Econometric Reviews, 34(1–2):32–55, 2015

ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474938.2014.944467

Factor Model Forecasts of Exchange Rates

Charles Engel[1], Nelson C. Mark[2], and Kenneth D. West[1]

[1]*Department of Economics, University of Wisconsin, Madison, Wisconsin, USA*
[2]*Department of Economics, University of Notre Dame, Notre Dame, Indiana, USA*

We construct factors from a cross-section of exchange rates and use the idiosyncratic deviations from the factors to forecast. In a stylized data generating process, we show that such forecasts can be effective even if there is essentially no serial correlation in the univariate exchange rate processes. We apply the technique to a panel of bilateral U.S. dollar rates against 17 Organisation for Economic Co-operation and Development countries. We forecast using factors, and using factors combined with any of fundamentals suggested by Taylor rule, monetary and purchasing power parity models. For long horizon (8 and 12 quarter) forecasts, we tend to improve on the forecast of a "no change" benchmark in the late (1999–2007) but not early (1987–1998) parts of our sample.

Keywords Exchange rates; Factor models; Forecasting.

JEL Classification F31; F37.

1. INTRODUCTION

In predictions of floating exchange rates between countries with roughly similar inflation rates, a random walk model works very well. The random walk forecast is one in which the (log) level of the nominal exchange rate is predicted to stay at the current log level; equivalently, the forecast is one of "no change" in the exchange rate. This forecast works well at various horizons, from one day to three years. It does well in the following sense: the out of sample mean squared (or mean absolute) error in predicting exchange rate movements generally is about the same, and often smaller, than that of models that use "fundamentals" data on variables such as money, output, inflation, productivity, and interest rates. Classic references are Meese and Rogoff (1983a,b); a recent update is Cheung et al. (2005).

Whether or not this stylized regularity is bad news for economic theory is unclear. Some economists think the regularity is very bad news. Bacchetta and van Wincoop

Address correspondence to Kenneth West, Department of Economics, University of Wisconsin, 1180 Observatory Dr., Madison, WI 53706, USA; E-mail: kdwest@wisc.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lecr.

(2006, p. 552) describe it as "...the major weakness of international macroeconomics." On the other hand, Engel and West (2005) argue that a near random walk is expected under certain conditions.[1]

Whether or not one thinks the empirical finding of near random walk behavior is bad news for economic theory, it is of interest to try to tease out connections (if any) between a given exchange rate and other data. A small literature has used panel data techniques to forecast exchange rates, finding relatively good success (Mark and Sul, 2001; Rapach and Wohar, 2004; Groen, 2005; Engel et al., 2008). A very large literature has found that factor models do a good job forecasting basic macro variables.[2] The present paper predicts exchange rates, via factor models, in the context of panel data estimation, and compares the predictions to those of a random walk via root mean squared prediction error.

The panel consists of quarterly data on 17 bilateral U.S. dollar exchange rates with Organisation for Economic Co-operation and Development (OECD) countries, 1973–2007. We construct factors from the exchange rates. We take the literature on predicting exchange rates to suggest that the exchange rate series *themselves* have information that is hard to extract from observable fundamentals. This information might be hard to extract because standard measures of fundamentals (e.g., money supplies and output) are error ridden, or because we simply lack any direct measures of nonstandard fundamentals such as risk premia or noise trading.[3]

We compare four different forecasting models to a benchmark model that makes a "no-change" forecast—the random walk model. One of our four models uses factors but no other variables to forecast. The other three use factors along with some measures of observable fundamentals. The three measures of observable fundamentals are as follows: (1) those of a "Taylor rule" model; (2) those of a monetary model; and (3) deviations from purchasing power parity (PPP). Our measure of forecasting performance is root mean squared prediction error (root MSPE).

On balance, these models have lower MSPE than does a random walk model for long (8 and 12 quarter) horizon predictions over the late part of our forecasting sample (1999–2007). These differences, however, are usually not significant at conventional levels. Predictions that span the entire two decades (1987–2007) or the early part (1987–1998) of our forecast sample generally have higher MSPE than does a no-change forecast. (Different samples involve different currencies, because of the introduction of the Euro in 1999.) The basic factor model and the factor model supplemented by PPP fundamentals

[1]The Engel and West (2005) argument is *not* that a random walk is produced by an efficient market; indeed the simple efficient markets model implies that exchange rate changes are predicted by interest rate differentials and any variables correlated with interest rate differentials. Rather, the argument relates to the behavior of an asset price that is determined by a present value model with a discount factor near 1.

[2]See Stock and Watson (2006). We use "factor" to refer to a data generating process driven by factors, even if the estimation technique involves principal components.

[3]See Diebold et al. (1994) for another attempt, with methodology very different from ours, to predict exchange rates using a cross section of exchange rates.

do best. We recognize that the good performance in the recent period may be ephemeral. But we are hopeful that our approach will prove useful in other datasets.

We close this introduction with some cautions. First, we make no attempt to justify or defend the use of out of sample analysis. We and others have found such analysis useful and informative. But we recognize that some economists might disagree. Second, judgment (sometimes rather arbitrary) has been used at various stages, so we are not (yet) proposing a completely replicable strategy. Third, our exercise is not "true" out of sample. For example, revised rather than real time data are used in some specifications. We use revised data because we are using out of sample analysis as a model evaluation tool, and the models presume that the best available data are used. More importantly, perhaps, our exercise is not true out of sample because we have relied on research that has already examined exchange rates during parts of our forecasting sample. The most pertinent reference is Engel et al. (2008), which used similar data, spanning 1973–2005 (vs. 1973–2007). Finally, we limit ourselves to simple linear models; papers such as Bulut and Maasoumi (2012) suggest that such models miss essential features of exchange rate data.

Section 2 presents a stylized model that illustrates analytically why our approach might predict well. Section 3 describes our empirical models, Section 4 our data and forecast evaluation techniques. Section 5 presents empirical results, Section 6 robustness checks, and Section 7 concludes. An appendix includes some algebraic details. Some additional appendices, available on request, present detailed empirical results omitted from the paper to save space.

2. WHY A FACTOR MODEL MAY FORECAST WELL

In this section, we present a factor model and a simple data generating process that motivates its use.

Our basic presumption is that the deviation of the exchange rate from a measure of central tendency will help predict subsequent movements in the exchange rate. Algebraically, let

$$s_{it} = \text{log of exchange rate in country } i \text{ in period } t, \tag{2.1}$$

$$z_{it} = \text{measure of central tendency defined below.}$$

For concreteness, we note that s_{it} is measured as log(foreign currency units/U.S. dollar), though that is not relevant to the present discussion. Algebraically, our basic presumption is that for a horizon h, $s_{it+h} - s_{it}$ can be predicted by $z_{it} - s_{it}$, maybe using two different measures of z in a single regression.

Many papers have relied on the same presumption (that $s_{it+h} - s_{it}$ can be predicted by $z_{it} - s_{it}$). For example, Mark (1995) sets z_{it} in accordance with the "monetary model," so that z_{it} depends on money supplies and output levels; Molodtsova and Papell (2008) set z_{it} in accordance with a "Taylor rule" model, so that z_{it} depends on the exchange

rate, inflation rates, output gaps, and parameters of monetary policy rules; Engel et al. (2008) set z_{it} in accordance with PPP, so that z_{it} depends on price levels. Some papers (see references above) have used these specifications of z_{it} in the context of panel data. Our twist is to construct one measure of z_{it} from factors estimated from the panel of exchange rates.

To exposit the idea, consider the following example. Suppose that the ith exchange rate follows the process

$$s_{it} = F_{it} + v_{it}. \tag{2.2}$$

Here, F_{it} is the effect the factor has on currency i; in a one factor model, for example $F_{it} = \delta_i f_{1t}$ where f_{1t} is the factor and δ_i is the factor loading for currency i. The idiosyncratic shock v_{it} is uncorrelated with F_{it}. For simplicity, make as well some further assumptions not required in our empirical work, namely, that F_{it} follows a random walk and that v_{it} is independent and identically distributed (i.i.d.):

$$F_{it} = F_{it-1} + \varepsilon_{it}, \quad \varepsilon_{it} \sim \text{i.i.d. } (0, \sigma_\varepsilon^2), \quad v_{it} \sim \text{i.i.d. } (0, \sigma_v^2), \quad E\varepsilon_{it} v_{is} = 0 \text{ all } t, s. \tag{2.3}$$

An i subscript is omitted from the variances σ_ε^2 and σ_v^2 for notational simplicity.

Then $\Delta s_{it} = \varepsilon_{it} + v_{it} - v_{it-1}$ and the univariate process followed by Δs_{it} is clearly an MA(1), say

$$\Delta s_{it} = \eta_{it} + \theta \eta_{it-1}, \quad E\eta_{it}^2 \equiv \sigma_\eta^2, \quad |\theta| < 1. \tag{2.4}$$

Here, η_{it} is the Wold innovation in Δs_{it}. The variance of η_{it} and the value of θ can be computed in straightforward fashion from the values of σ_ε^2 and σ_v^2.[4]

Let us compare population forecasts of Δs_{it+1} using the factor model (2.2), the MA(1) model (2.4), and a random walk model. As above, let "MSPE" denote "mean squared prediction error." Unless otherwise stated, in this section MSPE refers to a population rather than sample quantity. (This contrasts to the discussion of our empirical work below, in which MSPE refers to a sample quantity.) To forecast using (2.2), observe that $\Delta s_{it+1} = \Delta F_{it+1} + \Delta v_{it+1} = \varepsilon_{it+1} + v_{it+1} - v_{it} \Rightarrow E_t \Delta s_{it+1} = -v_{it} \equiv F_{it} - s_{it} \Rightarrow$

$$\text{forecast error from factor model} = \varepsilon_{it+1} + v_{it+1}, \quad \text{MSPE}_{factor} = \sigma_\varepsilon^2 + \sigma_v^2. \tag{2.5}$$

The MSPE from the univariate model (2.4) is of course σ_η^2. With a little bit of algebra, and using the formula for σ_η^2 given in the previous footnote, it may be shown that $\sigma_\varepsilon^2 + \sigma_v^2 < \sigma_\eta^2$. Hence the factor model has a lower MSPE than the MA(1) model.

But the relevant issue is whether the improvement (i.e., the fall) in the MSPE is notable, for a plausible data generating process. A plausible data generating process (DGP) would

[4]Specifically, let $\gamma = \sigma_\varepsilon^2 + 2\sigma_v^2$ denote the variance of Δs. Then $\sigma_\eta^2 = 0.5[\gamma + (\gamma^2 - 4\sigma_v^4)^{1/2}]$, $\theta = -\sigma_v^2/\sigma_\eta^2$.

be one in which there is very little serial correlation in Δs_{it}. Put differently, if the DGP is such that the MSPE from the MA(1) model is essentially the same as that from a random walk model, is it still possible that the MSPE from the factor model is substantially smaller than that of the random walk?

In our empirical work, we use Theil's U-statistic to compare (sample) MSPEs relative to that of a random walk. These are square roots of the following ratio: sample MSPE alternative forecast/sample MSPE forecast of no change. The forecast error of the random walk model is the actual change in $\Delta s_{it} = \varepsilon_{it} + v_{it} - v_{it-1}$; the corresponding population MSPE is $\sigma_\varepsilon^2 + 2\sigma_v^2$. In the context of the present section (population rather than MSPEs), define population U-statistics as

$$\mathrm{U}_{factor} = [(\sigma_\varepsilon^2 + \sigma_v^2)/(\sigma_\varepsilon^2 + 2\sigma_v^2)]^{1/2}, \qquad \mathrm{U}_{MA} = [\sigma_\eta^2/(\sigma_\varepsilon^2 + 2\sigma_v^2)]^{1/2}. \tag{2.6}$$

For select values of the first order autocorrelation of Δs_{it} (which is approximately the MA parameter θ introduced in (2.4)), these are as follows:

$corr(\Delta s_{it}, \Delta s_{it-1})$	−0.01	−0.02	−0.03	−0.04	−0.05	−0.10,	(2.7a)
U_{MA}	0.99995	0.9998	0.9996	0.9992	0.9987	0.9949,	(2.7b)
U_{factor}	0.995	0.990	0.985	0.980	0.975	0.949.	(2.7c)

The factor model improves on the moving average model by an order of magnitude. For example, when the first order autocorrelation of Δs_{it} is −0.10, the population root MSPE for the MA model is only 0.5% lower than for the random walk (because 0.9949 is about 0.5% smaller than 1), while the population root MSPE for the factor model is about 5% lower than the random walk.

We further note that the population values for U_{MA} in (2.7b) are so near 1 that in practice, for any of the values of the autocorrelations, one would not be surprised if sampling error in estimation of an MA(1) model led to sample U-statistics above 1. Hence, we view the figures in the table as consistent with the well-established finding that no univariate model predicts better than a random walk. But clearly the factor model has the potential to predict better, even accepting the point that in practice the best univariate model is a random walk.

In the simple DGP consisting of (2.2) and (2.3), whether a factor model will have lower MSPE than a model that uses information not only on exchange rates but also on fundamentals such as prices and output depends on whether the additional variables help pin down v_{it}. To allow for this possibility, our empirical work combines factors with observable fundamentals, as discussed in the next section.

3. EMPIRICAL MODELS

We use models with one, two or three factors. We will use the three factor model for illustration. The one and two factor models are analogous. In the three factor model,

we first estimate a set of three factors and factor loadings from the exchange rates. For currency i, $i = 1, \ldots, 17$, the model is

$$\begin{aligned} s_{it} &= \text{constant} + \delta_{1i} f_{1t} + \delta_{2i} f_{2t} + \delta_{3i} f_{3t} + v_{it} \\ &\equiv \text{constant} + F_{it} + v_{it}. \end{aligned} \tag{3.1}$$

The factors (the f's) are unobserved I(1) variables. Here and throughout, we do not attempt to test for unit roots in the factors or any other variable for that matter. See Bai (2004) on estimation of factor models with unit root data.

Let $F_{it} = \delta_{1i} f_{1t} + \delta_{2i} f_{2t} + \delta_{3i} f_{3t}$. We aim to use (estimates of) F_{it} to forecast s_{it}. In contrast to much work with factors, the factors are *not* constructed from a set of additional variables. For example, in Groen's (2006) work on exchange rates, factors are constructed from data on real activity and prices, and take the place of traditional measures of real and nominal activity. We take the literature on predicting exchange rates to suggest that the exchange rate series *themselves* may have low frequency information on common trends that is hard to extract from observable fundamentals. This low frequency information might be buried in standard, but noisy, measures of fundamentals such as relative money supplies and relative outputs. Or this information might be embedded in nonstandard measures of fundamentals that are sufficiently persistent that they function in part to drive common trends; examples are persistent risk premia or persistent noise trading. Put differently, we use factors to parsimoniously capture co-movements of exchange rates that are not well-captured by observable fundamentals. Crudely, we posit that a weighted average of the log levels of exchange rates represents a central tendency for the log level a given exchange rate, and use this weighted average to help forecast.

Mechanics are as follows. We assume that the factors component soaks up a common unit root component in the s's. That is, we assume that $F_{it} - s_{it}$ is stationary and may be useful in predicting (stationary) future changes in s_{it}. We do not attempt to test stationarity of v_{it}; our selection of number of factors was based on presumed limitations of a panel of cross-section dimension 17.[5] The factors f_{1t}, f_{2t}, and f_{3t} are uncorrelated by construction. We normalize the f's to have mean zero and unit variance.

So this first stage produces a time series for $\hat{f}_{1t}$, $\hat{f}_{2t}$, and $\hat{f}_{3t}$ and factor loadings, $\hat{\delta}_{1i}$, $i = 1, \ldots, 17$; $\hat{\delta}_{2i}$, $i = 1, \ldots, 17$; $\hat{\delta}_{3i}$, $i = 1, \ldots, 17$. In our simplest specification, the measure of central tendency z_{it} that was introduced in (2.1) is

$$z_{it} = \hat{\delta}_{1i} \hat{f}_{1t} + \hat{\delta}_{2i} \hat{f}_{2t} + \hat{\delta}_{3i} \hat{f}_{3t} \equiv \widehat{F}_{it}. \tag{3.3}$$

[5]In principle, one could use techniques to determine cointegrating rank to determine the number of I(1) factors and the factor loadings. We take results such as Ho and Sørensen (1996) to indicate that the finite sample performance of such techniques is likely to be poor, when the cross-section dimension is 17.

In this simplest specification, with $\widehat{F}_{it} \equiv \hat{\delta}_{1i}\hat{f}_{1t} + \hat{\delta}_{2i}\hat{f}_{2t} + \hat{\delta}_{3i}\hat{f}_{3t}$, for quarterly horizons $h = 1, 4, 8$, and 12 we use a standard panel data estimator (least squares with dummy variable) to estimate and forecast. For example, with a horizon of $h = 4$ quarters, we estimate

$$s_{it+4} - s_{it} = \alpha_i + \beta(\widehat{F}_{it} - s_{it}) + u_{it+4}, \tag{3.4}$$

where α_i is a fixed effect for country i. We then use $\hat{\alpha}_i$ and $\hat{\beta}$ to predict (some details below). Here, and in all of our specifications, there is no "time effect" (in the jargon of panel data): the factors are dynamic versions of time effects (we hope).

Our three other specifications combine factors with observables. The other three specifications are all of the form

$$s_{it+h} - s_{it} = \alpha_i + \beta(\widehat{F}_{it} - s_{it}) + \gamma(z_{it} - s_{it}) + u_{it+h}, \tag{3.5}$$

for three different z_{it}'s. Let country 0 refer to the USA. The three z_{it}'s are as follows:

$$\text{Taylor rule:} \quad z_{it} = 1.5(\pi_{it} - \pi_{0t}) + 0.5(\tilde{y}_{it} - \tilde{y}_{0t}) + s_{it}; \quad \pi = \text{inflation}, \quad \tilde{y} = \text{output gap};$$
$$z_{it} - s_{it} = 1.5(\pi_{it} - \pi_{0t}) + 0.5(\tilde{y}_{it} - \tilde{y}_{0t}); \tag{3.6}$$

$$\text{monetary model:} \quad z_{it} = (m_{it} - m_{0t}) - (y_{it} - y_{0t}); \quad m = \ln(\text{money}), \quad y = \ln(\text{output}); \tag{3.7}$$

$$\text{PPP model:} \quad z_{it} = p_{it} - p_{0t}; \quad p = \ln(\text{price level}). \tag{3.8}$$

The Taylor rule model builds on the recently developed view that interest rates rather than money supplies are the instrument of monetary policy. Expositions may be found in Benigno and Benigno (2006), Engel and West (2006), and Mark (2008). The monetary model was for many years the workhorse of international monetary economics; see, for example the textbook exposition in Mark (2001) or the abbreviated summary in Engel and West (2005). The PPP model (3.6) presumes convergence of price levels.

4. DATA AND FORECASTING EVALUATION

We use quarterly data, 1973:1–2007:4, with the out of sample period beginning in 1987:1. The basic data source is International Financial Statistics, supplemented on occasion by national sources. Exchange rates are end of quarter values of the U.S. dollar vs. the currencies of 17 OECD countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Japan, Italy, Korea, Netherlands, Norway, Spain, Sweden, Switzerland, and the United Kingdom. (See below on how we handled conversion to the Euro in 1999:1.) The price level is the Consumer Price Index (CPI) in the last month of the quarter; output is industrial production in last month of the quarter; money is M1

(with some complications); the output gap is constructed by HP detrending, computed recursively, using only data from periods prior to the forecast period. Because some of the data appeared to display seasonality, we seasonally adjusted prices, output and money by taking a four quarter average of the log levels before doing any empirical work. For example, the price level in country i is $p_{it} = 1/4[\log(\mathrm{CPI}_{it}) + \log(\mathrm{CPI}_{it-1}) + \log(\mathrm{CPI}_{it-2}) + \log(\mathrm{CPI}_{it-3})]$.

To explain the mechanics of our forecasting work, let us illustrate for the four quarter horizon ($h = 4$), for the first forecast, and for the model that uses only factors but not additional observable fundamentals. As depicted in (4.1) below, we use data from 1973:1 to 1986:4 to estimate factors and factor loadings, and construct $\widehat{F}_{it}$ for $i = 1, \ldots, 17$.

```
--------------data used to estimate factors--------------
----data used to estimate panel regression----
|_____________________________________________|_____|_____|
                                              85:4  86:4  87:4
```
(4.1)

We then use right hand side data from 1973:1 to 1985:4 to estimate panel data regression

$$s_{it+4} - s_t = \alpha_i + \beta(\widehat{F}_{it} - s_{it}) + u_{it+4}, \quad t = 1973:1, \ldots, 1985:4. \tag{4.2}$$

We use 1986:4 data to predict the 4 quarter change in s:

$$\text{Prediction of } (s_{i,1987:4} - s_{i,1986:4}) = \hat{\alpha}_i + \hat{\beta}(\widehat{F}_{i,1986:4} - s_{i,1986:4}). \tag{4.3}$$

We then add an observation to the end of the sample, and repeat.

As is indicated by this discussion, the recursive method is used to generate predictions: observations are added to the end of the estimation sample, so that the sample size used to estimate factors and panel data regressions grows. The direct (as opposed to iterated) method is used to make multiperiod predictions. The estimation technique is maximum likelihood, assuming normality.

An analogous setup is used for other horizons and for models with observable fundamentals.

For a given date, factors and r.h.s. variables are identical across horizons: for given t, the same values of $\widehat{F}_{it} - s_{it}$ are used. However, the l.h.s. variable is different (h period difference in s_t), and regression samples are smaller for larger h. This means that the regression coefficients $(\hat{\alpha}_i, \hat{\beta})$ and predictions vary with h.

For the 9 non-Euro currencies (Australia, Canada, Denmark, Japan, Korea, Norway, Sweden, Switzerland, and the United Kingdom), we report "long sample" forecasting statistics for a 1987–2007 sample. For all 17 currencies, we report "early" sample forecasting statistics for a 1987–1998 sample. For the 9 non-Euro currencies and the Euro, we report "late sample" forecasting statistics for a 1999–2007 sample. Early sample statistics involve forecasts whose forecast base begins in 1986:4 and ends in 1998:4. Towards the end of the early sample, the forecast occurs in the pre-Euro era, while the

TABLE 1
List of Currencies and Models

A. Number of Quarterly Observations in Prediction Sample

	Prediction sample	*Horizon h*			
		1	4	8	12
Long Sample	$(1986:4+h)-2007:4$	84	81	77	73
Early Sample	$(1986:4+h)-(1998:4+h)$	49	49	49	49
Late Sample	$(1999:1+h)-2007:4$	35	32	28	24

B. Currencies

Long sample	$N=9$	Australia, Canada, Denmark, Japan, Korea, Norway, Sweden, Switzerland, United Kingdom
Early sample	$N=17$	The long sample countries plus Austria, Belgium, Finland, France, Germany, Italy, Netherlands and Spain
Late Sample	$N=10$	The long sample countries plus the Euro

C. Models

$\widehat{F}_{it}-s_{it}$	$\widehat{F}_{it}$ is the estimated factor component of currency i, estimated from 17 currencies in each sample (with identical Euro values appearing post-1998 for the 8 Euro area currencies).
$\widehat{F}_{it}-s_{it}+\text{Taylor}$	Taylor rule fundamentals (3.6) also included as a regressor
$\widehat{F}_{it}-s_{it}+\text{Monetary}$	Monetary model fundamentals (3.7) also included as a regressor
$\widehat{F}_{it}-s_{it}+PPP$	Purchasing power parity fundamentals (3.8) also included as a regressor

Notes: 1. The sample period for estimation of models runs from 1973:1 to the forecast base. Models are estimated recursively. Factors are estimated using $N=17$ currencies, for all samples. 2. In the late sample, Euro area forecasts are made by averaging forecasts from the 8 Euro countries. 3. Long horizon forecasts are made using the direct method.

realization occurs during the Euro era. We rescaled Euro area currencies so that there was no discontinuity. See Table 1 for the exact number of forecasts for each sample and horizon, as well as a summary listing of models.

In all samples (long, early and late), we use data from all 17 countries to construct factors and panel data estimates. For post-1999 data, the left hand side variable in both factor and panel data estimation is identical for all 8 Euro area countries. But because all samples include some pre-1999 data in estimation, there are differences across countries in estimates of the factor $\widehat{F}_{it}$, and of course the measures of prices, output and money used in the PPP, monetary and Taylor rule models. This means that the forecasts are different for the Euro countries. We construct a Euro forecast by simple averaging of the 8 different forecasts.

Our measure of forecast performance is root mean squared prediction error (RMSPE). (Here and through the rest of the paper, all references to MSPE and RMSPE refer to sample rather than population values.) We compute Theil's U-statistic, the ratio of the RMSPE from each of our models to the RMSPE from a random walk model.

We summarize results by reporting the median (across 17 countries) of the U-statistic, and the number of currencies for which the ratio is less than one (since a value less than one means our model had smaller RMSPE than did a random walk). Individual currency results are available on request.

A U-statistic of 1 indicates that the (sample) RMSPEs from the factor model and from the random walk are the same. As argued by Clark and West (2006, 2007), this is evidence *against* the random walk model. If, indeed, a random walk generates the data, then the factor model introduces spurious variables into the forecasting process. In finite samples, attempts to use such variables will, on average, introduce noise that inflates the variability of the forecasting error of the factor model. Hence, under a random walk null, we expect sample U-statistics greater than 1, even though that null implies that population ratios of RMSPEs are 1.

We report 10% level one sided hypothesis tests on H_0: RMSPE(our model) = RMSPE(random walk) against H_A: RMSPE(our model) < RMSPE(random walk). (Here, "our model" refers to any one of the four models given in (3.4) or (3.6)–(3.8): factor model, factor model plus Taylor rule, factor model plus monetary, or factor model plus PPP.) These hypothesis tests are conducted in accordance with Clark and West (2006), who develop a test procedure that accounts for the potential inflation of the factor model's RMSPE noted in the previous paragraph. Of course, with many currencies (17, in our early sample), it is very possible that one or more test statistics will be significant even if none in fact predict better than a random walk. We guarded against this possibility by testing H_0: RMSPE(our model) = RMSPE(random walk) for all currencies against H_A: RMSPE(our model) > RMSPE(random walk) for at least one currency, using the procedure in Hubrich and West (2010).[6] This statistic, however, rarely had a *p*-value less than 0.10. We therefore do not report it, to keep down the number of figures reported.

5. EMPIRICAL RESULTS

For the largest sample used (1973–2007), Fig. 1 plots the estimates of the three factors, while Table 2 presents the factor loadings. The factor loadings in Table 2 are organized so that the first block of six currencies (Austria, ... ,Switzerland) includes currencies in the one-time German mark area. The next block of four currencies (Finland, ... ,Spain) are

[6]Although we report ratios of MSPEs, the Clark and West (2006) and Hubrich and West (2008) tests work off arithmetic difference of MSPEs, evaluating whether this difference is statistically different from zero. The null hypothesis is that the random walk generates the data. These tests begin by adjusting the MSPE difference to account for noise that is present in the alternative (the nonrandom walk model) under the null hypothesis. For Clark and West (2006), the standard Diebold–Mariano–West (DMW) statistic is then computed for the adjusted MSPE difference. For Hubrich and West (2009), a parametric bootstrap is executed, under the assumption of normality. We did 10,000 repetitions in this bootstrap. See West (1996, 2006) for basic theory and further discussion.

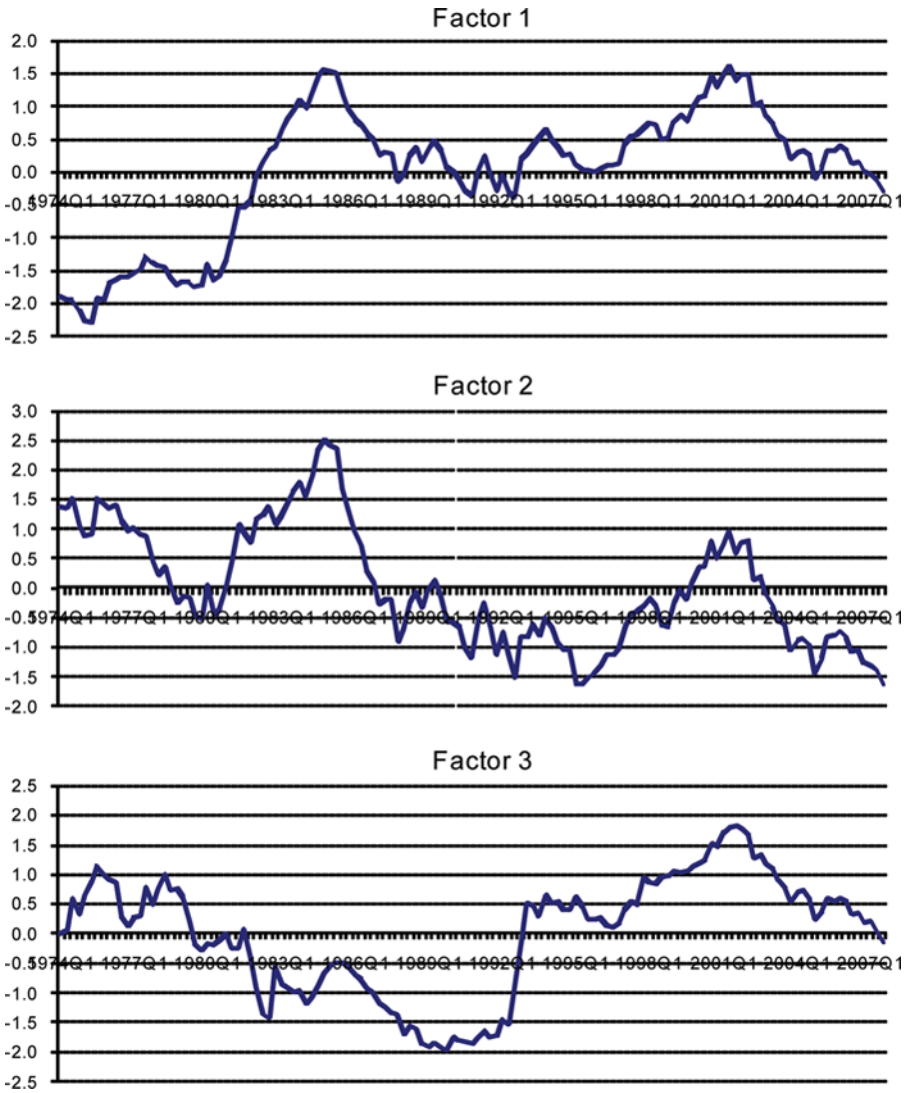

FIGURE 1 Factors, 1973–2007 sample.

Euro area countries not included in the first block. The final block (Australia,...,UK) lists the seven remaining countries.

The factor loadings suggest that the second factor reflects a central tendency of countries in the former German mark area. (If the factor loading on the second factor was zero for countries not in the mark area, then this second factor would literally be a weighted average of countries in the mark area (see Stock and Watson, 2006). The

TABLE 2
Factor Loadings, 1973–2007 Sample

	$\hat{\delta}_{1i}$	$\hat{\delta}_{2i}$	$\hat{\delta}_{3i}$
Austria	−0.06	1.00	0.02
Belgium	0.53	0.83	−0.11
Denmark	0.70	0.68	−0.16
Germany	−0.03	1.00	0.01
Netherlands	0.08	1.00	−0.02
Switzerland	−0.31	0.93	−0.03
Finland	0.88	0.23	0.34
France	0.85	0.50	−0.16
Italy	0.97	−0.14	0.15
Spain	0.98	−0.08	0.12
Australia	0.87	−0.31	0.16
Canada	0.80	−0.11	0.29
Japan	−0.55	0.78	−0.16
Korea	0.84	−0.27	0.28
Norway	0.95	0.19	0.12
Sweden	0.96	−0.07	0.22
UK	0.84	0.13	−0.01

Notes: 1. The fitted model is $s_{it} = \text{const.} + \hat{\delta}_{1i}\hat{f}_{1t} + \hat{\delta}_{2i}\hat{f}_{2t} + \hat{\delta}_{3i}\hat{f}_{3t} + \hat{v}_{it} \equiv \hat{F}_{it} + \hat{v}_{it}$; $\hat{f}_{1t}$, $\hat{f}_{2t}$ and $\hat{f}_{3t}$ are estimated factors.

coefficients are not zero on all non-mark countries, so the second factor is only roughly an average of mark countries.) By similar logic, the first factor seems to represent an average of everybody except countries in the former German mark area. The third factor is hard to label.

Of course this breakdown is not precise. Denmark's factor loading on what we have labeled the "mark" factor is smaller than is Japan's (0.68 vs. 0.78), and its factor loading on the first factor is, in absolute value, larger than Japan's (0.70 vs. −0.55).

Tables 3 and 4 present some forecasting results. We present in these tables summaries of results over all currencies. We present the median U-statistic across the currencies in the sample, the number of U-statistics less than 1 and the number of t-statistics greater than 1.282. (Recall that a U-statistic less than 1 means that the model's had a lower MSPE than did a random walk.) Currency-by-currency results are available on request.

Table 3 presents results for $r = 2$ factors, both for the model that uses only factors, and for the models that also include observable fundamentals. To read the table, consider the entry at the top of the table for model $= \widehat{F}_{it} - s_{it}$, sample $= 87 - 07$. The figure of "1.003" for "median U" and horizon "$h = 1$" means that of the 9 currencies, half had U-statistics above 1.003, and half had U-statistics below 1.003. The figures of "1(0)" immediately

TABLE 3
Two Factor ($r = 2$) Results

Model	Sample/No. Currencies	Statistic	Horizon h: 1	4	8	12
$\widehat{F}_{it} - s_{it}$	long/$N = 9$	median U	1.003	1.008	1.056	1.108
		#U < 1 or ($t > 1.282$)	1(0)	4(0)	4(0)	4(0)
$\widehat{F}_{it} - s_{it} + \text{Taylor}$	long/$N = 9$	median U	1.010	1.047	1.089	1.129
		#U < 1 or ($t > 1.282$)	1(0)	0(0)	1(0)	4(0)
$\widehat{F}_{it} - s_{it} + \text{Monetary}$	long/$N = 9$	median U	1.010	1.071	1.202	1.474
		#U < 1 or ($t > 1.282$)	3(2)	3(2)	3(2)	3(3)
$\widehat{F}_{it} - s_{it} + \text{PPP}$	long/$N = 9$	median U	1.003	0.996	0.953	0.938
		#U < 1 or ($t > 1.282$)	3(0)	5(0)	6(2)	5(0)
$\widehat{F}_{it} - s_{it}$	early/$N = 17$	median U	1.001	1.006	1.049	1.164
		#U < 1 or ($t > 1.282$)	6(0)	7(0)	4(0)	3(0)
$\widehat{F}_{it} - s_{it} + \text{Taylor}$	early/$N = 17$	median U	1.012	1.048	1.086	1.156
		#U < 1 or ($t > 1.282$)	1(0)	2(0)	1(0)	3(0)
$\widehat{F}_{it} - s_{it} + \text{Monetary}$	early/$N = 17$	median U	0.996	1.012	1.116	1.216
		#U < 1 or ($t > 1.282$)	10(3)	8(4)	7(3)	6(4)
$\widehat{F}_{it} - s_{it} + \text{PPP}$	early/$N = 17$	median U	0.999	0.983	1.027	1.128
		#U < 1 or ($t > 1.282$)	9(0)	13(1)	5(1)	3(0)
$\widehat{F}_{it} - s_{it}$	late/$N = 10$	median U	1.009	1.014	0.934	0.835
		#U < 1 or ($t > 1.282$)	3(1)	3(0)	7(0)	8(3)
$\widehat{F}_{it} - s_{it} + \text{Taylor}$	late/$N = 10$	median U	1.010	1.035	0.979	0.836
		#U < 1 or ($t > 1.282$)	2(1)	2(0)	6(0)	8(2)
$\widehat{F}_{it} - s_{it} + \text{Monetary}$	late/$N = 10$	median U	1.013	1.034	0.978	1.105
		#U < 1 or ($t > 1.282$)	3(1)	4(1)	6(3)	5(3)
$\widehat{F}_{it} - s_{it} + \text{PPP}$	late/$N = 10$	median U	1.006	1.000	0.891	0.727
		#U < 1 or ($t > 1.282$)	4(0)	5(0)	8(0)	9(5)

Notes: 1. Table 1 defines the long, early, and late sample periods, lists the currencies in each sample, and describes the models. 2. The U-statistic is = (RMSPE Model/RMSPE random walk); U < 1 means that the model had a smaller MSPE than did a random walk model; "median U" presents the median value of this ratio across 9, 17, or 10 currencies. "#U < 1" gives the number of currencies for which U < 1, a number that can range from 0 to the number of currencies N. 3. t is test of H_0: $U = 1$ (equality of RMSPEs) against one-sided H_A: $U < 1$ (RMSPE Model is smaller), using the Clark and West (2006) procedure. The number of currencies in which this test rejected equality at the 10 percent level is given in the ($t > 1.282$) entry.

below the figure of "1.003" means that only 1 of the 9 U-statistics was below 1, and that 0 of the t-statistics rejected the null of equal MSPE at the 10% level.[7]

One's eyes (or at least our eyes) are struck by the preponderance of median U-statistics that are above 1. In the long sample, 13 of the 16 the median U-statistics are above one (the three exceptions are for $\widehat{F}_{it} - s_{it} + \text{PPP}$ for the $h = 4$, 8 and 12 quarter horizons). In the early sample, it is again the case that 13 of the 16 the medians are above one

[7]The fact the median U was 1.003 but only 1 U-statistic was below 1 of course means that the U-statistics were tightly clustered near 1.

TABLE 4
Results for $\widehat{F}_{it} - s_{it}$, Varying Number of Factors r

			Horizon h			
No. of Factors®	*Sample/No. Currencies*	*Statistic*	1	4	8	12
1	long/$N = 9$	median U	1.011	1.037	1.083	1.154
		#U < 1 or ($t > 1.282$)	1(0)	1(0)	1(0)	1(0)
2	long/$N = 9$	median U	1.003	1.008	1.056	1.108
		#U < 1 or ($t > 1.282$)	1(0)	4(0)	4(0)	4(0)
3	long/$N = 9$	median U	1.003	0.996	0.996	1.038
		#U < 1 or ($t > 1.282$)	3(0)	5(0)	5(1)	4(0)
1	early/$N = 17$	median U	0.996	0.969	0.995	1.103
		#U < 1 or ($t > 1.282$)	9(3)	10(3)	9(1)	3(0)
2	early/$N = 17$	median U	1.001	1.006	1.049	1.164
		#U < 1 or ($t > 1.282$)	6(0)	7(0)	4(0)	3(0)
3	early/$N = 17$	median U	1.000	0.995	1.000	1.130
		#U < 1 or ($t > 1.282$)	10(1)	9(0)	9(0)	3(0)
1	late/$N = 10$	median U	1.021	1.081	1.179	1.294
		#U < 1 or ($t > 1.282$)	1(0)	1(0)	1(0)	1(1)
2	late/$N = 10$	median U	1.009	1.014	0.934	0.835
		#U < 1 or ($t > 1.282$)	3(1)	3(0)	7(0)	8(3)
3	late/$N = 10$	median U	1.008	1.020	0.953	0.822
		#U < 1 or ($t > 1.282$)	3(1)	3(0)	8(0)	8(1)

Notes: 1. The results for $r = 2$ in this table repeat, for convenience, those for $\widehat{F}_{it} - s_{it}$ in Table 3. 2. See notes to Table 3.

(the exceptions in this case being $\widehat{F}_{it} - s_{it} + \text{PPP}$ for h=1 and 4, and $\widehat{F}_{it} - s_{it} + \text{monetary}$ for $h = 1$). In the late sample, however, 9 of the 16 medians are above 1, with the models doing consistently better than a random walk (median $U < 1$) at 8 and 12 quarter horizons. Note in particular that in this sample, 8 of the 10 U-statistics were below 1 for $\widehat{F}_{it} - s_{it}$ and $\widehat{F}_{it} - s_{it} + \text{monetary}$, and 9 of 10 were below 1 for $\widehat{F}_{it} - s_{it} + \text{PPP}$.

Table 4 illustrates how varying the number of factors affects the simplest model, that of $\widehat{F}_{it} - s_{it}$; results for models that include Taylor rule, monetary or PPP fundamentals are similar. The table indicates that for the long sample, $r = 3$ performs a little better and the $r = 1$ model a little worse than does the $r = 2$ model presented in Table 3. In the early sample, the $r = 2$ model is the worst performing; the $r = 1$ model is the best performing. In the late sample, the $r = 2$ and $r = 3$ models perform similarly, with the $r = 1$ model performing distinctly more poorly than either of the other models.

To depict visually what underlies a U-statistic of various values, let us focus on the United Kingdom, $r = 3$ factors, model $= \widehat{F}_{it} - s_{it}$, long sample. The U-statistics happen to be 1.003 ($h = 1$), 0.996 ($h = 4$), 0.979 ($h = 8$), and 0.969 ($h = 12$). (These U-statistics, as well as other individual currency U-statistics discussed below, are not reported in any table.) A scatter plot of the recursive estimates of $\widehat{F}_{it} - s_{it}$, and of the subsequent h-quarter

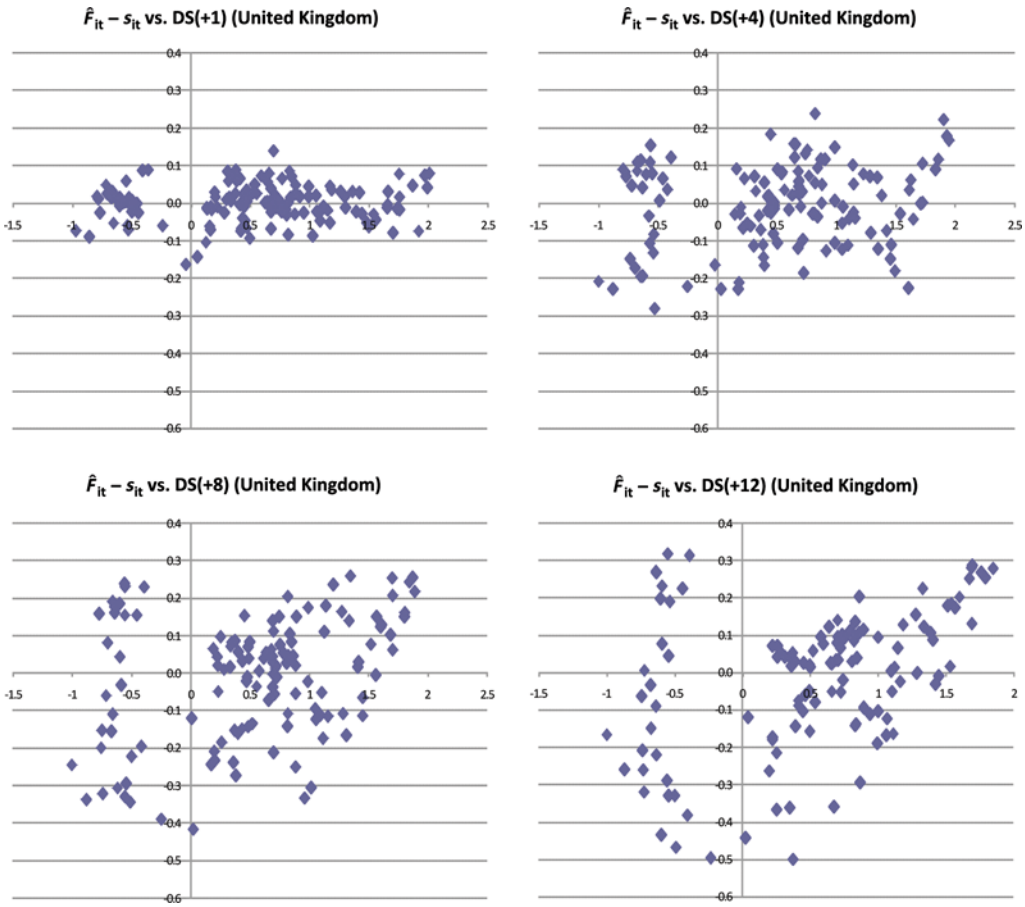

FIGURE 2 Forecast errors and realizations, U.K., $r = 3$ factors, model $= \widehat{F}_{it} - s_{it}$, long sample.

change in the exchange rate is in Fig. 2. The values of 0.979 and 0.968 for the U-statistics for $h = 8$and $h = 12$ imply a reduction in RMSPE relative to a no-change forecast of about 2–3%. Despite the seemingly small reduction, the figures for $h = 8$ and $h = 12$ depict an unambiguously positive relation between the deviation from the factor $\widehat{F}_{it} - s_{it}$ and the subsequent change in the exchange rate. On the other hand, there clearly is a lot of variation in a relation that is positive on average.

Our predictions fared especially poorly for the Japanese yen, which generally had one of the highest U-statistics in each sample and model. For example, the U-statistics for Japan, $r = 2$ factors, model $= \widehat{F}_{it} - s_{it}$, long sample were 1.008 ($h = 1$), 1.051 ($h = 4$), 1.085 ($h = 8$), and 1.160 ($h = 12$). That the yen does not quite fit into the same mold as the other currencies in our study is perhaps suggested by the large negative weight

of -0.55 for the yen on the first factor (see Table 2). In the late sample, continental currencies (Denmark, Norway, Sweden, Switzerland) and, to a lesser extent, the Euro were generally well predicted by our models. For example, the figures for the Euro for $r = 2$ factors, model $= \widehat{F}_{it} - s_{it}$, late sample were 1.009 ($h = 1$), 1.015 ($h = 4$), 0.939 ($h = 8$), and 0.816 ($h = 12$).

Over all specifications and horizons (1, 2 and 3 factors; long, early and late samples; horizons of 1, 4, 8 and 12 quarters), only the $\widehat{F}_{it} - s_{it} + \text{PPP}$ model had median U-statistics less than 1 in over 50 percent of the forecasts.

To further check the sensitivity of our results to particular sample periods, Figure 3 graphs recursively computed U-statistics for the $h = 1$ and $h = 8$ horizons, $r = 2$ factors, model $= \widehat{F}_{it} - s_{it}$, for the United Kingdom and Japan, long sample, and the Euro, late sample. The initial value in the graphs—1987:1 ($h = 1$) or 1988:4 ($h = 8$) for the U.K. and Japan, 1999:2 ($h = 1$) or 2001:1 ($h = 8$) for the Euro—is computed from a single observation. The number of observations used in computing the U-statistics increases through the sample with the number of observations used to compute the final value in 2007:4 given in the relevant entries of panel A of Table 1: 84 ($h = 1$) and 77 ($h = 8$) for the UK and Japan, and 35 ($h = 1$) and 28 ($h = 8$) for the Euro. The final values in the graphs, in 2007:4, is the one reported in the tables and the text above. For example, 0.939 for Euro, $h = 8$, is the final value for the Euro in the $h = 8$ graph. Note that the vertical scale is different for the $h = 1$ and $h = 8$ graphs.

Of course, the initial values in the graphs fluctuate quite a bit. But once a couple of years worth of observations have been accumulated, the values settle down. We see that the figures reported in the tables and text and discussed above are representative: apart from start up values computed from few observations, there is no apparent sensitivity to sample. In the $h = 1$ graph, U-statistics consistently are near 1, and generally are above 1. This indicates that for one quarter ahead forecasts, the average squared value of the forecast from the factor model generally is slightly above that of a random walk model. In the $h = 8$ graph, we see that the poor performance of our factor model that we noted above for Japan obtains for the whole sample; the modestly good performance that we noted for the United Kingdom also obtains for most of the sample; and the good performance for the Euro obtains consistently once the effects of initial observations have been averaged out.

6. ROBUSTNESS

We checked the robustness of these results in a number of dimensions.

1. We estimated by principal components rather than by maximum likelihood. Overall, results were comparable, with one technique doing a little better (occasionally, a lot better) in one specification and the other doing a little better (occasionally, a lot better) in other specifications. We also used the British pound rather than the U.S.

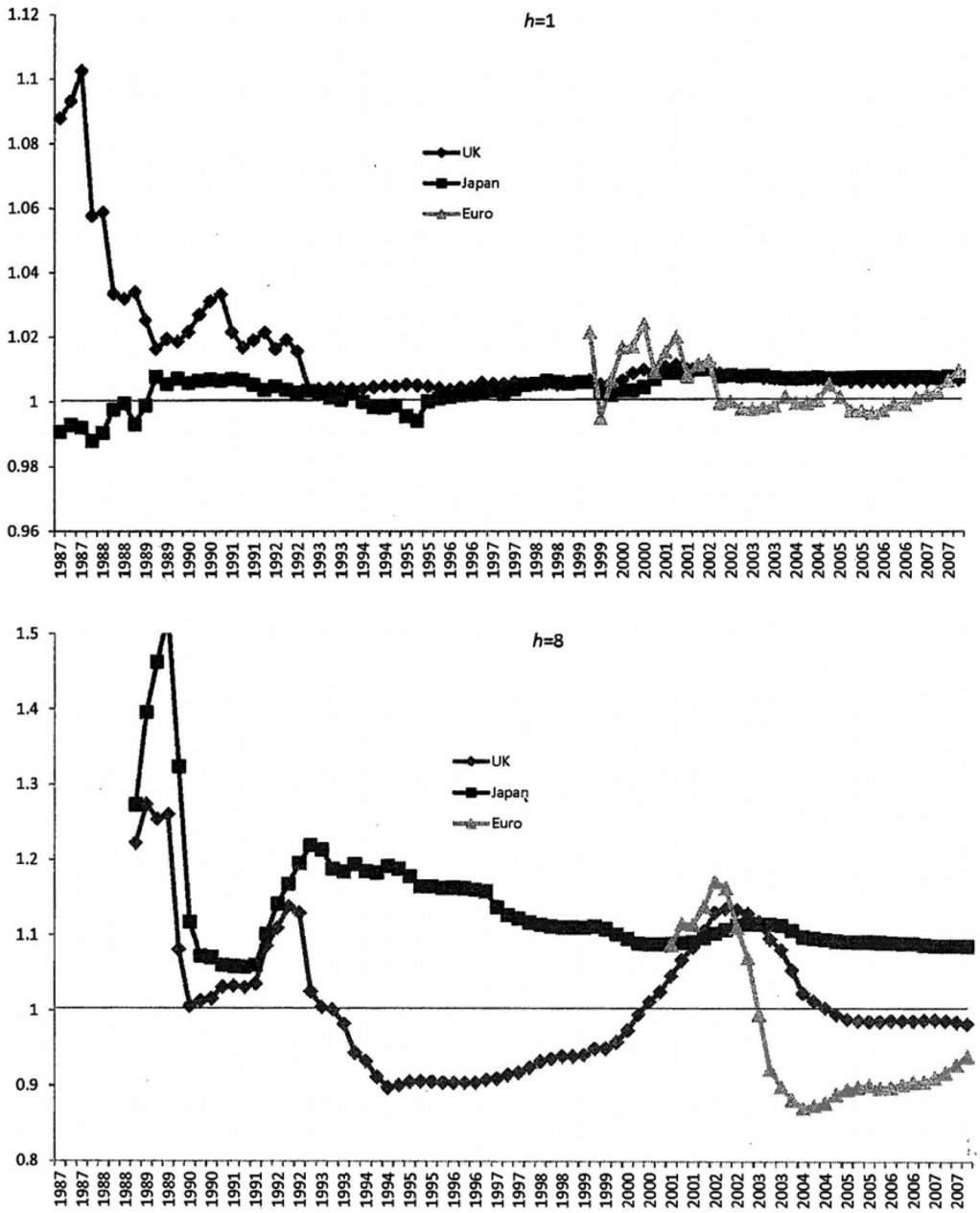

FIGURE 3 Recursive U-statistics, $r = 2$ factors, model $= \widehat{F}_{it} - s_{it}$, U.K., Japan, and Euro.

dollar as the base currency. Estimation was by maximum likelihood. Here, results were comparable for the early sample, somewhat worse for the long and late samples.

A detailed summary of the robustness checks is in an appendix available on request. To illustrate, let us take two lines from Table 3, and present analogous results from principal components estimation, and from estimation with the British pound as the base currency. These lines are chosen because they are representative:

	$h=1$	$h=4$	$h=8$	$h=12$	
$\widehat{F}_{it} - s_{it}$, early/$N=17$, maximum likelihood, U.S. dollar (Table 3)	6(0)	7(0)	4(0)	3(0)	(6.1a)
$\widehat{F}_{it} - s_{it}$, early/$N=17$, principal components, U.S. dollar	4(2)	4(3)	7(4)	8(2)	(6.1b)
$\widehat{F}_{it} - s_{it}$, early/$N=17$, maximum likelihood, British pound	6(1)	6(1)	4(2)	4(0)	(6.1c)
$\widehat{F}_{it} - s_{it} + \text{PPP}$, late/$N=10$, maximum likelihood, U.S. dollar (Table 3)	4(0)	5(0)	8(0)	9(5)	(6.2a)
$\widehat{F}_{it} - s_{it} + \text{PPP}$, late/$N=10$, principal components, U.S. dollar	3(3)	3(3)	7(2)	9(4)	(6.2b)
$\widehat{F}_{it} - s_{it} + \text{PPP}$, late/$N=10$, maximum likelihood, British pound	3(0)	1(0)	1(0)	1(0)	(6.2c)

We see in (6.1a) and (6.1b) that in terms of the number of U-statistics less than one, principal components improves over maximum likelihood at $h=12$ (8 versus 3 U-statistics less than 1), while the converse is true at $h=4$ (4 versus 7 U-statistics less than 1). Results for the British pound in (6.1c) similarly are better at some horizons, worse at other horizons. We see in (6.2a) and (6.2b) that principal components and maximum likelihood generate very similar numbers, while (6.2c) illustrates that with the British pound as the base currency, results degrade for the late sample.

2. We also computed a utility based comparison of our factor models relative to the random walk. Our approach is stimulated by that of West et al. (1993), who consider alternative models for conditional volatility in contrast to our comparison of models for conditional means. We consider an investor with a one period mean-variance utility function, allocating wealth between U.S. and foreign one period debt that is nominally riskless in own currency. Suppose that a given one of our factor models produces higher expected utility than does the random walk. We ask, what fraction of wealth would the investor be willing to give up to use our model rather than a random walk to forecast exchange rates? Of course, if the random walk forecasts better, we ask the same question, but in our tables present the result with a negative sign. We let $\overline{U}_F$ and $\overline{U}_{RW}$ denote utility gains from use of the factor and random walk models, cautioning the reader that the $\overline{U}$ here is *not* related to the "U" in Theil's U.

Details are in the Appendix. Interest rate data on government debt were obtained from Datastream, last day of the quarter. We calibrate our mean-variance utility function so that it implies a coefficient of relative risk aversion of 1 at the initial wealth

TABLE 5

A. Wealth Sacrifice to use Higher Utility Model, for $\widehat{F}_{it} - s_{it}$, Horizon $h = 1$, Two Factor Models

		model			
Sample/No. Currencies	*Statistic*	$\widehat{F}_{it} - s_{it}$	Taylor+ $\widehat{F}_{it} - s_{it}$	Mon+ $\widehat{F}_{it} - s_{it}$	PPP+ $\widehat{F}_{it} - s_{it}$
long/$N = 9$	median sacrifice $\mid \overline{U}_F > \overline{U}_{RW}$	19	4	179	51
	median sacrifice $\mid \overline{U}_{RW} > \overline{U}_F$	–65	–134	–479	–112
	$\#\overline{U}_F > \overline{U}_{RW}$	3	1	3	4
early/$N = 17$	median sacrifice $\mid \overline{U}_F > \overline{U}_{RW}$	147	74	583	173
	median sacrifice $\mid \overline{U}_{RW} > \overline{U}_F$	–176	–120	–56	–1572
	$\#\overline{U}_F > \overline{U}_{RW}$	12	6	9	15
late/$N = 10$	median sacrifice $\mid \overline{U}_F > \overline{U}_{RW}$	110	92	280	95
	median sacrifice $\mid \overline{U}_{RW} > \overline{U}_F$	–21	–143	–1031	–196
	$\#\overline{U}_F > \overline{U}_{RW}$	2	2	5	4

B. Results for $\widehat{F}_{it} - s_{it}$, Monthly data, Two Factor Models

		Horizon h			
Sample/No. Currencies	*Statistic*	1	12	24	36
long/$N = 9$	median U	1.002	1.005	0.975	0.974
	#U < 1 or ($t > 1.282$)	2(0)	3(0)	6(3)	6(5)
early/$N = 17$	median U	1.002	0.997	1.058	1.302
	#U < 1 or ($t > 1.282$)	6(0)	9(0)	3(1)	4(1)
late/$N = 10$	median U	1.002	1.006	0.979	0.985
	#U < 1 or ($t > 1.282$)	3(0)	4(1)	6(2)	6(3)

C. Full Sample Estimates of $\widehat{\beta}$ for $\widehat{F}_{it} - s_{it}$, Two Factor Models

		Horizon h			
Sample/No. Currencies	*Statistic*	1	4	8	12
long/$N = 9$	$\widehat{\beta}$	–0.008	–0.011	–0.040	–0.081
	std. error	(0.003)	(0.004)	(0.012)	(0.020)
early/$N = 17$	$\widehat{\beta}$	–0.009	–0.015	–0.047	–0.088
	std. error	(0.005)	(0.003)	(0.009)	(0.014)
late/$N = 10$	$\widehat{\beta}$	–0.007	–0.010	–0.041	–0.081
	std. error	(0.004)	(0.004)	(0.011)	(0.018)

Notes: 1. Consider a risk averse investor, with coefficient of relative risk aversion of 1, who uses either a factor or random walk model to allocate his wealth across two assets whose returns are nominally safe when measured in own currency. In Panel A, $\overline{U}_F > \overline{U}_{RW}$ means the factor model delivers higher expected utility. The "median sacrifice" reports the fraction of wealth expressed in annualized basis points that such an investor would be willing to give up to use the factor rather than random walk model (when $\overline{U}_F > \overline{U}_{RW}$) or the random walk rather than the factor model (when $\overline{U}_{RW} > \overline{U}_F$). (The $\overline{U}$ used here bears no relation to the "U" in Theil's U that is referenced in panel B of this Table and elsewhere in the paper.) See section 6 of the paper for additional detail. 2. Panel B presents estimates for monthly data comparable to the estimates in the $\widehat{F}_{it} - s_{it}$ lines in Table 3. 3. Panel C presents estimates of β from Eq. (3.5) in the case that $\gamma \equiv 0$ (i.e., the model is $\widehat{F}_{it} - s_{it}$) and the number of factors is two.

level. We answer "what fraction of wealth would the investor give up" in terms of annualized basis points. Results for one quarter ahead forecasts and two factor models are in Table 5A. Comparable results for the mean squared error criterion are in the $h = 1$ column that runs down Table 3. The utility based and mean squared criteria perform similarly in terms of whether a factor model performs better. For example, for the long sample, Table 5A indicates that the factor based model is preferred by the utility criterion in 11 $(= 3 + 1 + 3 + 4)$ of the 36 comparisons; we see in the $h = 1$ column in the top four rows of Table 3 that the comparable figure is 8 of 36 for the mean squared error criterion. For both criteria, the factor models are preferred in a larger fraction of comparisons in the early than in the long or late samples.

By both criteria, performance differences generally are small. Of the 24 performance fees in Table 5A, all but 5 are less than 200 basis points in absolute value, a comparison that is relevant since management fees typically run around 200 basis points. This is consistent with the finding that at a one quarter horizon, the estimates of Theil's U were generally very close to 1.

We conclude that by both statistical and utility based criteria, the differences between factor models and the random walk are small at a one quarter horizon.

3. We repeated the mean squared error comparison using monthly data, for two factor models, and horizons of 1, 12, 24, and 36 months. Results are in Table 5B. Comparable quarterly figures are in the three "$\widehat{F}_{it} - s_{it}$" lines in Table 3. Results are qualitatively similar for the two frequencies. The factor model does especially well at the long horizons in the late sample; performance differences are very small at shorter (1 and 12 month) horizons.
4. Finally, in Table 5C we report point estimates and standard errors for the slope coefficient β in (3.5), for the model $F_{it} - s_{it}$. Qualitatively, the results align with those of our out of sample tests, in that t-statistics tend to increase with the horizon. However, all but two of the t-statistics are significant at the five percent level (the exceptions being the early and late samples, $h = 1$). Thus, as is often the case, there is more significance for a predictor with in-sample than with out-of-sample evidence. We interpret this as an endorsement of our decision to focus on out of sample analysis: we otherwise might have been unduly optimistic about the performance of our factor model.

7. CONCLUSIONS

This first pass at extracting factors from the cross-section of exchange rates yielded mixed results. Results for late samples (1999–2007) were promising, at least for horizons of 8 or 12 quarters. With occasional exceptions for models that relied not only on factors but PPP fundamentals as well, other results suggested no ability to improve on a "no-change," or random walk, forecast.

Late samples allow larger sample sizes for estimation of factors. While that may be part of the reason for good results for late samples, that is not a sufficient condition for good results because our robustness checks found that the when the British pound is the base currency, late samples perform worse than early samples. Indeed, it remains to be seen whether our results for late samples are spurious. In any event, the framework here can be extended in a number of ways. It would be desirable to allow different slope coefficients across currencies, to allow more flexible specification of parameters in monetary and Taylor rule models, and to use a data dependent method of selecting the number of factors. Such extensions are priorities for future work. It would also be desirable to compare our predictions to, not only a random walk model, but to other models that have been compared to the random walk in earlier studies.

APPENDIX

In this Appendix, we describe the utility based calculation presented in Table 5A. We begin with some notation. We drop the i subscript from the exchange rate s_t and other variables for simplicity. Define

$$\hat{s}_t = \text{forecast of } s_{t+1}; \quad \hat{s}_t = s_t \text{ for random walk,}$$
$$\hat{s}_t = \text{factor model forecast for factor model;} \tag{A.1}$$

$$\hat{\sigma}_t^2 = \text{variance of } s_{t+1} \text{ as of time } t, \text{ computed as } t^{-1}\sum_{j=2}^{t}(s_j - s_{j-1})^2 \text{ for both models;} \tag{A.2}$$

$$R_{t+1}, R_{t+1}^* = \text{nominal return on one period nominal riskless debt in the U.S. and abroad;} \tag{A.3}$$

$$\theta_{t+1} = R_{t+1}^* - R_{t+1} - (s_{t+1} - s_t) = \text{ex-post return differential;} \tag{A.4}$$

$$\hat{\theta}_t = R_{t+1}^* - R_{t+1} - (\hat{s}_t - s_t) = \text{ex-ante return differential.} \tag{A.5}$$

Let us tentatively assume that $\hat{\theta}_t$ is positive, i.e., that when we use a given model for $\hat{s}_t$, the expected return is higher abroad than in the U.S. The goal of a U.S. investor with initial wealth W is to

$$\max_f E_t U_{t+1} = E_t[W_{t+1} - 0.5\gamma W_{t+1}^2] \qquad \text{s.t.}$$
$$W_{t+1} = W[f(R_{t+1}^* - \Delta s_{t+1}) + (1-f)R_{t+1}] = W(f\theta_{t+1} + R_{t+1}). \tag{A.6}$$

In solving for the optimal f, call it f^*, set $E_t\theta_{t+1} = \hat{\theta}_t$, $E_t\theta_{t+1}^2 = \hat{\theta}_t^2 + \hat{\sigma}_t^2$. Then,

$$f^* = \frac{1 - \gamma W R_{t+1}}{\gamma W}\frac{\hat{\theta}_t}{\hat{\theta}_t^2 + \hat{\sigma}_t^2}. \tag{A.7}$$

Plug f^* back into $E_t U_{t+1}$. Rearrange, and the result is

$$E_t U_{t+1} = [c_t + d_t E_t u_{t+1}(\theta_{t+1}, \theta_{t+1}^2, \hat{\theta}_t, \hat{\sigma}_t^2)]W, \tag{A.8}$$

$$c_t = R_{t+1} - .5\gamma W R_{t+1}^2, \qquad d_t = (1 - \gamma W R_{t+1})^2/\gamma W,$$

$$u_{t+1}(\cdot) = \frac{\hat{\theta}_t}{\hat{\theta}_t^2 + \hat{\sigma}_t^2}\left(\theta_{t+1} - .5\frac{\hat{\theta}_t}{\hat{\theta}_t^2 + \hat{\sigma}_t^2}\theta_{t+1}^2\right),$$

$$E_t u_{t+1}(\cdot) = \frac{\hat{\theta}_t}{\hat{\theta}_t^2 + \hat{\sigma}_t^2}\left(E_t\theta_{t+1} - .5\frac{\hat{\theta}_t}{\hat{\theta}_t^2 + \hat{\sigma}_t^2}E_t\theta_{t+1}^2\right).$$

We compute the total utility of a U.S. investor using one of our models as the sum over t of $c_t + d_t u_{t+1}$ in those periods in which $\hat{\theta}_t$ is positive, i.e., we compute the utility gains for a U.S. investor only in those quarters in which the foreign return is expected to be higher than the U.S. return (otherwise, the U.S. investor puts all wealth into R_{t+1} because this is a safe asset). We compute the total utility of a foreign investor (with signs of returns reversed) over those periods in which $\hat{\theta}_t$ is negative. We average the two utilities over P periods of predictions to get an average utility based measure of the quality of one of the models.

Let $\overline{U}_F$ and $\overline{U}_{RW}$ be the average utility measures that result from a factor model and the random walk model. We report the fraction of wealth that our investor would be willing to give up to use the higher utility model, expressed at an annualized rate, in basis points. When $\overline{U}_F > \overline{U}_{RW}$, Table 5A reports $40,000 \times (1 - \overline{U}_{RW}/\overline{U}_F)$. When $\overline{U}_{RW} > \overline{U}_F$, the table reports $-40,000 \times (1 - \overline{U}_F/\overline{U}_{RW})$. The factor of 4 converts quarterly to annual points, while the factor of 10,000 converts to basis points.

For quadratic utility (A.6), the coefficient of relative risk is $\gamma W/(1 - \gamma W)$. We fix this value at 1.

ACKNOWLEDGMENTS

We thank Wallice Ao, Roberto Duncan, Lowell Ricketts, and Mian Zhu for exceptional research assistance; the editor, two anonymous referees and seminar audiences at the European Central Bank and the IMF for helpful comments. The Additional Appendix that is referenced in the paper is available on request from the authors.

FUNDING

We thank the National Science Foundation for financial support.

REFERENCES

Bacchetta, P., van Wincoop, E. (2006). Can information heterogeneity explain the exchange rate determination puzzle? *American Economic Review* 96(3):552–576.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics* 122(1):137–183.

Benigno, G., Benigno, P. (2006). Exchange rate determination under interest rate rules. Manuscript, London School of Economics.

Bulut, L., Maasoumi, E. (2012). Predictability and specification in models of exchange rate determination. Forthcoming in *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis (In honor of Halbert White)*. New York: Springer.

Cheung, Y.-W., Chinn, M., Garcia Pascual, A. (2005). Empirical exchange rate models of the nineties: Are any fit to survive? *Journal of International Money and Finance* 24:1150–1175.

Clark, T. E., West, K. D. (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics* 135(1–2):155–186.

Clark, T. E., West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* 138(1):291–311.

Diebold, F. X., Gardeazabal, J., Yilmaz, K. (1994). On cointegration and exchange rate dynamics. *Journal of Finance* XLIX:727–735.

Engel, C., West, K. D. (2005). Exchange rates and fundamentals. *Journal of Political Economy* 113:485–517.

Engel, C., West, K. D. (2006). Taylor rules and the Deutschemark-Dollar real exchange rate. *Journal of Money, Credit and Banking* 38:1175–1194.

Engel, C., Mark, N. M., West, K. D. (2008). Exchange rate models are not as bad as you think. In: Acemoglu, D., Rogoff, K., Woodford, M., eds. *NBER Macroeconomics Annual, 2007*. Chicago: University of Chicago Press, pp. 381–443.

Groen, J. J. J. (2005). Exchange rate predictability and monetary fundamentals in a small multi-country panel. *Journal of Money, Credit and Banking* 37:495–516.

Groen, J. J. J. (2006). Fundamentals based exchange rate prediction revisited. Manuscript, Bank of England.

Hubrich, K., West, K. D. (2010). Forecast comparisons for small nested model sets. *Journal of Applied Econometrics* 25:574–594.

Ho, M. S., Sørensen, B. E. (1996). Finding cointegration rank in high dimensional systems using the Johansen test: An illustration using data based Monte Carlo simulations. *Review of Economics and Statistics* 726–732.

Mark, N. A. (1995). Exchange rates and fundamentals: Evidence on long-horizon predictability. *American Economic Review* 85:201–218.

Mark, N. A. (2001). *International Macroeconomics and Finance: Theory and Empirical Methods*. New York: Blackwell.

Mark, N. A. (2008). Changing monetary policy rules, learning and real exchange rate dynamics. Manuscript, University of Notre Dame.

Mark, N. A., Sul, D. (2001). Nominal exchange rates and monetary fundamentals: Evidence from a small post-Bretton Woods sample. *Journal of International Economics* 53:29–52.

Meese, R. A., Rogoff, K. (1983a). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics* 14:3–24.

Meese, R. A., Rogoff, K. (1983b). The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification. In: Frenkel, J. A., ed. *Exchange Rates and International Macroeconomics*. Chicago: University of Chicago Press.

Molodtsova, T., Papell, D. (2008). Taylor rules with real-time data: A tale of two countries and one exchange. *Journal of Monetary Economics* 55:S63–S79.

Rapach, D. E., Wohar, M. E. (2004). Testing the monetary model of exchange rate determination: A closer look at panels. *Journal of International Money and Finance* 23(6):841–865.

Stock, J. H., Watson, M. W. (2006). Forecasting with many predictors. In: Elliott, G., Granger, C. W. J., Timmermann, A., eds. *Handbook of Economic Forecasting*. Vol. 1. Amsterdam: Elsevier, pp. 515–550.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64:1067–1084.

West, K. D. (2006). Forecast evaluation. In: Elliott, G. Granger, C. W. J., Timmerman, A. eds. *Handbook of Economic Forecasting*. Vol. 1. Amsterdam: Elsevier, pp. 100–134.

West, K. D., Edison, H. J., Cho, D. (1993). A utility based comparison of some models of exchange rate volatility. *Journal of International Economics* 35.