CONDITIONAL PREDICTIVE ABILITY OF EXCHANGE RATES IN LONG RUN REGRESSIONS*

PREDICTIBILIDAD DEL TIPO DE CAMBIO: UNA EVALUACION CONDICIONAL CON REGRESIONES DE LARGO PLAZO

PABLO M. PINCHEIRA**
Central Bank of Chile

Abstract

In this paper we evaluate exchange rate predictability using a framework developed by Giacomini and White (2006). This new framework tests for conditional predictive ability rather than unconditional predictive ability, which has been the standard approach. Using several shrinkage based forecasting methods, including new methods proposed here, we evaluate conditional predictability of five bilateral exchange rates at differing horizons. Our results indicate that for most currencies a random walk would not be the optimal forecasting method in a real time forecasting exercise, at least for some predictive horizons. We also show that our proposed shrinkage methods in general perform on par with Bayesian shrinkage and ridge regressions, and sometimes they even perform better.

Keywords: *Exchange rate predictability, conditional predictive ability, Bayesian shrinkage, ridge regression, forecast evaluation.*

JEL Classification: *C22, C53, E37, F31.*

Resumen

En este documento evaluamos la predictibilidad de algunos tipos de cambio usando un nuevo enfoque desarrollado por Giacomini y White (2006). Lo novedoso de este enfoque es la realización de tests de predictibilidad condicionales no sólo incondicionales como ha sido la regla hasta ahora. Usando varias técnicas predictivas de reducción de parámetros, incluyendo algunos nuevos métodos presentados en este documento, se evalúa la predictibilidad condicional de cinco tipos de cambio bilaterales con respecto al dólar estadounidense, considerando distintos horizontes de predicción. Nuestros resultados indican que, para la mayoría de las monedas el camino aleatorio no sería el mejor predictor en un ejercicio de predicción en tiempo real, al menos para ciertos horizontes. También mostramos que, en general, los métodos de reducción de parámetros propuestos en este documento son tan buenos predictores como los métodos de reducción de parámetros tradicionales, y a veces incluso mejores.

Palabras clave: *Predictibilidad de tipos de cambio, evaluación predictiva condicional, contracción bayesiana, regresión ridge, evaluación capacidad predictiva.*

Clasificación JEL: *C22, C53, E37, F31.*

1. INTRODUCTION

One of the most striking contributions in the exchange rate literature is the well known result of Meese and Rogoff (1983a,b). Using a variety of linear exchange rate models, these authors showed that no economic model was able to consistently display improved forecast accuracy over a simple random walk model. This result was shown to be robust across different exchange rates and predictive horizons.

Later on, improved methodological techniques showed some results that partially overturned this seminal work. Some evidence of predictability is shown in Chinn and Meese (1995), Mark (1995), MacDonald and Marsh (1997), McCracken and Sapp (2005), and Clark and West (2006). Nevertheless, this evidence is still weak and no conclusive result on exchange rate predictability has been shown.

These improved methodological techniques are partly based upon the development of econometric strategies for forecast comparison under general loss functions. West (1996) and Diebold and Mariano (1995) established the basic econometric framework under which out-of-sample tests of predictive ability are carried out.

An important observation needs to be made. When engaging in tests of predictive ability there are two major questions that might be addressed. One is a question about theory. Namely, tests of predictive ability are used as instruments to test an economic theory. The second question is an empirical question seeking to find a profitable

forecasting method irrespective of any theoretical implications about the underlying data generating process. These two questions are not equivalent. In particular it is possible to show that even when the null hypothesis of no predictability is rejected, it is likely that a forecasting method based upon the rejected null model will outperform some forecasting methods based upon the alternative model.

This distinction is analyzed in depth in Giacomini and White (2006). They argue that the framework for out-of-sample predictive ability testing, developed by West (1996) and Diebold and Mariano (1995), might not be useful or appropriate for an applied forecaster trying to assess which of two competing forecasting methods will provide more accurate forecasts in the future. They propose an alternative approach that claims to be more relevant to economic forecasters.

The main distinction between the two approaches is twofold. First, Giacomini and White (2006) focus their analysis on conditional expectations of forecasts, while West (1996) and Diebold and Mariano (1995) focus on unconditional expectations. According to this distinction we will call the Giacomini and White (2006) approach the conditional approach, and that of West (1996) and Diebold and Mariano (1995) the unconditional approach. This difference is relevant for a forecaster that is highly interested in finding the best forecast for the next relevant period instead of a forecast that is the best on average. Second, the conditional approach is concerned with the whole "forecasting method" rather than just with the theoretical model used to generate forecasts, which is the main object of interest of the unconditional approach. The "forecasting method" is a much more general notion than the forecasting model because it includes the model, its estimation technique, the size of estimation and forecasting windows, and in general all the elements of the forecasting method that could possible affect its future predictive ability performance.

The recent literature that has partially overturned the result of Meese and Rogoff (1983a,b) has built on the unconditional approach to draw inference about exchange rate predictability. In consequence, it may be totally feasible that even for those currencies, models and horizons for which predictability is found, forecasts from these models may be outperformed by a simple random walk strategy in a real-time forecasting exercise. Little or no research has addressed the evaluation of conditional predictive ability for exchange rates.

To fill this gap, in this paper we perform tests of conditional predictive ability for several exchange rates, using a variety of shrinkage based forecasting methods based upon models of interest parity. Besides this contribution, we also introduce a new shrinkage estimation approach aimed at improving forecast accuracy under quadratic loss.

The rest of the paper is organized as follows: Section 2 further develops the conditional predictive ability approach and its differences with the unconditional approach. The relevant econometric environment is presented in Section 3[1]. Section 4 displays a description of the model and different estimation techniques that are used

---

[1]   Sections 2 and 3 are based upon Giacomini and White (2006).

to build the different "forecasting methods". Empirical results are reported for five bilateral exchange rates in Section 5. Section 6 concludes.

2. CONDITIONAL VERSUS UNCONDITIONAL TESTING FRAMEWORK

To correctly illustrate the main differences between the conditional and unconditional approaches, we consider two competing parametric forecasting models for the conditional expectation of a scalar time series y_{t+1}. We denote the forecasts from these two models as $y_{t+1}^1(\beta_1)$ and $y_{t+1}^2(\beta_2)$ where β_1 and β_2 denotes population parameters of the two competing models. For a given loss function $L = L\left(y_{t+1}, y_{t+1}^i(\beta_i)\right)$, $i = 1,2$ the unconditional approach suggests a test of equal forecast accuracy as follows

$$H_0 : E\left[L\left(y_{t+1}, y_{t+1}^1(\beta_1)\right) - L\left(y_{t+1}, y_{t+1}^2(\beta_2)\right)\right] = 0 \tag{1}$$

whereas the conditional approach suggests the following testing strategy

$$H_0 : E\left[L\left(y_{t+1}, y_{t+1}^1(\beta_{t1})\right) - L\left((y_{t+1}, y_{t+1}^2(\beta_{t2})\right) | \Psi_t\right] = 0 \ a.s. \ for \ all \ t \geq 0 \tag{2}$$

where $\hat{\beta}_{t1}$ and $\hat{\beta}_{t2}$ denote parameter estimates of β_1 and β_2 with information up until time t. The implementation of the conditional approach relies on the fact that (2) is equivalent to

$$E\left[h_t\left(L\left(y_{t+1}, y_{t+1}^1(\hat{\beta}_{t1})\right) - L\left(y_{t+1}, y_{t+1}^2(\hat{\beta}_{t2})\right)\right)\right] = 0$$

for all Ψ_t –measurable function h_t.

Some of the differences between the two approaches are evident. First, the unconditional approach asks a question directly involving the true unknown parameters of the competing models, whereas the conditional approach asks a direct question involving only estimates of those parameters[2]. When focusing on the true population parameters, the unconditional approach is implicitly testing the appropriateness of a model to correctly approximate the true data generating process. However, it is clear that even the true model might yield poor forecasts in the presence of parameter

---

2   We should point out that depending on the estimation method, a direct question involving only estimates of the population parameters defining the underlying data generating process will also impose some restrictions over these parameters. The main difference between the unconditional and conditional approach reduces to the fact that restrictions on the population parameters are in general different.

uncertainty, and clearly some "false" models have the chance to outperform the correct model in this context. In this regard, the use of known parameter estimates in the conditional approach might be more useful to determine which model will provide more accurate forecasts in a real time forecasting application. This is because in the conditional approach testing and future forecast accuracy now both depend upon the same magnitudes ($\hat{\beta}_{t1}$ and $\hat{\beta}_{t2}$), whereas in the unconditional approach testing focuses on the true population parameters but forecast accuracy is measured using $\hat{\beta}_{t1}$ and $\hat{\beta}_{t2}$. Second, Giacomini and White (2006) argue that a null hypothesis established as (1) can be interpreted as saying that, on average, the two models provide equal forecast accuracy. This information might not be very useful for a forecaster that needs to know which model provides the best forecast for tomorrow given information available today. The conditional null hypothesis seems a better choice for this scenario.

Some other differences are subtle. In particular we want to emphasize that when the conditional null hypothesis is stated in terms of the estimates of the true population parameters, this null is implicitly imposing restrictions on those population parameters, on the size of the estimation window and also on the ridge or shrinkage factors that may be used for estimation[3]. In other words, whereas the unconditional null only imposes a typically simple restriction on the parameters of the models, the conditional null imposes a restriction involving these parameters, the estimation sample size and the shrinkage factor used for estimation. This is important because the choice of estimation sample size and shrinkage factor may have an impact on the size of the conditional test[4]. As there is no empirical guide about how to choose these two magnitudes, we recommend caution when interpreting the results using conditional tests.

Further differences are also worth mentioning. For instance, the unconditional approach relies on stationarity assumptions, whether the unconditional approach relies on a more general assumption of heterogeneity. Besides, the conditional approach applies to both nested and non-nested models. On the contrary, the unconditional approach, originally established for only non-nested models, needs to make significant adjustments when models are nested, McCracken (2004). Further differences are described in detail in Giacomini and White (2006).

3. ECONOMETRIC ENVIRONMENT

Consider a scalar time series process with general term denoted by y_t and the set of information available until time t denoted by Ψ_t. We want to build τ-step-ahead forecasts for this scalar time series based upon information available until time t. We have two different methods to build τ-step-ahead forecasts for the relevant time

---

[3] We are extremely grateful to Kenneth West for making this point.
[4] In some cases there is a unique choice of sample size and shrinkage factor for which the conditional test is correctly sized.

series y_t. These methods provide two different forecasts denoted by $f_{R,t}$ and $g_{R,t}$. We will further assume that these forecasts are built from estimates of parametric models, so we can express

$$f_{R,t} = f_{R,t}\left(\hat{\gamma}_{R,t}\right)$$

$$g_{R,t} = g_{R,t}\left(\hat{\theta}_{R,t}\right)$$

The R subscript means that the forecasts are constructed using at most the last R sample observations available until time t. This strategy is well known as a rolling estimation window of maximum size R.

We will be using two models to build our forecasts:

$$\text{Model 1:} \qquad y_{t+1} = e_{t+1} \qquad\qquad\qquad (3)$$

$$\text{Model 2:} \qquad y_{t+1} = X_{t+1}'\beta + e_{t+1} \qquad\qquad (4)$$

where X_{t+1}' is a vector of exogenous random variables and e_{t+1} is a zero mean martingale difference series meaning that $E\left(e_{t+1} \mid \Psi_t\right) = 0$. The optimal forecast under quadratic loss is 0 for Model 1 and $X_{t+1}'\beta$ for Model 2. Therefore we propose the following forecasting methods

$$f_{R,t} = 0$$

$$g_{R,t} = X_{t+1}'\hat{\beta}_{R,t}^i$$

where $\hat{\beta}_{R,t}^i$ represents a rolling estimate of the unknown parameter β using rolling window size R, information available up to time t and estimation method i.

Forecast evaluation is carried out simulating an out-of-sample exercise. One has $T + 1$ observations of y_{t+1} and X_{t+1}'. The first R observations are used for the first estimation. Therefore the first τ-step-ahead forecast is built at time R and compared with the realization $y_{R+\tau}$. The second forecast is obtained using the last R observations available for estimation. This forecast is compared with the realization $y_{R+1+\tau}$. We iterate like this until the $T + 2 - \tau - R$ forecast is built again using the last R observations available for estimation. This forecast is compared with the realization y_{T+1}. We generate a total of P_τ forecasts, with P_τ satisfying $R + (P_\tau - 1) + \tau = T + 1$. So

$$P_\tau = T + 2 - \tau - R$$

These forecasts are evaluated using a loss function $L_{t+\tau}\left(y_{t+\tau}, g_{R,t}\right)$ depending on both the forecasts and the realization of the data. We will focus our analysis in a quadratic loss function. Then we test the following null hypothesis

$$H_0 : E\left[L_{t+\tau}\left(y_{t+\tau}, f_{R,t} \right) - L_{t+\tau}\left(y_{t+\tau}, g_{R,t} \right) \mid \Psi_t \right] = 0 \ a.s. \ for \ all \ t \geq 0$$

The implementation of the conditional approach relies on the fact that the null hypothesis is equivalent to

$$H_0 : E\left[h_t \left(L_{t+\tau}\left(y_{t+\tau}, f_{R,t} \right) - L_{t+\tau}\left(y_{t+\tau}, g_{R,t} \right) \right) \right] = 0$$
$$\doteq E\left[h_t \Delta L_{R,t+\tau} \right] = 0 \ a.s. \ for \ all \ t \geq 0$$

for all ψ_t –measurable function h_t.

We first select our preferred choice of a $q \times 1$ test function h_t to construct the relevant statistics that are described next.

3.1. One-Step-Ahead Conditional Test

When $\tau = 1$, the sequence $h_t \Delta L_{R,t+\tau}$ is a martingale difference sequence if the null is true. Giacomini and White (2006) propose the following statistic for the test of equal conditional predictive ability

$$T_{P_\tau,R}^h = P_\tau \left(\bar{Z}'_{P_\tau,R} \, \hat{\Omega}_{P_\tau}^{-1} \bar{Z}_{P_\tau,R} \right) \tag{5}$$

where

$$\bar{Z}_{P_\tau,R} = \frac{1}{P_\tau} \sum_{t=R}^{T} Z_{R,t+1}$$

$$Z_{P_\tau,R} = h_t \Delta L_{R,t+1}$$

$$\hat{\Omega}_{P_\tau} = \frac{1}{P_\tau} \sum_{t=R}^{T} Z_{R,t+1} Z'_{R,t+1}$$

Giacomini and White (2006) give conditions under which the asymptotic distribution of $T_{P_\tau,R}^h \mid H_0$ is Chi-square[5]

$$T_{P_\tau,R}^h \mid H_0 \xrightarrow{D} \chi_q^2 \ as \ P_\tau \to \infty.$$

---

[5] These are summarized in the Appendix.

3.2. Multi-Step Conditional Test

When $\tau > 1$ Giacomini and White (2006) propose the following statistic for the test of equal conditional predictive ability

$$T^h_{P_\tau,R,\tau} = P_\tau (\bar{Z}'_{P_\tau,R} \tilde{\Omega}^{-1}_{P_\tau} \bar{Z}_{P_\tau,R}) \tag{6}$$

where

$$\bar{Z}_{P_\tau,R} = \frac{1}{P_\tau} \sum_{t=R}^{T-\tau} Z_{R,t+\tau+1}$$

$$Z_{R,t+\tau+1} = h_t \Delta L_{R,t+\tau}$$

and $\tilde{\Omega}_{P_\tau}$ is a HAC estimate of the variance of $Z_{R,t+\tau+1}$ computed according to Newey and West (1987).

Giacomini and White (2006) give conditions under which the asymptotic distribution of $T^h_{P_\tau,R,\tau} \mid H_0$ is Chi-square[6]

$$T^h_{P_\tau,R,\tau} \mid H_0 \xrightarrow{D} \chi^2_q \ as \ P_\tau \to \infty .$$

3.3. A Forecasting Decision Rule

Assume that we are able to carry out a test of conditional predictive ability and we are also able to reject the null hypothesis. We then need to decide how to build a forecast for time $T + 2$. Rejection of the null hypothesis gives statistical evidence indicating that one forecasting method is more accurate than the other, and that the test function h_{T+1} contains useful information for the determination of the best forecasting method. Giacomini and White (2006) propose the following decision rule:

1. Pick a threshold level c.

2. Regress $\Delta L_{R,t+\tau}$ on h_t over the out-of-sample period to obtain the regression coefficient $\hat{\alpha}$.

3. Pick the forecast $g_{R,T+1}$ if $h'_t \hat{\alpha} > c$ and choose f otherwise.

---

[6]   These are summarized in the Appendix.

Giacomini and White (2006) also propose an indicator to evaluate the number of times this decision rule would have chosen forecast method g over f or the other way around

$$I_{c,P_\tau} = \frac{1}{P_\tau} \sum_{t=R}^{T+1-\tau} 1\{h_t' \hat{\alpha} > 0\} \qquad (7)$$

where

$$1\{h_t' \hat{\alpha} > 0\} = \left\{ \begin{array}{l} 1 \ if \ h_t' \hat{\alpha} > 0 \\ 0 \ if \ h_t' \hat{\alpha} \leq 0 \end{array} \right\}.$$

We will implement this same indicator in our empirical application.

4. FORECASTING METHODS AND MODELS

4.1. Derivation of a Forecasting Shrinkage Estimator

In this subsection we derive a new shrinkage estimator for the parameter of a linear regression model. Interestingly, this new shrinkage estimator provides a natural interpretation for the matrix of perturbations typically used in ridge regressions.

Let us assume that $\{y_t, X_t, e_t\}_{t=1}^{\infty}$ is a sequence satisfying the following expression

$$y_{t+1} = X_{t+1}' \beta_0 + \varepsilon_{t+1} \qquad (8)$$

where now β_0 is the true value of the parameter of the model and

$$E[e_{t+1} \mid \Psi_t] = 0$$

where $\{\Psi_t\}$ represents a filtration such that Ψ_t is the sigma-field generated by current and past Xs and es.

$$\Psi_t = \sigma\{X_{t+1}, e_t, X_t, e_{t-1}, X_{t-1}, e_{t-2}, ...\}$$

A traditional OLS and ridge estimators for β_0 are given by the following expression:

$$\hat{\beta}_{OLS} = \left[\sum_{t=0}^{R-1} (X_{t+1} X_{t+1}') \right]^{-1} \sum_{t=0}^{R-1} (y_{t+1} X_{t+1})$$

$$\hat{\beta}_{Ridge} = \left[\sum_{t=0}^{R-1} \left(X_{t+1} X'_{t+1} \right) + \lambda I_{k \times k} \right]^{-1} \sum_{t=0}^{R-1} \left(y_{t+1} X_{t+1} \right)$$

where λ is called the ridge factor. In principle the right or optimal ridge factor is unknown and needs to be estimated. We propose a natural approach to this problem: an approach that uses the context of out-of-sample model evaluation and has the same variance reduction advantages of traditional ridge regressions.

We will assume that the number of observations available $\left\{ y_{t+1}, X_{t+1} \right\}_{t=0}^{T}$ is $T + 1 = P + R$, where R is the size of the estimation window and P is the size of the prediction window. In this case we want to find an estimate of β_0 by solving the following problem:

$$\min_{\{\beta\}} L\left(\beta; \beta_0 \right) = E\left(y_{t+1} - X'_{t+1} \beta \right)^2 \tag{9}$$

To build our first forecasts we only use R observations of our sample, and we want to engage in a one-step-ahead prediction exercise that not only minimizes an in-sample estimate of the loss function, but also a combination of an in-sample estimate and an out-of-sample estimate of the loss function.

Notice that we could rewrite the problem (9) as

$$\min_{\{\beta\}} E\left[E\left(y_{t+1} - X'_{t+1} \beta \right)^2 | \Psi_R \right].$$

The expectation in (9) can be estimated as follows

$$E\left[E\left[\left(y_{t+1} - X'_{t+1} \beta \right)^2 | \Psi_R \right] \right] \approx \frac{1}{N} \sum_{t=0}^{N} E\left[\left(y_{t+1} - X'_{t+1} \beta \right)^2 | \Psi_R \right].$$

Let us consider now $N \gg R$. We have that

$$E\left[E\left[\left(y_{t+1} - X'_{t+1} \beta \right)^2 | \Psi_R \right] \right] = \sum_{t=0}^{R-1} E\left[\left(y_{t+1} - X'_{t+1} \beta \right)^2 | \Psi_R \right]$$
$$+ \sum_{t=R}^{N} E\left[\left(y_{t+1} - X'_{t+1} \beta \right)^2 | \Psi_R \right]$$

The terms inside the first expectation on the right hand side are a function of terms belonging to the information set Ψ_R so we could write the previous expression as

$$\sum_{t=0}^{R-1} E\left[\left(y_{t+1} - X_{t+1}'\beta\right)^2 | \Psi_R\right] = \sum_{t=0}^{R-1}\left(y_{t+1} - X_{t+1}'\beta\right)^2 +$$

$$\sum_{t=R}^{N} E\left[\left(y_{t+1} - X_{t+1}'\beta\right)^2 | \Psi_R\right]$$

Furthermore the second term on the right hand side can be decomposed as follows:

$$\sum_{t=0}^{N} E\left[\left(y_{t+1} - X_{t+1}'\beta\right)^2 | \Psi_R\right] = \sum_{t=R}^{N} E\left[\left(X_{t+1}'\beta_0 + e_{t+1} - y_{t+1} - X_{t+1}'\beta\right)^2 | \Psi_R\right]$$

$$= \sum_{t=R}^{N} E\left[\left(X_{t+1}'(\beta_0 - \beta) + e_{t+1}\right)^2 | \Psi_R\right]$$

$$= \sum_{t=R}^{N} E\left[(\beta_0 - \beta)'X_{t+1}'X_{t+1}(\beta_0 - \beta) + e_{t+1}^2 + 2X_{t+1}'(\beta_0 - \beta)e_{t+1} | \Psi_R\right]$$

Notice that

$$\sum_{t=R}^{N} E\left[2X_{t+1}'(\beta_0 - \beta)e_{t+1} | \Psi_R\right] = 0$$

because for all $\tau > 0$

$$E\left[X_{R+\tau}'(\beta_0 - \beta)e_{R+\tau} | \Psi_R\right] = E\left[E\left[X_{R+\tau}'(\beta_0 - \beta)e_{R+\tau} | \Psi_{R+\tau-1}\right] | \Psi_R\right]$$

$$= E\left[X_{R+\tau}'(\beta_0 - \beta)E\left[e_{R+\tau} | \Psi_{R+\tau-1}\right] | \Psi_R\right]$$

$$= 0 \ for \ all \ \tau > 0$$

therefore

$$\sum_{t=0}^{N} E\left[\left(y_{t+1} - X_{t+1}'\beta\right)^2 | \Psi_R\right] = \sum_{t=R}^{N} E\left[(\beta_0 - \beta)'X_{t+1}'X_{t+1}(\beta_0 - \beta) + e_{t+1}^2 | \Psi_R\right]$$

so finally

$$\sum_{t=0}^{N} E\left[\left(y_{t+1} - X_{t+1}'\beta\right)^2 | \Psi_R\right] = \sum_{t=0}^{R-1}\left(y_{t+1} - X_{t+1}'\beta\right)^2 +$$

$$\sum_{t=R}^{N} E\left[(\beta_0 - \beta)'X_{t+1}'X_{t+1}(\beta_0 - \beta) + e_{t+1}^2 | \Psi_R\right]$$

taking derivatives with respect to β we finally have that β satisfies

$$\left[2\sum_{t=0}^{R-1}\left(X_{t+1}X_{t+1}'\right) + 2\sum_{t=R}^{N} E\left[X_{t+1}X_{t+1}' \mid \Psi_R\right]\right]\beta =$$

$$2\sum_{t=0}^{R-1}\left(y_{t+1}X_{t+1}\right) + 2\sum_{t=R}^{N} E\left[X_{t+1}'X_{t+1} \mid \Psi_R\right]\beta_0 - \sum_{t=R}^{N}\frac{\partial}{\partial\beta}E\left[e_{t+1}^2 \mid \Psi_R\right].$$

We define our statistic β by replacing the unknowns in the expressions above by sample estimates:

$$\left[2\sum_{t=0}^{R-1}\left(X_{t+1}X_{t+1}'\right) + 2\sum_{t=R}^{N} \hat{E}\left[X_{t+1}X_{t+1}' \mid \Psi_R\right]\right]\tilde{\beta} =$$

$$2\sum_{t=0}^{R-1}\left(y_{t+1}X_{t+1}\right) + 2\sum_{t=R}^{N} \hat{E}\left[X_{t+1}'X_{t+1} \mid \Psi_R\right]\hat{\beta}_0 - \sum_{t=R}^{N}\frac{\partial}{\partial\beta}\hat{E}\left[e_{t+1}^2 \mid \Psi_R\right].$$

We will assume that the conditional expectation of the square of the perturbations is independent of the parameter β. Furthermore, we are interested in a shrinkage estimator to obtain benefits from variance reduction. This leads us to pick $\beta_0 = 0$. Therefore, with appropriate assumptions of identification, we propose the following estimator:

$$\tilde{\beta} = \left[2\sum_{t=0}^{R-1}\left(X_{t+1}X_{t+1}'\right) + 2\sum_{t=R}^{N} \hat{E}\left[X_{t+1}X_{t+1}' \mid \Psi_R\right]\right]^{-1} \sum_{t=0}^{R-1}\left(y_{t+1}X_{t+1}\right) \qquad (10)$$

This estimator is similar to the ridge estimator presented earlier. We need to be precise about two elements of this new estimator: the choice of N, and the expectation formation $\hat{E}\left[X_{t+1}X_{t+1}' \mid \Psi_R\right]$.

For the later we propose to estimate a VAR(p) model on the regressors X_{t+1}. Usual model selection criteria may be followed. For the choice of N we propose three strategies:

1. $N = R$. The idea here is to impose the fact that when forecasting with rolling OLS regressions we are only imposing an in-sample minimization of the loss function. However the evaluation of the forecast accuracy involves comparing the forecast with the unknown predicted value. We try to overcome this situation with this new estimator.

2. $N = T + 1$. In this case we are imposing that our estimate will minimize a combination of the in-sample loss function and an estimate of the future loss

function that will be used to evaluate forecast accuracy. The problem with this scheme is that now $\tilde{\beta} = \tilde{\beta}(R,P)$. For many applications this is not a problem, but for the application we are interested in here, dependence from P may render a degenerate distribution for our loss function comparisons.

3. $N = \lambda R$, where $\lambda > 1$ gives us an approximation of the importance that the forecaster gives to the out-of-sample minimization versus the in-sample minimization.

Finally, we want to present a particular case in which there is only one regressor and this is a constant.

4.1.1. Example

Consider the original model in which X'_{t+1} is a constant, and the vector e_{t+1} is *i.i.d.* with homoscedastic shocks $E\left(e_{t+1}^2 \mid \Psi_T\right) = \sigma^2$. We could rewrite our model as

$$y_{t+1} = \beta_0 + e_{t+1}$$

In this particular case we have that

$$\frac{\partial}{\partial \beta} \hat{E}\left[e_{t+1}^2 \mid \Psi_R \right] = 0 \ \textit{for all} \ t > T$$

$$\sum_{t=R}^{N} \hat{E}\left[X_{t+1} X'_{t+1} \mid \Psi_R \right] = N - R + 1$$

$$\sum_{t=0}^{R-1} \left(X_{t+1} X'_{t+1} \right) = R$$

so, our estimate is

$$\tilde{\beta} = \frac{1}{N+1} \sum_{t=0}^{R-1} y_{t+1} + \frac{N-R+1}{N+1} \hat{\beta}_0$$

If we choose $\tilde{\beta}_0 = 0$ then we get a shrunken *OLS* for an arbitrary *N*:

$$\tilde{\beta}_1 = \frac{R}{N+1} \hat{\beta}_{OLS}$$

Furthermore, if we choose $N = T$, we have

$$\tilde{\beta}_2 = \frac{R}{P+R}\hat{\beta}_{OLS}.$$

Finally, if we choose $N = \lambda R - 1$, where $\lambda > 1$, then we have

$$\tilde{\beta}_3 = \frac{1}{\lambda}\hat{\beta}_{OLS}.$$

In our empirical application we will implement three of these variations. First, we will use (10) with a choice of N given by $N = \lambda R$, where $\lambda > 1$. Second, we will implement the full shrinkage approach given by $\tilde{\beta}_3$ and finally, we will also implement $\tilde{\beta}_1$.

4.2. Forecasting Methods

In this subsection we introduce the model and estimation strategies that are used for the implementation of the conditional tests. Our target is to build forecasts of the log difference of five US dollar bilateral exchange rates using monthly data. We analyze the cases of Canada, Japan, Switzerland, the UK and Chile[7]. We want to evaluate the conditional predictive ability of six different forecasting methods based upon an uncovered interest parity model, and compare their predictive ability with a forecasting method based upon a random walk model. All of the six methods basically posit that exchange rate returns are predicted by two regressors: a constant and the one-month interest differential. The forecasts are constructed according to Mark (1995) using the following equation:

$$s_{t+\tau} - s_t = \alpha_\tau + x_t\gamma_\tau + e_{t+\tau,t}$$

and we will denote

$$\beta_\tau = vec(\alpha_\tau, \gamma_\tau)$$

$$X'_{t+1} = (1, x_t)$$

---

[7] The data from Canada, Japan, Switzerland, and the UK were generously provided by Todd Clark and correspond to the same database used in Clark and West (2006). Interest rates correspond to 1-month Eurocurrency deposit rates, taking an average of bid and ask rates at London close. Monthly time series are formed as the last daily rate of each month. Data was obtained from Global Insight's FACS database. We obtained the data for Chile from the International Financial Statistics. This time we use the discount rates as measures of interest rates.

where s_t represents nominal exchange rate at time t and x_t represents interest rate differential.

The only difference between these six forecasting methods is the way the parameters are estimated. We will denote $\hat{\beta}^i_{t,\tau}$ as the estimate of β_τ using method i and the information available up until time t. A description of the different estimation methods follows next. For simplicity the analysis is written assuming $\tau = 1$.

The six different estimation approaches that we use have the following two features in common. First, all of these estimation approaches are rolling with an estimation window of the same size, R. Second, all six estimation techniques can be summarized by the following general expression

$$\hat{\beta}^i_{t,\tau=1} = \left[\sum_{s=t-R+1}^{t} \left(X_s X_s' \right) + M_i \right]^{-1} \sum_{s=t-R+1}^{t} \left(y_s X_s \right), \ i = 1,...,6$$

where M_i is a real matrix that truly identifies each of the proposed methods. The choice of M_i is described next:

1. Rolling OLS (OLS). The choice of M_1 is given by:

$$M_1 = 0$$

    so the unknown parameter β is estimated via OLS using the last R available observations. Therefore $\hat{\beta}^1_{t,\tau=1}$ satisfies:

$$\left[\sum_{s=t-R+1}^{t} \left(X_s X_s' \right) \right]_i \hat{\beta}^1_{t,\tau=1} = \sum_{s=t-R+1}^{t} \left(y_s X_s \right)$$

2. Rolling Bayesian Shrinkage (Bayesian): The choice of M_2 is given by:

$$M_2 = \hat{\sigma}^2 V^{-1}$$

    where $\hat{\sigma}^2$ is the standard deviation of the residuals of a regression between y_{t+1} and y_t. V is the diagonal variance-covariance matrix for the prior distribution of β. We set

$$\alpha \sim N\left(0,10^8\right)$$

$$\gamma \sim N\left(0, \left(\varpi\lambda\left(\hat{\sigma}_y / \hat{\sigma}_x\right)\right)^2 \right)$$

where $\hat{\sigma}_y$ represents the sample standard deviation of the dependent variable (exchange rate returns) and $\hat{\sigma}_x$ represents the sample variance of the interest rate differential variable. We also need to provide *a priori* values for the hyperparameters λ and ϖ. Following Litterman (1986) we use $\lambda = \varpi = 0.2$.

3. Deterministic Rolling Out-of-Sample OLS (Det OOS-OLS): The choice of M_3 is given by:

$$M_3 = (\mu - 1) \sum_{s=t-R+1}^{t} \left(X_s X_s' \right), \ \mu \geq 1$$

with a choice of μ given by

$$\mu = \frac{R + N - t}{R}, \ t = R, R+1, ..., T+1, \ N = N(R) \gg T+1$$

4. Full Shrinkage Approach (Full): The choice of M_4 is given by:

$$M_3 = (\mu - 1) \sum_{s=t-R+1}^{t} \left(X_s X_s' \right), \ \mu > 1$$

where the parameter $\mu - 1 > 0$ is arbitrarily big. In our empirical application we set $\mu = 20$. It is easy to see that

$$\hat{\beta}_{t,\tau=1}^4 = \frac{\hat{\beta}_{t,\tau=1}^1}{\mu}$$

5. Rolling Out-of-Sample OLS (OOS-OLS): The choice of M_5 is given by:

$$\sum_{s=t+1}^{N(R)} E\left(X_s X_s' \mid \Psi_{t+1} \right)$$

where the expectation $E\left(X_s X_s' \mid \Psi_{t+1} \right)$ is estimated fitting a VAR(p) model over the vector X_{t+1}, $t = R, ..., R + P_\tau - 1$. In our empirical application we use prior information about the process of the interest rate differential. Following Clark and West (2006) we fit an AR(1) model.

6. Rolling Ridge Regression (Ridge): The choice of M_6 is given by:

$$M_6 = \lambda I_{k \times k}$$

where the ridge parameter $\lambda > 0$ is set to $\lambda = 20$, and k is the number of variables in the regression.

We will use these methods to evaluate conditional predictive ability of several bilateral exchange rates in the next section.

5. EMPIRICAL RESULTS

In this section we present results for a number of tests of conditional predictive ability for five bilateral exchange rates. We analyze the cases of Canada, Japan, Switzerland, the UK and Chile. For these countries we take the series of 1-, 2-, 4-, 6-, 8-, 12- and 16-months ahead forecast errors to conduct tests of conditional predictive ability using a quadratic loss function. For Canada, Japan, Switzerland and the UK we set $R = 120$ and $P = T + 1 - R$, where $T + 1$ is the total number of observations. For Chile we set $R = 36$, $P = 108$. Our main goal is to evaluate if any of our forecasting methods may outperform the random walk. For each country we run a total of 126 tests of the following form

$$H_0 : E\left[h_{it}\left(y_{t+\tau}^2 - \left(y_{t+\tau} -, g_{R,t}^{j,k}\right)^2 \right)\right] = 0, \ i = 1,...,3, \ j = 1,...,6, \ k \in \Theta$$

$$\Theta = \{1,2,4,6,8,12,16\}$$

where subscript i denotes the type of test function used in the analysis, j denotes the estimation technique used to obtain parameter estimates of the model and k denotes the 7 horizons used in the analysis according to the following description:

$$h_{1t} = 1$$

$$h_{2t} = vec\left(1, \Delta L_{R,t}\right)$$

$$h_{3t} = vec\left(1, \left(x_t^2\right)^{1/5}, \Delta L_{R,t}\right)$$

$j = 1$ means traditional OLS estimation, $j = 2$ means Bayesian shrinkage estimation, $j = 3$ means deterministic out-of-sample OLS, $j = 4$ means a full shrinkage procedure, $j = 5$ means Out-of-Sample OLS, and $j = 6$ means a ridge regression. All these methods are described in the previous section. Finally, $k = 1, 2, 4, 6, 8, 12, 16$, denotes the horizon of the analysis.

In case of rejection of the null hypothesis we also implement the decision rule in (7) to evaluate which method would have been selected. We also report the percentage gain (loss) in Mean Square Prediction Error (MSPE) for all the predictions.

Tables 1-5 in the appendix show p-values of the tests of conditional predictive ability for each of the five countries. Tables 6-10 in the appendix show the percentage gain (loss) in MSPE of each forecasting method with respect to the random walk. We analyze our results according to the following criteria. First, we simply want to know whether it is possible to beat the random walk in a real-time forecasting exercise under quadratic loss. Second, in case evidence of predictability is found, we want to know how predictability varies along different predictive horizons. Third, in case evidence of predictability is found, we would like to know how the conditional tests perform better than the unconditional tests. Finally, we would like to know the size of the improvement in predictability should any evidence of predictability be found, and we would also like to identify the best estimation method to carry out a real time exchange rate forecasting exercise.

5.1. Predictability

Tables 1-5 shows p-values for the null of equal predictive ability. A minus sign indicates that the decision rule in (7), suggests using the random walk as a forecasting method, while a plus sign indicates the decision rule points to the corresponding alternative approach. Tables 1-4 show that evidence of predictability is found for Canada, Chile, Japan and Switzerland, as some p-values are lower than the 10% significance level and (7) suggests using the corresponding alternative model to build forecasts. No predictability evidence is found for the UK, as every time the null of equal predictability is rejected, the decision rule in (7) suggests the use of the random walk over any other alternative approach considered.

5.2. Predictability Horizons

According to Tables 1-4, evidence of predictability is found for Canada at the 1-, 2- and 4-month ahead forecast horizon. For Chile and Switzerland, evidence of predictability is found at every single considered horizon. For Japan, evidence of predictability is only found at the 4-, 6- and 8-month ahead forecast horizon.

Differing from Mark (1995), we do not see long-term predictability dominating over short term predictability. In fact, we see that every time there is long time

predictability (Chile and Switzerland) there is also short term predictability. In this respect these results are consistent with those of McCracken and Sapp (2005) and those of Kilian (1999).

5.3. Conditional and Unconditional Predictability

Tables 1-5 report three panels of p-values. Each panel corresponds to a different testing function h_{it}, $i = 1, \ldots, 3$. We remark here that the first testing function is $h_{1t} = 1$. In this case the conditional approach of Giacomini and White (2006) reduces to an unconditional approach in which the true unknown value of the parameters is replaced by sample estimates. We compare results from the first panel ($h_{1t} = 1$) with results obtained using more general testing functions. These results are shown in panels 2 and 3, labeled $h2$ and $h3$ in Tables 1-5. We compare whether rejection in panels 2 and 3 is encompassed by rejection in panel 1. In case rejections in panel 2 and 3 provide new information, we attempt to check for robustness of these rejections by comparing suggestions from the decision rule (7) and the sign of the difference in MSPE[8].

Table 1 shows that for Canada there is no new information from the truly conditional panels 2 and 3, as any rejection of the null of equal predictive ability in panels 2 and 3 is also found in the "unconditional" panel 1. Quite the contrary happens with Chile. For this country, the "unconditional" approach shows no rejection whatsoever. The "truly" conditional panels, however, show a number of rejections that are consistent with Table 7 in terms of choosing the forecasting model displaying the lowest MSPE. Out of 28 rejections, there is only 1 "mistake"[9]. For Japan the conditional panels add two new "correct" rejections in favor of the random walk forecasting method. For Switzerland, the conditional panels add 11 new and "correct" rejections whereas for the UK the conditional panels add two new and "correct" rejections.

Overall, we see that conditioning seems to help in getting statistically significant information about conditional predictive ability. Notice that with a simple unconditional approach, (panel 1) fewer rejections of the null of equal conditional predictive ability would have occurred.

---

[8]   We have three tests for each forecasting method and predictive horizon. We make a forecasting decision if the number of rejections in favor of one method outnumbers the number of rejections in favor of the competing method.

[9]   For two month ahead forecasts and OOS-OLS forecasting method, the use of the testing function h3 jointly with the decision rule suggests choosing the OOS-OLS forecasting method over the random walk, yet the MSPE of the random walk is lower than that of its competing forecasting method. We label this situation as a mistake. It only happens once.

5.4. Predictability Size and Best Method

Tables 6-10 show percentage gains (losses) in out-of-sample MSPE between the random walk and the corresponding alternative forecasting method. A negative value means that the random walk displays lower out-of-sample MSPE whereas a positive value means that the corresponding alternative forecasting method is more accurate in terms of quadratic loss.

Once the null of equal conditional predictive ability is rejected against the random walk, we care about the size of the out-of-sample MSPE. We measure this as the percentage gain in MSPE over the random walk. From Tables 6-10 we see that MSPE percentage gains range from 0% to 2.5% with an average gain of 0.76%. Even though gains are mild, statistical rejection suggests that they are also systematic.

In terms of choosing a forecasting method we consider two variables: power and predictive accuracy. In other words, a forecasting method is good if in a testing environment it yields a powerful test and, if used in a predictive exercise, its accuracy is high. For the first point of view we see that the full shrinkage approach is the most powerful method as it accounts for the 43% of all the rejections. The rest of the shrinkage procedures provide roughly around the same number of rejections. Notice that via direct OLS estimation there is no rejection whatsoever. It is worth mentioning that excluding the case of Chile, the only three methods providing rejection are the three proposed shrinkage methods: full shrinkage, Det OOS-OLS and OOS-OLS.

In terms of forecast accuracy, the full shrinkage approach performs poorly as it provides percentage MSPE gains ranging between 0% and 0.3%. Much higher MSPE gains can be obtained with Det OOS-OLS and OOS-OLS, methods which give improvements up to 1.2% and 2.1% respectively.

In summary we confirm that shrinkage methods are more appropriate than simple OLS estimation to provide both more powerful tests of conditional predictive ability and more accurate forecasts. We also showed that the three proposed shrinkage methods perform well and sometimes much better than their considered competitors.

6. DISCUSSION

This paper evaluates exchange rate predictability using a new conditional framework developed by Giacomini and White (2006). Instead of testing an economic theory, this framework is more appropriate for an applied forecaster trying to assess which of two competing forecasting methods will provide more accurate forecasts in the future.

We use six different forecasting methods, based upon a model of interest parity, to test the null of equal conditional predictive ability when the benchmark forecasting

method is a random walk. We consider seven different predictive horizons to perform a total of 126 tests for each bilateral exchange rate corresponding to Canada, Chile, Japan, the UK and Switzerland.

Our results indicate that all bilateral exchange rates, with the exception of the British pound, display statistically significant evidence of conditional predictability against the random walk, at least for some small group of predictive horizons. Furthermore, our results reveal that conditional predictive ability is more frequently found at shorter or medium horizons rather than at longer horizons.

This is interesting because it coincides with results showed by McCracken and Sapp (2005) and Kilian (1999). We emphasize again that our question and testing framework are different than those in previous papers. We are trying to detect exchange rate predictability from a forecaster point of view and we are not directly interested in testing economic theory.

We also provide evidence indicating that shrinkage methods are more appropriate than simple OLS estimation in providing both more powerful tests of conditional predictive ability and more accurate forecasts. Similarly, we show that the three proposed shrinkage methods perform well and sometimes much better than their considered competitors.

We have made a number of assumptions to obtain our results. For instance, all of our forecasting methods are based upon the simple interest parity model. We have also chosen priors, testing functions, ridge factors, loss functions, and forecasting and estimation windows size, among other variables. A natural extension of this paper should relax some of these assumptions. The consideration of more models, more estimation techniques and the use of bootstrap critical values are left for future research.

TABLE 1

P-VALUES FOR CANADIAN EXCHANGE RATE CONDITIONAL PREDICTIVE ABILITY TESTS

	Horizon	(1) OLS	(2) Bayesian	(3) Det OOS-OLS	(4) Full	(5) OOS-OLS	(6) Ridge
h1	1	0.29	0.55	0.07+	0.04+	0.24	0.64
	2	0.39	0.68	0.15	0.04+	0.38	0.72
	4	0.52	0.74	0.19	0.06+	0.49	0.75
	6	0.79	0.78	0.52	0.26	0.70	0.77
	8	0.92–	0.79	0.82	0.58	0.88	0.78
	12	0.87	0.80	0.80	0.46	0.91–	0.77
	16	0.99–	0.84	0.87	0.67	0.98–	0.80
h2	1	0.68	0.62	0.16	0.09+	0.53	0.71
	2	0.61	0.50	0.38	0.10+	0.80	0.51
	4	0.70	0.49	0.68	0.20	0.95	0.50
	6	0.58	0.45	0.87	0.52	0.86	0.47
	8	0.26	0.42	0.52	0.73	0.53	0.44
	12	0.36	0.42	0.58	0.90	0.27	0.45
	16	0.02–	0.32	0.51	0.73	0.06–	0.40
h3	1	0.85	0.77	0.28	0.18	0.69	0.88
	2	0.80	0.70	0.50	0.20	0.88	0.69
	4	0.76	0.68	0.75	0.34	0.88	0.67
	6	0.76	0.63	0.93	0.72	0.92	0.63
	8	0.40	0.57	0.66	0.79	0.71	0.59
	12	0.52	0.58	0.68	0.90	0.40	0.61
	16	0.04–	0.45	0.71	0.89	0.08–	0.53

Notes:
1.   See Section 4.2 for a description of the six forecasting methods that are used, besides the random walk. We have labeled these methods as OLS (meaning estimation is done via OLS), Bayesian (meaning estimation is done via Bayesian Shrinkage), Det OOS-OLS (meaning estimation is done via a deterministic Out-of-Sample OLS), Full (meaning estimation is done by shrinking OLS estimates by a big number), OOS-OLS (meaning estimation is done via Out-of-Sample OLS) and Ridge (meaning estimation is done via a ridge regression).
2.   A plus sign (+) is used to denote rejections at 10% in favor of the alternative method.
3.   A minus sign (–) is used to denote rejections at 10% in favor of the random walk.
4.   Rejection in favor of either method is determined using the decision rule described in Section 3.3.
5.   First column shows the three different test functions used in the analysis. See beginning of Section 5 for a description of these three testing functions.
6.   Second column shows the different horizons considered in the analysis in a monthly basis.
7.   In each cell we report *p*-values for a null of equal conditional predictive ability.
8.   Data range: 1980:01–2003:10 for Canada and Japan, 1975:01–2003:10 for Switzerland and UK and 1993:4–2005:4 for Chile.
9.   For Canada, Japan, Switzerland, and UK we set *R* = 120. For Chile we set *R* = 36.

TABLE 2

P-VALUES FOR CHILEAN EXCHANGE RATE CONDITIONAL PREDICTIVE ABILITY TESTS

Horizon		(1) OLS	(2) Bayesian	(3) Det OOS-OLS	(4) Full	(5) OOS-OLS	(6) Ridge
h1	1	0.66	0.46	0.48	0.19	0.28	0.30
	2	0.70	0.47	0.69	0.24	0.52	0.31
	4	0.59	0.46	0.71	0.24	0.36	0.31
	6	0.77	0.48	0.57	0.28	0.64	0.31
	8	0.50	0.44	0.41	0.21	0.32	0.31
	12	0.53	0.46	0.73	0.36	0.37	0.30
	16	0.63	0.46	0.85	0.59	0.65	0.31
h2	1	0.36	0.41	0.31	0.34	0.53	0.34
	2	0.02–	0.01+	0.14	0.02+	0.11	0.01+
	4	0.04–	0.01+	0.05–	0.10+	0.17	0.01+
	6	0.02–	0.01+	0.09–	0.04+	0.32	0.01+
	8	0.10	0.01+	0.24	0.09+	0.26	0.01+
	12	0.07–	0.03+	0.08–	0.04+	0.13	0.02+
	16	0.09–	0.03+	0.36	0.74	0.09–	0.02+
h3	1	0.09–	0.09+	0.05+	0.04+	0.10+	0.09+
	2	0.02+	0.01+	0.11	0.01+	0.08+	0.01+
	4	0.04+	0.01+	0.04+	0.04+	0.07+	0.01+
	6	0.04+	0.01+	0.18	0.08+	0.33	0.01+
	8	0.15	0.01+	0.32	0.16	0.26	0.01+
	12	0.05+	0.02+	0.13	0.09+	0.10+	0.02+
	16	0.09+	0.02+	0.24	0.37	0.04+	0.02+

Notes:
See notes on Table 1.

TABLE 3

P-VALUES FOR JAPANESE EXCHANGE RATE CONDITIONAL PREDICTIVE ABILITY TESTS

	Horizon	(1) OLS	(2) Bayesian	(3) Det OOS-OLS	(4) Full	(5) OOS-OLS	(6) Ridge
h1	1	0.70	0.62	0.20	0.13	0.60	0.59
	2	0.65	0.58	0.43	0.16	0.34	0.55
	4	0.47	0.47	0.36	0.09+	0.13	0.43
	6	0.41	0.47	0.22	0.05+	0.12	0.44
	8	0.45	0.50	0.19	0.04+	0.07+	0.48
	12	0.61	0.63	0.50	0.18	0.49	0.60
	16	0.62	0.61	0.44	0.14	0.69	0.60
h2	1	0.78	0.11	0.57	0.38	0.60	0.13
	2	0.73	0.23	0.95	0.53	0.91	0.25
	4	0.92	0.27	0.88	0.38	0.52	0.28
	6	0.97	0.27	0.66	0.22	0.42	0.29
	8	0.97	0.29	0.65	0.21	0.27	0.30
	12	0.80	0.33	1.00	0.62	0.70	0.34
	16	0.58	0.32	0.47	0.27	0.41	0.33
h3	1	0.65	0.06–	0.33	0.26	0.13	0.07–
	2	0.64	0.21	0.57	0.38	0.28	0.21
	4	0.78	0.20	0.38	0.16	0.23	0.19
	6	0.72	0.21	0.25	0.11	0.15	0.20
	8	0.75	0.24	0.18	0.01+	0.14	0.22
	12	0.53	0.27	0.37	0.27	0.13	0.25
	16	0.42	0.26	0.26	0.28	0.45	0.24

Notes:
See notes on Table 1.

TABLE 4

P-VALUES FOR SWISS EXCHANGE RATE CONDITIONAL PREDICTIVE ABILITY TESTS

Horizon	(1) OLS	(2) Bayesian	(3) Det OOS-OLS	(4) Full	(5) OOS-OLS	(6) Ridge
h1						
1	0.55	0.34	**0.08+**	**0.04+**	0.22	0.29
2	0.53	0.42	0.29	**0.10+**	0.14	0.38
4	0.55	0.39	0.40	0.13	0.22	0.37
6	0.53	0.41	0.39	0.14	0.28	0.40
8	0.53	0.44	0.47	0.21	0.38	0.43
12	0.69	0.53	0.74	0.43	0.69	0.53
16	0.86	0.59	0.72	0.40	0.95−	0.64
h2						
1	0.60	0.49	0.15	0.10+	0.30	0.30
2	0.25	0.37	0.09+	0.04+	0.44	0.26
4	0.22	0.33	0.08+	0.03+	0.38	0.22
6	0.27	0.38	0.11	0.06+	0.72	0.27
8	0.53	0.48	0.17	0.09+	0.95	0.40
12	0.86	0.67	0.12	0.09+	0.81	0.71
16	0.42	0.56	0.10−	0.12	0.25	0.59
h3						
1	0.67	0.25	0.10+	0.07+	0.36	0.16
2	0.17	0.21	0.09+	0.07+	0.31	0.15
4	0.24	0.22	0.10+	0.06+	0.31	0.16
6	0.28	0.26	0.10+	0.08+	0.56	0.20
8	0.52	0.35	0.08+	0.06+	0.74	0.30
12	0.80	0.43	0.10	0.06+	0.92	0.49
16	0.61	0.34	0.08−	0.04+	0.40	0.48

Notes:
See notes on Table 1.

TABLE 5

P-VALUES FOR BRITISH EXCHANGE RATE CONDITIONAL PREDICTIVE ABILITY TESTS

Horizon		(1) OLS	(2) Bayesian	(3) Det OOS-OLS	(4) Full	(5) OOS-OLS	(6) Ridge
h1	1	0.90	0.95–	0.66	0.51	0.68	0.84
	2	0.88	0.96–	0.79	0.57	0.77	0.87
	4	0.94–	0.99–	0.91–	0.76	0.87	0.97–
	6	0.92–	0.99–	0.85	0.68	0.85	0.98–
	8	0.85	0.99–	0.74	0.51	0.75	0.94–
	12	0.97–	0.97–	0.73	0.44	0.91–	0.96–
	16	0.99–	0.98–	0.89	0.7	0.95–	0.99–
h2	1	0.26	0.10	0.45	0.45	0.60	0.36
	2	0.34	0.17	0.34	0.23	0.39	0.40
	4	0.25	0.02–	0.22	0.24	0.24	0.17
	6	0.33	0.02–	0.43	0.39	0.32	0.11
	8	0.38	0.04–	0.34	0.23	0.28	0.26
	12	0.10–	0.09–	0.67	0.52	0.37	0.10
	16	0.03–	0.06–	0.44	0.43	0.28	0.04–
h3	1	0.42	0.17	0.62	0.66	0.80	0.53
	2	0.28	0.31	0.23	0.35	0.47	0.58
	4	0.04–	0.03–	0.01–	0.02–	0.12	0.29
	6	0.12	0.01–	0.10–	0.27	0.21	0.21
	8	0.36	0.04–	0.26	0.36	0.36	0.41
	12	0.16	0.10	0.73	0.72	0.58	0.11
	16	0.05–	0.06–	0.26	0.42	0.46	0.05–

Notes:
See notes on Table 1.

TABLE 6

PERCENTAGE GAIN (LOSS) IN MEAN SQUARE PREDICTION ERROR FOR CANADA

Horizon	(1) OLS	(2) Bayesian	(3) Det OOS-OLS	(4) Full	(5) OOS-OLS	(6) Ridge
1	1.70	-0.21	1.20	0.28	1.40	-0.48
2	0.55	-0.61	1.80	0.22	0.47	-0.67
4	-0.10	-0.85	1.60	0.20	0.03	-0.76
6	-1.20	-1.00	-0.10	0.07	-0.65	-0.84
8	-2.30	-1.10	-1.50	-0.03	-1.40	-0.88
12	-2.10	-1.20	-1.50	0.01	-1.50	-0.88
16	-3.90	-1.30	-2.40	-0.07	-2.30	-0.97

Notes:
1. See notes on Table 1.
2. Each number in each cell represents the percentage gain or loss in Mean Square Prediction Error of each alternative forecasting model with respect to that of the random walk. For instance, in column (2) and horizon 1 there is a 1.70 value. This means that the out-of-sample mean square prediction error of the forecasting model estimated with OLS is 1.70% lower than the out-of-sample mean square prediction error of the random walk model.
3. Light shaded cell: It means that the conditional approach with testing function $h = 1$ correctly rejected the null of equal conditional predictive ability. By "correct" we mean that the sign of the Mean Square Prediction Error Percentage gain (loss) is consistent with our decision forecasting rule.
4. Dark shaded cell: It means that the conditional approach with testing function $h2$ or $h3$ correctly rejected the null of equal conditional predictive ability. By "correct" we mean that the sign of the Mean Square Prediction Error Percentage gain (loss) is consistent with our decision forecasting rule.

TABLE 7

PERCENTAGE GAIN (LOSS) IN MEAN SQUARE PREDICTION ERROR FOR CHILE

Horizon	(1) OLS	(2) Bayesian	(3) Det OOS-OLS	(4) Full	(5) OOS-OLS	(6) Ridge
1	–2.70	0.57	0.14	0.28	2.50	1.80
2	–3.40	0.45	–1.70	0.20	–0.23	1.80
4	–1.50	0.57	–1.50	0.17	1.30	1.80
6	–6.10	0.35	–0.55	0.17	–1.60	1.80
8	0.08	0.82	0.95	0.29	2.20	1.80
12	–0.46	0.61	–2.30	0.11	1.40	1.90
16	–2.30	0.57	–3.40	–0.07	–1.70	1.90

Notes:

1. See notes on Table 6.

2. In this table only the darker shaded cell represents the only inconsistent results between our testing strategy and Mean Square Prediction error percentage gains. In this case we have that the tests indicate that there is rejection of the null of equal conditional predictive ability in favor of the forecasting method estimated with OOS-OLS. However, the Out-of-sample MSPE is lower for the random walk model.

TABLE 8

PERCENTAGE GAIN (LOSS) IN MEAN SQUARE PREDICTION ERROR FOR JAPAN

Horizon	(1) OLS	(2) Bayesian	(3) Det OOS-OLS	(4) Full	(5) OOS-OLS	(6) Ridge
1	−2.00	−0.60	0.70	0.22	−0.20	−0.42
2	−1.50	−0.45	0.36	0.19	0.58	−0.25
4	0.30	0.18	0.79	0.25	1.70	0.34
6	0.90	0.16	1.60	0.32	1.90	0.30
8	0.50	0.00	1.90	0.34	2.10	0.12
12	−1.00	−0.79	−0.03	0.18	0.05	−0.57
16	−1.00	−0.71	0.33	0.24	−0.82	−0.56

Notes:
See notes on Table 6.

TABLE 9

PERCENTAGE GAIN (LOSS) IN MEAN SQUARE PREDICTION ERROR FOR SWITZERLAND

Horizon	(1) OLS	(2) Bayesian	(3) Det OOS-OLS	(4) Full	(5) OOS-OLS	(6) Ridge
1	–0.49	0.61	1.20	0.35	0.48	0.78
2	–0.27	0.28	1.10	0.30	0.89	0.41
4	–0.47	0.41	0.55	0.27	0.72	0.48
6	–0.24	0.33	0.58	0.26	0.46	0.37
8	–0.26	0.22	0.14	0.21	0.21	0.23
12	–1.30	–0.09	–1.40	0.04	–0.27	–0.10
16	–3.60	–0.33	–1.30	0.06	–1.10	–0.45

Notes:
See notes on Table 6.

TABLE 10

PERCENTAGE GAIN (LOSS) IN MEAN SQUARE PREDICTION ERROR FOR UNITED KINGDOM

Horizon	(1) OLS	(2) Bayesian	(3) Det OOS-OLS	(4) Full	(5) OOS-OLS	(6) Ridge
1	–4.40	–1.70	–0.32	–0.01	–1.30	–1.00
2	–5.10	–2.00	–1.70	–0.04	–2.50	–1.30
4	–4.30	–1.80	–2.80	–0.15	–2.50	–1.50
6	–3.80	–1.70	–2.30	–0.11	–2.20	–1.40
8	–2.90	–1.20	–1.40	–0.01	–1.50	–1.10
12	–1.90	–1.00	–1.20	0.03	–1.10	–0.92
16	–2.40	–1.10	–2.20	–0.10	–1.30	–0.99

Notes:
See notes on Table 6.

APPENDIX

A.1. Theoretical Appendix

Theorem 1 *(Conditional Predictive Accuracy Test) For forecast horizon $\tau = 1$, $Z_{R,t+1} = h_t \Delta L_{R,t+1}$, maximum estimation window of size $R < \infty$ and $q \times 1$ test function sequence $\{h_t\}$ suppose:*

1. $\{y_t, X_t\}$, $\{h_t\}$ *are mixing sequences with ϕ of size $-r / (2r - 1)$, $r \geq 1$, or α of size $-r / (r - 1)$, $r > 1$.*

2. $E\left|Z_{R,t+1,i}\right|^{2(r+\delta)} < C$ *for some $\delta > 0$, $i = 1, \ldots, q$ and for all t.*

3. $\Omega_{P_\tau} = \dfrac{1}{P_\tau} \sum\limits_{t=R}^{T} E\left[Z_{R,t+1} Z'_{R,t+1} \right]$ *is uniformly positive definite.*

    Then under the null of equal conditional predictive accuracy

$$T_{P_\tau,R}^h \mid H_0 \xrightarrow{\ D\ } \chi_q^2 \text{ as } P_\tau \to \infty$$

   where $T_{P_\tau,R}^h$ is given by (5).

Proof *See Giacomini and White (2006).* ∎

Theorem 2 *(Multi-Step Conditional Predictive Accuracy Test) For given forecast horizon $\tau > 1$, $Z_{R,t+\tau} = h_t \Delta L_{R,t+\tau}$ maximum estimation window of size $R < \infty$ and $q \times 1$ test function sequence $\{h_t\}$ suppose:*

1. $\{y_t, X_t\}$, $\{h_t\}$ *are mixing sequences with ϕ of size $-r / (2r - 2)$, $r \geq 2$, or α of size $-r / (r - 2)$, $r > 2$.*

2. $E\left|Z_{R,t+1,i}\right|^{(r+\delta)} < C < \infty$ *for some $\delta > 0$, $i = 1, \ldots, q$ and for all t.*

3. $E\tilde{\Omega}_{P_\tau}$ *is uniformly positive definite, where $\tilde{\Omega}_{P_\tau}$ is a HAC estimate of the variance of $Z_{R,t+\tau+1}$.*

    Then under the null of equal conditional predictive accuracy

$$T_{P_\tau,R,\tau}^h \mid H_0 \xrightarrow{\ D\ } \chi_q^2 \text{ as } P_\tau \to \infty$$

   where $T_{P_\tau,R,\tau}^h$ is given by (6).

Proof *See Giacomini and White (2006).* ∎

REFERENCES

CHINN, M. and R.A. MEESE (1995). "Banking on Currency Forecasts: How Predictable is Change in Money?", *Journal of International Economics* 38, pp. 161-178.

CLARK, T. and K. WEST (2006). "Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis", *Journal of Econometrics* 135, pp. 155-186.

DIEBOLD, F. and R. MARIANO (1995). "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics* 13, pp. 253-263.

GIACOMINI, R. and H. WHITE (2006). "Test of Conditional Predictive Ability", *Econometrica* 74, pp. 1545-1578.

KILIAN, L. (1999). "Exchange Rates and Monetary Fundamentals: What Do We Learn From Long-Horizon Regressions?", *Journal of Applied Econometrics* 14, pp. 491-510.

LITTERMAN, R.B. (1986). "Forecasting with Bayesian Vector Autoregressions - Five Years of Experience", *Journal of Business and Economic Statistics* 4, pp. 25-38.

MARK, N. (1995). "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability", *American Economic Review* 85, pp. 201-218.

MacDONALD, R. and I. MARSH (1997). "On Fundamentals and Exchange Rates: A Casselian Perspective", *Review of Economics and Statistics* 79, pp. 655-664.

McCRACKEN, M. (2004). "Asymptotics for Out-of-sample Tests of Causality". *Manuscript*, University of Missouri.

McCRACKEN, M. and S. SAPP (2005). "Evaluating the Predictability of Exchange Rates Using Long Horizon Regressions", *Journal of Money, Credit, and Banking* 37 (3), pp. 473-494.

MEESE, R. and K. ROGOFF (1983a). "Empirical Exchange Rate Models of the Seventies. Do They Fit Out-of-Sample?", *Journal of International Economics* 14, pp. 3-24.

MEESE, R. and K. ROGOFF (1983b). "The Out-of-Sample Failure of Empirical Exchange Rate Models?", in J. Frenkel (ed.), *Exchange Rates and International Macroeconomics*, pp. 67-105. Chicago: University of Chicago Press.

NEWEY, W. and K. WEST (1987). "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica* 55, pp. 703-708.

WEST, K. (1996). "Asymptotic Inference About Predictive Ability", *Econometrica* 64, pp. 1067-1084.

WEST, K. (2006). "Forecast Evaluation", in G. Elliot, C.W. Granger and A. Timmerman (eds.), *Handbook of Economic Forecasting*, Volume 1, Holland: Elsevier.