} \)” denote a sample covariance. Then \( \hat{\gamma}_1 = \text{cov}(\gamma_1, \eta_{t-1}) = (1 - b)\eta_{t-1} \). Thus

\[
\sqrt{T}(\hat{\gamma}_1 - \gamma_1^*) = \sqrt{T}[\text{cov}(\eta_{t-1}, 1) - \text{cov}(\eta_{t-1}^*, 1)]
\]

\[
= \sqrt{T}(1 - b)\text{cov}(\eta_{t-1}, 1)
\]

\[
= \sqrt{T}(1 - b)^2\text{cov}(\eta_{t-1}, 1)
\]

Oversimplifying somewhat, the Appendix shows that, as \( T \to \infty \), \( z_t \) and \( \eta_t \) behave very much like \( z_{t-1}^* \) and \( \eta_{t-1}^* \) - indeed, so much so that to derive the limiting value of the terms on the right-hand side, one can replace \( z_t \) with \( z_{t-1}^* \) and \( \eta_t \) with \( \eta_{t-1}^* \). Hence the first term converges to zero, the second term to \( \delta_1^{En}_j z_{t-2}^* = \delta_e \); the final term is \( (1 - b) \) times a term that converges to \( \delta_2^{En}_j z_{t-2}^* \) and hence converges to zero because \( 1 - b \to 0 \). The only nonzero quantity comes from the third term, which converges to \( \delta_3^{En}_j z_{t-1}^* \).

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

Theorem 3.2. For \( j > 0 \), \( \sqrt{T}(\hat{y}_{j,y} - \gamma_{j,y}^*) \to_p \delta \text{Eu}_j z_{t-1}^* \).

To interpret Theorem 3.1, let \( \hat{\jmath}_j = \hat{\gamma}_{j,y} \) and \( \hat{\jmath}_{j,y}^* = \gamma_{j,y}^* / \gamma_{j,y}^* \) denote the sample autocorrelations of \( \Delta y_t \) and \( \Delta y_t \). Suppose that \( \eta_t^* \) is i.i.d., implying that \( \sqrt{T}\hat{\jmath}_j \to_h N(0,1) \). Then

\[
\text{Theorem 3.1 implies } \hat{\jmath}_j \approx N(k_1(1 - b)/\sqrt{T}, 1/T), \quad k = \text{En}_j^* z_{t-1}^* \gamma_{j,y}^*.
\]

Thus, for any \( T \), when \( T(1 - b) = \delta \), the asymptotic approximation implies that \( \hat{\jmath}_j \) is centered \( k \) standard deviations away from zero. Let \( \gamma \) 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

\[
\text{prob}(\gamma < -1.645 - k) + \text{prob}(\gamma > 1.645 - k) = k = \text{En}_j^* z_{t-1}^*/\gamma_{j,y}^*.
\]

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 \( \Delta y_t = \beta u_{t-j} + \text{residual} \). With \( \gamma \) a standard normal variable, the asymptotic size of a (say) 0.10 t-test for \( H_0 : \beta = 0 \) in this regression is

\[
\text{prob}(\gamma < -1.645 - k) + \text{prob}(\gamma > 1.645 - k) = k = \text{Eu}_j z_{t-1}^*/(\text{var}(u_{t-1}) \gamma_{j,y}^*).
\]

Theorem 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 \( u_{t-1} = n_{t-1} - x_{t-1} \) and write \( u_{t-1} = \gamma_{t-1} - y_{t-1} = u_{t-2} + \text{const.} + x_{t-1} + u_{t-1} \); compute the numerator of (3.3) as \( \delta k \cdot (\text{variance}(x_t) + \text{covariance}(x_t, u_{t-1})) \) and similarly for the denominator.
6. Consider a generalization of assumption (2.8):
\[ T^0 (1 - b) = \delta, \quad 0 < \delta < \infty, \quad 0 < \theta. \]  
(3.4)

Then Theorems 3.1 and 3.2 continue to hold with \( \sqrt{T} \) replaced by \( T^0 \): for \( j > 0 \), \[ T^0 (\hat{y}_j - \hat{y}_j^*) \rightarrow_p \delta \mathbb{E} \epsilon_{t-j} \Delta z_{t-j}^+ \rightarrow_p \delta \mathbb{E} \epsilon_{t-j} z_{t-j}^+ \]. Thus estimates of correlations with lagged variables tend to zero: for any \( \theta > 0 \) in (3.4), \( \hat{y}_j \rightarrow 0 \) (\( j \neq 0 \)), \( \hat{y}_{-j} \rightarrow 0 \) (\( j > 0 \)). 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. 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 \( \Delta y_t \) 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:
\[ f_{t-1} = E_{t-1} y_{t}. \]  
(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 \( \Delta y_t^* \) defined in (2.7) above. Begin by noting that, with a little algebra, it may be shown that \( \Delta y_t = \Delta x_t + b \Delta z_t \). So the level variable consistent with this model and given presample values \( y_0, x_0 \) and \( z_0 \) for \( t \geq 0 \) is
\[ y_t = c_0 + x_t + b z_t, \quad c_0 \equiv y_0 - x_0 - 2 b z_0. \]  
(4.2)

I define \( y_t^* \) by dropping \( c_0 \) and setting \( b = 1 \):
\[ y_t^* = x_t + z_t. \]  
(4.3)

An arbitrary constant (including \( c_0 \)) can be added to the right-hand side of (4.3) without changing any of the case; 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 \( \Delta y_t \) on \( f_{t-1} - y_{t-1} \), will yield a coefficient near unity with high probability, for sufficiently large \( T \). Write these regressions as
\[ \Delta y_t = \text{constant} + \beta f_{t-1} - y_{t-1} + \text{residual}, \]  
(4.4)
\[ y_t = \text{constant} + \beta f_{t-1} - y_{t-1} + \text{residual}. \]  
(4.5)

Let \( \hat{\beta} \) and \( \hat{\beta}_t \) denote the least squares estimators of the slope coefficients.

Theorem 4.1. \( a \) \( \hat{\beta}_t \) \( \equiv N(0, \sqrt{T}^2) \), with \( V > 0 \) the asymptotic variance of the estimator of the slope coefficient in a least squares regression of \( \Delta y_t \) on a constant and \( z_{t-1}^+ \).

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

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, \( T (\hat{\beta}_t - 1) \) is nondegenerate asymptotically, with a nonstandard distribution: \( \hat{\beta}_t^* \rightarrow 1 \) 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 \( \beta_t \) is superconsistent. To see why, recall from Eqs. (4.2) and (4.3) that \( y_t = c_0 + x_t + b z_t, y_t^* = x_t + z_t \). From the point of view of the (I) variables \( y_t \) and \( y_t^* \), the finite variance discrepancy between \( b z_t \) and \( z_t \) is very small for \( b \) near 1, so small that part (b) of the theorem is implied.

Part (a) suggests a possible reason that estimates of \( \hat{\beta}_t \) are not clustered near unity: the estimator is inconsistent. (Note that the quantity that is asymptotically normal is \( \hat{\beta}_t - 1 \) rather than the usual \( \sqrt{T} (\hat{\beta}_t - 1) \).) The reason for the inconsistency is that, under condition (2.8), the forward rate \( f_{t-1} \) approaches \( y_{t-1} \) as \( T \rightarrow \infty \). 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 \( \Delta y_t^* \) on \( z_{t-1}^+ \), observe that \( f_{t-1} - y_{t-1} = (1 - b) z_{t-1} + \mu \).

For large \( T \), a regression of \( \Delta y_t \) on a constant and \( z_{t-1} + \mu \) will behave like that of a regression of \( \Delta y_t^* \) on \( z_{t-1}^+ \) and hence imply a variance \( V \) for the regression coefficient; the actual regression is of \( \Delta y_t \) on \( (1 - b) z_{t-1} + \mu \), with the factor of \( (1 - b) \) contributing the factor of \( 1/b^2 \). The variance \( V/\sqrt{T}^2 \) of the limiting normal distribution of \( \beta_t - 1 \) is larger when \( \delta \) is smaller: the “closer” is to 0 (in the sense of small \( \delta \)), the noisier the slope coefficient. Though not stated in the theorem, the conventional t-test of \( H_0: \beta_t = 1 \) is well-behaved asymptotically, despite the inconsistency in estimation of \( \beta_t \).

The fact that the limiting (nondegenerate) distribution is centered around 1 means that Theorem 4.1(a) does not rationalize a systematic tendency for \( \beta_t \) 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 \( \beta_t \approx 1 \). 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 \( \Delta y_t \) is stationary. Maynard and Phillips (2001) develop an asymptotic theory that derives the implications for \( \beta_t \). That theory also implies that a regression of \( y_{t-1} - y_{t-2} \) on \( f_{t-1} - y_{t-2} \) will also produce a slope coefficient that
converges in probability to zero.\textsuperscript{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
\[ y_t - y_{t-1} = \text{constant} + \beta m (y_{t-1} - y_{t-2}) + \text{residual}. \] (4.6)

Let \( \beta_m \) be the least squares estimator of \( \beta_m \).

**Theorem 4.2.** \( \sqrt{T} (\hat{\beta}_m - 1) \sim N(0, V) \), with \( V > 0 \) the asymptotic variance 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 \( N(0,1) \).

Let \( \beta_1 \) be the least squares estimator of \( \beta_1 \).

Four comments.
1. **Theorem 4.2** and a generalization of **Theorem 4.1** obtain when one maintains (3.4), i.e., \( T^{1/2} (1 - b) = \delta \) for arbitrary \( \theta > 0 \). The generalization of **Theorem 4.1** is
\[ T^{1/2} (\hat{\beta}_1 - 1) \sim N(0, V/b^2). \] (4.7)

For \( \theta < \frac{1}{2} \), one implication of (4.7) is that \( \hat{\beta}_1 \rightarrow p 1 \); for \( \theta > \frac{1}{2} \), one implication of (4.7) is that \( \hat{\beta}_1 \rightarrow \beta_1 \) becomes increasingly disperse as \( T \) grows.

The \( \theta = \frac{1}{2} \) result stated above, without reference to \( \theta \) or \( \delta \), under each of the setups described in **Theorem 4.2**, holds as stated above, without reference to \( \theta \) or \( \delta \), under each of the setups described in 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 \( \Delta x_t \) was as follows:
\[ \Delta x_t = \Delta x_{t-1} + u_{t-1}. \] (5.1a)
\[ \Delta x_{t-1} = \varphi \Delta x_{t-2} + u_{t-2} + u_{t-1} - u_{t-2}. \] (5.1b)
\[ e_t = (u_{t-1}, u_{t-2}, y_{t-1}) \sim i.i.d. N(0, \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)). \] (5.1c)

It may be shown that the implied process for \( z_t \) (and thus for \( f_t - y_t = (1 - b)z_t \)) 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 component allows the first one or two autocorrelations of \( \Delta x_t \) and \( f_t - y_t \) to be well below \( \varphi \), even if \( \varphi \) 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 \( \Delta x_t \) does not vary with \( T \). But moments of \( \Delta y \) and \( y - \hat{y} \) do vary with \( T \). Table 3 and Table 3 describe one sense in which a random walk better approximates \( \Delta y_t \) as \( T \) grows with \( \sqrt{T} (1 - b) \) held fixed: the first-order autocorrelation of \( \Delta y_t \) falls from 0.08 to 0.03 to 0.01; the cross-correlation between \( \Delta y_t \) and \( f_t - y_t \) 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
Table 3
Population moments, baseline DGP.

<table>
<thead>
<tr>
<th>$T$</th>
<th>$\Delta y_t$, $\Delta x_{t-1}$</th>
<th>$f_t - y_t$, $f_{t-1} - y_{t-1}$</th>
<th>$\Delta y_t$, $\Delta y_{t-1}$</th>
<th>$\Delta y_t$, $\Delta x_{t-1}$</th>
<th>$\sigma_{\Delta y_t}/\sigma_{f_t - y_t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.49</td>
<td>0.95</td>
<td>0.08</td>
<td>0.11</td>
<td>6.2</td>
</tr>
<tr>
<td>400</td>
<td>0.49</td>
<td>0.95</td>
<td>0.08</td>
<td>0.05</td>
<td>12.4</td>
</tr>
<tr>
<td>1600</td>
<td>0.49</td>
<td>0.95</td>
<td>0.01</td>
<td>0.04</td>
<td>24.8</td>
</tr>
</tbody>
</table>

Notes:
1. The model is $y_t = (1 - b) \sum_{j=1}^{\infty} b^j E x_{t+j}$, with $x_{t+j} = E_{t+j} y_t$.
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

<table>
<thead>
<tr>
<th>$T$</th>
<th>$\hat{\rho}_1$</th>
<th>$\hat{\rho}_{1,y}$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_{1,y}$</th>
<th>$\hat{\beta}_n$</th>
<th>$\hat{\beta}_{n,y}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>-0.12</td>
<td>0.04</td>
<td>0.20</td>
<td>-0.15</td>
<td>0.03</td>
<td>0.21</td>
</tr>
<tr>
<td>400</td>
<td>-0.06</td>
<td>0.03</td>
<td>0.11</td>
<td>-0.08</td>
<td>0.03</td>
<td>0.13</td>
</tr>
<tr>
<td>1600</td>
<td>-0.03</td>
<td>0.01</td>
<td>0.05</td>
<td>-0.04</td>
<td>0.02</td>
<td>0.07</td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Notes:
1. The model and DGP are described in notes to Table 3.
2. The “$\infty$” lines give asymptotic values, computed analytically. All other results are based on 5000 repetitions.
3. $\hat{\rho}_1$ is the first-order autocorrelation of $\Delta y_t$; $\hat{\rho}_{1,y}$ is the correlation between $\Delta y_t$ and $\Delta x_{t-1}$; $\hat{\beta}_1$, $\hat{\beta}_{1,y}$ and $\hat{\beta}_n$, $\hat{\beta}_{n,y}$ are slopes in linear projections of $y_t - y_{t-1}$ on $f_{t-1} - y_{t-1}$, $y_t$ on $f_{t-1} - y_{t-1}$, 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 $\Delta y_t$ 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 $H_0 : \text{slope coefficient} = 1$ 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.

The main effect of increasing $T$ is to increase the standard deviation of $\Delta y_t$ relative to $f_{t-1} - y_{t-1}$.

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 $\Delta y_t$ and the correlation between $\Delta y_t$ and $\Delta x_{t-1}$ 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 $\Delta y_t$ and $\Delta x_{t-1}$, computed in accordance with (3.2) and (3.3) and reported in the line labelled “$\infty$”, 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 $\Delta y_t$ 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 $\hat{\beta}_1$ 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 “$\infty$” 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 $\hat{\beta}_1 = 0.36$ 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 $\hat{\beta}_1$. Suppose counterfactually that right-hand-side variable $f_t - y_t$ followed an AR(1) with parameter $\rho$ and i.i.d. normal innovation $u_t \equiv f_t - y_t = \rho (f_{t-1} - y_{t-1}) + u_t$. Then the results of Stambaugh (1999) apply, and $E \hat{\beta}_1 - 1 \approx -[\text{cov}(u_t, \eta_t)/\text{var}(u_t) (1 + 3 \rho)/T]$. In the DGPs in the simulations, $f_t - y_t$ does not follow an AR(1). But let $\rho$ be the first-order autocorrelation of $f_t - y_t$, define $u_t \equiv f_t - y_t - \rho (f_{t-1} - y_{t-1})$, and compute an approximate analytical bias as $-\text{cov}(u_t, \eta_t)/\text{var}(u_t) (1 + 3 \rho)/T$ 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.
Table 5
Simulation results, alternative DGPs, $T = 100$.

A. Population moments

<table>
<thead>
<tr>
<th>Correlations</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta x_t, \Delta x_{t-1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_t - y_t, f_{t-1} - y_{t-1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta y_t, \Delta y_{t-1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta y_t, f_{t-1} - y_{t-1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta y_t, \Delta y_{t-1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{y2} / \sigma_{y1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.49</td>
<td>0.95</td>
<td>0.08</td>
<td>0.16</td>
<td>0.11</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.40</td>
<td>0.77</td>
<td>0.13</td>
<td>0.18</td>
<td>0.09</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.32</td>
<td>0.94</td>
<td>0.02</td>
<td>0.06</td>
<td>0.03</td>
<td>15.8</td>
<td></td>
</tr>
</tbody>
</table>

B. Quantiles of point estimates

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP</td>
<td>$\hat{\beta}_f$</td>
<td>$\hat{\beta}_f$</td>
<td>$\hat{\beta}_m$</td>
<td>$\hat{\beta}_m$</td>
<td>$\hat{\beta}_m$</td>
<td>$\hat{\beta}_m$</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.50</td>
<td>0.95</td>
<td>0.05</td>
<td>0.50</td>
<td>0.95</td>
</tr>
<tr>
<td>A</td>
<td>$-0.12$</td>
<td>$0.04$</td>
<td>$0.20$</td>
<td>$-0.15$</td>
<td>$0.03$</td>
<td>$0.21$</td>
</tr>
<tr>
<td>B</td>
<td>$-0.08$</td>
<td>$0.10$</td>
<td>$0.27$</td>
<td>$-0.15$</td>
<td>$0.06$</td>
<td>$0.23$</td>
</tr>
<tr>
<td>C</td>
<td>$-0.15$</td>
<td>$0.01$</td>
<td>$0.16$</td>
<td>$-0.17$</td>
<td>$-0.01$</td>
<td>$0.15$</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_1 = 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{1,ys} = 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.200</td>
<td>0.100</td>
<td>0.100</td>
<td>0.100</td>
<td>0.100</td>
<td>0.100</td>
</tr>
<tr>
<td>$\beta_1 = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_m = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

The values for the median $\hat{\beta}_f$ are repeated from panel A in Table 4, and the bias is computed by subtraction from 1. The column $\hat{\beta}_m < 0$ gives the percentage of the 5000 samples in which the estimate of $\beta_m$ 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 $\hat{\beta}_f$ tends to be below one.

Columns (5) and (6) in Table 4 consider the regression of $y_t$ on $f_{t-1}$ and $y_{t-1}$ on $f_{t-2}$ and $y_{t-2}$. 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 simulations 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 $\Delta y_t$, 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 $\hat{\beta}_f$, with downward bias. (3) Good performance of estimates and hypothesis tests on $\hat{\beta}_f$ and $\hat{\beta}_m$.

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 → ∞. 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., Z = T^{-1} \sum z_{t-1}. For given b, 0 < b ≤ 1, and for ε, defined in (2.2), write

\[ z_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \quad z_t^* = \sum_{j=0}^{\infty} a_j^* \varepsilon_{t-j}, \]  

(A.1)

where a_0 and a_j^* are 1 × n, the “b” subscript in a_j emphasizes dependence of the moving average weights on b, and a_j^* ≡ a_{j1}. For Ψ and α defined in (2.2), define the 1 × n vectors ψ_j and (ψ b) as

\[ \psi_j = \alpha \Psi \varepsilon_j, \quad (\psi b) \equiv \alpha^\prime \sum_{j=0}^{\infty} b_j \psi_j \equiv \sum_{j=0}^{\infty} b_j \psi_j. \]  

(A.2)

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

Lemma A.1. (a) For a_j, a_j^*, ψ_j, and ψ (b) defined in (A.1) and (A.2),

\[ a_j = \sum_{j=0}^{\infty} b_j' \psi_{j+1+i}, \quad a_j^* = \sum_{j=0}^{\infty} \psi_{j+1+i} b_j. \]  

Proof. (a) For Δw_t defined in (2.2) and for 0 < b ≤ 1, let q_t = \sum_{j=0}^{\infty} a_j E_t(Δw_{t+j} − EΔw_t) = (say) \sum_{j=0}^{\infty} a_j E_t \psi_{t+j}. Then A_j = \sum_{j=0}^{\infty} 0 b_j \psi_{t+j}. Since z_t = \alpha E_t q_t, we have a_j ≡ 0 \alpha A_j b_j, which is the desired result.

(b) Follows from (E_t − E_{t-1})Δw_{t+j} = ψ_j \varepsilon_{t-j}, ∆x_t = α' Δw_t, and the fact that η_t = \sum_{j=0}^{\infty} b_j (E_t − E_{t-1}) Δx_{t+j}.

Lemma A.2. (a) \sum_{j=0}^{\infty} |j| |ψ_j| < ∞. (b) For θ defined in (3.4),

\[ T^θ(|ψ (b) − ψ (1)|) = \sum_{j=0}^{\infty} (b - 1) + b + \cdots + b^{-1} |ψ_j| = \sum_{j=0}^{\infty} [(b - 1) + b + \cdots + b^{-1}] |ψ_j| \]  

(b) follows from (A.2).

Proof. (a) For \sum_{j=0}^{\infty} |j| |ψ_j| < ∞. For θ defined in (3.4),

\[ T^θ(|ψ (b) − ψ (1)|) = \sum_{j=0}^{\infty} (b - 1) + b + \cdots + b^{-1} |ψ_j| \]  

From Lemma A.2(c), the follows immediately under (2.4a) from Hansen (1970, p. 210) and under (2.4b) from White (1984, p. 47). But the same references apply as well to moments involving z_j. Under (2.4a) (c) implies that z_j’s moving average weights are absolutely summable, which makes Hansen (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 ε_t (Hansen and Sargent, 1981), implying that z_t also obeys the mixing and moment conditions of (2.4b).

Lemma A.3. For \ell = 1, 2, 3, 4, let the scalar q_{\ell} denote either z_\ell or an element of e_\ell. Let p denote a fourfold product of z and of elements of e, at arbitrary dates, t, t − m, t − j and t − k : \[ p = q_{\ell} e_{\ell-1} q_{\ell-1} e_{\ell-2} q_{\ell-2} e_{\ell-3} k \]  

for any \ell, where \ell = \ell; 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

\[ \mathbb{E} \left[ E_t z_{t-\ell} e_{t-\ell} e_{t-k} \right] = \mathbb{E} \left[ \sum_{j=0}^{\infty} \sum_{j=0}^{\infty} \sum_{j=0}^{\infty} \sum_{j=0}^{\infty} a_j a_j^* \psi_j \psi_j^* \right] \]  

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

Lemma A.4. (a) Let q_{\ell} = (e_{\ell-1}, Δw_{\ell-1}, z_\ell^*). Then, for any j, \[ T−1 \sum_{\ell=j+1}^{\infty} q_{\ell} q_{\ell-1} \rightarrow p \mathbb{E} q_{\ell} q_{\ell-1}. \]  

(b) For θ defined in (3.4), \[ T^{-1/θ} \sum_{\ell=j+1}^{\infty} z_{\ell}^* e_{\ell-1}^* \rightarrow \mathbb{E} z_{\ell}^* e_{\ell-1}^*. \]  

Proof. (a) For moments involving ε_t and Δw_t but not z_j, the result follows immediately under (2.4a) from Hansen (1970, p. 210) and (2.4b) from White (1984, p. 47). But the same references apply as well to moments involving z_j. Under (2.4a) (c) implies that z_j’s moving average weights are absolutely summable, which makes Hansen (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 ε_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 \[ T^{-\frac{1}{θ}} \sum_{\ell=j}^{\infty} z_{\ell}^* e_{\ell-1} \]  

converges in mean square and hence in probability to zero. Hence it suffices to show \[ T^{-\frac{1}{θ}} \sum_{\ell=j}^{\infty} z_{\ell}^* e_{\ell-1} \rightarrow p 0. \]  

We have \[ T^{-\frac{1}{θ}} \sum_{\ell=j}^{\infty} (z_{\ell} - z_{\ell-\ell}) e_{\ell-1} \rightarrow p 0. \]  

The terms in braces are pairwise uncorrelated. Assume ε_t is a scalar for notational simplicity. Then the expectation of the square is \[ T^{-2θ} \sum_{\ell=j}^{\infty} (z_{\ell} - z_{\ell-\ell})^2 e_{\ell-1}^2 \right| = T^{-2θ} \mathbb{E} (z_{\ell} - z_{\ell-\ell})^2 e_{\ell-1}^2 \rightarrow 0 \]  

by θ > 0 and Lemma A.2(e).

(c) I will illustrate for given j ≥ 0. Assume ε_t is a scalar for simplicity. We wish to show \[ T^{-1} \sum_{\ell=j+1}^{\infty} (z_{\ell} - z_{\ell-\ell}) e_{\ell-1} \rightarrow p 0. \]  

We have

\[ T^{-1} \sum_{\ell=j+1}^{\infty} (z_{\ell} - z_{\ell-\ell}) e_{\ell-1} = \sum_{\ell=j}^{\infty} \left[ \sum_{k=0}^{\infty} (a_k - a_0^*) e_{\ell-k} \right] e_{\ell-1} \]  

\[ = \sum_{k=0}^{\infty} (a_k - a_0^*) \sum_{\ell=j+1}^{\infty} \left[ \sum_{k=0}^{\infty} (a_k - a_0^*) e_{\ell-k} \right] e_{\ell-1} \]  

\[ + \sum_{k=0}^{\infty} \sum_{\ell=j+1}^{\infty} (a_k - a_0^*) e_{\ell-1} e_{\ell-k} \]  

for any j, T−1 \sum_{\ell=j+1}^{\infty} z_{\ell}^* e_{\ell-1}^* \rightarrow p E z_{\ell}^* e_{\ell-1}^*.
The first term converges in probability to zero because, for given
$k, 0 \leq k \leq j, (1) T^{-1} \sum_{t=j+1}^{T} e_{t-k} e_{t-j} \rightarrow P E e_{t-k} e_{t-j} (0 = \text{if } j \neq k, = E e_{t-k}^2 \text{ if } j = k), \text{and (ii) } a_{t-k} \rightarrow 0$. For the second term, the
$(T-j)$ terms $e_{t-k} \sum_{k=1}^{j} a_{t-k} e_{t-k} = \text{pairwise uncorrelated and the}
\text{log used in part (b) of this lemma delivers the result.}$

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

Lemma A.5. (a) $T^{-\frac{1}{2}} \sum_{t=1}^{T} f(t) \rightarrow N(0, E f(t)^2)$, $T^{-\frac{1}{2}} \sum_{t=1}^{T} z_{t}^2 \rightarrow \bar{z}_{t}^2 \rightarrow N(0, E (z_{t}^2 - \bar{z}_{t}^2)^2)$. (b) $T^{-\frac{1}{2}} \sum_{t=1}^{T} z_{t} \rightarrow T^{-\frac{1}{2}} \sum_{t=1}^{T} z_{t} \rightarrow \bar{z}_{t}$.

Proof. (a) Follows from White (1984, p. 124).
(b) Assume $\hat{e}_{t}$ is a scalar for notational simplicity. We have $T^{-\frac{1}{2}} \sum_{t=1}^{T} z_{t} \rightarrow \bar{z}_{t}^2 \rightarrow T^{-\frac{1}{2}} \sum_{t=1}^{T} z_{t} \rightarrow T^{-\frac{1}{2}} \sum_{t=1}^{T} \hat{z}_{t} - \bar{z}_{t}^2 \hat{z}_{t} - \bar{z}_{t}^2 \rightarrow (T^{-\frac{1}{2}} \sum_{t=1}^{T} \hat{z}_{t} - \bar{z}_{t}^2 \hat{z}_{t} - \bar{z}_{t}^2)$.

Lemma A.6. Let $q_{t} = (\hat{e}_{t}, \Delta w_{t}, z_{t}^2, \bar{z}_{t}^2)$, $q_{t} \rightarrow (\hat{e}_{t}, \Delta w_{t}, z_{t}, \bar{z}_{t})$. Then, for any $j$, $T^{-1} \sum_{t=1}^{T} q_{t} \rightarrow N(0, E q_{t}^2)$.

Proof. Follows from Lemma A.4 and algebra similar to that used to
prove 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., $\hat{y}_{t} = T^{-1} \sum_{t=1}^{T} \Delta y_{t}$, $\partial y_{t}=\Delta y_{t}$, $\partial y_{t}=\Delta y_{t}$, terms relating to the sample means of $\Delta y_{t}$ and $\Delta y_{t}$ can be handled by an argument such as that below. For $\theta$ defined in
Lemma A.2(b), and $T^{-1} \sum_{t=1}^{T} e_{t-k} e_{t-j} \rightarrow 0$ by
Lemma A.6. Hence this first term converges in probability to zero.

The second term converges in probability to zero, because $T^{\delta} (1 - b) = \delta$, and, by Lemma A.6, $T^{-1} \sum_{t=1}^{T} \epsilon_{t-j} \rightarrow P E \epsilon_{t-j} (0)$. By Lemma A.6, the third term converges in probability to $\delta E \epsilon_{t-j} \rightarrow 0$. The fourth term converges in probability to zero, because $T^{\delta} (1 - b)^2 \rightarrow 0$ and $T^{-1} \sum_{t=1}^{T} |z_{t} \rightarrow 0|, and, from Lemma A.5, $T^{-1} \sum_{t=1}^{T} \bar{z}_{t} \rightarrow \bar{z}_{t}$.

References


