Banco Central de Chile
Documentos de Trabajo

Central Bank of Chile
Working Papers

N° 376

Noviembre 2006

SHRINKAGE BASED TESTS OF THE MARTINGALE DIFFERENCE HYPOTHESIS

Pablo Pincheira

Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Agustinas 1180
Teléfono: (56-2) 6702475; Fax: (56-2) 6702231

Documento de Trabajo N° 376 **Working Paper N° 376**

SHRINKAGE BASED TESTS OF THE MARTINGALE DIFFERENCE HYPOTHESIS

Pablo Pincheira
Economista senior
Gerencia de Investigación Económica
Banco Central de Chile

Resumen

En este documento definimos una familia de pruebas para testear la hipótesis nula de que una serie de tiempo sea una Martingala en Diferencia (MD). La definición de esta familia de pruebas se basa en el principio de reducción de parámetros. Estas pruebas de hipótesis tienen en común que el rechazo de la hipótesis nula implica que el modelo alternativo, ajustado por un factor de reducción, necesariamente provee proyecciones con menor Error Cuadrático Medio (ECM) que el modelo que supone la hipótesis nula. Esta metodología generaliza la mayoría de los tests existentes pues ellos sólo comparan los errores de predicción del modelo nulo y el alternativo, sin ser este último modificado vía un proceso de reducción de parámetros. Observamos que los tests derivados de acuerdo a este principio de reducción tienen en general un mejor comportamiento en muestra pequeña. Esto ocurre pues nuestros tests se benefician implícitamente con la menor varianza de los estimadores de parámetros reducidos. Finalmente ilustramos el uso de nuestros tests con una aplicación a la predicción de tipos de cambio.

Abstract

In this paper we define a family of tests for the Martingale Difference Hypothesis (MDH) based upon a shrinkage principle. Tests within this family are such that rejection of the null implies that forecasts from the alternative model, adjusted by a shrinkage factor, will display lower Mean Square Prediction Error (MSPE) than forecasts from the null model. This generalizes most previous tests which compare forecast errors of one model, the null, to errors of the plain alternative model, not allowing for shrinkage. We argue that tests derived from this shrinkage approach display in general better small sample properties than MSPE based tests of the MDH. This occurs because the shrinkage based tests implicitly consider the reduced variance benefits of shrinkage estimators. Finally, we illustrate the use of our tests in an empirical application within the exchange rate literature.

I thank Steven Durlauf for all his support and guidance, Kenneth West for multiple meetings and greatly appreciated contributions, Todd Clark for assistance with the codes, tables, data and helpful comments. Wonderful comments and meetings with Salvador Navarro, Randy Fortenbery, Bruce Hansen, Charles Engel, Jack Porter, Menzie Chinn, Ananth Seshadri and Roberto Rigobon are gratefully appreciated. I also thank Andrea Bentancor, Jose Rodrigues Neto, Rogerio Mazali, Hisatoshi Tanaka, Ivan Canay, Jason Wu, Federico Diez, Santiago Oliveros, Jonathan Hore and Deokwoo Nam, Alan Spearot and Katie Howard for helpful comments (or assistance). I thank Sandra Stark and participants of GEN-BUS 765 for a fruitful discussion. All errors are mine.
E-mail: ppincheira@bcentral.cl.

1 Introduction

The economic literature has usually used a martingale model as a benchmark to test for predictability. In the context of asset prices, for instance, the martingale model posits that the best forecast of tomorrow's price is today's price. In other words, the martingale model assumes that future expected returns, given information available today, are zero. This condition is known as the Martingale Difference Hypothesis (MDH).

While the simple MDH is generally rejected when the econometrician engages in conventional in-sample analysis, it is indeed a difficult benchmark to beat when an out-of-sample approach is followed. The seminal paper of Meese and Rogoff (1983) is a classical example of this problem in the context of the exchange rate literature. This is sometimes interpreted as an indication that in-sample analysis is affected by overfitting or data mining problems and therefore should be disregarded. While the conflicting results from the in-sample and out-of-sample approaches are not entirely clear, Inoue and Kilian (2003) emphasize the higher power of in-sample strategies over out-of-sample analysis. According to this argument, out-of-sample tests of the MDH would fail to reject the null of no predictability mainly due to the lack of power of these tests. Therefore, it is essential to derive out-of-sample tests of the MDH that display power improvements with respect to their competitors. In this paper we mainly pursue this goal: deriving out-of-sample tests of the MDH with improved small sample properties, including power properties.

Despite some critics, a vast literature has primarily used out-of-sample Mean Square Prediction Errors (MSPE) as a leading measure of loss when testing the MDH. Different specifications like the mean absolute error, mean prediction error and direction of change, among others, have also received some attention. See McCracken and West (2002), Chinn and Meese (1991), Cheung, Chinn and Garcia (2002), Patton and Timmermann (2003) and Giacomini and White (2003).

A key point to consider when testing the MDH is that traditional statistical methods of comparing predictive accuracy, as those presented by West (1996) and Diebold and Mariano (1995), are not adequate. The reason for this is that the models under analysis are nested, see West (1996). McCracken (2004) and Clark and McCracken (2003) show that when comparing nested models, direct application of traditional methods may result in tests of inadequate size. In particular, Clark and West (2005a) show that traditional comparisons of MSPE render a test with low power and questionable size.

Two interesting alternatives to construct proper statistical methods of comparing predictive accuracy when models are nested are presented by McCracken (2004) and Clark and West (2005a). In the first paper, the asymptotic distribution of a t-type statistic, comparable to that suggested by West (1996) and Diebold and Mariano (1995), is derived. While the distribution is non standard, tables of asymptotically correct critical values are provided. In the second paper, Clark and West (2005a) make an important observation about the null hypothesis of equal MSPE when models are nested. They show that under the null the sample distribution of the difference in MSPE is not zero. Instead, the sample MSPE from the null model is expected to be smaller than that of the alternative model. They propose an adjusted procedure that takes into account this finding and demonstrate via simulations that this adjusted procedure is well-sized and relatively powerful in small and moderate samples. In addition they show that the procedure works well in an empirical implementation within the exchange rate literature.

A third alternative to properly test for model adequacy when models are nested is provided by Chao, Corradi and Swanson (2000). They provide a test of Granger causality that displays good size and power properties in small samples when applied to the MDH. See Clark and West (2005a).

Despite these interesting alternatives to properly test the MDH, this field of research is still not closed. From a theoretical point of view, we have already mentioned that it is desirable to derive more powerful tests in small and moderate samples and, more generally, to derive tests with better small sample properties. We also would like to have a test with a direct economic interpretation that allows us, for example, to obtain positive returns when using a given trading rule. In addition, we would like to be able to construct better forecasts than those from the martingale difference model when the null hypothesis is rejected. Finally, we also would like to find more empirical evidence against the MDH either in the exchange rate literature or in a more general economic environment.

Following this last motivation, we focus here on deriving tests of the MDH by exploring whether the alternative model may or may not be used to construct more accurate forecasts under quadratic loss. Our tests differ from other tests in that they explicitly search for gains in forecast accuracy. This is in opposition, for instance, to orthogonality tests that shed no light on the construction of a more accurate forecast, yet may have good small sample properties.

We proceed by defining a family of tests for the MDH based upon a shrinkage principle. In other words, tests within this family are such that rejection of the null implies that forecasts from the alternative model, adjusted by a shrinkage factor, will display lower MSPE than forecasts from the null model. This generalizes most previous tests which compare forecast errors of one model, the null, to forecast errors of the plain alternative model, not allowing for shrinkage. We argue that tests derived from this generalization allow us to construct tests of the MDH that display good small sample properties. By this generalization of MSPE comparisons, tests of model adequacy based on forecast accuracy become useful decision rules when an economic agent is concerned about forecasting: if a test rejects the null then we should be able to obtain more accurate forecasts from the alternative model. Interestingly, one of the tests we derive is a direction of change test. We show that under mild assumptions this test is closely related to MSPE based tests.

In summary, our procedure allows us to provide the following contributions. First, we show that our shrinkage based tests display, in general, better small sample properties that traditional tests based upon plain comparisons of MSPE. In particular, we show that a new test, called Max-MSPE-Adjusted, displays good size and, at least in some relevant contexts, more power than a set of competitors. Second, our tests are useful forecasting selection tools, meaning that rejection of the null ensures the existence of a shrinkage factor for which shrunken forecasts from the alternative model will display lower MSPE than forecasts from the null model.

The rest of the paper is organized as follows: Section 2 introduces our shrinkage approach; Section 3 derives some tests from the shrinkage approach; Section 4 describes the experimental design and delivers the simulation results; Section 5 presents an empirical application within the exchange rate literature; Section 6 briefly displays four appealing extensions for future research; and Section 7 concludes.

2 The Shrinkage Approach

In this section we introduce a modification of traditional comparisons of MSPE. The purpose is the derivation of tests leading to the potential construction of more accurate forecasts when the null model is rejected. We will then see that this approach leads to tests with good small sample properties. Our approach is in between (or linking) the literature of conditional and unconditional tests of predictive ability, see Giacomini and White (2003). These authors 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 is claimed to be more relevant to economic forecasters. With our approach we try to connect both worlds: we are trying to test a theory by evaluating whether it is possible for this theory to provide more accurate forecasts in a real-time forecasting exercise using several shrinkage strategies.

Now we will set up the econometric context of our analysis. We will use an environment similar to that in Clark and West (2005a).

2.1 Econometric Context

Consider two simple models for a scalar stationary time series y_{t+1} :

$$\begin{aligned} Model\,1 \quad (null) \quad &: \quad y_{t+1} = e_{t+1} && (1) \\ Model\,2 \quad (alternative) \quad &: \quad y_{t+1} = \widetilde{y}_{t+1}(X_{t+1}, \beta) + e_{t+1} && (2) \end{aligned}$$

where X_{t+1} is a vector of stationary and exogenous random variables and e_{t+1} is a zero mean martingale difference meaning that

$$E(e_{t+1}|\mathfrak{F}_t) = 0 \tag{3}$$

where $\{\mathfrak{F}_t\}$ represents a filtration such that $\mathfrak{F}_t$ is the sigma-field generated by current and past X's and e's.

$$\mathfrak{F}_t = \sigma\left\{X_{t+1}, e_{t,} X_{t,} e_{t-1}, X_{t-1}, e_{t-2}...\right\}$$

Notice that we are using the index $t+1$ to denote exogenous variables known at time t. Thus X_{t+1} is a vector containing known variables at time t. Notice also that we will use the notation

$E(e_{t+1}|\mathfrak{F}_t)$ or $E_t(e_{t+1})$. The alternative model posits that the conditional expectation of y_{t+1} with respect to the filtration $\mathfrak{F}_t$ only depends in the vector X_{t+1} and an unknown parameter β :

$$E(y_{t+1}|\mathfrak{F}_t) = \widetilde{y}_{t+1}(X_{t+1}, \beta) \tag{4}$$

where $\widetilde{y}_{t+1}(X_{t+1}, \beta)$ denotes some function of X_{t+1} and β. For instance, in a linear model we will have

$$\widetilde{y}_{t+1}(X_{t+1}, \beta) = X'_{t+1}\beta$$

For simplicity we will refer to the conditional expectation in (4) by $\widetilde{y}_{t+1}$. We further impose the condition

$$\widetilde{y}_{t+1}(X_{t+1}, 0) = 0 \tag{5}$$

This condition ensures nestedness.

According to this condition, the null hypothesis may be written in terms of a restriction over the vector of parameters: $\beta = 0$ against the alternative $\beta \neq 0$. The mean square prediction error (MSPE) for both models is given below

$$\begin{aligned} MSPE1 \ \ (null) \ &: \ E(e_{t+1}^2) = E(y_{t+1}^2) && (6) \\ MSPE2 \ \ (alternative) \ &: \ E(e_{t+1}^2) = E(y_{t+1} - \widetilde{y}_{t+1}(X_{t+1}, \beta))^2 && (7) \end{aligned}$$

And the difference of this MSPE is given by

$$\begin{aligned} \Delta MSPE \ &\equiv \ MSPE1 \ \ - MSPE2 && (8) \\ &= \ E(y_{t+1}^2) - E(y_{t+1} - \widetilde{y}_{t+1}(X_{t+1}, \beta))^2 && (9) \\ &= \ 2E(y_{t+1}\widetilde{y}_{t+1}) - E(\widetilde{y}_{t+1})^2 && (10) \\ &= \ E(\widetilde{y}_{t+1})^2 && (11) \end{aligned}$$

Under the null, the population mean square error of both models is equal because $\beta = 0$ and $\widetilde{y}_{t+1} = 0$, but under the alternative it is positive because $\widetilde{y}_{t+1} \neq 0$:

$$\begin{aligned} H_0 \ &: \ \Delta MSPE = E(\widetilde{y}_{t+1})^2 = 0 && (12) \\ H_A \ &: \ \Delta MSPE = E(\widetilde{y}_{t+1})^2 > 0 && (13) \end{aligned}$$

So, under the alternative we expect the MSPE of the true model to be lower than the MSPE of the wrong null model. This result leads us to focus on one sided tests.

This clear distinction between the null and the alternative does not hold anymore when we work with sample analogs. Following Clark and West (2005a) the sample difference in MSPE under the null is negative, indicating that the null model performs better than the alternative model. Under the alternative, in turn, the sample difference in MSPE has an ambiguous sign. To see this result, we need to introduce some notation first.

We will focus our analysis on one step ahead forecasts. One has $T + 1$ observations, from which the last P are used for predictions and $R = T + 1 - P$ are used for the initial estimation of the parameters. $\widehat{\beta}_t$ denotes a generic estimate of β with information available until time t. In general, the estimation scheme may be either fixed, rolling or recursive. The fixed scheme is one in which $\widehat{\beta}_t$ is estimated only once using the first R data points. The rolling scheme updates the estimate of $\widehat{\beta}_t$ using the last R observations. The recursive scheme also updates the estimate of $\widehat{\beta}_t$, but this time using all available information until time t. That is to say, in the recursive scheme the estimation sample increases with t. We will work with the rolling scheme, partly because it is appropriate when working with time series that may have experienced breaks, and partly because we do not want parameter uncertainty to vanish asymptotically.[1] We will use the following abbreviation: $\widetilde{y}_{t+1}$ will denote the population conditional expectation whereas $\widehat{y}_{t+1}$ will denote its sample analog.

Let us define the sample analog of $\Delta MSPE$ as follows:

$$\begin{aligned}
\Delta\widehat{MSPE} &= \widehat{MSPE}_1 - \widehat{MSPE}_2 \\
&= \frac{1}{P}\sum_{t=R}^{T}(y_{t+1})^2 - \frac{1}{P}\sum_{t=R}^{T}(y_{t+1} - \widehat{y}_{t+1}(X_{t+1}, \widehat{\beta}_t))^2 \\
&= \frac{2}{P}\sum_{t=R}^{T} y_{t+1}\widehat{y}_{t+1}(X_{t+1}, \widehat{\beta}_t) - \frac{1}{P}\sum_{t=R}^{T}(\widehat{y}_{t+1}(X_{t+1}, \widehat{\beta}_t))^2
\end{aligned}$$

Under the null, y_{t+1} is a zero mean martingale difference, so $y_{t+1} = e_{t+1}$ and as shown by Clark and West (2005a)

$$E(y_{t+1}\widehat{y}_{t+1}(X_{t+1}, \widehat{\beta}_t)) = E(e_{t+1}\widehat{y}_{t+1}(X_{t+1}, \widehat{\beta}_t)) = 0$$

Thus, under the null we should have $E\left(\Delta\widehat{MSPE}\right) < 0$:

$$\begin{aligned}
E\left(\Delta\widehat{MSPE} \mid H_0\right) &= \frac{2}{P}\sum_{t=R}^{T} E(e_{t+1}\widehat{y}_{t+1}|\ H_0) - \frac{1}{P}\sum_{t=R}^{T} E((\widehat{y}_{t+1})^2|\ H_0) \\
&= -\frac{1}{P}\sum_{t=R}^{T} E((\widehat{y}_{t+1})^2|\ H_0) < 0 \\
&= -E((\widehat{y}_{t+1})^2|\ H_0) < 0
\end{aligned}$$

Under the alternative we have in turn: $y_{t+1} = \widetilde{y}_{t+1}(X_{t+1}, \beta) + e_{t+1}$ therefore

$$\begin{aligned}
E(y_{t+1}\widehat{y}_{t+1}) &= E(\widetilde{y}_{t+1}\widehat{y}_{t+1}) + E(e_{t+1}\widehat{y}_{t+1}) \\
&= E(\widetilde{y}_{t+1}\widehat{y}_{t+1})
\end{aligned}$$

[1] Working with the fixed sheme is also consistent with not vanishing parameter uncertainty, but it is probably inefficient. The resursive scheme, however, is conflicting with not vanishing uncertainty. The increasing precision of the recursive estimation due to the use of additional data might reduce the variance of the estimate, making it converge to the population parameter. This intuition is given by West (2005b).

Thus, under the alternative we have $E\left(\Delta\widehat{MSPE}\right) \lesseqgtr 0$:

$$\begin{aligned}
E\left(\Delta\widehat{MSPE} \mid H_A\right) &= E\left(\frac{2}{P}\sum_{t=R}^{T} y_{t+1}\widehat{y}_{t+1} - \frac{1}{P}\sum_{t=R}^{T}(\widehat{y}_{t+1})^2 \mid H_A\right) \\
&= \frac{2}{P}\sum_{t=R}^{T} E(\widetilde{y}_{t+1}\widehat{y}_{t+1} |\ H_A) - \frac{1}{P}\sum_{t=R}^{T} E((\widehat{y}_{t+1})^2 |\ H_A) \\
&= 2E(\widetilde{y}_{t+1}\widehat{y}_{t+1} \mid H_A) - E((\widehat{y}_{t+1})^2 |\ H_A) \lesseqgtr 0
\end{aligned}$$

We will use this result recurrently in what follows.

2.2 Shrinkage Comparisons, Population Moments

Let us consider the following auxiliary function $f : \mathbb{R}_+ \to \mathbb{R}$

$$f(s) = MSPE_1 - MSPE_2(s)$$

where

$$\begin{aligned}
MSPE_1 &= E(y_{t+1}^2) \\
MSPE_2(s) &= E(y_{t+1} - \widetilde{y}_{t+1}(s))^2
\end{aligned}$$

and $\widetilde{y}_{t+1}(s)$ is a continuously differentiable perturbation of the conditional expectation (4) such that $\widetilde{y}_{t+1}(1) = \widetilde{y}_{t+1}$. More precisely, we are interested in a particular scaling or shrinkage transformation as follows:

$$\widetilde{y}_{t+1}(s) = \frac{\widetilde{y}_{t+1}}{s} \tag{14}$$

Notice that with this scaling transformation typical mean square prediction errors between models 1 and 2 are captured by evaluating the auxiliary function in the value $s = 1$, therefore

$$f(1) = MSPE_1 - MSPE_2$$

Under the null, $\beta = 0$, so $\widetilde{y}_{t+1} = 0$, and $f(s) = 0$ for all $s > 0$. Under the alternative, however, $f(s) > 0 \Longleftrightarrow s > \frac{1}{2}$ as we can see below

$$\begin{aligned}
f(s) &= E(y_t^2) - E(y_t - \widetilde{y}_{t+1}(s))^2 \\
&= 2E(y_t\widetilde{y}_{t+1}(s)) - E(\widetilde{y}_{t+1}(s))^2 \\
&= 2E(\frac{y_t\widetilde{y}_{t+1}}{s}) - \frac{1}{s^2}E(\widetilde{y}_{t+1})^2
\end{aligned}$$

Therefore, under the alternative

$$\begin{aligned}
f(s) &= 2E(\frac{y_{t+1}\widetilde{y}_{t+1}}{s}) - \frac{1}{s^2}E(\widetilde{y}_{t+1})^2 \\
&= \frac{2}{s}E(\widetilde{y}_{t+1})^2 + \frac{2}{s}E(e_{t+1}\widetilde{y}_{t+1}) - \frac{1}{s^2}E(\widetilde{y}_{t+1})^2 \\
&= \frac{2}{s}E(\widetilde{y}_{t+1})^2 - \frac{1}{s^2}E(\widetilde{y}_{t+1})^2
\end{aligned} \tag{15}$$

thus, we have that

$$f(s) > 0 \Longleftrightarrow s > \frac{1}{2} \tag{16}$$

In particular, we expect $f(1) > 0$ which is simply to say that under the alternative, the MSPE of the wrong model should be greater than that of the correct model.

2.3 Shrinkage Comparisons, Sample Analogs

Let us analyze what happens when instead of considering the theoretical population moments, we take their feasible sample analogs. Define the sample analog of $f(s)$ as follows:

$$\begin{aligned}
\widehat{f}(s) &= \widehat{MSPE}_1 - \widehat{MSPE}_2(s) \\
&= \frac{1}{P}\sum_{t=R}^{T}(y_{t+1})^2 - \frac{1}{P}\sum_{t=R}^{T}(y_{t+1} - \frac{\widehat{y}_{t+1}}{s})^2 \\
&= \frac{2}{P}\sum_{t=R}^{T}\frac{y_{t+1}\widehat{y}_{t+1}}{s} - \frac{1}{P}\sum_{t=R}^{T}(\frac{\widehat{y}_{t+1}}{s})^2
\end{aligned}$$

Under the null y_{t+1} is a zero mean martingale difference, so

$$E(y_{t+1}\widehat{y}_{t+1}) = E(e_{t+1}\widehat{y}_{t+1}) = 0 \tag{17}$$

and

$$E\left(\widehat{f}(s) \mid H_0\right) = -E((\frac{\widehat{y}_{t+1}}{s})^2 | \ H_0) < 0 \tag{18}$$

Thus, under the null we should have $E\left(\widehat{f}(s)\right) < 0$ for all $s > 0$.

Under the alternative we have in turn: $y_{t+1} = \widetilde{y}_{t+1} + e_{t+1}$, with $\widetilde{y}_{t+1} \neq 0$. Therefore, under the alternative we might have $E\left(\widehat{f}(s)\right) \neq 0$ for some $s > 0$:

$$\begin{aligned}
E\left(\widehat{f}(s) \mid H_A\right) &= E\left(\frac{2}{P}\sum_{t=R}^{T}\frac{y_{t+1}\widehat{y}_{t+1}}{s} - \frac{1}{P}\sum_{t=R}^{T}(\frac{\widehat{y}_{t+1}}{s})^2 \mid H_A\right) \\
&= \frac{2}{P}\sum_{t=R}^{T}E(\frac{\widetilde{y}_{t+1}\widehat{y}_{t+1}}{s} \mid H_A) - \frac{1}{P}\sum_{t=R}^{T}E(\frac{\widehat{y}_{t+1}}{s})^2 | \ H_A) \\
&= 2E(\frac{\widetilde{y}_{t+1}\widehat{y}_{t+1}}{s} \mid H_A) - E(\frac{\widehat{y}_{t+1}}{s})^2 | \ H_A)
\end{aligned}$$

so

$$E\left(\widehat{f}(s) \mid H_A\right) > 0 \Longleftrightarrow 0 < \frac{1}{s} < \frac{2E(\widetilde{y}_{t+1}\widehat{y}_{t+1} \mid H_A)}{E(\widehat{y}_{t+1}^2|H_A)} \tag{19}$$

Notice that from the theoretical point of view it is irrelevant whether improved forecasts are achieved when $s = 1$ or when $s \neq 1$. All that matters is the existence of a positive s for which the alternative model displays a lower MSPE[2]. Based upon this fact we will consider

[2] We emphasize here that all along the document we are using the shrinkage factor s with the only purpose to identify the relevant condition ensuring predictability. We are not trying here to estimate s. We are just saying that if a given condition holds true, then using a big enough shrinkage factor we should be able to get inproved forecasts over the null model. Furthermore, we are not arguing here about the size of the improvement. It can be either large or small, but surely it will be an improvement, meaning that the null hypothesis cannot be true.

that a decision rule for choosing the alternative model over the null model for forecasting purposes is of the form:

$$E(\widetilde{y}_{t+1}\widehat{y}_{t+1} \mid H_A) > \frac{1}{s}, s > 0 \tag{20}$$

The existence of a positive shrinkage factor s such that (19) holds true hinges upon the following condition:

$$2E(\widetilde{y}_{t+1}\widehat{y}_{t+1} \mid H_A) > 0$$

Notice that departing from the population case in (16), it might no longer be true that

$$\frac{2E(\widetilde{y}_{t+1}\widehat{y}_{t+1} \mid H_A)}{E(\widehat{y}^2_{t+1}|H_A)} = 2$$

therefore it might no longer be true that

$$E\left(\widehat{f}(s) \mid H_A\right) > 0 \Longleftrightarrow s > \frac{1}{2}$$

and in particular we should not expect

$$E\left(\widehat{MSPE}_1 - \widehat{MSPE}_2 | \; H_A\right) = E\left(\widehat{f}(1) \mid H_A\right) > 0$$

Figures 1 and 2 in the Appendix show that paying attention to $s = 1$ might be too restrictive to find evidence for the alternative model because the alternative and the null's MSPE may only be distinguishable (display different sign) for greater values of s. Figure 1 depicts the "ideal case" in which the MSPE difference under the alternative is positive for every $s \gtrapprox 1$ (actually for every $s \gtrapprox 0.5$). Figure 2 depicts a situation in which the MSPE difference under the alternative is positive only for $s \gtrapprox 3$. Both cases are independent draws of a simulated exercise following a true data generating process different from a zero mean martingale difference. Whereas in the first case one might be prone to reject the null hypothesis because the alternative model has lower MSPE, in the second case one might incorrectly be prone not to reject the null.

Notice that our approach uses more information than just the value of our function $\widehat{f}(s)$ at $s = 1$. When exploring a wider region we will be able to judge rejection of the null when a shrunken alternative model displays lower MSPE. Put differently, should we be able to reject the null we will know that for some shrinkage factor, s, our shrunk comparisons of MSPE will be positive. This means that we could potentially construct better forecasts or use our shrunken forecast as a decision rule in a trading strategy with more success than when using the zero forecast of the null model.

Before deriving our tests, we extend the analysis presented here in the next subsection to the case in which we allow different rates of shrinkage for different additive components of the conditional expectation of the alternative model.

2.4 The Multidimensional Case, Populations Moments

So far we have focused on the specific case of a univariate function $f : \mathbb{R}_+ \to \mathbb{R}$. In this section we seek further generality considering the multivariate definition of our auxiliary function f.

Let us assume that the conditional expectation of our alternative model can be decomposed into k additive components:

$$\widetilde{y}_{t+1}(X_{t+1}, \beta) = \widetilde{y}_{1,t+1}(X_{t+1}, \beta) + \widetilde{y}_{2,t+1}(X_{t+1}, \beta) + ... + \widetilde{y}_{k,t+1}(X_{t+1}, \beta)$$

Let us use the notation

$$\widetilde{y}_{t+1}(\overrightarrow{s}) = \frac{\widetilde{y}_{1,t+1}(X_{t+1}, \beta)}{s_1} + \frac{\widetilde{y}_{2,t+1}(X_{t+1}, \beta)}{s_2} + ... + \frac{\widetilde{y}_{k,t+1}(X_{t+1}, \beta)}{s_k} \tag{21}$$

and for sample analogs

$$\widehat{y}_{t+1}(\overrightarrow{s}) = \frac{\widehat{y}_{1,t+1}(X_{t+1}, \beta)}{s_1} + \frac{\widehat{y}_{2,t+1}(X_{t+1}, \beta)}{s_2} + ... + \frac{\widehat{y}_{k,t+1}(X_{t+1}, \beta)}{s_k} \tag{22}$$

where

$$\overrightarrow{s} = (s_1, ..., s_k)$$

Consider the function $f : \mathbb{R}^k_+ \to \mathbb{R}$

$$\begin{aligned} f(\overrightarrow{s}) &= MSPE_1 - MSPE_2(\overrightarrow{s}) \\ &= E(y_t^2) - E(y_t - \widetilde{y}_{t+1}(\overrightarrow{s}))^2 \\ &= 2E(y_t \widetilde{y}_{t+1}(\overrightarrow{s})) - E(\widetilde{y}_{t+1}(\overrightarrow{s}))^2 \end{aligned}$$

Under the null we have again that $\beta = 0$, so $\widehat{y}_{t+1}(\overrightarrow{s}) = f(s_1, ..., s_k) = 0$ for all $s_i > 0$, $i = 1...k$. Under the alternative, however, $\beta \neq 0$, so we expect $f(s_1, ..., s_k) \neq 0$ for vectors $(s_1, ..., s_k) \in \mathbb{R}^k_+$ satisfying the following condition:

$$2E(y_{t+1}\widetilde{y}_{t+1}(\overrightarrow{s})) \neq E(\widetilde{y}_{t+1}(\overrightarrow{s}))^2 \tag{23}$$

The condition $f(s_1, ..., s_k) \neq 0$ is satisfied at least for every vector $(s_1, ..., s_k) \in \mathbb{R}^k_+$ such that $s_i = s_j = s \in \mathbb{R}_+/\{\frac{1}{2}\}$ for all $i, j \in \{1, ..., k\}$ $i \neq j$. Similarly, the condition $f(s_1, ..., s_k) > 0$ will be satisfied at least for every vector $(s_1, ..., s_k) \in \mathbb{R}^k_+$ such that $s_i = s_j = s > \frac{1}{2}$. Therefore we have that the rejection set

$$C_{H_A} = \left\{ \overrightarrow{s} = (s_1, ..., s_k) \in \mathbb{R}^k_+ \text{ such that } f(s_1, ..., s_k) > 0, \text{ given } H_A \text{ is true} \right\}$$

satisfies $C_{H_A} \neq \phi$ and that $(s_1, ..., s_k) = (1, ..., 1) \in C_{H_A}$ which means that the set C_{H_A} includes the case where the MSPE of the alternative model is smaller than the MSPE of the null model.

2.5 The Multidimensional Case, Sample Analogs

The sample analog of function f is given by

$$\widehat{f}(\overrightarrow{s}) = \frac{2}{P}\sum_{t=R}^{T} y_{t+1}\widehat{y}_{t+1}(\overrightarrow{s}) - \frac{1}{P}\sum_{t=R}^{T}(\widehat{y}_{t+1}(\overrightarrow{s}))^2$$

Under the null y_{t+1} is a zero mean martingale difference, so we should have $E\widehat{f}(s_1, ..., s_k) < 0$ for all $(s_1, ..., s_k) \in \mathbb{R}_+^k$:

$$\begin{aligned} E\left(\widehat{f}(\overrightarrow{s}) \mid H_0\right) &= \frac{2}{P}\sum_{t=R}^{T} E(e_{t+1}\widehat{y}_{t+1}(\overrightarrow{s})| \; H_0) - \frac{1}{P}\sum_{t=R}^{T} E((\widehat{y}_{t+1}(\overrightarrow{s}))^2| \; H_0) \\ &= -E((\widehat{y}_{t+1}(\overrightarrow{s}))^2| \; H_0) < 0 \end{aligned}$$

Now, we are interested in the rejection set

$$\widehat{C}_{H_A} = \left\{\overrightarrow{s} = (s_1, ..., s_k) \in R_+^k \text{ such that } E\left(\widehat{f}(\overrightarrow{s}) \mid H_A\right) \geq 0\right\}$$

A little algebra shows that under the alternative we have

$$E\left(\widehat{f}(\overrightarrow{s}) \mid H_A\right) = 2E\left(\sum_{i=1}^{k} y_{t+1}\frac{\widehat{y}_{i,t+1}}{s_i} \; |H_A\right) - E\left(\left(\sum_{i=1}^{k}\frac{\widehat{y}_{i,t+1}}{s_i}\right)^2 \mid H_A\right)$$

This last expression has the chance to be non-negative for some $\overrightarrow{s} \in \mathbb{R}_+^k$ if and only if the first term on the right-hand-side is non-negative, that is to say, if the following condition holds true

$$\exists \overrightarrow{s} \in \mathbb{R}_+^k \colon E\left(\sum_{i=1}^{k} y_{t+1}\frac{\widehat{y}_{i,t+1}}{s_i} \; |H_A\right) \geq 0 \tag{24}$$

which in turn has a chance to be non-negative if $\exists\, i \in \{1, ..., k\}$ such that $E\left(y_{t+1}\widehat{y}_{i,t+1} \; |H_A\right) \geq 0$.

It is straightforward to show that

$$\widehat{C}_{H_A} \neq \phi \Longleftrightarrow E(y_{t+1}\widehat{y}_{i,t+1}) = \max_{j\; \in\{1,...k\}}\{E(y_{t+1}\widehat{y}_{i,t+1})\} \geq 0$$

The following proposition states and proves this result.

Proposition 1 *If* $\exists\; i \in \{1, ..., k\}$ *such that* $E\left(y_{t+1}\widehat{y}_{i,t+1} \; |H_A\right) > 0$ *then there is a vector* $\overrightarrow{s} = \overrightarrow{\overline{s}} \in \mathbb{R}_+^k \colon E\left(\sum_{i=1}^{k}\frac{y_{t+1}\widehat{y}_{i,t+1t}^{i}}{\overline{s}_i} \; |H_A\right) > 0$ *and it is possible to find* $h \in \mathbb{R}_+$ *such that* $E\left(\widehat{f}(h\overrightarrow{\overline{s}}) \mid H_A\right) > 0$.

Proof. Straightforward (see Appendix). ∎

The key idea here is again to relax the restriction $\overrightarrow{s} = \overrightarrow{1}$. This constraint might be too restrictive because the alternative and the null's MSPE may be only distinguishable for values of $\overrightarrow{s}$ different from $\overrightarrow{1}$ in small and moderate samples.

In the next subsection we give another interpretation of our approach. We will see that our tests might be used to test the null of no diversification gains.

2.6 Testing for Combination Gains

In light of the "Combination of Forecasts" literature, it is possible to give another interpretation of our shrinkage procedure. Several authors have pointed out that the combination of forecasts may improve forecast accuracy under quadratic loss. We see that in our context with two competing nested models, the null of no predictability implies the null of no combination gains.

It is straightforward to notice that our shrinkage approach searches over different combinations between the null model and the alternative, seeking significant combination gains. In fact we could write the shrunken forecast as a combination of the forecasts of the two models:

$$\widetilde{y}_{t+1}(s) = \frac{\widetilde{y}_{t+1}}{s} = 0 * \frac{(s-1)}{s} + \frac{\widetilde{y}_{t+1}}{s}$$

Therefore the existence of a shrunken forecast displaying a significantly lower MSPE is evidence of combination gains and hence supports the alternative model. When sample moments are used, any statistically significant gain from combining with positive weights should be considered evidence supporting the alternative model.

To see this clearly notice that our auxiliary function $f(s)$ could be written as follows[3]:

$$f(s) = \sigma_1^2 - \sigma_C^2(\frac{s-1}{s})$$

Combination gains means that $f(s) > 0$ for some $s > 0$, no combination gains means $f(s) \leq 0$ for all $s > 0$.

Similarly, when working with sample analogs we have

$$\widehat{f}(s) = \widehat{\sigma}_1^2 - \widehat{\sigma}_C^2(\frac{s-1}{s})$$

where

$$\begin{aligned} \widehat{\sigma}_1^2 &= \frac{1}{P}\sum_{t=R}^{T}(y_{t+1})^2 = \frac{1}{P}\sum_{t=R}^{T}(e_{1t+1})^2 \\ \widehat{\sigma}_C^2(\frac{s-1}{s}) &= \frac{1}{P}\sum_{t=R}^{T}(y_{t+1} - \frac{\widehat{y}_{t+1}}{s})^2 \end{aligned}$$

[3]

$$\begin{aligned} f(s) &= E(y_{t+1}^2) - E(y_{t+1} - \frac{\widetilde{y}_{t+1}}{s})^2 \\ &= E(y_{t+1}^2) - E[\frac{s-1}{s}(y_{t+1} - 0) + \frac{1}{s}(y_{t+1} - \widetilde{y}_{t+1})]^2 \\ &= E(e_{1t+1}^2) - E[\frac{s-1}{s}(e_{1t+1}) + \frac{1}{s}(e_{2t+1})]^2 \\ &= \sigma_1^2 - \sigma_C^2(\frac{s-1}{s}) \end{aligned}$$

Therefore under the null we have

$$E\widehat{f}(s) = E[\widehat{\sigma}_1^2 - \widehat{\sigma}_C^2(\frac{s-1}{s})] = -E(\frac{\widehat{y}_{t+1}}{s})^2 < 0$$

and as mentioned earlier no combination gains are possible.

Under the alternative, however,

$$E\widehat{f}(s) > 0 \text{ for some } s > 0$$

and combination gains are possible.

In summary, we are claiming that direct comparisons of out-of-sample MSPE between nested models might not be the most appropriate way to test the MDH. The reason being that direct MSPE comparisons neglect benefits from combining, benefits that are only accomplished under the alternative. We will show in the simulation section that accounting for the possibility of combination gains may also increase the power of the tests, at least in the context of our simulations.

3 Derivation of the Tests

In this section and the next we will assume for simplicity a linear alternative model. That is to say, we will assume a model of the following shape:

$$\widetilde{y}_{t+1} = E(y_{t+1}|X_{t+1}) = X_{t+1}'\beta$$

The existence of an s satisfying condition (19) hinges on the sign of

$$E(y_{t+1}X_{t+1}'\widehat{\beta}_t)$$

Should this term be positive then such an s will exist. On the other hand, the existence of an $h\overrightarrow{s}$ satisfying condition (35) (see the Appendix) hinges on the sign of

$$\max_{j\ \in\{1,\ldots k\}}\left\{E(y_{t+1}x_{t+1}^j\widehat{\beta}_t^j)\right\}$$

These results lead us to focus on the sign of the following estimators:

$$\frac{1}{P}\sum_{t=R}^{T} y_{t+1}X_{t+1}'\widehat{\beta}_t \tag{25}$$

or

$$\max_{j\ \in\{1,\ldots k\}}\left\{\frac{1}{P}\sum_{t=R}^{T}(y_{t+1}x_{t+1}^j\widehat{\beta}_t^j)\right\} \tag{26}$$

We first propose to build tests based in these two statistics (25) and (26). Clark and West (2005a) derive an asymptotically normal test based on a statistic like (25). In this sense we will include their test as a shrinkage based test because it is possible to derive the same statistic following our shrinkage analysis. Clark and West (2005a) called their statistic

"MSPE-Adjusted". We will follow the same tradition and call our statistic in (26) "Max-MSPE-Adjusted".

Notice that our analysis reveals that the relevant information for testing the MDH is summarized in the sign of (25) and (26), rather than in all the information that might be possibly enclosed in the statistics (25) and (26). In other words, when testing the martingale hypothesis the magnitude of (25) and (26) is irrelevant; only the sign is relevant because we can construct a forecast with lower MSPE than the random walk model as long as (25) or (26) are positive and not only when they are positive and large with respect to their variance.

Second, the previous remark leads us to build a sign test based[4] on (25). While the proper construction of a sign test based on (25) requires some further assumptions, we think that for the applications of interest here, these assumptions are mild and provide a test with interesting properties. Arguments for preferring sign tests instead of mean tests are given by Diebold and Mariano (1995); Reineke, Baggett, and Elfessi (2003); and Brown and Ibragimov (2005). The first authors remark that sign tests are powerful. The second authors provide conditions under which a standard sign test has more power than the standard t-student test. In particular they show that deviations from normality toward leptokurtic distributions result in a standard sign test being more powerful that the standard t-test. Provided that series of returns in finance and economics are typically leptokurtic, it is our belief that the results of Reineke, Baggett, and Elfessi (2003) may be extended to the environment we are interested in for this paper. Finally, Brown and Ibragimov (2005) stress the possibility of building exact sign tests rather than asymptotic tests. Exact tests are especially useful when the number of observations is small. We will see later that a sign test based on (25) corresponds to a direction of change test. This test was applied apologetically by Cheng et.al. (2002). The authors argue that one could criticize them for "changing the rules of the game". But what we see here is that the linkage between a MSPE test and a direction of change test is close. We will come back to this point later.

Finally we will present an induced test. This is a test searching all over the domain of the function f. Therefore, we are interested in a statistic like

$$\underset{s\in\mathbb{R}_+}{Sup}\ \sqrt{P}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}$$

where

$$\widehat{V}(s)=4\widehat{V}(y_{t+1}X_{t+1}'\widehat{\beta}_t)+\frac{1}{s^2}\widehat{V}((X_{t+1}'\widehat{\beta}_t)^2)-\frac{4}{s}\widehat{Cov}(y_{t+1}X_{t+1}'\widehat{\beta}_t,(X_{t+1}'\widehat{\beta}_t)^2)$$

For simplicity we will focus in the unidimensional version of our function f. We call this statistic "Sup MSPE".

In the following subsections we formally introduce these tests and some of their properties, but first we need to introduce some assumptions.

[4]Simulations indicate that a sign test based on (26) is also appealing to work with. For now we just focus on a signed test based on (25).

3.1 Assumptions

The following assumptions will be used in what follows

$$E(e_{t+1}X'_{t+1}\widehat{\beta}_t)^2 = \sigma_t^2 > 0 \text{ s.t } \frac{1}{P}\sum_{t=1}^{P}\sigma_t^2 \to \sigma^2 > 0 \tag{27}$$

$$E|e_{t+1}X'_{t+1}\widehat{\beta}_t|^r < \infty \text{ for some } r > 2 \text{ and all } t \tag{28}$$

$$\frac{1}{P}\sum_{t=1}^{P}(e_{t+1}X'_{t+1}\widehat{\beta}_t)^2 \xrightarrow{\Pr} \sigma^2 \tag{29}$$

These are standard assumptions to ensure asymptotic normality of martingale difference sequences. See White (1984).

3.2 MSPE-Adjusted (MSPE-Adj)

As previously noted, the MSPE-Adj test was proposed by Clark and West (2005a). The context in which this test was developed is one in which the size of the test is the relevant target in the analysis. In other words, Clark and West (2005a) claim that the MSPE-Adj test has better size, even in small and moderate samples than traditional tests of MSPE comparisons. Intuitively this test shows good size because it does not take into account a term that introduces noise into its forecasts by estimating a parameter vector that under the null should be zero.

In this subsection we show that the same MSPE-Adj test can be derived using our shrinkage approach. Therefore we provide an alternative interpretation for this test.

Remark:

$$\lim_{s\to\infty} P^{1/2}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}} = P^{1/2}\frac{\frac{2}{P}\sum_{t=R}^{T} y_{t+1}X'_{t+1}\widehat{\beta}_t}{\sqrt{4\widehat{V}(y_{t+1}X'_{t+1}\widehat{\beta}_t)}}$$

where

$$\widehat{f}(s) = \frac{2}{P}\sum_{t=R}^{T} y_{t+1}X'_{t+1}\widehat{\beta}_t - \frac{1}{P}\sum_{t=R}^{T}\frac{(X'_{t+1}\widehat{\beta}_t)^2}{s}$$

$$\widehat{V}(s) = 4\widehat{V}(y_{t+1}X'_{t+1}\widehat{\beta}_t) + \frac{1}{s^2}\widehat{V}((X'_{t+1}\widehat{\beta}_t)^2) - \frac{4}{s}\widehat{Cov}(y_{t+1}X'_{t+1}\widehat{\beta}_t, (X'_{t+1}\widehat{\beta}_t)^2)$$

The proof is straightforward.

This remark shows that the Clark and West (2005a) test explores the asymptotic behavior of a normalized version of the limit of our univariate auxiliary function f. They search for evidence for the alternative model far away from the factor $s = 1$. They search for evidence at the limit when the shrinkage factor goes to infinity. The intuition for the good behavior of this test is straightforward. It is clear that the existence of $\overline{s}$ such that $f(\overline{s}) > 0$ implies that

$$f(s) > 0 \quad \text{for all } s > \overline{s}$$

therefore exploring the behavior of $f(s)$ in the limit when s goes to infinity is the safest way to check for the existence of $\overline{s}$ such that $f(\overline{s}) > 0$.

3.3 Max-MSPE-Adjusted

Let us focus now on the following statistic:

$$Max - MSPE - Adjusted: \max_{i\, \in\{1,...k\}} \left\{ \frac{1}{P} \sum_{t=R}^{T} y_{t+1} x_{t+1}^{i} \widehat{\beta}_t^{i} \right\}$$

where $x_{t+1}^{i} \widehat{\beta}_t^{i}$ are the components of the estimated conditional mean $X_{t+1}' \widehat{\beta}_t$.

Consider the k-dimensional vector u:

$$u = \left(\frac{1}{P^{1/2}} \sum_{t=R}^{T} y_{t+1} x_{t+1}^{1} \widehat{\beta}_t^{1}, ..., \frac{1}{P^{1/2}} \sum_{t=R}^{T} y_{t+1} x_{t+1}^{k} \widehat{\beta}_t^{k} \right)$$

Under the null and assumptions (27), (28) and (29)

$$u \rightarrow_A N(0, V); \quad V \in M_{k \times k}$$

We propose finding a consistent estimate $\widehat{V}$ of V and then proper asymptotic critical values for a one-sided test based on the "Max-MSPE-Adjusted" statistic via simulations.

3.4 Sup MSPE

Consider an infinite induced test as follows:

$$\underset{s \in \mathbb{R}_+}{Sup} \sqrt{P} \frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}} \tag{30}$$

where

$$\widehat{V}(s) = 4\widehat{V}(y_{t+1} X_{t+1}' \widehat{\beta}_t) + \frac{1}{s^2} \widehat{V}((X_{t+1}' \widehat{\beta}_t)^2) - \frac{4}{s} \widehat{Cov}(y_{t+1} X_{t+1}' \widehat{\beta}_t, (X_{t+1}' \widehat{\beta}_t)^2)$$

It turns out that a test (one-sided) based on Sup MSPE has the same asymptotic critical values as a test based upon the MSPE-Adjusted statistic. It is easy to understand the linkage between MSPE-Adjusted and our induced test. We already saw that when looking for evidence of a region in which $\widehat{f}(s)$ is positive, we only need to look at the limiting behavior when s goes to infinity. That is equivalent to looking over all the domain of the function f. Proposition 2 summarizes this result.

Proposition 2 *Under the null and assumptions (27), (28) and (29)*

$$\lim_{P\to\infty} P\left(\left[\underset{s>0}{Sup} P^{1/2}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}\right]\leq u\right) \to \Phi(u) \quad for \quad all\ u>0$$

where Φ *denotes a standard normal distribution function and*

$$\widehat{f}(s) = \frac{2}{P}\sum_{t=R}^{T} y_{t+1}X'_{t+1}\widehat{\beta}_t - \frac{1}{P}\sum_{t=R}^{T}\frac{(X'_{t+1}\widehat{\beta}_t)^2}{s}$$

$$\widehat{V}(s) = 4\widehat{V}(y_{t+1}X'_{t+1}\widehat{\beta}_t) + \frac{1}{s^2}\widehat{V}((X'_{t+1}\widehat{\beta}_t)^2) - \frac{4}{s}\widehat{Cov}(y_{t+1}X'_{t+1}\widehat{\beta}_t, (X'_{t+1}\widehat{\beta}_t)^2)$$

Proof. See the Appendix. ■

We stress here that this result only holds for one sided tests because the negative tail of (30) is different from that of the MSE-Adjusted test.

Via simulations we observe that both the MSPE-Adj and the Sup MSPE statistics are very similar under the alternative, at least for sample sizes relevant to our applications. This happens because most of the time the statistic (30) reaches a maximum when s goes to infinity. In other words the probability of reaching an interior solution is low.

Therefore, we see that in one-sided tests we should expect the MSPE-Adjusted statistic to have similar behavior to the Sup MSPE test given by (30). Due to this fact, results for the Sup MSPE test are omitted and we only display results for the MSPE-Adj test.

3.5 Sign Tests

As mentioned earlier, our derivations show that evidence for the alternative model could be found whenever a statistic like (25) or (26) is positive. With this idea in mind, we suggest the implementation of a sign test over (25). The proper application of a sign test in this context requires one further assumption. Now we will not only assume that e_{t+1} represents a zero mean martingale difference, but also a zero median martingale difference. That is to say, we will assume:

$$E(e_{t+1}|\mathfrak{F}_t) = 0 \quad (31)$$
$$m(e_{t+1}|\mathfrak{F}_t) = 0 \quad (32)$$

where $m(e_{t+1}|\mathfrak{F}_t)$ represents the conditional median of e_{t+1}. This implies assumption (32) can be rewritten as

$$P(e_{t+1}>0|\mathfrak{F}_t) = P(e_{t+1}<0|\mathfrak{F}_t) = \frac{1}{2}$$

Notice that a test like the MSPE-Adjusted is a test over the following expectation

$$E(y_{t+1}X'_{t+1}\widehat{\beta}_t)$$

Notice also that under the null given by assumption (32) and assuming that $X'_{t+1}\widehat{\beta}_t = 0$ with probability zero, the following median is zero:

$$m(e_{t+1}X'_{t+1}\widehat{\beta}_t) = 0$$

This is because

$$P(e_{t+1}X'_{t+1}\widehat{\beta}_t > 0) =$$

$$\begin{array}{rcl} P(e_{t+1}X'_{t+1}\widehat{\beta}_t & > & 0|X'_{t+1}\widehat{\beta}_t > 0)P(X'_{t+1}\widehat{\beta}_t > 0) + \\ P(e_{t+1}X'_{t+1}\widehat{\beta}_t & > & 0|X'_{t+1}\widehat{\beta}_t < 0)P(X'_{t+1}\widehat{\beta}_t < 0) + \\ P(e_{t+1}X'_{t+1}\widehat{\beta}_t & > & 0|X'_{t+1}\widehat{\beta}_t = 0)P(X'_{t+1}\widehat{\beta}_t = 0) \end{array}$$

So

$$P(e_{t+1}X'_{t+1}\widehat{\beta}_t > 0) =$$

$$\begin{array}{rcl} P(e_{t+1} & > & 0|X'_{t+1}\widehat{\beta}_t > 0)P(X'_{t+1}\widehat{\beta}_t > 0) + \\ P(e_{t+1} & < & 0|X'_{t+1}\widehat{\beta}_t < 0)P(X'_{t+1}\widehat{\beta}_t < 0) \end{array}$$

and

$$P(e_{t+1}X'_{t+1}\widehat{\beta}_t > 0) = \frac{1}{2}$$

There is sufficient theoretical and empirical evidence supporting the conditional symmetry of some macroeconomic and financial data returns, so it is our belief that there is a number of cases for which testing (32) is equivalent to test the MDH. See Clark and West (2005a, 2005b), Tsay (2002), and Diebold and Mariano (1995).

Under assumption (32) it is possible to apply an asymptotically normal test as well as an exact test for the MDH.

Let us consider the following statistic:

$$sign(y_{t+1}X'_{t+1}\widehat{\beta}_t) = \left\{ \begin{array}{c} 1 \ if \ y_{t+1}X'_{t+1}\widehat{\beta}_t > 0 \\ 0 \ if \ y_{t+1}X'_{t+1}\widehat{\beta}_t = 0 \\ -1 \ if \ y_{t+1}X'_{t+1}\widehat{\beta}_t < 0 \end{array} \right\}$$

We can state the following proposition:

Proposition 3 *Under the null and assumption (32) the sequence* $sign(y_{t+1}X'_{t+1}\widehat{\beta}_t)$ *form a martingale-difference sequence with respect to the filtration* $\mathfrak{F}_t$.

Proof. See Appendix ■

This proposition enables us to use standard theorems of asymptotic normality for martingale sequences that require mild assumptions. See Hamilton (1994).

Notice that under the alternative we expect

$$E(sign(y_{t+1}X'_{t+1}\widehat{\beta}_t)|\mathfrak{F}_t) > 0$$

because if we knew the true parameter β,we would expect

$$\begin{aligned}
& E(sign(y_{t+1}X'_{t+1}\beta)|F_t) = \\
= \quad & E(sign((X'_{t+1}\beta)^2 + e_{t+1}X'_{t+1}\beta)|F_t) \\
= \quad & P((X'_{t+1}\beta)^2 + e_{t+1}X'_{t+1}\beta > 0|F_t) - P((X'_{t+1}\beta)^2 + e_{t+1}X'_{t+1}\beta < 0|F_t)
\end{aligned}$$

but a little algebra shows that

$$\begin{aligned}
P((X'_{t+1}\beta)^2 + e_{t+1}X'_{t+1}\beta &> 0|\mathfrak{F}_t) > \frac{1}{2} \\
P((X'_{t+1}\beta)^2 + e_{t+1}X'_{t+1}\beta &< 0|\mathfrak{F}_t) < \frac{1}{2}
\end{aligned}$$

so,

$$\begin{aligned}
& E(sign(y_{t+1}X'_{t+1}\beta)|F_t) = \\
= \quad & P((X'_{t+1}\beta)^2 + e_{t+1}X'_{t+1}\beta > 0|F_t) - P((X'_{t+1}\beta)^2 + e_{t+1}X'_{t+1}\beta < 0|F_t) > 0
\end{aligned}$$

While asymptotic normality is a nice property, the next proposition, based on Ibragimov and Brown (2005), shows that under assumption (32) we can also construct an exact test.

For completeness we give a proof in the Appendix.

Proposition 4 *Consider the following sequences*

$$\left\{sign(y_{t+1}X'_{t+1}\widehat{\beta}_t)\right\}_{t\geq 1}, \{\varepsilon_{t+1}\}_{t\geq 0}$$

where $\{\varepsilon_{t+1}\}_{t\geq 0}$ *denotes an i.i.d. sequence of symmetric Bernoulli random variables independent of* $sign(y_{t+1}X'_{t+1}\widehat{\beta}_t)$ *and the information available until time* t, $\mathfrak{F}_t$. *Then, under the null and assumption (32), the sequence*

$$\eta_{t+1} = sign(y_{t+1}X'_{t+1}\widehat{\beta}_t) + \varepsilon_{t+1}I(y_{t+1}X'_{t+1}\widehat{\beta}_t = 0)$$

forms an i.i.d. sequence of Bernoulli random variables taking values in $\{1, -1\}$ *,where* I *denotes an indicator function.*

Proof. *See the Appendix.* ■

Once we have proved that $\{\eta_{t+1}\}$ is an independent sequence of symmetric Bernoulli random variables we know that the sum of these variables follows an exact binomial distribution:

Remark: *Under the null and assumption (32) we have that*

$$\frac{\sum_{t=R}^{T}\eta_{t+1} + P}{2} \rightarrow B(P, \frac{1}{2})$$

We notice here that a test based on $sign(y_{t+1}X'_{t+1}\widehat{\beta}_t)$ is a direction of change test, so under symmetry of $y_{t+1}X'_{t+1}\widehat{\beta}_t$ or under assumption (32), we have that a test based on MSPE-Adjusted is equivalent to a direction of change test, and therefore they are both shrinkage based tests[5] derived from traditional comparisons of MSPE. This is an important remark. We see that under assumption (32), rejection of the direction of change test implies the existence of a shrinkage factor for which the alternative model displays lower MSPE. In this sense a direction of change based test is closely related to MSPE comparisons because it also searches for diversification gains. We will be using asymptotically normal critical values for our sign test. This test is denoted by Sign-N.

The next section presents the experimental design as well as the main results of the simulations. These results show the behavior of the statistics and tests developed in this paper.

4 Experimental Design and Simulation Results

4.1 Experimental Design

Following Clark and West (2005a) we use Monte Carlo simulations for two multivariate data-generating processes (DGPs) to evaluate the finite-sample properties of the statistics previously presented in this paper. The idea is to compare small sample properties of the shrinkage based tests (Sign-N, Max-MSPE-Adjusted and MSPE-Adjusted) with those of some benchmark tests already available in the literature. We consider three benchmark tests: the traditional Diebold and Mariano (1995) test that we call MSPE-Normal; a test of MSPE differences that uses McCracken (2004) critical values and that we call MSPE-McCracken; and a test introduced by Chao, Corradi, and Swanson (2000) that we call CCS. The first two tests, MSPE-Normal and MSPE-McCracken, are tests that compare the MSPE between the null and the alternative model. The CCS test is a projection test that has been reported to be powerful, see Clark and West (2005a).

We use DGPs following Clark and West (2005a). These DGPs are calibrated to match common features of exchange rate series for which the martingale difference is a sensible null hypothesis. Basically the DGPs used are variations of the following process:

$$\begin{aligned} y_{t+1} &= \beta x_t + e_{t+1} \\ x_t &= 0.95x_{t-1} + v_t; \\ e_{t+1} &= N(0,1); v_{t+1} = N(0,\sigma_v^2) \end{aligned}$$

with $E(e_{t+1}|\mathfrak{F}_t) = 0$, $E(v_{t+1}|\mathfrak{F}_t) = 0$ and $var(e_{t+1}) = 1$. DGP 1 is calibrated to match exchange rate features based on interest parity so we will have $var(v_t) = \sigma_v^2$ (with $\sigma_v = 0.025$) and $corr(e_t, v_t) = 0$. We set $\beta = -2$ in experiments evaluating power and $\beta = 0$ in experiments evaluating size. DGP 2 is the same as DGP 1 except that we assume that e_{t+1}

[5] Both tests can be seen as tests of the null of no diversification gains as well.

has a $t(2)$ distribution displaying fat tails. We assume data generated from homoskedastic draws of their respective distributions.

Estimation always includes a constant term in each regression. We explore the performance of rolling schemes for a number of sample sizes $(T+1)$ and decompositions of the sample in the estimation window (size R) and the prediction window (size $P, (T+1 = R+P)$). We run simulations for the following sample sizes: $R = 35$, 120 and 240; $P = 48, 109, 166, 226, 480$ and 700. All results are displayed in tables in the appendix and in the following section.

4.2 General Results

In this subsection we discuss aggregate results of our experiments. More details can be found in tables in the Appendix. Table 1 below shows averages over both DGP's and different values of P and R for the following categories: Empirical Size, Raw Power, and Size-Adjusted-Power.

Table 1
Averages of Small Sample Properties

(1)	(2) Empirical Size	(3) Raw Power	(4) Size-Adjusted Power
Max-MSPE-Adjusted	*0.096*	*0.399*	*0.383*
Sign-N	*0.098*	*0.347*	*0.334*
MSPE-Adjusted	*0.078*	*0.377*	*0.397*
MSPE-Normal	*0.005*	*0.064*	*0.369*
MSPE-McCracken	*0.068*	*0.324*	*0.369*
CCS	*0.110*	*0.430*	*0.402*

Notes:

1. See sections 3.1, 3.2 and 3.4 for the definition of the shrinkage based tests Sign-N, Max-MSPE-Adjusted, and MSPE-Adjusted. See section 4.1 for a description of the benchmark tests MSPE-Normal, MSPE-McCracken, and CCS.
2. For all the experiments nominal size is set to 10%.

Column 2 shows results on empirical size when the nominal size of the tests is 10%. We see that, on average, Sign-N and Max-MSPE-Adjusted display the closest empirical size to the nominal size. Notice also that the worst performances on empirical size are those corresponding to both MSPE based tests. Column 3 shows results on power. The best or highest values correspond to CCS and Max-MSPE-Adjusted. On the other hand, the worst values are those of the MSPE-Normal and MSPE-McCracken tests. The last column shows results on size-adjusted power. All the tests display similar size-adjusted power although the lowest average size-adjusted-power corresponds to the Sign-N test. In tables in the Appendix it can be seen that the size-adjusted-power of the Sign-N depends dramatically on the particular DGP. Therefore, averages results for this Sign-N test should be taken with caution.

4.3 Results on Size

Results on size, power and size adjusted power are attached in tables in the Appendix. Some remarks about empirical size follow below:

1) Max-MSPE-Adjusted: It is a little undersized on average. It has an average size of 9.6%. It also moves in a narrow region. For instance, its minimum value is 8.5% and the maximum is 11%. It shows no relationship whatsoever with R and P. There is a slight tendency to have a higher size in fat tail distributions.

2) Sign-N: It is mildly undersized on average. It has an average size of 9.8%. It has the best average size of all the statistics considered in this paper. As in the previous statistic, it covers a narrow range of values from a minimum size of 8.6% to a maximum of 11.2%. The variation of size with respect to P is not clear. The little variation that the size of this statistic displays with R is striking. For instance, in DGP2 for $P = 109$, the empirical size is 8.9%, 8.6% and 8.6% for R given by 35, 120 and 240 respectively. It seems also to be very robust with respect to different distributional assumptions of the errors. Basically its behavior is similar in both DGPs.

3) MSPE-Adjusted: In both DGPs the MSPE-Adjusted is uniformly undersized. Its maximum empirical size is 9.6%, (the minimum is 6.3%) and its average empirical size over all the experiments is 7.8%. We notice that there is no clear pattern of variation with the parameters P and R. The empirical size is, however, much closer to the correct nominal size in the experiments with fat tails.

4) MSPE-Normal: In both DGPs traditional MSPE is uniformly extremely undersized. Its maximum empirical size is 3.8%, its average is 0.5%. We notice that the empirical size decreases with P and increases with R. There is no significant difference between the results with fat tails and normal tails.

5) MSPE-McCracken: It is undersized on average with an average size of 6.8%. It is not uniformly undersized, however. For both DGPs it sometimes reaches sizes over 10%. It shows an increasing pattern with R and a decreasing pattern with P, with a few exceptions. It is not evident how the assumption of fat tails affects the size of this test.

6) CCS: It is oversized on average. Its average size is 11%. It is not uniformly oversized however. The maximum size is 15.6% and the minimum is 8.9%. It shows a decreasing trend with respect to P (with a few exceptions) and no relationship whatsoever with respect to R. Recall that R is not a factor in the determination of this test. This test is also highly oversized in DGPs with fat tails for small values of P.

4.4 Results on Power

On average we observe that CCS presents the highest raw power followed by Max-MSPE-Adjusted, then by MSPE-Adjusted, then by the Sign test, then by MSPE-MCracken and finally by MSPE-Normal. All tests, with the exception of CCS, show higher raw-power for higher values of R. We recall that CCS is neutral to changes in R because there is no estimated parameter in this statistic. All tests, with the exception of MSPE-Normal and occasionally MSPE-McCracken, show increasing raw-power with the number of forecasts P. Most of the time the raw-power of MPSE-Normal decreases along with P. An intuition for this is given in Clark and West (2005a). In general, fat tails have a negative effect on raw-power for all the tests. Similarly, in most cases Max-MSPE-Adjusted has higher raw-power than MSPE-Adjusted.

We also notice that in the following circumstances our shrinkage statistics have the highest raw-power: For DGP 1, Max-MSPE-Adjusted display the highest power, for all cases with $R = 240$ and $R = 120$ when $P \leq 226$. For DGP 2, Sign-N display the highest power for $R = 120$ and $P = 166$, 226 and $P = 700$. For $R = 240$ when $P \geq 109$.

Regarding size adjusted power we want to make a few observations. The drastic difference in raw-power observed between MSPE-Normal and the rest of the statistics now decreases. MSPE-Normal now has relatively similar size adjusted power compared to its competitors. CCS still has more size-adjusted power for low R and high P. We also observe that Sign tests display the highest size-adjusted-power in DGP2 for $R \geq 120$ (with only one exception, $R = 120$, $P = 700$).

5 Empirical Application

In this section we study the behavior of our statistics using monthly forecasts of five US dollar bilateral exchange rates. We analyze the cases of Canada, Japan, Switzerland, U.K, and Chile[6]. While the null model corresponds to a zero mean martingale difference for the percentage change in exchange rates, the alternative model posits that this percentage change is explained by two regressors: a constant and the one-month interest differential[7]. The data from Canada, Japan, Switzerland, and U.K. were generously provided by Todd Clark and correspond to the same database used in Clark and West[8] (2005a). We obtained the data for Chile from the International Financial Statistics. This time we use the discount rates as measures of interest rates.

Using rolling regressions estimated by OLS we engage in two empirical exercises. First, we assume that the number of observations used for the first estimation (R) as well as the number of predictions (P) are fixed. We follow Clark and West (2005a) to choose R relatively small with respect to P. For Canada, Japan, Switzerland, and U.K. we set $R = 120$ and $P = T + 1 - R$, where $T + 1$ is the total number of observations. For Chile we set $R = 36, P = 108$. Then we compute our statistics and we analyze whether the tests are able

[6] Eventhough the MDH may not be the most efficient benchmark to use in the absence of pure flotation, we still think it is a simple an interesting framework to compare with. The obvious exception being variants of fixed exchange regimes.

[7] See Clark and West (2005a) for a discussion about exchange rate models based on interest parity.

[8] 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.

to reject the null of a MDH at three different significance levels: 1%, 5% and 10%. The tests under consideration are: MSPE-Adjusted, Max-MSPE-Adjusted, MSPE with McCracken critical values, Sign-N, and CCS.

Second, we analyze how robust the results from the first empirical exercise are when we slightly change the parameters R and P. We consider a symmetric variation of roughly 10% of the data around P to check the percentage of rejections that we obtain in a neighborhood of $P = 166$ for Canada and Japan, $P = 226$ for Switzerland and U.K, and $P = 108$ for Chile.

Results are displayed in Tables 2 and 3. First, we clearly see how difficult is to reject the null hypothesis using a traditional comparison of MSPEs. This traditional comparison uses standard normal critical values. We see in column (5) that none of the values obtained for these countries would allows us to reject the martingale difference model using standard normal critical values. We already mentioned that standard normal critical values are not appropriate for this application. Use of the correct asymptotic critical values tabulated by McCracken (2004) leads to an improvement. Now the null model is rejected in two out of five countries (Canada and Chile at a 5% and 10% significance level respectively. See Column 6). Similar results are displayed by the CCS, MSPE-Adjusted, and Max-MSPE-Adjusted statistics. The CCS test rejects the null model for two countries: Chile, at a 5% significance level, and Japan, at a 10% significance level. The MSPE-Adjusted statistic rejects the null for Canada and Switzerland at a 5% significance value, and is close to rejection for Japan and Chile. The Max-MSPE-Adjusted statistic rejects the null for Chile at a 5% significance level, and for Switzerland at a 10% significance level. Finally, the sign test, Sign-N displays the strongest results. This test rejects the null in four out of five countries: it rejects the null in Canada at a 1% level, in Chile at a 10% level, in Japan at a 5% level, and in Switzerland at a 1% significance level. The null is never rejected for the U.K.

Table 2
Forecasts of Monthly Changes in U.S. Dollar Exchange Rates

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
country	**Max MSPE-Adj**	**Sign-N**	**MSPE Adjusted**	**MSPE Normal**	**MSPE McCracken**	**CCS**	**f(s1,s2)**
Canada	0.07	15	0.14	0.04	0.04		0.05
		(6.44)	(0.08)	(0.08)	(0.08)		s1=1.392
	0.89	2.33***	1.77**	0.54	0.54**	1.94	s2=1.531
Chile	5.88E-05	7.0	4.00E-05	-5.34E-06	-5.34E-06		17.4E-06
		(5.20)	(3.42E-05)	(3.48E-05)	(3.48E-05)		s1=44.750
	6.11E-04**	1.35*	1.17	-0.15	-0.15*	7.25**	s2=2.238
Japan	0.36	11	0.53	-0.23	-0.23		0.09
		(6.44)	(0.43)	(0.44)	(0.44)		s1=2.854
	4.64	1.71**	1.23	-0.52	-0.52	5.02*	s2=2.854
Switzerland	1.03	20	0.90	-0.06	-0.06		0.21
		(7.52)	(0.48)	(0.48)	(0.48)		s1=2.134
	15.49*	2.66***	1.87**	-0.13	-0.13	1.72	s2=2.134
U.K.	0.18	1	0.00	-0.43	-0.43		0.0057
		(7.52)	(0.33)	(0.34)	(0.34)		s1=1000
	2.76	0.13	0.00	-1.28	-1.28	0.77	s2=20

*Rejection at the 10% significance level.
**Rejection at the 5% significance level.
***Rejection at the 1% significance level.

Notes:

1. See sections 3.1, 3.2 and 3.4 for the definition of the shrinkage based tests Sign-N, Max-MSPE-Adjusted and MSPE-Adjusted. See section 4.1 for a description of the benchmark tests MSPE-Normal, MSPE-McCracken and CCS.

2. Rejections at 1% (***), 5% (**) and 10% (*) level of significance.

3. Standard Deviations in parenthesis.

4. For each country and each statistic we report three numbers. First, we report the value of the statistic without normalization. Next, we report its standard deviation and finally we report the value of the normalized statistic to be compared with critical values. For instance, in the case of Canada and Sign -N test the value of the statistic without normalization is 15. Its standard deviation is 6.44 and the ratio between both is 2.33.

5. The last column displays the difference between the MSPE of the null model and the MSPE of the alternative model using different shrinkage factors.

6. Data range: 1980:01-2003:10 for Canada and Japan, 1975:01-2003:10 for Switzerland and U.K. and 1993:4-2005:4 for Chile.

In this empirical application we can group our shrinkage based tests in two families. First, a family including the MSPE-Adjusted and Max-MSPE-Adjusted tests. This family has a similar rate of rejection within itself and when compared with the rest of the asymptotically correct tests in the literature (CCS and MSPE-McCracken). Second, a family that includes the sign test that shows more rejection than its competitors. As explained in section 3, we do not need to focus on the relative size of a statistic like (25), but only on its sign. As long as a statistic like (25) is positive we know that a shrunken alternative model might be more useful than the martingale difference model to correctly predict changes in exchange rates.

We also notice that in all five cases the Max-MSPE-Adjusted statistic is positive. This implies that it would be possible to construct forecasts from the alternative model displaying lower MSPE with a big enough shrinkage factor. Column (8) shows the value of the difference in MSPE when the alternative model is shrunk by some specific factors. With these shrinkage factors the alternative model is more accurate than the null model for all countries including the U.K.

Table 3 explores the robustness of previous results with respect to changes in the number of observations used for prediction and estimation. Results seem to be robust in a neighborhood of P and R with just a couple of exceptions. As we can see from Table 3, anytime the null is rejected, it is also rejected for most of the different combinations of P and R, the only exception being Switzerland with the Max-MSPE-Adjusted statistic where the rejection rate is rarely low. On the other hand, almost every time the null is not rejected, it is also not rejected for most of the different combinations of P and R. The only exception to this is the case of Chile with the MSPE-Adjusted statistic, where there is a rejection rate of about 0.5.

Table 3
Forecasts of Monthly Changes in U.S. Dollar Exchange Rates
Percentage of Rejection at 10% level

(1)	(2)	(3)	(4)	(5)	(6)
country	**Max MSPE-Adj**	**Sign-N**	**MSPE Adjusted**	**MSPE McCracken**	**CCS**
Canada	*0.00*	*1.00*	*0.80*	*0.80*	*0.00*
	0.89	*2.33****	*1.77***	*0.54***	*1.94*
Chile	*0.73*	*0.94*	*0.45*	*0.73*	*0.73*
	*6.11E-04***	*1.35**	*1.17*	*-0.15*	*7.25***
Japan	*0.00*	*0.77*	*0.29*	*0.00*	*0.89*
	4.64	*1.71***	*1.23*	*-0.52*	*5.02**
Switzerland	*0.34*	*0.77*	*0.64*	*0.28*	*0.00*
	*15.49**	*2.66****	*1.87***	*-0.13*	*1.72*
U.K.	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	2.76	*0.13*	*0.00*	*-1.28*	*0.77*

*Rejection at the 10% significance level.
**Rejection at the 5% significance level.
***Rejection at the 1% significance level.

Notes:

1. See sections 3.1, 3.2 and 3.4 for the definition of the shrinkage based tests Sign-N, Max-MSPE-Adjusted and MSPE-Adjusted. See section 4.1 for a description of the benchmark tests MSPE-Normal, MSPE-McCracken and CCS.

2. Rejections at 1% (***), 5% (**) and 10% (*) level of significance.

3. For each country and each statistic we report two numbers. First we report the rejection rate of the null at the 10% level of significance. Next, we report the value of the normalized statistic presented in Table 2. For instance, in the case of Canada and Sign -N test we see that in all the combinations of R and P analyzed the null was rejected at the 10% level. Finally we report the value 2.33 which is the value of the normalized statistic obtained in Table 2.

Our tests, based on a shrinkage perturbation of traditional MSPE comparisons, provide interesting and sound evidence of the predictability of exchange rates in 4 out of 5 countries. Aside from this statistical evidence, our approach also allows us to find an interpretation for our rejections: rejection means the existence of a positive shrinkage factor for which the MSPE difference between the null and the shrunk alternative favors the alternative. Put differently, rejection means statistically significant evidence of diversification gains. This result would allow us to find, at least theoretically, a better forecast in terms of MSPE than the martingale difference model anytime we reject the null.

6 Extensions

6.1 Forecast Accuracy

We have claimed that rejecting the null ensures the existence of a positive shrinkage factor s, such that shrunken forecasts from the alternative display lower MSPE than the null. The question about estimation of the shrinkage factor, however, remains open and further research in this direction should be done.

6.2 Granger Causality Tests

Our shrinkage procedure may be extended to construct tests in more general contexts beyond those of the MDH. We consider here a brief extension to test for Granger causality. Following Clark and West (2005b) consider the models

$$\begin{aligned} Model\,1 \quad (null) \quad &: \quad y_{t+1} = X'_{t+1}\beta + e_{1t+1} && (33) \\ Model\,2 \quad (alternative) \quad &: \quad y_{t+1} = X'_{t+1}\beta + Z'_{t+1}\gamma + e_{2t+1} && (34) \end{aligned}$$

where $E(e_{1t+1}|\mathcal{F}_t) = E(e_{2t+1}|\mathcal{F}_t) = 0$, and we have assumed for simplicity that $EX_{t+1}Z'_{t+1} = 0$. Working at the population level we could define the following auxiliary function $f : \mathbb{R}_+ \to \mathbb{R}$

$$\begin{aligned} f(s) &= E(y_{t+1} - X'_{t+1}\beta)^2 - E((y_{t+1} - X'_{t+1}\beta) - \frac{Z'_{t+1}\gamma}{s})^2 \\ &= 2E(y_{t+1} - X'_{t+1}\beta)\frac{Z'_{t+1}\gamma}{s} - E\left[\frac{Z'_{t+1}\gamma}{s}\right]^2 \end{aligned}$$

Under the null, $\gamma = 0$, so $f(s) = 0$ for all $s > 0$. Under the alternative, however, $\gamma \neq 0$ and we have

$$f(s) > 0 \Leftrightarrow E(y_{t+1} - X'_{t+1}\beta)Z'_{t+1}\gamma > 0$$

The existence of different behavior under the null and the alternative suggests that it would be possible to construct a test based upon the sample analog of

$$E(y_{t+1} - X'_{t+1}\beta)Z'_{t+1}\gamma$$

Of course further analysis is required to give a formal definition of a test. Our only intention is to show that this is a natural extension of our shrinkage based tests, which seems to be a promising area for future research.

6.3 Equivalent Measure

Our tests are derived based on a robustness principle. Basically, we use the fact that under the null model, a deterministic perturbation over a conditional expectation would leave the MSPE differential unchanged. Under the alternative we expected to find some change. A natural extension for future research is to explore the robustness of the MSPE differential to stochastic perturbations. For instance, let us assume that we have a set of regressors in the vector X_{t+1}, and a positive stochastic perturbation

$$0 < m_t \in \mathcal{F}_t$$

Notice that this positive stochastic factor does not affect the distinctive features under the null and under the alternative of the population analog of a statistic like (25).

In fact, we have

$$\begin{aligned} E(y_{t+1}X_{t+1}\beta m_t | H_0) &= E(e_{t+1}X_{t+1}\beta m_t) = 0 \\ E(y_{t+1}X_{t+1}\beta m_t | H_A) &= E((X_{t+1}\beta)^2 m_t) > 0 \end{aligned}$$

Preliminary simulations and empirical applications with different stochastic perturbations, not included for the sake of brevity, show that some stochastic perturbations boost the empirical power of the tests presented here, sometimes drastically. This preliminary evidence shows that there is some room for research on this topic.

6.4 Combination of Tests

Combination of forecasts has been proven useful in the exchange rate literature to outperform the random walk model in forecasting comparisons under a quadratic loss (see Clemen (1989)). Combining strategies have been reported as having excellent predictive behavior by several authors including Wright (2003) and Avramov (2002), who independently showed the predictive power of Bayesian Model Averaging as a combining tool. We have also stressed throughout the paper that diversification gains may arise when combining different forecasts. A natural question to ask is the following: is it possible to obtain diversification gains when combining different tests of the MDH? The intuition is as follows: assume that we have two statistics T_1 and T_2 that under the null are normally distributed

$$T_i \rightsquigarrow N(0, V_i) \quad i = 1, 2$$

Consider the following combined statistic

$$T^C = \omega T_1 + (1 - \omega) T_2$$

Then we have

$$T^C \rightsquigarrow N(0, V(\omega))$$

and we could pick ω to minimize $V(\omega)$. If this is so, and diversification gains prevail, the combined statistic will have narrower confidence intervals. It seems worthwhile to analyze whether these narrower confidence intervals will be associated with gains in power. An extension like this is also left for future research.

7 Conclusion

We define a family of tests for the Martingale Difference Hypothesis based upon a shrinkage principle. In other words, tests within this family are such that rejection of the null implies that forecasts from the alternative model, adjusted by a shrinkage factor, will display lower Mean Square Prediction Error than forecasts from the null model. In particular we introduce a new test called Max-MSPE-Adjusted that displays good power and size properties.

We also show that already known tests, like the direction of change test and the MSPE-Adjusted test, can be derived from our shrinkage procedure. These shrinkage tests explore the ability of the alternative model to produce more accurate forecasts than the null model. In other words, rejecting the null with these tests implies that shrunken forecasts from the alternative model should display lower MSPE than the null model. In this sense we show that direction of change tests are closely related to MSPE comparisons.

We explore via simulations the small sample behavior of shrinkage based tests along with other tests already known in the literature. Following Clark and West (2005a), we use a bivariate DGP calibrated to match a simple model of exchange rates based upon interest parity. Surprisingly, we find that tests based on direct MSPE comparisons are outperformed by other tests in terms of power properties. We argue that these MSPE based tests might not be as powerful as other tests of the MDH because they neglect benefits from combination and therefore do not use all the information about potential combination gains. This result questions the traditional use of MSPE comparisons, at least in the context we have used them here.

Simulation results indicate that the sign test (direction of change) and the Max-MSPE-Adjusted display the closest empirical size to nominal size. In terms of power, the Max-MSPE-Adjusted and the CCS test seem to outperform the rest of the tests for normal tails, whereas for thick tails the Sign test tends to outperform the rest of the tests.

We illustrate the use of our tests in an application within the exchange rate literature. Using monthly series of bilateral exchange rates and an alternative model based on interest parity, the Sign-N test (direction of change) rejects the MDH for Canada, Chile, Japan, and Switzerland, sometimes at extremely tight critical values. The MSPE-Normal test is unable to reject the null for all five countries, and the rest of the tests reject the null for only two countries. For the U.K, however, no test rejects the null. For all five cases analyzed, however, it is possible to find a shrinkage factor for which the alternative renders improved forecasts over the null model. Therefore, this application suggests that, at least in the context of our simulations, the Sign-N test is more appropriate than other tests to detect diversification gains. These results seem to be robust to the size of the estimation and forecasting windows, as long as P is relatively big with respect to R.

Several interesting extensions are suggested for future research. First, the estimation of an optimal shrinkage factor to improve forecast accuracy is still an open question. Second, the extension of our shrinkage based tests to more general environments of Granger causality is promising but not straightforward. Third, we claim that the use of a stochastic perturbation might boost the empirical power of some of the tests. Finally we ask whether a suitable combination of tests may yield power gains when testing the MDH.

8 Appendix

8.1 Figures

Figure 1
Auxiliary Function f(s)

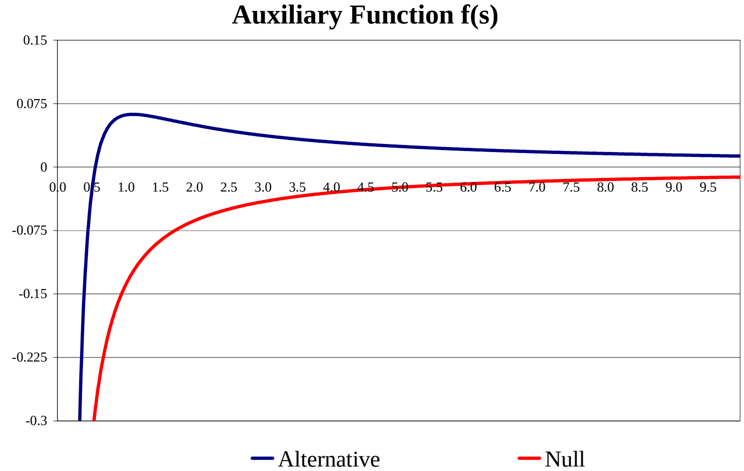
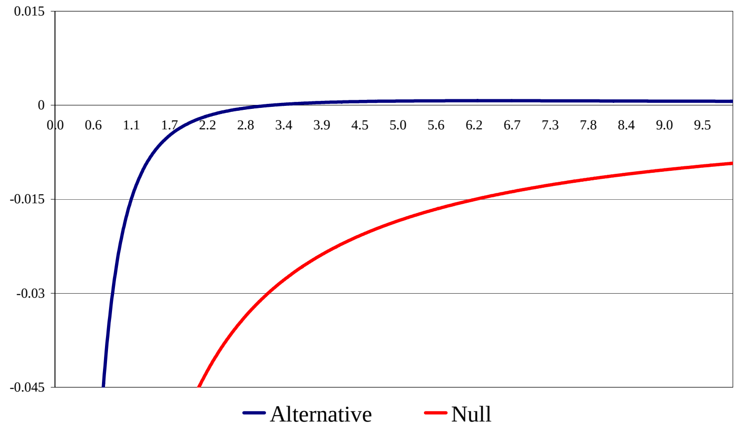


Figure 2
Auxiliary Function f(s)



8.2 Simulation Results

8.2.1 Results on Empirical Size

Table 4
Empirical Size DGP 1
R=35, Size = 10%, beta = -2, Gaussian errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.094*	*0.100*	*0.105*	*0.100*	*0.095*	*0.102*
Sign-N	*0.102*	*0.092*	*0.093*	*0.101*	*0.087*	*0.109*
MSPE-Adjusted	*0.075*	*0.075*	*0.083*	*0.076*	*0.081*	*0.089*
MSPE-Normal	*0.002*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
MSPE-McCracken	*0.072*	*0.048*	*0.040*	*0.031*	*0.018*	*0.016*
CCS	*0.128*	*0.101*	*0.098*	*0.094*	*0.089*	*0.095*

Table 5
Empirical Size DGP 1
R=120, Size = 10%, beta = -2, Gaussian Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.095*	*0.088*	*0.088*	*0.085*	*0.098*	*0.099*
Sign-N	*0.099*	*0.090*	*0.094*	*0.107*	*0.095*	*0.107*
MSPE-Adjusted	*0.075*	*0.063*	*0.067*	*0.067*	*0.074*	*0.082*
MSPE-Normal	*0.022*	*0.006*	*0.002*	*0.001*	*0.000*	*0.000*
MSPE-McCracken	*0.097*	*0.086*	*0.078*	*0.076*	*0.066*	*0.059*
CCS	*0.127*	*0.107*	*0.102*	*0.100*	*0.098*	*0.092*

Table 6
Empirical Size DGP 1
R=240, Size = 10%, beta = -2, Gaussian Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.091*	*0.089*	*0.090*	*0.086*	*0.090*	*0.092*
Sign-N	*0.098*	*0.092*	*0.092*	*0.101*	*0.093*	*0.109*
MSPE-Adjusted	*0.076*	*0.070*	*0.067*	*0.064*	*0.069*	*0.071*
MSPE-Normal	*0.038*	*0.017*	*0.011*	*0.005*	*0.001*	*0.000*
MSPE-McCracken	*0.103*	*0.095*	*0.093*	*0.095*	*0.083*	*0.080*
CCS	*0.128*	*0.113*	*0.104*	*0.098*	*0.097*	*0.095*

Table 7
Empirical Size DGP 2
R=35, Size = 10%, beta = -2, t(2) Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.098*	*0.104*	*0.110*	*0.104*	*0.102*	*0.099*
Sign-N	*0.097*	*0.089*	*0.094*	*0.104*	*0.089*	*0.112*
MSPE-Adjusted	*0.082*	*0.085*	*0.090*	*0.090*	*0.095*	*0.096*
MSPE-Normal	*0.002*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
MSPE-McCracken	*0.064*	*0.046*	*0.038*	*0.034*	*0.032*	*0.041*
CCS	*0.150*	*0.119*	*0.117*	*0.111*	*0.107*	*0.105*

Table 8
Empirical Size DGP 2
R=120, Size = 10%, beta = -2, t(2) Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.096*	*0.094*	*0.094*	*0.097*	*0.102*	*0.100*
Sign-N	*0.096*	*0.086*	*0.097*	*0.104*	*0.092*	*0.105*
MSPE-Adjusted	*0.084*	*0.076*	*0.078*	*0.078*	*0.084*	*0.087*
MSPE-Normal	*0.018*	*0.004*	*0.001*	*0.001*	*0.000*	*0.000*
MSPE-McCracken	*0.088*	*0.081*	*0.072*	*0.068*	*0.056*	*0.050*
CCS	*0.153*	*0.118*	*0.111*	*0.113*	*0.099*	*0.098*

Table 9
Empirical Size DGP 2
R=240, Size = 10%, beta = -2, t(2) Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.099*	*0.092*	*0.095*	*0.095*	*0.095*	*0.100*
Sign-N	*0.102*	*0.086*	*0.097*	*0.105*	*0.097*	*0.111*
MSPE-Adjusted	*0.086*	*0.080*	*0.079*	*0.078*	*0.076*	*0.081*
MSPE-Normal	*0.032*	*0.014*	*0.009*	*0.005*	*0.000*	*0.000*
MSPE-McCracken	*0.106*	*0.091*	*0.089*	*0.093*	*0.076*	*0.070*
CCS	*0.156*	*0.118*	*0.114*	*0.108*	*0.103*	*0.099*

Notes:

1.See sections 3.1, 3.2 and 3.4 for the definition of the shrinkage based tests Sign-N, Max-MSPE-Adjusted and MSPE-Adjusted. See section 4.1 for a description of the benchmark tests MSPE-Normal, MSPE-McCracken and CCS.

2. For all the experiments, nominal size is set to 10%.

3. See section 4.1 for the definition of the DGPs. Notice that DGP 2 differs from DGP 1 in that the distribution of the dependent variable perturbations displays fat tails.

8.2.2 Results on Power

Table 10
Raw Power DGP 1
R=35, Size = 10%, beta = -2, Gaussian Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.228*	*0.312*	*0.367*	*0.423*	*0.556*	*0.660*
Sign-N	*0.192*	*0.230*	*0.275*	*0.326*	*0.422*	*0.540*
MSPE-Adjusted	*0.201*	*0.284*	*0.357*	*0.407*	*0.591*	*0.708*
MSPE-Normal	*0.015*	*0.004*	*0.003*	*0.001*	*0.000*	*0.000*
MSPE-McCracken	*0.178*	*0.225*	*0.264*	*0.281*	*0.388*	*0.486*
CCS	*0.258*	*0.375*	*0.492*	*0.602*	*0.876*	*0.966*

Table 11
Raw Power DGP 1
R=120, Size = 10%, beta = -2, Gaussian Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.314*	*0.467*	*0.568*	*0.640*	*0.849*	*0.929*
Sign-N	*0.237*	*0.301*	*0.395*	*0.461*	*0.619*	*0.741*
MSPE-Adjusted	*0.279*	*0.413*	*0.514*	*0.593*	*0.816*	*0.906*
MSPE-Normal	*0.093*	*0.083*	*0.087*	*0.085*	*0.104*	*0.114*
MSPE-McCracken	*0.265*	*0.405*	*0.500*	*0.578*	*0.787*	*0.874*
CCS	*0.256*	*0.378*	*0.497*	*0.598*	*0.883*	*0.963*

Table 12
Raw Power DGP 1
R=240, Size = 10%, beta = -2, Gaussian Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.353*	*0.537*	*0.655*	*0.749*	*0.917*	*0.968*
Sign-N	*0.260*	*0.355*	*0.435*	*0.529*	*0.710*	*0.832*
MSPE-Adjusted	*0.333*	*0.500*	*0.611*	*0.703*	*0.895*	*0.957*
MSPE-Normal	*0.150*	*0.179*	*0.199*	*0.219*	*0.312*	*0.409*
MSPE-McCracken	*0.307*	*0.439*	*0.557*	*0.665*	*0.878*	*0.946*
CCS	*0.257*	*0.381*	*0.494*	*0.612*	*0.880*	*0.964*

Table 13
Raw Power DGP 2
R=35, Size = 10%, beta = -2, t(2) Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.140*	*0.150*	*0.151*	*0.166*	*0.169*	*0.183*
Sign-N	*0.144*	*0.153*	*0.168*	*0.207*	*0.232*	*0.286*
MSPE-Adjusted	*0.120*	*0.135*	*0.136*	*0.156*	*0.172*	*0.195*
MSPE-Normal	*0.003*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
MSPE-McCracken	*0.092*	*0.071*	*0.062*	*0.059*	*0.059*	*0.075*
CCS	*0.203*	*0.195*	*0.208*	*0.236*	*0.336*	*0.415*

Table 14
Raw Power DGP 2
R=120, Size = 10%, beta = -2, t(2) Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.168*	*0.186*	*0.213*	*0.213*	*0.277*	*0.295*
Sign-N	*0.168*	*0.199*	*0.231*	*0.265*	*0.322*	*0.404*
MSPE-Adjusted	*0.148*	*0.160*	*0.178*	*0.184*	*0.245*	*0.274*
MSPE-Normal	*0.035*	*0.015*	*0.007*	*0.005*	*0.001*	*0.000*
MSPE-McCracken	*0.140*	*0.153*	*0.153*	*0.154*	*0.172*	*0.178*
CCS	*0.194*	*0.200*	*0.213*	*0.235*	*0.335*	*0.403*

Table 15
Raw Power DGP 2
R=240, Size = 10%, beta = -2, t(2) Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.179*	*0.212*	*0.241*	*0.255*	*0.321*	*0.362*
Sign-N	*0.184*	*0.224*	*0.264*	*0.304*	*0.393*	*0.476*
MSPE-Adjusted	*0.160*	*0.189*	*0.207*	*0.224*	*0.281*	*0.327*
MSPE-Normal	*0.064*	*0.042*	*0.032*	*0.026*	*0.010*	*0.006*
MSPE-McCracken	*0.159*	*0.173*	*0.195*	*0.214*	*0.249*	*0.279*
CCS	*0.194*	*0.196*	*0.220*	*0.236*	*0.340*	*0.409*

1. See notes on tables above.

8.2.3 Results on Size Adjusted Power

Table 16
Size Adjusted Power DGP 1
R=35, Nominal Size = 10%, beta = -2, Gaussian Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.160*	*0.287*	*0.332*	*0.354*	*0.533*	*0.637*
Sign-N	*0.127*	*0.230*	*0.275*	*0.284*	*0.422*	*0.513*
MSPE-Adjusted	*0.174*	*0.324*	*0.377*	*0.410*	*0.603*	*0.718*
MSPE	*0.157*	*0.319*	*0.389*	*0.418*	*0.620*	*0.734*
CCS	*0.150*	*0.357*	*0.484*	*0.569*	*0.874*	*0.966*

Table 17
Size Adjusted Power DGP 1
R=120, Nominal Size = 10%, beta = -2, Gaussian Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.374*	*0.487*	*0.577*	*0.638*	*0.847*	*0.922*
Sign-N	*0.237*	*0.301*	*0.395*	*0.417*	*0.619*	*0.721*
MSPE-Adjusted	*0.324*	*0.473*	*0.575*	*0.625*	*0.846*	*0.914*
MSPE	*0.269*	*0.415*	*0.535*	*0.596*	*0.829*	*0.907*
CCS	*0.216*	*0.344*	*0.482*	*0.565*	*0.880*	*0.964*

Table 18
Size Adjusted Power DGP 1
R=240, Nominal Size = 10%, beta = -2, Gaussian Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.503*	*0.639*	*0.728*	*0.777*	*0.932*	*0.972*
Sign-N	*0.260*	*0.355*	*0.435*	*0.479*	*0.710*	*0.815*
MSPE-Adjusted	*0.386*	*0.551*	*0.665*	*0.732*	*0.919*	*0.967*
MSPE	*0.298*	*0.435*	*0.555*	*0.625*	*0.888*	*0.953*
CCS	*0.216*	*0.343*	*0.470*	*0.574*	*0.877*	*0.964*

Table 19
Size Adjusted Power DGP 2
R=35, Nominal Size =10%, beta = -2,t(2) Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.122*	*0.116*	*0.126*	*0.103*	*0.126*	*0.136*
Sign-N	*0.144*	*0.153*	*0.168*	*0.173*	*0.232*	*0.262*
MSPE-Adjusted	*0.135*	*0.150*	*0.153*	*0.145*	*0.169*	*0.189*
MSPE	*0.129*	*0.142*	*0.140*	*0.133*	*0.149*	*0.149*
CCS	*0.129*	*0.157*	*0.190*	*0.196*	*0.325*	*0.406*

Table 20
Size Adjusted Power DGP 2
R=120, Nominal Size = 10%, beta = -2, t(2) Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.140*	*0.143*	*0.158*	*0.154*	*0.209*	*0.242*
Sign-N	*0.168*	*0.199*	*0.231*	*0.226*	*0.322*	*0.379*
MSPE-Adjusted	*0.160*	*0.170*	*0.201*	*0.197*	*0.262*	*0.298*
MSPE	*0.143*	*0.161*	*0.184*	*0.176*	*0.232*	*0.277*
CCS	*0.135*	*0.156*	*0.185*	*0.192*	*0.315*	*0.409*

Table 21
Size Adjusted Power DGP 2
R=240, Nominal Size = 10%, beta = -2, t(2) Errors

	P=48	P=109	P=166	P= 226	P=480	P=700
Max-MSPE-Adjusted	*0.160*	*0.183*	*0.205*	*0.197*	*0.271*	*0.304*
Sign-N	*0.184*	*0.224*	*0.264*	*0.267*	*0.393*	*0.452*
MSPE-Adjusted	*0.165*	*0.199*	*0.235*	*0.236*	*0.307*	*0.353*
MSPE	*0.146*	*0.171*	*0.204*	*0.202*	*0.287*	*0.329*
CCS	*0.124*	*0.157*	*0.197*	*0.192*	*0.324*	*0.398*

1. See notes on tables above

8.3 Proof of Proposition 1

Proposition: If $\exists\, i \in \{1, ..., k\}$ such that $E\left(y_{t+1}\widehat{y}_{i,t+1}\ |H_A\right) > 0$ then there is a vector $\overrightarrow{s} = \overrightarrow{\overline{s}} \in \mathbb{R}^k_+ : E\left(\sum\limits_{i=1}^{k} \frac{y_{t+1}\widehat{y}^i_{i,t+1t}}{\overline{s}_i}\ |H_A\right) > 0$ and it is possible to find $h \in \mathbb{R}_+$ such that $E\left(\widehat{f}(h\overrightarrow{\overline{s}})\ |\ H_A\right) > 0$.

Proof. Without loss of generality let us assume $E\left(y_{t+1}\widehat{y}_{1,t+1}\ |H_A\right) > 0$. Pick $\overline{s}_1$ small enough and components of $\overline{s}_{-1}$ big enough so that $E\left(\sum\limits_{i=1}^{k} \frac{y_{t+1}\widehat{y}^i_{i,t+1t}}{\overline{s}_i}\ |H_A\right) > 0$. Now we need to pick $h \in \mathbb{R}_+$ according to:

$$\begin{aligned}
E\left(\widehat{f}(h\overrightarrow{\overline{s}})\ |\ H_A\right) &> 0 \Longleftrightarrow \\
\frac{2}{h}E\left(\sum_{i=1}^{k} \frac{y_{t+1}\widehat{y}_{i,t+1}}{\overline{s}_i}\ |H_A\right) - \frac{1}{h^2}E\left(\left(\sum_{i=1}^{k} \frac{\widehat{y}_{i,t+1}}{\overline{s}_i}\right)^2\ |\ H_A\right) &> 0 \Longleftrightarrow \\
2E\left(\sum_{i=1}^{k} \frac{y_{t+1}\widehat{y}_{i,t+1}}{\overline{s}_i}\ |H_A\right) - \frac{1}{h}E\left(\left(\sum_{i=1}^{k} \frac{\widehat{y}_{i,t+1}}{\overline{s}_i}\right)^2\ |\ H_A\right) &> 0 \Longleftrightarrow \\
\frac{2E\left(\sum\limits_{i=1}^{k} \frac{y_{t+1}\widehat{y}_{i,t+1}}{\overline{s}_i}\ |H_A\right)}{E\left(\left(\sum\limits_{i=1}^{k} \frac{\widehat{y}_{i,t+1}}{\overline{s}_i}\right)^2\ |\ H_A\right)} &> \frac{1}{h} \qquad (35)
\end{aligned}$$

Because $E\left(\left(\sum\limits_{i=1}^{k} \frac{\widehat{y}_{i,t+1}}{\overline{s}_i}\right)^2\ |\ H_A\right) > 0$, restriction (35) is well defined. Therefore for h big enough we will have $0 < E\left(\widehat{f}(h\overrightarrow{\overline{s}})\ |\ H_A\right)$. ∎

8.4 Proof of Proposition 2

Proposition: Under the null and assumptions (27), (28) and (29)

$$\lim_{P\to\infty} P\left(\left[\underset{s>0}{Sup} P^{1/2} \frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}\right] \leq u\right) \to \Phi(u) \text{ for all } u > 0$$

where Φ denotes a standard normal distribution function and

$$\widehat{f}(s) = \frac{2}{P}\sum_{t=R}^{T} y_{t+1}X'_{t+1}\widehat{\beta}_t - \frac{1}{P}\sum_{t=R}^{T}\frac{(X'_{t+1}\widehat{\beta}_t)^2}{s}$$

$$\widehat{V}(s) = 4\widehat{V}(y_{t+1}X'_{t+1}\widehat{\beta}_t) + \frac{1}{s^2}\widehat{V}((X'_{t+1}\widehat{\beta}_t)^2) - \frac{4}{s}\widehat{Cov}(y_{t+1}X'_{t+1}\widehat{\beta}_t, (X'_{t+1}\widehat{\beta}_t)^2)$$

Proof. For brevity we will use the following notation:

$$\begin{aligned} n_{t+1} &= y_{t+1}X'_{t+1}\widehat{\beta}_t; \quad d_{t+1} = (X'_{t+1}\widehat{\beta}_t)^2 \\ \overline{n} &= \frac{1}{P}\sum_{t=R}^{T} y_{t+1}X'_{t+1}\widehat{\beta}_t; \quad \overline{d} = \frac{1}{P}\sum_{t=R}^{T}(X'_{t+1}\widehat{\beta}_t)^2 \end{aligned}$$

$$\begin{aligned} f_1 &= \frac{2}{P}\sum_{t=R}^{T} y_{t+1}X'_{t+1}\widehat{\beta}_t \\ \overline{d} &= \frac{1}{P}\sum_{t=R}^{T}(X'_{t+1}\widehat{\beta}_t)^2 \end{aligned}$$

and

$$\begin{aligned} \widehat{V}(s) &= 4V_1 + \frac{1}{s^2}V_2 - \frac{4}{s}V_3 \\ V_1 &= \widehat{V}(y_{t+1}X'_{t+1}\widehat{\beta}_t) = \widehat{V}(n_{t+1}) \\ V_2 &= \widehat{V}((X'_{t+1}\widehat{\beta}_t)^2) = \widehat{V}(d_{t+1}) \\ V_3 &= \widehat{Cov}(y_{t+1}X'_{t+1}\widehat{\beta}_t, (X'_{t+1}\widehat{\beta}_t)^2) = \widehat{Cov}(n_{t+1}, d_{t+1}) \end{aligned}$$

Now notice that given $\{y_{t+1,}y_t, ..., y_{1;}X_{t+1,}X_t, ..., X_{1;}\}$ we have that

$$\lim_{s\to\infty}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}} = \frac{2\overline{n}}{\sqrt{4V_1}} = \frac{\frac{2}{P}\sum_{t=R}^{T} y_{t+1}X'_{t+1}\widehat{\beta}_t}{\sqrt{4\widehat{V}(y_{t+1}X'_{t+1}\widehat{\beta}_t)}} \tag{36}$$

$$\lim_{s\to 0}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}} = -\frac{\overline{d}}{\sqrt{V_2}} = -\frac{\frac{1}{P}\sum_{t=R}^{T}\frac{(X'_{t+1}\widehat{\beta}_t)^2}{s}}{\sqrt{\widehat{V}((X'_{t+1}\widehat{\beta}_t)^2)}} < 0 \tag{37}$$

These results and the fact that given $\{y_{t+1,}y_t, ..., y_{1;}X_{t+1,}X_t, ..., X_1\}$ the function $\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}$ is continuous and differentiable for $s > 0$, show that this function is bounded. Let us assume that this function has a local interior maximum $s^* > 0$. This local solution satisfies:

$$s^* = \frac{2\overline{n}V_2 - \overline{d}V_3}{2(2\overline{n}V_3 - \overline{d}V_1)} > 0$$

It is also possible to show that

$$\frac{d}{ds}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}} = \left\{ \begin{array}{lll} < 0 & if & s > s^* \\ > 0 & if & s < s^* \end{array} \right\}$$

which means that s^* is the unique interior local maximum.

If there is no local interior solution, then the supremum must be found either at (36) or at (37).

More generally we can characterize $\underset{s>0}{Sup}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}$ as follows:

$$\underset{s>0}{Sup}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}=\left\{\begin{array}{c} \frac{\widehat{f}(s^*)}{\sqrt{\widehat{V}(s^*)}} \;\; if \;\; 2\overline{n}V_2-\overline{d}V_3>0 \;\text{ and } 2\overline{n}V_3-\overline{d}V_1>0 \\ \max\{-\frac{\overline{d}}{\sqrt{V_2}},\frac{2\overline{n}}{\sqrt{4V_1}}\} if \;\; 2\overline{n}V_2-\overline{d}V_3<0 \;\text{ and } 2\overline{n}V_3-\overline{d}V_1<0 \\ \max\{-\frac{\overline{d}}{\sqrt{V_2}},\frac{2\overline{n}}{\sqrt{4V_1}}\} if \;\; (2\overline{n}V_3-\overline{d}V_1)\neq 0 \;\text{ and}\frac{2\overline{n}V_2-\overline{d}V_3}{2\overline{n}V_3-\overline{d}V_1}\leq 0 \\ -\frac{\overline{d}}{\sqrt{V_2}} \;\; if \;\; 2\overline{n}V_2-\overline{d}V_3<0 \text{ and } \; 2\overline{n}V_3=\overline{d}V_1 \\ \frac{2\overline{n}}{\sqrt{4V_1}} if \;\; 2\overline{n}V_2-\overline{d}V_3>0 \text{ and } \; 2\overline{n}V_3=\overline{d}V_1 \\ \frac{2\overline{n}}{\sqrt{4V_1}} if \;\; 2\overline{n}V_2=\overline{d}V_3 \text{ and } \; 2\overline{n}V_3=\overline{d}V_1 \end{array}\right\}$$

So, we have a description of $\underset{s>0}{Sup}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}$ in terms of six different and disjoint sets, namely:

$$\begin{array}{c} A_1=\left\{2\overline{n}V_2-\overline{d}V_3>0 \;\; and \; 2\overline{n}V_3-\overline{d}V_1>0\right\} \\ A_2=\left\{2\overline{n}V_2-\overline{d}V_3<0 \;\; and \;\; 2\overline{n}V_3-\overline{d}V_1<0\right\} \\ A_3=\left\{(2\overline{n}V_3-\overline{d}V_1)\neq 0 \;\; and \;\; \frac{2\overline{n}V_2-\overline{d}V_3}{2\overline{n}V_3-\overline{d}V_1}\leq 0\right\} \\ A_4=\left\{2\overline{n}V_2-\overline{d}V_3<0 \;\; and \;\; 2\overline{n}V_3-\overline{d}V_1=0\right\} \\ A_5=\left\{2\overline{n}V_2-\overline{d}V_3>0 \;\; and \;\; 2\overline{n}V_3-\overline{d}V_1=0\right\} \\ A_6=\left\{\; 2\overline{n}V_2-\overline{d}V_3 \;\; =0 \; and \;\; 2\overline{n}V_3-\overline{d}V_1=0\right\} \end{array}$$

satisfying

$$\sum_{i=1}^{6}\Pr(A_i)=1$$

It turns out that under the null hypothesis, 4 of these 6 sets have asymptotic probability zero (A_1, A_4, A_5, A_6). To see this, first notice that

$$\begin{aligned} 2\overline{n}V_2-\overline{d}V_3 &\rightarrow \;_P 0 \\ 2\overline{n}V_3-\overline{d}V_1 &\rightarrow \;_P -E((X'_{t+1}\widehat{\beta}_t)^2)V(e_{t+1}X'_{t+1}\widehat{\beta}_t) \\ &= \;\; -E((X'_{t+1}\widehat{\beta}_t)^2)E((X'_{t+1}\widehat{\beta}_t)^2E(e_{t+1}^2|\mathcal{F}_t))<0 \end{aligned}$$

which, when using our parameterization, reduces to

$$2\overline{n}V_3-\overline{d}V_1\rightarrow_P -E((X'_{t+1}\widehat{\beta}_t)^2)^2<0$$

To abbreviate notation and make explicit the dependence of $2\overline{n}, V_3, \overline{d}, V_1$ and V_2 from P let us write:

$$\begin{aligned} Z_{1P} &= 2\overline{n}V_2-\overline{d}V_3\rightarrow_P 0 & (38) \\ Z_{2P} &= 2\overline{n}V_3-\overline{d}V_1\rightarrow_P \mu<0 & (39) \end{aligned}$$

We are interested in the following limits

$$\lim_{P\to\infty} \Pr(A_i), ..., i = 1, ..., 6$$

By definition, and following (38) and (39), we have that for all $\theta_1, \theta_2 > 0$,

$$\begin{aligned}
\lim_{P\to\infty} \Pr(|Z_{1P}| &> \theta_1) = 0 \\
\lim_{P\to\infty} \Pr(|Z_{2P} - \mu| &> \theta_2) = 0
\end{aligned}$$

Now

$$\begin{aligned}
0 \leq \Pr(A_1) &\leq \Pr(Z_{2P} > 0) \leq \Pr(Z_{2P} > -\mu) \leq \\
\Pr(Z_{2P} - \mu &> -2\mu) \leq \Pr(|Z_{2P} - \mu| > -2\mu)
\end{aligned}$$

taking the limit when P goes to infinity and considering $\theta_2 = -2\mu > 0$ we have

$$0 \leq \lim_{P\to\infty} \Pr(A_1) \leq \lim_{P\to\infty} \Pr(|Z_{2P} - \mu| > -2\mu) = 0$$

therefore

$$\lim_{P\to\infty} \Pr\left(A_1\right) = 0$$

Now consider $i = 4, 5, 6$.

$$\begin{aligned}
\Pr(A_i) &\leq \Pr(Z_{2P} = 0) \leq \Pr(Z_{2P} > 2\mu) = \\
\Pr(Z_{2P} - \mu &> -\mu) \leq \Pr(|Z_{2P} - \mu| > -\mu)
\end{aligned}$$

taking the limit when P goes to infinity and considering $\theta_2 = -\mu > 0$ we have

$$0 \leq \lim_{P\to\infty} \Pr(A_i) \leq \lim_{P\to\infty} \Pr(|Z_{2P} - \mu| > -\mu) = 0$$

therefore

$$\lim_{P\to\infty} \Pr\left(A_i\right) = 0, ..., i = 4, 5, 6$$

From which we obtain that

$$\lim_{P\to\infty} \Pr\left(A_2 \cup A_3\right) = 1$$

Let us denote

$$\Theta = A_1 \cup A_4 \cup A_5 \cup A_6$$

Now notice that

$$0 \leq \Pr\left(\left[\underset{s>0}{Sup} P^{1/2} \frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}\right] \leq u | \Theta\right) \Pr\left(\Theta\right) \leq \Pr\left(\Theta\right)$$

therefore

$$\lim_{P\to\infty}\Pr\left(\left[\underset{s>0}{Sup}P^{1/2}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}\right]\leq u|\Theta\right)\Pr(\Theta)=0$$

On the other hand, consider $u\in\mathbb{R}^{+}$

$$\begin{aligned}\Pr\left(\left[\underset{s>0}{Sup}P^{1/2}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}\right]\leq u|A_2\cup A_3\right) &= \Pr\left(P^{1/2}\max\{-\frac{\overline{d}}{\sqrt{V_2}},\frac{2\overline{n}}{\sqrt{4V_1}}\}\leq u\right)\\ &= \Pr\left(P^{1/2}\frac{2\overline{n}}{\sqrt{4V_1}}\leq u\right)\end{aligned}$$

Therefore, for every $u\in\mathbb{R}^{+}$:

$$\Pr\left(\left[\underset{s>0}{Sup}P^{1/2}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}\right]\leq u\right)=\Pr\left(\left[\underset{s>0}{Sup}P^{1/2}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}\right]\leq u|\Theta\right)\Pr(\Theta)+$$

$$+\Pr\left(\left[\underset{s>0}{Sup}P^{1/2}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}\right]\leq u|A_2\cup A_3\right)\Pr(A_2\cup A_3)$$

So

$$\lim_{P\to\infty}\Pr\left(\left[\underset{s>0}{Sup}P^{1/2}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}\right]\leq u\right)=$$

$$\begin{aligned}\lim_{P\to\infty}\Pr\left(\left[\underset{s>0}{Sup}P^{1/2}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}\right]\leq u|A_2\cup A_3\right)\lim_{P\to\infty}\Pr(A_2\cup A_3) &=\\ \lim_{P\to\infty}\Pr\left(\left[\underset{s>0}{Sup}P^{1/2}\frac{\widehat{f}(s)}{\sqrt{\widehat{V}(s)}}\right]\leq u|A_2\cup A_3\right) &=\\ \lim_{P\to\infty}\Pr\left(P^{1/2}\max\{-\frac{\overline{d}}{\sqrt{V_2}},\frac{2\overline{n}}{\sqrt{4V_1}}\}\leq u\right) &=\\ \lim_{P\to\infty}\Pr\left(P^{1/2}\frac{2\overline{n}}{\sqrt{4V_1}}\leq u\right) &= \Phi(u)\end{aligned}$$

■

8.5 Proof of Proposition 3

Proposition: Under the null and assumption (32) the sequence $sign(y_{t+1}X'_{t+1}\widehat{\beta}_t)$ form a martingale-difference sequence with respect to the filtration F_t.

Proof.

$$\begin{aligned} E(sign(y_{t+1}X'_{t+1}\widehat{\beta}_t)|\mathfrak{F}_t) &= E(sign(e_{t+1}X'_{t+1}\widehat{\beta}_t)|\mathfrak{F}_t) = \\ P(e_{t+1}X'_{t+1}\widehat{\beta}_t &> 0|\mathfrak{F}_t) - P(e_{t+1}X'_{t+1}\widehat{\beta}_t < 0|\mathfrak{F}_t) \end{aligned}$$

but

$$P(e_{t+1}X'_{t+1}\widehat{\beta}_t > 0|\mathfrak{F}_t) = \left\{ \begin{array}{c} P(e_{t+1} > 0|\mathfrak{F}_t) = \frac{1}{2}\ if\ X'_{t+1}\widehat{\beta}_t > 0 \\ P(e_{t+1} < 0|\mathfrak{F}_t) = \frac{1}{2}\ if\ X'_{t+1}\widehat{\beta}_t < 0 \\ P(0 > 0|\mathfrak{F}_t) = 0\ if\ \ X'_{t+1}\widehat{\beta}_t = 0 \end{array} \right\}$$

$$P(e_{t+1}X'_{t+1}\widehat{\beta}_t < 0|\mathfrak{F}_t) = \left\{ \begin{array}{c} P(e_{t+1} < 0|\mathfrak{F}_t) = \frac{1}{2}\ if\ X'_{t+1}\widehat{\beta}_t > 0 \\ P(e_{t+1} > 0|\mathfrak{F}_t) = \frac{1}{2}\ if\ X'_{t+1}\widehat{\beta}_t < 0 \\ P(0 < 0|\mathfrak{F}_t) = 0\ if\ \ X'_{t+1}\widehat{\beta}_t = 0 \end{array} \right\}$$

Therefore

$$E(sign(e_{t+1}X'_{t+1}\widehat{\beta}_t)|\mathfrak{F}_t) = 0$$

and we conclude that, under the null, the sequence $sign(e_{t+1}X'_{t+1}\widehat{\beta}_t)$ is a martingale difference sequence with respect to the filtration F_t. ■

8.6 Proof of Proposition 4 (based on Ibragimov and Brown (2005))

Proposition: Let us consider the following sequences $\left\{sign(y_{t+1}X'_{t+1}\widehat{\beta}_t)\right\}_{t\geq 1}$, $\{\varepsilon_{t+1}\}_{t\geq 0}$ where $\{\varepsilon_{t+1}\}_{t\geq 0}$ denotes an i.i.d. sequence of symmetric Bernoulli random variables independent of $sign(y_{t+1}X'_{t+1}\widehat{\beta}_t)$ and the information available until time t, F_t. Then, under the null and assumption (32), the sequence $\eta_{t+1} = sign(y_{t+1}X'_{t+1}\widehat{\beta}_t) + \varepsilon_{t+1}I(y_{t+1}X'_{t+1}\widehat{\beta}_t = 0)$ forms an i.i.d. sequence of Bernoulli random variables taking values in $\{1, -1\}$,where I denotes an indicator function.

Proof. First notice that η_{t+1} takes only values in$\{1, -1\}$. Let us denote $\widetilde{P}(A) = P(A|F_t)$,then we have:

$$\begin{aligned} \widetilde{P}(\eta_{t+1} &= 1) = \widetilde{P}(\eta_{t+1} = 1|I = 1)\widetilde{P}(I = 1) + \widetilde{P}(\eta_{t+1} = 1|I = 0)\widetilde{P}(I = 0) \\ &= \widetilde{P}(\varepsilon_{t+1} = 1)\widetilde{P}(I = 1) + \widetilde{P}(sign(e_{t+1}X'_{t+1}\widehat{\beta}_t) > 0)\widetilde{P}(I = 0) \end{aligned}$$

Now notice that $\widetilde{P}(\varepsilon_{t+1} = 1) = 0.5$. Besides we have that:

$$\widetilde{P}(I = 1) = P(e_{t+1}X'_{t+1}\widehat{\beta}_t = 0|\mathfrak{F}_t) = \left\{ \begin{array}{lll} 1 & \text{if} & X'_{t+1}\widehat{\beta}_t = 0 \\ 0 & \text{if} & X'_{t+1}\widehat{\beta}_t \neq 0 \end{array} \right\}$$

$$\widetilde{P}(I = 0) = P(e_{t+1}X'_{t+1}\widehat{\beta}_t \neq 0|\mathfrak{F}_t) = \left\{ \begin{array}{lll} 0 & \text{if} & X'_{t+1}\widehat{\beta}_t = 0 \\ 1 & \text{if} & X'_{t+1}\widehat{\beta}_t \neq 0 \end{array} \right\}$$

$$\widetilde{P}(sign=1)=P(e_{t+1}X_{t+1}'\widehat{\beta}_t>0|\mathfrak{F}_t)=\left\{\begin{array}{ccc} 0 & \text{if} & X_{t+1}'\widehat{\beta}_t=0 \\ 0.5 & \text{if} & X_{t+1}'\widehat{\beta}_t\neq 0 \end{array}\right\}$$

this implies that

$$\widetilde{P}(\eta_{t+1}=1)=\left\{\begin{array}{ccc} 0.5 & \text{if} & X_{t+1}'\widehat{\beta}_t=0 \\ 0.5 & \text{if} & X_{t+1}'\widehat{\beta}_t\neq 0 \end{array}\right\}=0.5$$

therefore it is clear that the sequence η_t is a sequence of symmetric Bernoulli random variables. Let us show now that they are indeed independent. First notice that this sequence forms a martingale difference sequence:

$$\begin{aligned} E(\eta_{t+1}|\mathfrak{F}_t) &= E(sign(e_{t+1}X_{t+1}'\widehat{\beta}_t)|\mathfrak{F}_t)+E(\varepsilon_t I(e_{t+1}X_{t+1}'\widehat{\beta}_t=0)|\mathfrak{F}_t) \\ &= 0+E(I(e_{t+1}X_{t+1}'\widehat{\beta}_t=0)|\mathfrak{F}_t)E(\varepsilon_t) \\ &= 0 \end{aligned}$$

Now consider the following scalars

$$1\leq l_1\leq l_2\leq ...\leq l_k, k=2,3...$$

then we have

$$E\left(\eta_{l_1}\eta_{l_2}...\eta_{l_k}\right)=E\left(\eta_{l_1}\eta_{l_2}...\eta_{l_{k-1}}E(\eta_{l_k}|\mathfrak{F}_{l_k-1})\right)=0$$

Notice that we could rewrite $I(e_{t+1}X_{t+1}'\widehat{\beta}_t=h_{t+1})$ as follows

$$I(\eta_{t+1}=h_{t+1})=\frac{(1+\eta_{t+1}h_{t+1})}{2}$$

where $h_{t+1}\in\{1,-1\}$Therefore we have that for all

$$1\leq j_1\leq j_2\leq ...\leq j_m, m=2,3...$$

and any $h_{j_k}\in\{1,-1\}, k=1,2,...,m$

$$\begin{aligned} P(\eta_{j_1} &= h_{j_1},...,\eta_{j_m}=h_{j_m})=EI(\eta_{j_1}=h_{j_1})I(\eta_{j_2}=h_{j_2})...I(\eta_{j_m}=h_{j_m}) \\ &= \frac{E(1+\eta_{j_1}h_{j_1})(1+\eta_{j_2}h_{j_2})...(1+\eta_{j_m}h_{j_m})}{2^m} \\ &= \frac{1}{2^m} \\ &= P(\eta_{j_1}=h_{j_1})P(\eta_{j_2}=h_{j_2})...P(\eta_{j_m}=h_{j_m}) \end{aligned}$$

given that any expectation involving products of η_{l_k} terms is zero. ■

9 References

Anatolyev S. and A. Gerco (2003) "A Trading Approach to Testing for Predictability" Unpublished manuscript, New Economic School.

Avramov D. (2002) “Stock return predictability and model uncertainty”, Journal of Financial Economics 64, 423-458.

Brown D. and R. Ibragimov (2005) “Sign Tests for Dependent Observations and Bounds for Path-Dependent Options”. Yale ICF Working Paper No. 05-19. September 2005.

Campbell J., Andrew L and A.C MacKinlay (1997) “The Econometric of Financial Markets”. Princeton University Press.

Chao J., Corradi V. and N. Swanson (2000) “An Out of Sample Test for Granger Causality”, Macroeconomic Dynamics 5, 598-620.

Chinn, M., and Meese, R.A (1991) “Banking on Currency Forecasts: Is Change in Money Predictable?” Unpublished manuscript, University of California, Berkeley, Graduate School of Business.

Cheung, Y., Chinn M. and Garcia A. (2002) “Empirical Exchange Rate Models of the Nineties : Are Any Fit to Survive?” NBER Working Paper No. w9393.

Clark T. and M. McCracken (2003) “Evaluating Long Horizon Forecasts” manuscript University of Missouri..

Clark T. and K. West (2005a) “Using Out-Of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis”. Forthcoming Journal of Econometrics.

Clark T. and K. West (2005b) “Approximately Normal Tests for Equal Predictive Accuracy in Nested Models”. Manuscript University of Wisconsin.

Clemen R. (1989) “Combining Forecasts: A Review and Annotated Bibliography”. International Journal of Forecasting 5 559-583.

Diebold F. and R. Mariano (1995). “Comparing Predictive Accuracy”, Journal of Business and Economic Statistics,13, 253-263.

Efron, B. and Morris, C (1975) “Data Analysis Using Stein’s Estimator and its Generalizations”. Journal of the American Statistical Association, 70, 311-319.

Giacomini R. and H. White (2003) “Test of Conditional Predictive Ability”. Manuscript, University of California, San Diego.

Hamilton J. (1994) “Time Series Analysis”, Princeton University Press.

Ibragimov R. and D. Brown (2005) “Sign Tests for Dependent Observations and Bounds for Path-Dependent Options”. Cowles Foundation Discussion Paper No 1518, Yale University.

Inoue A. and L. Kilian (2004) “How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation”. Manuscript University of Michigan.

Inoue A. and L. Kilian (2003) “In-Sample or Out-of-Sample Tests of Predictability? Which one Should We Use? ”. Manuscript University of Michigan.

James, W., and Stein, C. (1961). “Estimation with quadratic loss”. Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1, 311-319.

McCracken M. (2004) “Asymptotics for Out of Sample Tests of Causality”. Manuscript, University of Missouri..

McCracken M. and K. West (2002) “Inference About Predictive Ability”. 299-321 in M. Clements and D. Hendry (eds) Companion to Economic Forecasting, Oxford: Blackwell (2002).

Meese R. and K. Rogoff (1983) “Empirical Exchange Rate Models of the Seventies. Do They Fit Out-of-Sample?” Journal of International Economics, 14, pp 3-24.

Miller D. and D. Williams (2003) “Shrinkage Estimators of Time Series Seasonal Factors and their Effect on Forecasting Accuracy”. International Journal of Forecasting, 19, (2003) 669-684.

Patton A. and A. Timmermann (2003) “Properties of Optimal Forecasts”. Manuscript University of California, San Diego.

Reineke D, Baggett J. and A. Elfessi (2003) “A Note on the Effect of Skewness, Kurtosis, and Shifting on One-Sample t and Sign Tests”. Journal of Statistics Education Volume 11, Number 3.

Skouras, S. (2001) “Financial returns and efficiency as seen by an artificial technical analyst”. Journal of Economic Dynamics and Control 25, 213-244.

Timmermann A. (2004) “Forecast Combinations”. Forthcoming in Handbook of Economic Forecasting.

Tsay (2002) “Analysis of Financial Time Series”. Wiley Series in Probability and Statistics.

West, K.(1996) “Asymptotic Inference About Predictive Ability”. Econometrica 64, 1067-1084.

West, K.(2005) “Forecast Evaluation”.Manuscript, University of Wisconsin-Madison.

White, H. (1984) “Asymptotic Theory for Econometricians” in Economic theory, econometrics and mathematical economics. academic Press Inc.

Wright J. (2003) “Bayesian Model Averaging and Exchange Rate Forecasts”. Board of Governors of the Federal Reserve System, International Finance Discussion Papers. Number 779. September 2003.

Documentos de Trabajo Banco Central de Chile	Working Papers Central Bank of Chile
NÚMEROS ANTERIORES	PAST ISSUES

La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: www.bcentral.cl/esp/estpub/estudios/dtbc. Existe la posibilidad de solicitar una copia impresa con un costo de $500 si es dentro de Chile y US$12 si es para fuera de Chile. Las solicitudes se pueden hacer por fax: (56-2) 6702231 o a través de correo electrónico: bcch@bcentral.cl.

Working Papers in PDF format can be downloaded free of charge from: www.bcentral.cl/eng/stdpub/studies/workingpaper. Printed versions can be ordered individually for US$12 per copy (for orders inside Chile the charge is Ch$500.) Orders can be placed by fax: (56-2) 6702231 or e-mail: bcch@bcentral.cl.

DTBC-375 Octubre 2006
Real Dollarization, Financial Dollarization, and Monetary Policy
Alain Ize y Eric Parrado

DTBC-374 Agosto 2006
Trade Liberalization, Price Distortions, and Resource Reallocation
Roberto Álvarez y Matías Braun

DTBC-373 Agosto 2006
Openness Can be Good for Growth: The Role of Policy Complementarities
Roberto Chang, Linda Kaltani y Norman Loayza

DTBC-372 Agosto 2006
Market Reforms and Efficiency Gains in Chile
Raphael Bergoeing, Andrés Hernando y Andrea Repetto

DTBC-371 Agosto 2006
Sovereign Debt in the Americas: New Data and Stylized Facts
Kevin Cowan, Eduardo Levy-Yeyati, Ugo Panizza y Federico Sturzenegger

DTBC-370 Agosto 2006
Identifying Fiscal Policy Shocks in Chile and Colombia
Jorge E. Restrepo y Hernán Rincón

DTBC-369 Julio 2006
Bank Ownership and Lending Behavior
Alejandro Micco y Ugo Panizza

DTBC-368 Julio 2006
Inflation Targeting in Dollarized Economies
Leonardo Leiderman, Rodolfo Maino y Eric Parrado

DTBC-367 Junio 2006
El Banco Central de Chile en Comparación con los Bancos Centrales del Mundo
Marcelo Ochoa y Klaus Schmidt-Hebbel

DTBC-366 Junio 2006
An Alternative Measure of Core Inflation
Michael Pedersen

DTBC-365 Junio 2006
Chile's Economic Growth [El Crecimiento Económico de Chile]
Klaus Schmidt-Hebbel

DTBC-364 Mayo 2006
Metas de Inflación y el Objetivo de Pleno Empleo
José de Gregorio

DTBC-363 Mayo 2006
Skill Premium in Chile: Studying the Skill Bias Technical Change Hypothesis in The South
Francisco A. Gallego

DTBC-362 Mayo 2006
Setting the Operational Framework for Producing Inflation Forecasts
Jorge Canales-Kriljenko, Turgut Kisinbay, Rodolfo Maino y Eric Parrado

DTBC-361 Mayo 2006
Estimaciones de *NAIRU* para Chile
Jorge Restrepo

DTBC-360 Abril 2006
Central Bank Independence and Monetary Policymaking Institutions: Past, Present, and Future
Alex Cukierman

DTBC-359 Febrero 2006
The Consumption-Real Exchange Rate Anomaly: non-Traded Goods, Incomplete Markets and Distribution Services
Jorge Selaive y Vicente Tuesta