Banco Central de Chile
Documentos de Trabajo

Central Bank of Chile
Working Papers

N° 459

Febrero 2008

COMBINING TESTS OF PREDICTIVE ABILITY: THEORY AND EVIDENCE FOR CHILEAN AND CANADIAN EXCHANGE RATES

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° 459

Working Paper
N° 459

COMBINING TESTS OF PREDICTIVE ABILITY: THEORY AND EVIDENCE FOR CHILEAN AND CANADIAN EXCHANGE RATES

Pablo Pincheira
Gerencia de Investigación Económica
Banco Central de Chile

Resumen

En este trabajo nos concentramos en combinar estadísticos fuera de muestra para la Hipótesis de una Martingala en Diferencias, de modo de explorar si un nuevo estadístico combinado puede generar un *test* con mayor potencia asintótica. El supuesto de normalidad asintótica implica que se puede obtener mayor potencia al encontrar la ponderación óptima en un cuociente del tipo t. Desafortunadamente, esta ponderación óptima es degenerada cuando la hipótesis nula de no predictibilidad es verdadera. Para superar este problema se introduce una función de penalización que atrae la ponderación óptima al interior del conjunto factible de combinaciones. La nueva ponderación asociada a este problema de penalización está bien definida bajo la hipótesis nula, asegurando normalidad asintótica del *test* combinado resultante. Demostramos, por medio de simulaciones, que nuestra propuesta de combinación de *tests* muestra importantes ganancias en poder y un correcto tamaño empírico. De hecho, el nuevo *test* supera en desempeño a sus componentes individuales mostrando ganancias en poder de hasta 45%. Finalmente ilustramos el uso de nuestro *test* con una aplicación empírica que examina la predictibilidad de los retornos de tipo de cambio chileno y canadiense.

Abstract

In this paper we focus on combining out-of-sample test statistics of the Martingale Difference Hypothesis (MDH) to explore whether a new combined statistic may induce a test with higher asymptotic power. Asymptotic normality implies that more power can be achieved by finding the optimal weight in a combined t-ratio. Unfortunately, this optimal weight is degenerated under the null of no predictability. To overcome this problem we introduce a penalization function that attracts the optimal weight to the interior of the feasible combination set. The new optimal weight associated with the penalization problem is well defined under the null, ensuring asymptotic normality of the resulting combined test. We show, via simulations, that our proposed combined test displays important gains in power and good empirical size. In fact, the new test outperforms its single components displaying gains in power up to 45%. Finally, we illustrate our approach with an empirical application aimed at testing predictability of Chilean and Canadian exchange rate returns.

I thank Juan Díaz for excellent assistance and Andrea Bentancor for useful comments. All errors are mine. E-mail: ppinchei@bcentral.cl.

1 Introduction

A vast 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. This condition is known as the Martingale Difference Hypothesis (MDH) and it is closely related to the efficient market hypothesis.

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) argue that this conflict relies upon the higher power of in-sample over out-of-sample strategies. According to this argument, out-of-sample tests of the MDH would fail to reject the null of no predictability mainly due to the low power of these tests. It seems advisable then to move into the direction of constructing new out-of-sample tests of the MDH displaying power improvements with respect to their competitors. Several authors have recently engaged in this endeavour[1]. Despite their efforts, simulations shown by Clark and West (2006) indicates that there is still plenty of room for improvement in this ground.

Another branch of the literature has entirely focused in predictive accuracy, sometimes without even conducting inference about predictive ability. Generally speaking, this litera-

[1] See for instance the papers by Clark and West (2006, 2007), Pincheira (2006), Anatolyev and Gerko(2005) and Clark and McCracken (2001), among others.

ture looks for improvement in forecast accuracy under a given loss function. One of the most striking results in this regard is related to the combination of forecasts. A number of papers show empirically how the combination of forecasts from different sources is useful to generate a combined forecast with improved predictive properties, see Bates and Granger (1969), Clemen (1989) and Wright (2003). In general it is possible to build a new forecast as a linear combination of a set of given forecasts, so that this new forecast displays lower out-of-sample mean square prediction error than any of its components. While the success of combination of forecasts is widely known in the forecasting literature, a thorough and satisfactory explanation for this result is yet to come. Interesting attempts are provided by Clements and Hendry (2004) Timmermann (2006) and Smith and Wallis (2005) for example.

Inspired by this literature, we focus here on combining statistics used for out-of-sample tests of the MDH to explore whether a "combined test" may yield power improvements. This question seems to be unexplored yet and particularly relevant for the application we are interested here, application in which a number of tests are available and power gains are still required.

The rest of the paper is organized as follows: Section 2 explains in some detail the problem we are addressing here and the econometric context that we use. In section 3 we develop the combination strategy we present in this paper. Section 4 describes the experimental design. Section 5 delivers the simulation results when deterministic weights are used for combination. Section 6 show results when quasi-optimal weights are used for combination. In section 7 we implement our combination strategy in an empirical application within the exchange rate literature. Section 8 concludes.

2 Econometric Context

We will use an environment similar to that in Clark and West (2006). Consider two simple models for a scalar stationary time series y_{t+1} :

$$Model\,1 \quad (null) \quad : \quad y_{t+1} = e_{t+1} \tag{1}$$

$$Model\,2 \quad (alternative) \quad : \quad y_{t+1} = X_{t+1}^{T}\beta + e_{t+1} \tag{2}$$

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

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. 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 :

$$E(y_{t+1}|\mathfrak{F}_t) = X_{t+1}^{T}\beta \tag{3}$$

For simplicity we will refer to the conditional expectation in (3) by $\widetilde{y}_{t+1}$. We will also impose the condition

$$m(e_{t+1}|\mathfrak{F}_t) = 0 \tag{4}$$

when needed. This condition says that the perturbations are also a zero median martingale process.

We are interested in testing the null hypothesis $H_0 : \beta = 0$ against a local alternative $H_A : \beta = \beta_0 \neq 0$. Let us consider a test statistic Ts assumed to be useful for this purpose. We will focus in two properties of the test: size and power.

2.1 Size

The size of a test is the probability of rejecting the null when the null is the true model. Therefore, it asses the probability of making a mistake. This mistake is called the type I error. In other words, it is said that the size of a test is α if

$$\Pr(Ts \in R(\alpha)|H_0) = \alpha$$

where $R(\alpha)$ is called the "rejection region of level α".

The econometrician usually does not know the exact distribution of the test statistic Ts. This distribution has to be estimated in some way. In this paper we will use as an estimate, the asymptotic distribution of the test statistic Ts.

An important distinction has to be made: the difference between the nominal size and the empirical size of a test. The nominal size is given by the approximation of the distribution of the test statistic Ts. In case the approximation is standard normal, for instance, a rejection region given by

$$R = (-\infty, 1.96) \cup (1.96, \infty)$$

will be associated to a nominal size of 5%.

In empirical applications, however, and assuming that the null model is the true model, the number of times that the test statistic Ts would fall inside the rejection region would be typically different from the nominal size. The empirical size of the test results from the empirical distribution of the test statistic under the null model. In other words, when the distribution of the test statistic Ts is approximated by the empirical distribution, the rejection region given by

$$R = (-\infty, 1.96) \cup (1.96, \infty)$$

will be associated to an empirical size given by.

$$\widehat{\Pr}(Ts \in R | H_0) = \alpha_E \neq \alpha$$

where the hat is used to emphasize that the probability is estimated according to the empirical distribution.

Ideally the nominal and empirical size of a test will coincide. As long as the empirical size is lower than the nominal size, the test will be called "undersized". As long as the empirical size is higher than the nominal size, the test will be called "oversized". The econometrician is always looking for a test with correct size, that is to say, a test for which both empirical and nominal size are the same.

2.2 Power

The power of a test is the probability of rejecting the null when the alternative is the true model. Therefore, it is a measure of the probability of succeeding in the detection of the alternative model. In other words, the power of a test against a local alternative H_A is given by

$$\Pr(Ts \in R(\alpha)|H_A) = Power$$

where $R(\alpha)$ is called the "rejection region of level α".

In most cases a test may have high power against a particular alternative, but will have low power against other alternatives. To be precise in these respects the econometrician usually refers to the local power of a test.

The ideal test would have power equal 1, and in general a test with high power will be preferable against another test with lower power, everything else being the same.

Power is an important property. It is possible to think that the low success in beating the random walk benchmark in the exchange rate literature, might be in part explained by the low of power of out-of-sample tests of predictive ability.

When comparing the power of two different tests, attention should also be placed on the empirical size of the tests. If one of the tests under evaluation is undersized and the other is correctly sized, for instance, the former test might mistakenly look like having lower power simply because the comparison being made is not at all fair. In other words we would be comparing power of two tests at different significance levels, 10% and 5% for instance, clearly leading to an unfair comparison. Sometimes researchers deal with this problem evaluating both "raw power" and "size-adjusted-power". While "raw power" refers to an empirical measure of the power of a test, "size- adjusted- power " refers to a measure of power when the size of a test is artificially fixed to a desired level. While "size-adjusted-power" is a measure of power ensuring a fair comparison, it is only available through simulations and distant from real empirical applications.

3 Test Combination

Combination of forecasts has been proven useful in the forecasting literature to outperform the random walk model in forecasting comparisons under 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. The basic idea of combination is well articulated by Timmermann (2006) and different explanations of the benefits of combined forecasts are found in Clements and Hendry (2004), Timmermann (2006) and Smith and Wallis (2005) for example.

In this section we give arguments in favor of the construction of a combined test. Under asymptotic normality we claim that our combination approach may create a new test with higher asymptotic power.

We will assume that we have available two test statistics that are sample analogs of a moment that may be written according to the general form:

$$E(g_1(y_t)\ g_2(\widehat{y}_t)) \tag{5}$$

where g_1 and g_2 are real functions and $\widehat{y}_t$ is the forecast of the variable y_t. We will also assume that under the null of no predictability (5) is zero, whereas under the alternative is positive. Different choices of g_1 and g_2 will define different test statistics. Having this in mind, let us now present the construction of our combined test.

3.1 Construction of the Combined Test

Let us assume that we have a bidimensional ergodic and stationary stochastic process given by

$$U_t = \begin{pmatrix} H_t \\ Z_t \end{pmatrix}_{t=-\infty}^{t=\infty}$$

such that

$$E(U_t) = \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} ; \quad V(U_t) = V = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} ; \sigma_1, \sigma_2 > 0$$

Let us also assume that $\{U_t - \mu\}_{t=-\infty}^{t=\infty}$ is a zero mean martingale difference sequence.

We are interested in the following statistics

$$T_1 = \frac{1}{P} \sum_{t=1}^{P} H_t \ ; \ T_2 = \frac{1}{P} \sum_{t=1}^{P} Z_t \tag{6}$$

The connection between (6) and (5) is given by

$$\begin{aligned} H_t &= g_1(y_t)\ g_2(\widehat{y}_t) \\ Z_t &= \widetilde{g}_1(y_t)\ \widetilde{g}_2(\widehat{y}_t) \end{aligned}$$

so that T_1 and T_2 are sample analogs of (5).

Using traditional central limit results for martingales, see Hamilton (1994) or White

(2001), we have that under some regularity conditions

$$P^{-\frac{1}{2}}\sum_{t=1}^{P}(U_t-\mu)\xrightarrow{A}N(0,V)$$

and therefore the marginal random variables satisfy

$$\begin{aligned}\sqrt{P}(T_1-\mu_1) &= P^{-\frac{1}{2}}\sum_{t=1}^{P}(H_t-\mu_1)\xrightarrow{A}N(0,\sigma_1^2)\\ \sqrt{P}(T_2-\mu_2) &= P^{-\frac{1}{2}}\sum_{t=1}^{P}(Z_t-\mu_2)\xrightarrow{A}N(0,\sigma_2^2)\end{aligned}$$

We are interested in testing the null hypothesis of no predictability. Under this null, we expect $\mu_1=\mu_2=0$, therefore, under the null

$$\begin{aligned}\sqrt{P}T_1 &= P^{-\frac{1}{2}}\sum_{t=1}^{P}H_t\xrightarrow{A}N(0,\sigma_1^2)\\ \sqrt{P}T_2 &= P^{-\frac{1}{2}}\sum_{t=1}^{P}Z_t\xrightarrow{A}N(0,\sigma_2^2)\end{aligned}$$

Under the alternative of predictability we expect $\mu_1>0$ and $\mu_2>0$ and therefore we could use the following approximation

$$\begin{aligned}\frac{1}{P}\sum_{t=1}^{P}H_t &\rightsquigarrow N(\mu_1,\frac{\sigma_1^2}{P})\\ \frac{1}{P}\sum_{t=1}^{P}Z_t &\rightsquigarrow N(\mu_2,\frac{\sigma_2^2}{P})\end{aligned}$$

Consider now the following combination $Y_t(\omega)=\omega H_t+(1-\omega)Z_t$ with $\omega\in(0,1)$, and the following combined statistic:

$$T^C(\omega) = \frac{1}{P}\sum_{t=1}^{P} Y_t(\omega) = \omega T_1 + (1-\omega)T_2$$

Then we have that $\{Y_t(\omega) - \mu(\omega)\}_{t=-\infty}^{t=\infty}$ is a zero mean martingale difference sequence as well, with $\mu(\omega) = \omega\mu_1 + (1-\omega)\mu_2 = \omega(\mu_1 - \mu_2) + \mu_2$. Besides

$$\begin{aligned} \sigma^2(\omega) &\equiv Var(Y_t(\omega) - \mu(\omega)) = \omega^2\sigma_1^2 + (1-\omega)^2\sigma_2^2 + 2\omega(1-\omega)\sigma_{1,2} \\ \sigma_{1,2} &= Cov(H_t, Z_t) \end{aligned}$$

Therefore we have

$$\sqrt{P}(T^C(\omega) - \mu(\omega)) = P^{-\frac{1}{2}}\sum_{t=1}^{P}(Y_t(\omega) - \mu(\omega)) \xrightarrow{A} N(0, \sigma^2(\omega))$$

Under the null hypothesis of no predictability we have $\mu_1 = \mu_2 = 0$, and then $\mu(\omega) = 0$ for all $\omega \in (0,1)$. Therefore, under the null

$$\sqrt{P}T^C(\omega) = P^{-\frac{1}{2}}\sum_{t=1}^{P} Y_t(\omega) \xrightarrow{A} N(0, \sigma^2(\omega)) \text{ for all } \omega \in (0,1)$$

Under the alternative of predictability we expect $\mu_1 > 0$ and $\mu_2 > 0$ and therefore $\mu(\omega) > 0$ for all $\omega \in (0,1)$. Under the alternative we could use the following approximation

$$\sqrt{P}T^C(\omega) = P^{-\frac{1}{2}}\sum_{t=1}^{P} Y_t(\omega) \rightsquigarrow N(\sqrt{P}\mu(\omega), \sigma^2(\omega))$$

which is equivalent to

$$\frac{1}{P}\sum_{t=1}^{P} Y_t(\omega) \rightsquigarrow N(\mu(\omega), \frac{\sigma^2(\omega)}{P})$$

In summary we have that

$$\begin{aligned} H_0 &: T^{CN}(\omega) = \sqrt{P}\frac{\overline{Y}_t(\omega)}{\sigma(\omega)} \xrightarrow{A} N(0,1) \\ H_A &: T^{CN}(\omega) = \sqrt{P}\frac{\overline{Y}_t(\omega)}{\sigma(\omega)} \approx N(\mu_{AN}(\omega),1) \\ \mu_{AN}(\omega) &= \sqrt{P}\frac{\mu(\omega)}{\sigma(\omega)} \end{aligned}$$

At this stage we have said nothing about the selection of the combination weight. One could use a random number $\omega \in (0,1)$ or one could try to pick an optimal weight according to some criteria. In the next subsection we will look for an optimal combination weight in a very specific sense.

3.2 Asymptotic Power Maximization

The idea is now to choose ω to maximize the asymptotic power of the combined test. This asymptotic power is given by the following expression:

$$P_\alpha(\mu_{AN}(\omega)|H_A) = \Pr(T^{CN}(\omega) > t_\alpha|H_A) \tag{7}$$

where α is the nominal size of the test and t_α the corresponding α-quantile.

The next proposition shows that under asymptotic normality our maximization problem is simple.

Proposition 1 *Under asymptotic normality, maximization of power in (7) is equivalent to* *maximize the following combined t-type statistic*

$$\mu_{AN}(\omega) = \frac{\mu(\omega)}{\sigma(\omega)} \tag{8}$$
$$\omega \in [0,1]$$

Proof. See the appendix. ■

The next proposition gives the solution to our maximization problem.

Proposition 2 *Let us assume that we have a bidimensional ergodic and stationary stochastic process given by*

$$U_t = \begin{pmatrix} H_t \\ Z_t \end{pmatrix}_{t=-\infty}^{t=\infty}$$

such that

$$E(U_t) = \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \geq 0; \quad V(U_t) = V = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \; ; \quad \sigma_1, \sigma_2 > 0$$

Let us also assume that $\{U_t - \mu\}_{t=-\infty}^{t=\infty}$ *is a zero mean martingale difference sequence. Furthermore, assume that*

$$\rho^2 = \left(\frac{\sigma_{12}}{\sigma_1 \sigma_2}\right)^2 < 1$$

and that

$$\mu_1(\sigma_{12} - \sigma_2^2) + \mu_2(\sigma_{12} - \sigma_1^2) \neq 0$$

Then, the solution to the maximization problem is given by

$$\overline{\omega} = \arg\max_{\omega \in \Psi} \frac{\mu(\omega)}{\sigma(\omega)}$$

$$\begin{aligned} \Psi &= \{1,0\} \cup \Gamma \\ \Gamma &= \begin{bmatrix} \{\omega^*\} \quad if \;\; \omega^* \in [0,1] \\ \\ \phi \;\; if \;\; \omega^* \notin [0,1] \end{bmatrix} \end{aligned}$$

with

$$\omega^* = \frac{\mu_2 \sigma_{1,2} - \mu_1 \sigma_2^2}{\mu_1(\sigma_{1,2} - \sigma_2^2) + \mu_2(\sigma_{1,2} - \sigma_1^2)} \tag{9}$$

Proof. See the appendix ■

Proposition 2 is giving us simple mathematical conditions to ensure the existence of a combination scheme that maximizes power of the combined test. It is also not hard to find conditions under which ω^* is the unique solution for the maximization problem. As we can see from (9), the optimal weight is a function of unknown parameters and hence its value is also unknown. Nevertheless, a consistent estimate of the optimal weight is readily available by replacing populations moments by sample consistent moments. Let $\widehat{\omega}$ be a consistent estimate of ω^*. We could consider the following statistic

$$\sqrt{P}(\widehat{\omega}, 1-\widehat{\omega}) * \begin{pmatrix} T_{1t} - \mu_1 \\ \\ T_{2t} - \mu_2 \end{pmatrix} \xrightarrow{A} N(0, \begin{pmatrix} \omega^* \\ \\ 1-\omega^* \end{pmatrix}^T V \begin{pmatrix} \omega^* \\ \\ 1-\omega^* \end{pmatrix}) \tag{10}$$

For the result in (10) to apply we need convergence in probability of $\widehat{\omega}$ towards ω^*. Notice, however, that for an interior solution ω^* to exist we have assumed that at least one of the parameters μ_1 and μ_2 is different from zero. This assumption is clearly violated under the null hypothesis which imposes the restriction that both μ_1 and μ_2 are zero. Therefore, under the null ω^* is not defined and the asymptotic distribution of the statistic in (10) may not exist, or, in case of existence, may not be normal. In fact, numerical simulations show that the asymptotic distribution exists but it is not standard.

In the next subsection we develop our strategy to deal with the fact that the optimal weight is not defined under the null hypothesis. Our approach introduces a penalization function and a new and different objective function to maximize. This different objective function is supposed to be a good approximation of (8). For this purpose we recall that we have assumed $\mu_1 \geq 0$ and $\mu_2 \geq 0$. Under these conditions we have that

$$\mu_{AN}(\omega) = \frac{\mu(\omega)}{\sigma(\omega)} \geq 0$$

so the maximization of $\mu_{AN}(\omega)$ is equivalent to the maximization of $(\mu_{AN}(\omega))^2$.

Details of our strategy to deal with the fact that the optimal weight is not defined under the null hypothesis follow next.

3.3 Quasi-Maximization of the Asymptotic Power

Instead of engaging in the calculation of the correct asymptotic distribution in (10) we propose to maximize an objective function that is slightly different from the square of (8) and then to check the appropriateness of the approximation. In order to do so we introduce the following penalization function Pf:

$$\begin{aligned} Pf(w) &= \frac{1-(w^2+(1-w)^2)}{\sigma^2(\omega)} \\ &= \frac{2(w-w^2)}{\sigma^2(\omega)} \end{aligned}$$

which is depicted in figure 1. We notice that this penalization function is continuous and concave. Furthermore Pf is positive in the open set (0,1), it is equal to zero in the boundaries of the open set (0,1) and negative elsewhere. As we will see, these are important and useful properties. To see the importance of these properties we need to briefly revisit the theory behind the introduction of penalization functions. Let us assume that we are trying to maximize an objective function h over a feasible set F. This problem can be simply written as follows

$$\max_{x \in F} h(x) \tag{11}$$

In general, solving an optimization problem with restrictions is harder than solving an optimization problem without restrictions. It would be desirable then to transform a restricted problem into a problem without restrictions. Unfortunately, this is in general not possible. It is possible, however, to approximate the solution of a restricted problem by a sequence of solutions of unrestricted problems. These unrestricted problems need to "punish" the objective function h when taking on values outside the feasible set F, otherwise there is no guarantee that the sequence of solutions of the unrestricted problems will converge to a point inside the feasible region F. We "punish" h by adding a penalization function which is typically positive within the boundaries of the feasible region and negative outside. Let us call, for expositional purposes only, this penalization function by Ph. Instead of solving the "difficult" problem

(11) we could solve the following sequence of “easy” problems:

$$\max h(x) + \lambda_n P h(x) \tag{12}$$

where λ_n is a sequence of positive numbers converging to zero. Problem (12) is simpler than (11) in the sense that it has no restrictions whatsoever. Besides, if $\lambda_n > 0$ is small enough we could use the corresponding solution of (12) to approximate the solution of (11).

Inspired by this interesting approach we propose to approximate the solution of our problem

$$\max_{\omega \in [0,1]} (\mu_{PAN}(\omega))^2 = \frac{\mu^2(\omega)}{\sigma^2(\omega)} \tag{13}$$

by the solution of

$$\max_{\omega \in R} \frac{\mu^2(\omega)}{\sigma^2(\omega)} + \lambda P f(w) \tag{14}$$

When $\lambda = 0$ the objective functions of (13) and (14) are exactly the same. When λ is small, the problem in (14) is “close” to (13) and the penalization function makes costly for the solution in (14) to take values outside the feasible region in (13) by adding an extra negative burden. Figure 2 displays, for a given set of parameters including $\lambda = 0.05$, the proximity of the objective functions in (13) and (14). Notice that under the null $\mu^2(\omega) = 0$, therefore problem (14) would be solving the maximization of a non zero function, whereas the objective function in (13) would be exactly zero. This fact would allow the solution of (14) to be well defined even under the null. On the contrary, as we emphasized previously, the solution of (13) is not defined under the null. We will evaluate via simulations the usefulness of our approach.

Figure 1
Penalization Function

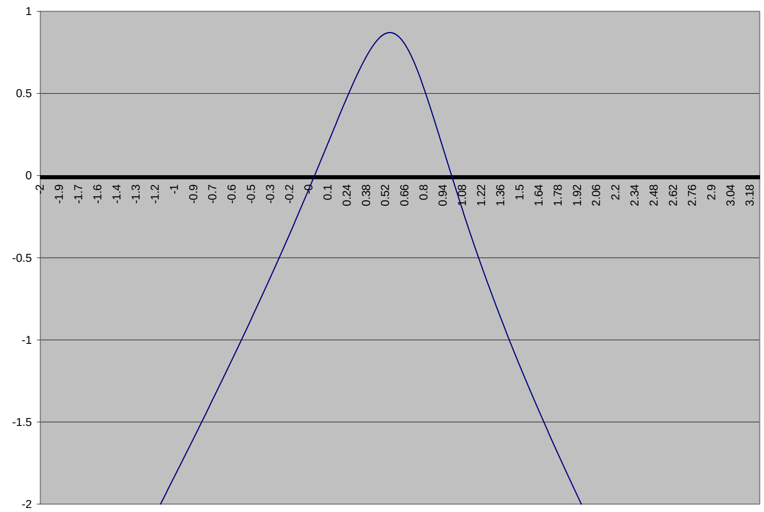
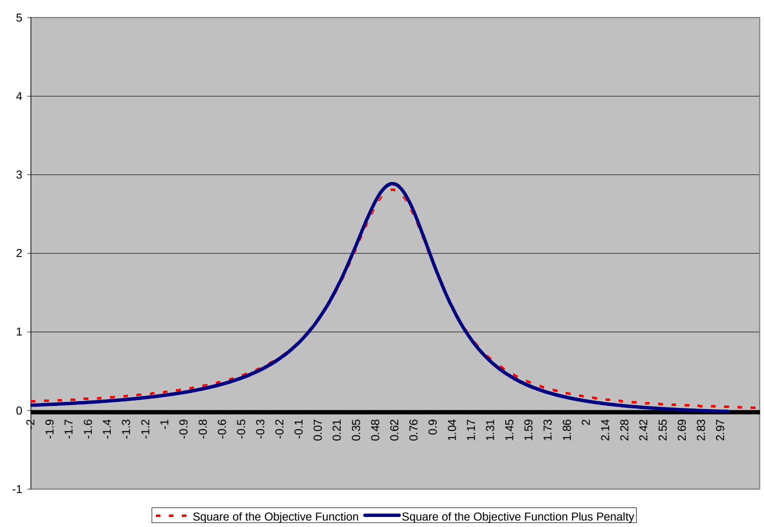


Figure 2
Objective Functions



The next proposition shows that only a mild assumption is required for all possible solutions of (14) to be well defined even when the null is true. This means that they are useful to construct an asymptotically normal combined test.

Proposition 3 *The critical points of*

$$\mu_{PAN}(\omega) = \frac{\mu^2(\omega)}{\sigma^2(\omega)} + \lambda Pf(w)$$
$$\lambda > 0$$

are given by

$$\omega_{1,2} = \frac{[(c^2-2\lambda)\sigma_2^2 - \sigma_T^2\mu_2^2] \pm \sqrt{[(c^2-2\lambda)\sigma_2^2 - \sigma_T^2\mu_2^2]^2 - 4d(\lambda)}}{2\left[\left[(\mu_1\mu_2 - \mu_2^2 + \lambda)\sigma_T^2\right] - (c^2-2\lambda)[\sigma_{12} - \sigma_2^2]\right]} \tag{15}$$

where

$$d(\lambda) = \left[(\mu_1\mu_2 - \mu_2^2 + \lambda)\sigma_T^2 - (c^2-2\lambda)[\sigma_{12} - \sigma_2^2]\right]\left[\mu_2^2\sigma_{12} - [\mu_1\mu_2 + \lambda]\sigma_2^2\right]$$
$$c = \mu_{1-}\mu_2$$
$$\sigma_T^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$$

In particular, and provided that $\sigma_1^2 \neq \sigma_2^2$ *, under the null these solutions reduce to*

$$\begin{matrix} \omega_1 \\ \omega_2 \end{matrix} = \begin{bmatrix} \frac{\sigma_2}{\sigma_1+\sigma_2} \\ \frac{\sigma_2}{\sigma_2-\sigma_1} \end{bmatrix}$$

Proof. Straightforward. ■

It is simple to check, under the null, that among all possible solutions of (14), ω_1 is the unique solution. This can be seen by noticing that

$$\begin{aligned} 0 &< \omega_1 = \frac{\sigma_2}{\sigma_1+\sigma_2} < 1 \\ \omega_2 &= \frac{\sigma_2}{\sigma_2-\sigma_1} > 1 \end{aligned}$$

therefore

$$\begin{aligned} \mu_{PAN}(\omega_2) &= \mu_{PAN}(\frac{\sigma_2}{\sigma_2-\sigma_1}) = \lambda Pf(\frac{\sigma_2}{\sigma_2-\sigma_1}) < 0 \\ \mu_{PAN}(\omega_1) &= \mu_{PAN}(\frac{\sigma_2}{\sigma_1+\sigma_2}) = \lambda Pf(\frac{\sigma_2}{\sigma_1+\sigma_2}) > 0 \end{aligned}$$

These last inequalities are obtained from the fact that the penalization function Pf is negative outside the interval [0,1] and it is positive within this interval.

Expression (15) shows us that ω_1 and ω_2 are continuous functions of (μ_1, μ_2). Given that ω_1 is the solution of (14) when $(\mu_1, \mu_2) = (0, 0)$, we expect ω_1 to be also the solution of (14) in a neighborhood around $(\mu_1, \mu_2) = (0, 0)$. We will use ω_1 as our proxy for the solution of the original problem (8). We will call this solution the quasi-optimal weight.

In empirical applications the econometrician may proceed in two different directions. As proposition 3 suggests, the econometrician may prefer to estimate the quasi-optimal weight and then to test the null using the estimated combined test. A different direction might be taken as well. Instead of looking for the quasi-optimal combination, the researcher might be interested in using a set of deterministic weights to construct a combined test to test the null. In case of rejection, the econometrician may well argue that these combined tests are particularly powerful in the direction of the relevant alternative hypothesis. The next proposition shows that this approach will not hurt.

Proposition 4 *Consider a bidimensional ergodic and stationary stochastic process given by*

$$U_t = \begin{pmatrix} H_t \\ \\ Z_t \end{pmatrix}_{t=-\infty}^{t=\infty}$$

defined as in proposition (2). Then

$$\begin{aligned} \mu_{AN}(\omega) &\geq \min\left\{\frac{\mu(1)}{\sigma(1)}, \frac{\mu(0)}{\sigma(0)}\right\} \\ \omega &\in (0,1) \end{aligned}$$

Proof. See the appendix ■

This last proposition is telling us that by using a set of weights that are not optimal we cannot do worse than just looking at the worst single test. In this sense combining is relatively costless.

As a final comment we see no further difficulty in extending the present analysis to a more general context. For instance, although it seem tedious, the extension to a general expression for the combination of an arbitrary finite number of tests seems theoretically straightforward.

3.4 A Couple of Caveats

Empirical size of the combined and single tests may differ. This difference should not be an important problem due to the fact that a combination of close-to-zero terms should remain close to zero. In other words, we should not expect important size distortions induced by combination. Nevertheless, it is important to remark that size properties may be different depending on the combination strategy. For instance, a combination strategy based upon the solution of an optimization problem like (14) could in principle induce higher size on the combined test compared to a combination scheme based upon the random choice of a deterministic weight. In fact, optimal weights, like those derived in previous sections, would tend to increase the objective function in (13) for both the null and alternative models. On the contrary, a random choice of a deterministic weight is expected to boost the objective function of (13) either for the null or for the alternative, but not necessarily for both models. In any case, simulations reported below, indicate that the size of the combined test is, in general, adequate.

For simplicity we made the assumption of a martingale difference sequence. This assumption is not strictly necessary, and our results may be extended to more general ergodic and stationary processes provided that the correct long run variances of the processes are considered.

4 Experimental Design

Following Clark and West (2006) we use Monte Carlo simulations based upon variations of a multivariate Data-Generating-Process (DGP) to compare small sample properties of our combined tests with those of their components. We consider three individual tests: the MSPE-Adjusted test proposed by Clark and West (2006), a Direction of Change (DC) test originally proposed by Diebold and Timmermannn (1992) and an Excess Profitability test (EP) proposed by Anatolyev and Gerko (2005). The combined tests we consider are the three possible pairwise combinations. Combination 1 is a convex linear combination of MSPE-

Adjusted and DC, Combination 2 is a convex linear combination of MSPE-Adjusted and EP, whereas Combination 3 is a convex linear combination of DC and EP. A brief overview of the three individual tests used for the combined tests follows next.

4.1 The Tests

- **MSPE-Adjusted**

The MSPE-Adjusted test was proposed by Clark and West (2006). They derive this test from typical comparisons of MSPE between a linear model and the null of a martingale difference sequence. Clark and West claim that the MSPE-Adjusted test has better size than traditional tests of MSPE comparisons when the models under evaluation are nested. 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. To see this we notice that the sample analog of the difference in MSPE between the two models considered in (1) and (2) is given by

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

Clark and West notice that the second term in the right hand side introduces a bias that does not vanish as P goes to ∞. They propose to build a test based upon the first term in the right hand side. Their test is given by

$$MSPE - Adjusted : P^{1/2} \frac{\frac{2}{P}\sum_{t=R}^{T} y_{t+1} X_{t+1}^{T} \widehat{\beta}_t}{\sqrt{4\widehat{V}(y_{t+1} X_{t+1}^{T} \widehat{\beta}_t)}}$$

- **Direction of Change Test**

This test was originally proposed by Diebold and Timmermannn (1992). It is a sign test aimed at evaluating the direction of change in the price of an asset. Following Pincheira (2006) we will use an adaptation of this test based upon the following statistic:

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

The test is given by

$$DC : P^{1/2} \frac{\frac{1}{P}\sum_{t=R}^{T} sign(y_{t+1} X_{t+1}^{T} \widehat{\beta}_t)}{\sqrt{\widehat{V}(sign(y_{t+1} X_{t+1}^{T} \widehat{\beta}_t))}}$$

- **The Excess Profitability Test**

This test was originally introduced by Anatolyev and Gerco (2005) as follows

$$EP \equiv \frac{A_T - B_T}{\sqrt{\widehat{V}(A_T - B_T)}} \quad (16)$$

$$A_T = \frac{1}{P} \sum_{t=R+1}^{T+1} r_t \quad (17)$$

$$B_T = \left(\frac{1}{P} \sum_{t=R+1}^{T+1} sign(\widehat{y}_t) \right) \left(\frac{1}{P} \sum_{t=R+1}^{T+1} y_t \right) \quad (18)$$

where

$$r_t = sign(\widehat{y}_t)y_t \tag{19}$$

The intuition behind the definition of the EP test relies on the fact that under the null of no predictability both A_T and B_T converge in probability to the same value, so they are asymptotically equal. r_t represents returns of the following trading rule: if the forecast is that the price of the asset will go up, then a buy signal is issued. If the forecasts is that the price of the asset will go down, then a sell signal is issued.

Under mild conditions, all three tests previously presented are asymptotically normal.

4.2 Data Generating Process

We use a DGP following Clark and West (2006). This DGP is calibrated to match common features of exchange rate series for which the martingale difference is a sensible null hypothesis. The DGP can be described as follows:

$$\begin{aligned} y_{t+1} &= \beta x_t + e_{t+1} \\ x_t &= 0.95x_{t-1} + u_t; \\ e_{t+1} &= t(\nu); \quad u_{t+1} = N(0, \sigma_u^2) \end{aligned}$$

with $E(e_{t+1}|\mathfrak{F}_t) = 0$, $E(u_{t+1}|\mathfrak{F}_t) = 0$ and $var(e_{t+1}) = 1$.

Our DGP is calibrated to match exchange rate features based on interest parity so we will have $var(u_t) = \sigma_u^2$ (with $\sigma_u = 0.025$) and $corr(e_t, u_t) = 0$. We set $\beta = -2$ in experiments evaluating power[2] and $\beta = 0$ in experiments evaluating size. We assume that e_{t+1} has a $t(\nu)$

[2] We are aware that $\beta = 1$ is the theoretical implication. Empirical estimations for a number of industrial countries, however, provide estimates around -2. See Clark and West (2006) for further details and Obstfeld and Rogoff (2002) for a thorough coverage of the forward premium puzzle.

distribution (t-student with ν degrees of freedom). Our central scenario sets $\nu = 6$ to simulate shocks with fat tails like those displayed by exchange rate returns. We also consider lower values of ν to asses the behavior of our tests as tails get thicker. We assume data generated from homoskedastic draws from their distributions.

We focus, for simplicity, only on one step ahead forecasts. One has a total of $T + 1$ $= P + R$ observations. The last P observations are used for predictions and R are used for the initial estimation of the vector of parameters. β_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 β_t is estimated only once using the first R observations. The rolling scheme updates the estimate of β_t using the last R observations. The recursive scheme also updates the estimate of β_t but using all available information until time t. That is to say, in the recursive scheme the estimation sample increases with t. Following Clark and West (2006) we will work with the rolling scheme which is particularly appropriate when one work with series that may have experienced breaks.

Estimation always includes a constant term in each regression. We explore the performance of our tests for a number of sample sizes $(T + 1)$ and decompositions of the sample into the estimation window (size R) and the prediction window (size P). We run simulations for the following sample sizes: $R = 100$ and 200; $P = 100, 150, 200$ and 250.

4.3 Experiments

We consider two major exercises. First we take a grid of possible deterministic combination weights ($\omega \in [0, 1]$). For every single weight ω in the grid we construct the three combined tests described previously. Then we compare, via Monte Carlo simulations, power and size adjusted power of the combined tests to power and size adjusted power of their components. We also pay attention to the empirical size of the combined tests. Results for the best combinations on average are displayed in tables in the following section.

Second, we evaluate the performance of the 3 combinations using 3 different specific weights: the quasi-optimal weight for the problem with a penalty function ($\omega = \omega_1$), the remaining root for the problem with a penalty function ($\omega = \omega_2$) and the simple mean ($\omega_3 = 0.5$). Results for all three combinations and all three weights are analyzed in terms of power, size adjusted power and empirical size.

5 Results with Deterministic Weights

In this section we discuss the main results of our experiments with deterministic weights. To motivate this section we first show three graphs displaying results on power, size-adjusted-power and empirical size for the MSPE-Adjusted test, the Direction of Change test and the combination of them using a number of deterministic weights. In the horizontal axis we have all the different weights used for combination. For each possible combination weight we run 1000 replications of the DGP to calculate empirical power, size-adjusted-power and empirical size of the tests. A weight $\omega = 0$ means that the combined test equals the MSPE-Adjusted test. On the contrary, $\omega = 1$ means that the combined test equals the Direction of Change test.

Figure 3 displays results on power, figure 4 displays results on size-adjusted-power and figure 5 on empirical size. Graphs correspond to simulation results obtained for a nominal size of 10%, an estimation window of size $R = 100$, a prediction window of size $P = 250$ and a parameter ν set at $\nu = 6$.

Figure 3
Power (10%)
(P=250, R=100, and t(nu=6))

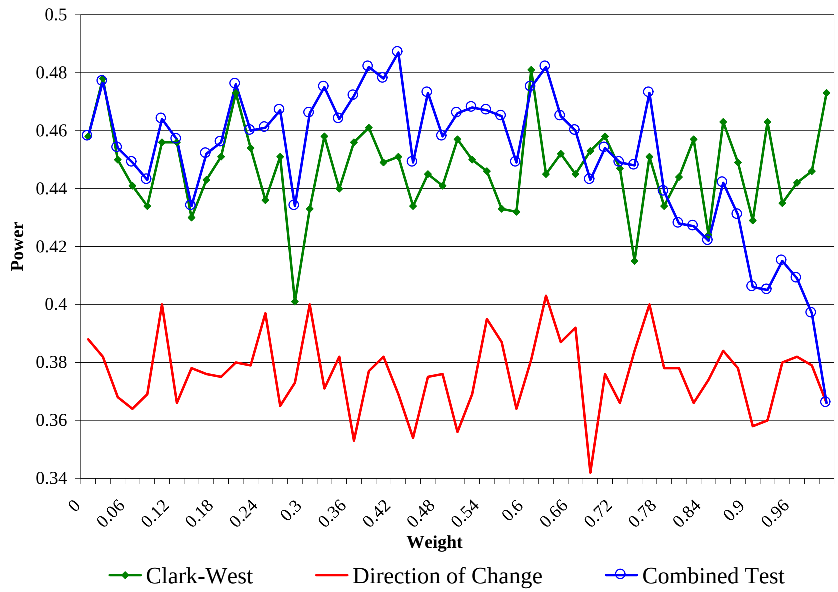


Figure 4
Size-Adjusted-Power (10%)
(P=250, R=100, and t(nu=6))

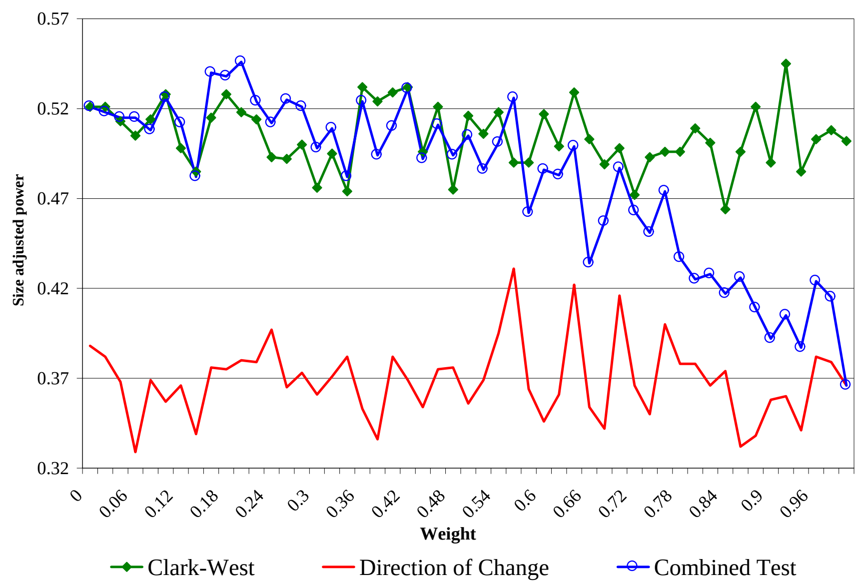
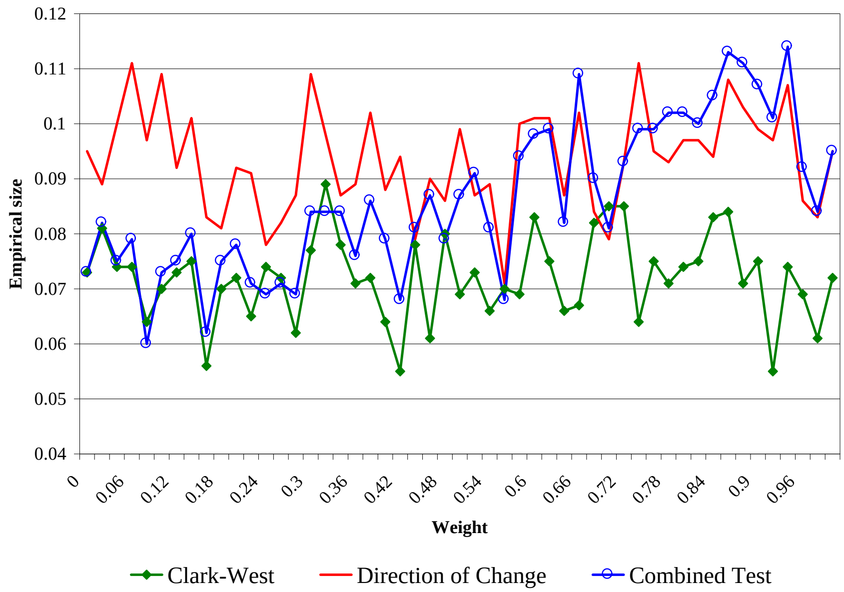


Figure 5
Empirical Size (10%)
(P=250, R=100, and t(nu=6))



From figures 3-4 we notice that the MSPE-Adjusted test has higher power and size-adjusted-power than the direction of change test, at least for the alternative model we are interested here. In terms of empirical size, the direction of change test is, in general, better sized than the MSPE-Adjusted test which is a little undersized.

Figure 3 shows a wide range of weights for which the combined test displays higher empirical power than both single tests. Actually, it is possible to obtain power gains for weights lower than 0.8. For higher weights the combined test is outperformed by the MSPE-Adjusted test. Figure 4 reveals that for a smaller region, the combined test displays higher size-adjusted-power as well. Finally, figure 5 shows that the empirical size of the combined test is adequate. General results, for all the experiments, are shown in tables in the next subsection.

5.1 Results on Power

In this subsection we show tables summarizing percentage gains in power and size-adjusted-power for a number of different parameter values. In the left panel of the tables we report gains associated to a nominal size of 5% whereas in the right panel we show gains associated to a nominal size of 10%. We see that power gains are far from negligible. Table 1 shows that percentage power gains range from 0% to 26.8%. On average, the combination of the MSPE-Adjusted and the Direction of Change tests outperforms other combinations. Interestingly, on average percentage gains are higher at the 5% rather than at the 10% significance level. In only two cases there are no gains whatsoever, that is to say, in most cases combination is fruitful.

Table 1
Percentage Power Gains

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Parameters			Size 5%			Size 10%		
			Combination 1	**Combination 2**	**Combination 3**	**Combination 1**	**Combination 2**	**Combination 3**
R	P	df	*(MSPE-Adj/DC)*	*(MSPE-Adj/EP)*	*(DC/EP)*	*(MSPE-Adj/DC)*	*(MSPE-Adj/EP)*	*(DC/EP)*
100	100	6	26.8%	7.5%	6.0%	12.2%	6.8%	8.7%
100	150	6	12.9%	6.0%	8.8%	10.6%	6.6%	7.1%
100	200	6	12.9%	7.2%	8.4%	10.1%	6.4%	7.8%
100	250	6	10.3%	8.0%	9.0%	8.3%	4.3%	4.8%
100	100	5	18.9%	7.8%	15.2%	14.3%	5.6%	10.2%
100	150	5	16.3%	7.0%	10.5%	11.2%	5.9%	6.5%
100	200	5	12.3%	7.4%	13.5%	12.0%	7.5%	6.1%
100	250	5	14.2%	7.3%	8.0%	11.2%	5.1%	8.9%
100	100	2	17.6%	9.6%	12.0%	4.5%	3.0%	6.2%
100	150	2	17.5%	10.0%	13.7%	0.0%	6.9%	0.0%
100	200	2	4.9%	6.2%	0.0%	13.1%	7.0%	10.1%
100	250	2	18.1%	11.5%	12.6%	10.3%	5.9%	5.3%
200	100	6	14.1%	7.6%	10.4%	9.0%	5.1%	4.7%
200	150	6	13.7%	8.5%	8.7%	8.9%	6.5%	3.9%
200	200	6	9.7%	4.3%	9.2%	8.8%	3.7%	6.3%
200	250	6	8.0%	3.9%	7.1%	7.6%	3.6%	6.7%
averages			**14.3%**	**7.5%**	**9.6%**	**9.5%**	**5.6%**	**6.4%**

Notes:

1. R is the size of the estimation window.
2. P is the size of the prediction window.
3. df means degrees of freedom.

5.2 Results on Size-Adjusted-Power

Table 2 below shows that combination yields even more important gains in terms of size-adjusted-power. Gains range from 0% to 37.1%. On average, the combination of the Excess Profitability test and the Direction of Change test outperforms other combinations with an average gain of 17.7% and 12.6% at the 5% and 10% significance levels, respectively. Again, percentage gains are higher at the 5% rather than at the 10% significance level. Now, in only one case there is no gain whatsoever, that is to say, in most cases combination is fruitful.

Table 2
Percentage Size-Adjusted-Power Gains

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Parameters			Size 5%			Size 10%		
			Combination 1	**Combination 2**	**Combination 3**	**Combination 1**	**Combination 2**	**Combination 3**
R	P	df	*(MSPE-Adj/DC)*	*(MSPE-Adj/EP)*	*(DC/EP)*	*(MSPE-Adj/DC)*	*(MSPE-Adj/EP)*	*(DC/EP)*
100	100	6	14.0%	3.6%	20.5%	9.4%	2.4%	7.2%
100	150	6	11.7%	3.4%	16.2%	6.2%	4.2%	8.3%
100	200	6	11.6%	3.8%	21.7%	7.6%	4.0%	12.0%
100	250	6	5.9%	2.6%	17.2%	7.3%	4.0%	7.1%
100	100	5	15.1%	3.4%	16.7%	8.4%	3.7%	11.1%
100	150	5	10.0%	5.1%	18.7%	11.9%	7.5%	10.3%
100	200	5	19.1%	9.7%	12.0%	8.2%	3.5%	15.9%
100	250	5	8.1%	5.4%	13.6%	11.2%	1.9%	11.5%
100	100	2	37.1%	6.8%	32.7%	22.6%	7.5%	30.1%
100	150	2	28.2%	19.5%	17.9%	22.3%	4.9%	19.8%
100	200	2	32.3%	8.0%	18.7%	20.1%	7.5%	19.4%
100	250	2	23.1%	5.2%	12.8%	16.8%	5.6%	11.5%
200	100	6	8.3%	4.4%	14.2%	4.5%	0.0%	12.7%
200	150	6	4.4%	2.4%	19.0%	4.9%	0.4%	8.7%
200	200	6	6.2%	4.5%	14.3%	5.2%	2.3%	8.8%
200	250	6	3.6%	0.0%	16.8%	4.3%	2.3%	7.9%
averages			**14.9%**	**5.5%**	**17.7%**	**10.7%**	**3.9%**	**12.6%**

Notes: See table 1.

5.3 Results on Size

Table 3 displays the empirical size of all three combined tests when combined at specific weights. We choose weights such that gains in power are maximized on average. In other words we pick the same weights used in table 1. The important feature of this table is to show that power gains are obtained from a correctly sized test. As we can see from table 3 below, all three combined tests display empirical sizes very close to the nominal size. On average, combination 1 and 2 look a little undersized whereas combination 3 is roughly correctly sized. Therefore we see that our combination strategy allows for gains in power that are not associated to size distortions.

Table 3
Empirical Size at Combinations

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Parameters			Size 5%			Size 10%		
			Combination 1	**Combination 2**	**Combination 3**	**Combination 1**	**Combination 2**	**Combination 3**
R	P	df	*(MSPE-Adj/DC)*	*(MSPE-Adj/EP)*	*(DC/EP)*	*(MSPE-Adj/DC)*	*(MSPE-Adj/EP)*	*(DC/EP)*
100	100	6	3.50%	4.50%	5.60%	10.10%	8.20%	10.40%
100	150	6	3.30%	3.30%	5.50%	7.90%	6.70%	9.90%
100	200	6	5.10%	3.00%	5.10%	6.60%	7.70%	8.70%
100	250	6	3.70%	4.70%	4.40%	9.90%	9.60%	9.60%
100	100	5	3.90%	4.20%	5.20%	8.20%	7.80%	10.30%
100	150	5	3.80%	4.70%	4.70%	8.60%	8.30%	9.40%
100	200	5	2.50%	3.60%	4.20%	7.30%	8.80%	12.00%
100	250	5	4.10%	4.50%	4.70%	7.80%	8.20%	10.10%
100	100	2	5.80%	4.20%	5.60%	9.40%	10.30%	9.70%
100	150	2	4.70%	3.90%	5.00%	9.80%	11.50%	9.80%
100	200	2	3.40%	3.80%	4.10%	9.70%	9.70%	9.70%
100	250	2	5.10%	3.80%	5.20%	6.90%	8.60%	10.40%
200	100	6	3.80%	5.30%	5.10%	8.70%	9.20%	10.20%
200	150	6	3.80%	4.60%	4.40%	6.80%	7.40%	10.60%
200	200	6	4.40%	3.10%	4.80%	9.50%	7.90%	9.50%
200	250	6	4.30%	3.20%	6.00%	7.10%	9.00%	10.50%
averages			**4.08%**	**4.03%**	**4.98%**	**8.39%**	**8.68%**	**10.05%**

Notes: See table 1.

6 Results with Quasi-Optimal Weights

In this subsection we show tables summarizing the performance of combined tests built with three different combination weights: the optimal weight for the problem with a penalization function (denoted by $w1$ and called quasi-optimal weight), the remaining root for the problem with a penalization function (denoted by $w2$) and the simple mean[3] (denoted in tables by $w3$). We run 5.000 simulations for $R = 100$ and 200; $P = 100, 150, 200$ and 250 and for $\nu = 6, 5$ and 2. We compute empirical size, percentage gains in power and percentage gains in size-adjusted-power for all three combinations and all three weights. In tables 4-5 we report averages and extreme values across all the exercises. Table 4 displays results for a nominal size of 5% whereas table 5 displays the results for a nominal size of 10%.

In terms of power, we see that the quasi-optimal weight ($w1$), displays the highest average gain for all combinations. Average gains are always positive and far from negligible ranging from 5.8% to 13.4%. We also would like to emphasize the highest gain of 45.2% obtained for this weight. The second best weight on average corresponds to the simple average ($w3$). The worst case is given by the remaining critical point of (14) which clearly shows that it does not correspond to a local maximum.

Results are less impressive in terms of size-adjusted-power. The best combination weight is, again, the quasi-optimal weight ($w1$). We further notice that positive gains on average are only obtained for Combinations 1 and 2, and they are below 5%. Highest gains, however, are in general important, peaking at 25.5%. The worst outcome on average is again displayed by the remaining critical point of (14).

In terms of empirical size, we notice that the quasi-optimal weight and the simple average induce tests with adequate size. In particular we notice that the quasi-optimal weight induces a correctly sized test for most of the simulations. Only for Combination 3 we detect a mild tendency to have an oversized test. Finally, the remaining critical point of (14) induces

[3] We worked with a penalty factor $\lambda = 0.05$.

important size distortions for some combinations. For instance, we see in table 5 that instead of the nominal size of 10%, Combination 2 with weight $w2$ displays an average empirical size of only 4.6%.

In summary we see that the quasi-optimal combination ($w1$) induces a test displaying important gains in power and adequate size, which proofs this type of combination fruitful.

Table 4
Small Sample Properties of Combined Tests with Quasi-Optimal Weights

Nominal Size is 5%

Properties of the Tests	Combination 1 *(MSPE-Adj/DC)*			Combination 2 *(MSPE-Adj/EP)*			Combination 3 *(DC/EP)*		
	w1	w2	w3	w1	w2	w3	w1	w2	w3
Results on Power									
Average Gains	10.0%	-79.7%	0.8%	13.4%	-82.1%	0.9%	9.9%	-77.9%	-1.3%
Highest Gain	27.0%	-51.0%	14.9%	45.2%	-77.5%	6.6%	30.1%	-56.7%	6.5%
Lowest Gains	2.4%	-92.1%	-17.0%	1.5%	-86.0%	-3.6%	3.7%	-90.4%	-21.3%
Results on Size Adjusted Power									
Average Gains	1.5%	-76.9%	-2.3%	-6.0%	-64.5%	-10.0%	2.9%	-74.7%	-0.7%
Highest Gain	25.5%	-54.5%	6.6%	-1.9%	-43.4%	5.8%	16.4%	-47.3%	7.4%
Lowest Gain	-9.8%	-87.5%	-12.5%	-7.5%	-73.9%	-14.9%	-4.8%	-89.1%	-23.5%
Results on Size									
Average Size	4.3%	3.5%	4.0%	4.8%	1.9%	4.4%	5.6%	4.5%	5.0%
Highest Size	5.1%	6.3%	4.5%	6.5%	2.4%	4.9%	5.9%	5.0%	5.4%
Lowest Size	3.9%	2.3%	3.4%	3.9%	1.1%	3.8%	5.3%	3.7%	4.0%

Table 5
Small Sample Properties of Combined Tests with Quasi-Optimal Weights

Nominal Size is 10%

Properties of the Tests	Combination 1 *(MSPE-Adj/DC)*			Combination 2 *(MSPE-Adj/EP)*			Combination 3 *(DC/EP)*		
	w1	w2	w3	w1	w2	w3	w1	w2	w3
Results on Power									
Average Gains	5.8%	-72.2%	0.6%	9.0%	-74.3%	1.4%	5.9%	-69.2%	-1.2%
Highest Gain	14.3%	-37.9%	10.0%	27.5%	-69.1%	5.6%	19.6%	-40.4%	4.9%
Lowest Gains	-7.0%	-86.7%	-26.5%	1.7%	-77.4%	-1.8%	-4.9%	-84.6%	-23.2%
Results on Size Adjusted Power									
Average Gains	1.1%	-68.4%	-1.7%	-3.7%	-55.6%	-7.2%	2.9%	-66.2%	-0.5%
Highest Gain	16.4%	-45.8%	6.6%	0.3%	-35.8%	1.4%	7.9%	-33.1%	6.4%
Lowest Gain	-6.0%	-79.2%	-7.7%	-8.7%	-63.4%	-15.1%	-4.1%	-84.3%	-17.7%
Results on Size									
Average Size	8.7%	7.4%	8.4%	9.3%	4.6%	9.0%	10.7%	9.4%	10.1%
Highest Size	10.3%	12.1%	9.1%	12.6%	5.4%	9.7%	11.8%	10.3%	10.6%
Lowest Size	7.7%	4.7%	7.7%	7.8%	3.8%	8.2%	10.0%	8.5%	9.6%

7 Empirical Application

In this section we study the behavior of our combining strategies using monthly forecasts for a couple of US dollar bilateral exchange rates. We analyze the cases of Canada and Chile. 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. The data from Canada was generously provided by Todd Clark and correspond to the same database used in Clark and West[4] (2006). We obtained the data for Chile from the International Financial Statistics. In this case we use the discount rates as measures of interest rates.

Using rolling regressions estimated by OLS we engage in the following empirical exercise: 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 (2006) to choose R relatively small with respect to P, so we choose $R \approx P/3$. For Canada we set $R = 95$ and $P = 191$. For Chile we set $R = 48$ and $P = 96$. Then we compute the MSPE-Adjusted test, the Direction of Change test and three different weights to combine these two test statistics. We use the same three weights denoted by $w1 - w3$ in previous sections. We then analyze whether these tests are able to reject the null of a MDH.

Table 6
Forecasts of Monthly Changes in U.S. Dollar Exchange Rates
Standardized Statistics

[4] 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.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Combinations		
Country	Sample Size	MSPE Adjusted	Direction of Change	Quasi-optimal weight (w1)	Remaining Root (w2)	w3= 1/2
Canada	R=95, P=191	*1.53**	*1.52**	*1.73***	*-0.02*	*1.68***
Chile	R=48, P=96	*1.23*	*1.22*	*1.35**	*0.01*	*1.23*

Notes:

1. Rejections at 5% (**) and 10% (*) level of significance.
2. Data range: 1980:01-2003:10 for Canada and 1993:4-2005:4 for Chile.
3. R is the size of the estimation window, P is the size of the prediction window.
4. Table display standardized statistics to be compare with critical values from a standard normal distribution.

Results are displayed in table 6. We clearly see how some combinations provide higher standardized statistics. Whereas in the case of Canada both the MSPE-Adjusted test and the Direction of Change test rejects the null of no predictability only at the 10% significance level, combinations with $w1$ and $w3$ allow rejection at the 5% significance level. In the case of Chile combination is even more helpful. In fact, none of the single tests, nor the simple average of them, are able to reject the null of no predictability at the 10% significance level, yet the quasi-optimal weight induces a test that does reject the null at a 10% significance level, providing evidence of predictability for the Chilean monthly exchange rate returns.

8 Conclusion

We have shown that the popular combination principle, which is used extensively in the forecasting literature, can be successfully extended to boost power in asymptotically normal tests of predictive ability.

Asymptotic normality implies that more power can be achieved simply by finding a combination that maximizes a combined t-ratio. This allows us to define an optimal combination weight. Unfortunately, this optimal weight is degenerated under the null of no predictability. To overcome this problem we introduce a penalization function that attracts the optimal weight to the interior of the feasible combination set. The new optimal weight associated with the penalization problem, that we call quasi-optimal weight, is well defined under the null, ensuring asymptotic normality of the resulting combined test

Using a simple data generating process of exchange rate returns based upon a model of interest parity, we show via simulations that the proposed quasi-optimal weight induces a test with adequate size and improved power. In fact, the new combined test may outperform its single components displaying gains in power up to 45%. We also show that this combination strategy, in general outperforms the simple average.

Finally, we illustrate how the combined tests may help to detect predictability in exchange rate returns for the cases of Chile and Canada. We see that our quasi-optimal weight, along with the simple average, induces a combined test that allows detection of predictability for Canadian exchange returns at the 5% significance level. This is important because individual tests only detect predictability at the 10% significance level. For Chile our combination strategy is even more fruitful. In this case the single tests and their simple average cannot reject the null of no predictability at the 10% level. When these single tests are combined with our proposed quasi-optimal weight, however, evidence of predictability is detected.

9 Appendix

9.1 Proof of Proposition 1

Proof. Under asymptotic normality we have

$$P_\alpha(\mu_{AN}(\omega)) = \Pr(T^{CN}(\omega) > t_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{t_\alpha}^{\infty} e^{-\frac{1}{2}(x-\mu_{AN}(\omega))^2} dx$$

using the usual change of variables

$$z = x - \mu_{AN}(\omega)$$

we have that

$$\begin{aligned} P_\alpha(\mu_{AN}(\omega)) &= \frac{1}{\sqrt{2\pi}} \int_{t_\alpha - \mu_{AN}(\omega)}^{\infty} e^{-\frac{1}{2}z^2} dz \\ &= 1 - \Phi(t_\alpha - \mu_{AN}(\omega)) \end{aligned}$$

where Φ denotes the distribution of a standard normal random variable. It is straightforward to check that

$$\frac{\partial P_\alpha(\mu_{AN}(\omega))}{\partial \mu_{AN}(\omega)} > 0$$

Therefore the maximization of (7) is equivalent to the maximization of the following objective function

$$\begin{aligned} \mu_{AN}(\omega) &= \sqrt{P}\frac{\mu(\omega)}{\sigma(\omega)} \\ \omega &\in [0, 1] \end{aligned}$$

which is also equivalent to maximize

$$\begin{aligned} \mu_{AN}(\omega) &= \frac{\mu(\omega)}{\sigma(\omega)} \\ \omega &\in [0, 1] \end{aligned}$$

■

9.2 Proof of Proposition 2

Proof. First of all we will check the continuity of μ_{AN} for $\omega \in [0, 1]$.In order to do that we recall that $\mu_{AN}(\omega)$ is given by the following expression:

$$\mu_{AN}(\omega) = \sqrt{P}\frac{\mu(\omega)}{\sigma(\omega)} = \sqrt{P}\frac{\omega\mu_1 + (1-\omega)\mu_2}{\sigma(\omega)}$$

where

$$
\begin{aligned}
\sigma^2(\omega) &\equiv \omega^2\sigma_1^2 + (1-\omega)^2\sigma_2^2 + 2\omega(1-\omega)\sigma_{1,2} \geq 0 \\
for\ \omega &\in [0,1]
\end{aligned}
$$

In the first place we will show that :σ_1, $\sigma_2 > 0$ and $\rho^2 < 1 \Longrightarrow \sigma^2(w) > 0$.

We have that

$$
\begin{aligned}
\sigma^2(\omega) &\equiv \omega^2\sigma_1^2 + (1-\omega)^2\sigma_2^2 + 2\omega(1-\omega)\sigma_{1,2} \\
&= \omega^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}) + 2\omega(\sigma_{1,2} - \sigma_2^2) + \sigma_2^2
\end{aligned}
$$

suppose

$$
\begin{aligned}
\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2} &= 0 \\
\sigma_1^2 + \sigma_2^2 &= 2\sigma_{1,2} \\
\sigma_1^2 + \sigma_2^2 &= \frac{2\sigma_{1,2}}{\sigma_1\sigma_2}\sigma_1\sigma_2 \\
\sigma_1^2 + \sigma_2^2 &= 2\rho\sigma_1\sigma_2
\end{aligned}
$$

we notice that $\rho = \frac{\sigma_{1,2}}{\sigma_1\sigma_2} > 0$ because $0 < \sigma_1^2 + \sigma_2^2 = 2\sigma_{1,2}$. Therefore

$$
\sigma_1^2 + \sigma_2^2 = 2\rho\sigma_1\sigma_2 < 2\sigma_1\sigma_2
$$

so

$$
\begin{aligned}
\sigma_1^2 + \sigma_2^2 &< 2\sigma_1\sigma_2 \\
\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2 &< 0 \\
(\sigma_1 - \sigma_2)^2 &< 0
\end{aligned}
$$

which is a clear contradiction. Therefore we must have

$$
\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2} \neq 0
$$

In case $\sigma_{1,2} = 0$ then $\rho = \frac{\sigma_{1,2}}{\sigma_1\sigma_2} = 0$ so

$$
\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2} = \sigma_1^2 + \sigma_2^2 > 0
$$

if $\sigma_{1,2} \neq 0$ then $\rho = \frac{\sigma_{1,2}}{\sigma_1\sigma_2} \neq 0$ and we have

$$
\begin{aligned}
(\sigma_1 - \sigma_2)^2 &\geq 0 \\
\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2 &\geq 0 \\
\sigma_1^2 + \sigma_2^2 &\geq 2\sigma_1\sigma_2 = \frac{2\sigma_{1,2}}{\rho} > 2\sigma_{1,2}
\end{aligned}
$$

this last step due to the assumption $\rho^2 < 1$.

Therefore

$$
\sigma_1^2 + \sigma_2^2 > 2\sigma_{1,2}
$$

and finally

$$\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2} > 0$$

This result implies that

$$\sigma^2(\omega) \equiv \omega^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}) + 2\omega(\sigma_{1,2} - \sigma_2^2) + \sigma_2^2$$

is a strictly convex quadratic function. As such $\sigma^2(\omega)$ admits a unique global minimum. To make sure this function is always positive we need to find its roots and check whether they are complex or real. Therefore we need the following discriminant to be negative

$$\begin{aligned} b^2 - 4ac &< 0 \\ b &= 2(\sigma_{1,2} - \sigma_2^2) \\ a &= (\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}) \\ c &= \sigma_2^2 \end{aligned}$$

We have that

$$\begin{aligned} b^2 - 4ac &= 4(\sigma_{1,2} - \sigma_2^2)^2 - 4(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2})\sigma_2^2 \\ &= 4(\sigma_{1,2}^2 + \sigma_2^4 - 2\sigma_{1,2}\sigma_2^2) - 4(\sigma_1^2\sigma_2^2 + \sigma_2^4 - 2\sigma_{1,2}\sigma_2^2) \\ &= 4(\sigma_{1,2}^2 - 2\sigma_{1,2}\sigma_2^2) - 4(\sigma_1^2\sigma_2^2 - 2\sigma_{1,2}\sigma_2^2) \\ &= 4(\sigma_{1,2}^2 - \sigma_1^2\sigma_2^2) \\ &= 4(\rho^2\sigma_1^2\sigma_2^2 - \sigma_1^2\sigma_2^2) \\ &= 4\sigma_1^2\sigma_2^2(\rho^2 - 1) < 0 \end{aligned}$$

therefore the quadratic form only admits complex roots. As a consequence $\sigma^2(\omega)$ is strictly positive and the ratio

$$\mu_{AN}(\omega) = \sqrt{P}\frac{\mu(\omega)}{\sigma(\omega)}$$

is a well defined ratio of two continuous functions, with a positive function in the denominator. Therefore $\mu_{AN}(\omega)$ is continuous over the compact set $[0, 1]$. By Weierstrass' theorem we can find $\overline{\omega} \in [0, 1]$ such that

$$\overline{\omega} = \underset{\omega \in [0,1]}{\arg\max} \mu_{AN}(\omega) \tag{20}$$

Step 2. If $\overline{\omega}$ is a solution of (20) then either $\overline{\omega} \in \{0, 1\}$ or $\overline{\omega}$ is interior. An interior solution for the maximization problem may be found in the set C of critical points of $\mu_{AN}(\omega)$:

$$\begin{aligned} C &= \left\{\omega \in R \text{ such that } \frac{d\mu_{AN}(\omega)}{d\omega} = 0\right\} \\ C &= \left\{\omega \in R \text{ such that } \sqrt{P}\left(\frac{\acute{\mu}(\omega)\sigma(\omega) - \mu(\omega)\acute{\sigma}(\omega)}{\sigma^2(\omega)}\right) = 0\right\} \end{aligned}$$

therefore, critical points satisfy

$$\frac{\acute{\mu}(\omega^*)\sigma(\omega^*) - \mu(\omega^*)\acute{\sigma}(\omega^*)}{\sigma^2(\omega^*)} = 0$$

or simply

$$\acute{\mu}(\omega^*)\sigma(\omega^*) = \mu(\omega^*)\acute{\sigma}(\omega^*)$$

Now, we have

$$\begin{aligned}
\acute{\mu}(\omega) &= \mu_1 - \mu_2 \equiv c \\
\sigma(\omega) &= \sqrt{f(\omega)} \\
f(\omega) &= \sigma^2(\omega) \\
&= \omega^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}) + 2\omega(\sigma_{1,2} - \sigma_2^2) + \sigma_2^2
\end{aligned}$$

$$\begin{aligned}
f'(\omega) &= 2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2})\omega + 2(\sigma_{1,2} - \sigma_2^2) \\
\sigma'(\omega) &= \frac{1}{2} f(\omega)^{-\frac{1}{2}} f'(\omega)
\end{aligned}$$

so

$$\begin{aligned}
\acute{\mu}(\omega^*)\sigma(\omega^*) &= \mu(\omega^*)\acute{\sigma}(\omega^*) \\
c\sigma(\omega^*) &= \mu(\omega^*)\frac{1}{2} f(\omega^*)^{-\frac{1}{2}} f'(\omega^*) \\
2cf(\omega^*)^{\frac{1}{2}} &= \mu(\omega^*) f(\omega^*)^{-\frac{1}{2}} f'(\omega^*) \\
2cf(\omega^*) &= \mu(\omega^*) f'(\omega^*)
\end{aligned}$$

Notice that $f(\omega)$ and $\mu(\omega)f'(\omega)$ are quadratic forms:

$$\begin{aligned}
2cf(\omega) &= 2c\omega^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}) + 4c\omega(\sigma_{1,2} - \sigma_2^2) + 2c\sigma_2^2 \\
&= A_1\omega^2 + A_2\omega + A_3 \\
A_1 &= 2c(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}) \\
A_2 &= 4c(\sigma_{1,2} - \sigma_2^2) \\
A_3 &= 2c\sigma_2^2
\end{aligned}$$

on the other hand

$$\begin{aligned}
\mu(\omega)f'(\omega) &= \left\{2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2})\omega + 2(\sigma_{1,2} - \sigma_2^2)\right\}\{\mu_1\omega + \mu_2(1-\omega)\} \\
&= \left\{2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2})\omega + 2(\sigma_{1,2} - \sigma_2^2)\right\}\{c\omega + \mu_2\} \\
&= B_1\omega^2 + B_2\omega + B_3 \\
B_1 &= 2c(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}) \\
B_2 &= 2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2})\mu_2 + 2c(\sigma_{1,2} - \sigma_2^2) \\
B_3 &= 2(\sigma_{1,2} - \sigma_2^2)\mu_2
\end{aligned}$$

Therefore, any critical point ω^* must satisfy

$$(A_1 - B_1)\omega^2 + (A_2 - B_2)\omega + A_3 - B_3 = 0$$

But

$$(A_1 - B_1) = 2c(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}) - 2c(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}) = 0$$

Furthermore

$$\begin{aligned}
(A_2 - B_2) &= 4c(\sigma_{1,2} - \sigma_2^2) - \left\{2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2})\mu_2 + 2c(\sigma_{1,2} - \sigma_2^2)\right\} \\
&= 2c(\sigma_{1,2} - \sigma_2^2) - 2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2})\mu_2 \\
&= 2\mu_1(\sigma_{1,2} - \sigma_2^2) - \left\{2\mu_2(\sigma_{1,2} - \sigma_2^2) + 2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2})\mu_2\right\} \\
&= 2\mu_1(\sigma_{1,2} - \sigma_2^2) - \left\{2\mu_2\sigma_{1,2} + 2(\sigma_1^2 - 2\sigma_{1,2})\mu_2\right\} \\
&= 2\mu_1(\sigma_{1,2} - \sigma_2^2) + 2\mu_2(\sigma_{1,2} - \sigma_1^2)
\end{aligned}$$

this last expression is different from zero by assumption. Besides we have

$$\begin{aligned}
(A_3 - B_3) &= 2c\sigma_2^2 - 2(\sigma_{1,2} - \sigma_2^2)\mu_2 \\
&= 2(\mu_1 - \mu_2)\sigma_2^2 - 2(\sigma_{1,2} - \sigma_2^2)\mu_2 \\
&= 2\left\{\mu_1\sigma_2^2 - \mu_2(\sigma_2^2 + \sigma_{1,2} - \sigma_2^2)\right\} \\
&= 2\left\{\mu_1\sigma_2^2 - \mu_2\sigma_{1,2}\right\}
\end{aligned}$$

therefore we have that the unique critical point ω^* satisfies

$$\omega^* = \frac{\mu_2\sigma_{1,2} - \mu_1\sigma_2^2}{\mu_1(\sigma_{1,2} - \sigma_2^2) + \mu_2(\sigma_{1,2} - \sigma_1^2)}$$

Therefore, the solution to the maximization problem is given by

$$\overline{\omega} = \arg\max_{\omega \in \Psi} \frac{\mu(\omega)}{\sigma(\omega)}$$

$$\begin{aligned}
\Psi &= \{1, 0\} \cup \Gamma \\
\Gamma &= \begin{bmatrix} \{\omega^*\} \text{ if } \omega^* \in [0,1] \\ \phi \text{ if } \omega^* \notin [0,1] \end{bmatrix}
\end{aligned}$$

■

9.3 Proof of Proposition 4

Proof. Let us suppose that $\omega \in (0, 1)$ is such that

$$\left(\sqrt{P}\right)^{-1} \mu_{AN}(\omega) = \frac{\mu(\omega)}{\sigma(\omega)} < \min\left\{\frac{\mu(0)}{\sigma(0)}, \frac{\mu(1)}{\sigma(1)}\right\}$$

we will also assume, without loss of generality, that

$$\frac{\mu_1}{\sigma_1} = \frac{\mu(0)}{\sigma(0)} \leq \frac{\mu_2}{\sigma_2} = \frac{\mu(1)}{\sigma(1)}$$

so

$$\mu_2 \geq \frac{\mu_1}{\sigma_1}\sigma_2$$

Therefore we have

$$\begin{aligned}
\left(\frac{\mu(\omega)}{\sigma(\omega)}\right)^2 &\geq \frac{\omega^2\mu_1^2+(1-\omega)^2\mu_2^2+2\omega(1-\omega)\mu_1\mu_2}{\omega^2\sigma_1^2+(1-\omega)^2\sigma_2^2+2\omega(1-\omega)\sigma_{1,2}} \\
&\geq \frac{\omega^2\mu_1^2+(1-\omega)^2\left(\frac{\mu_1}{\sigma_1}\sigma_2\right)^2+2\omega(1-\omega)\frac{\mu_1^2}{\sigma_1}\sigma_2}{\omega^2\sigma_1^2+(1-\omega)^2\sigma_2^2+2\omega(1-\omega)\sigma_{1,2}} \\
&\geq \frac{\mu_1^2}{\sigma_1^2}\left[\frac{\omega^2\sigma_1^2+(1-\omega)^2\sigma_2^2+2\omega(1-\omega)\sigma_2\sigma_1}{\omega^2\sigma_1^2+(1-\omega)^2\sigma_2^2+2\omega(1-\omega)\sigma_{1,2}}\right] \\
&\geq \frac{\mu_1^2}{\sigma_1^2}\left[\frac{\omega^2\sigma_1^2+(1-\omega)^2\sigma_2^2+2\omega(1-\omega)\sigma_2\sigma_1}{\omega^2\sigma_1^2+(1-\omega)^2\sigma_2^2+2\omega(1-\omega)\sigma_2\sigma_1\rho}\right] \\
&\geq \frac{\mu_1^2}{\sigma_1^2}\left[\frac{\omega^2\sigma_1^2+(1-\omega)^2\sigma_2^2+2\omega(1-\omega)\sigma_2\sigma_1}{\omega^2\sigma_1^2+(1-\omega)^2\sigma_2^2+2\omega(1-\omega)\sigma_2\sigma_1}\right] \\
&\geq \frac{\mu_1^2}{\sigma_1^2} = \min\left\{\left(\frac{\mu(0)}{\sigma(0)}\right)^2, \left(\frac{\mu(1)}{\sigma(1)}\right)^2\right\} > \left(\frac{\mu(\omega)}{\sigma(\omega)}\right)^2
\end{aligned}$$

which is a contradiction. Therefore combining will at least yield the minimum single outcome. ■

References

- Anatolyev S, A. Gerco 2005. A Trading Approach to Testing for Predictability. *Journal of Business and Economic Statistics* Vol 2, N°4 455-461.
- Avramov D. 2002. Stock return predictability and model uncertainty. *Journal of Financial Economics* 64, 423-458.
- Bates J.M, C. Granger. 1969. The combination of forecasts. *Operational Research Quarterly*, 20, 451-68.
- Brown D, Ibragimov R. 2005. Sign Tests for Dependent Observations and Bounds for Path-Dependent Options. Yale ICF Working Paper no. 05-19.
- Chao J, Corradi V, Swanson N. 2000. An Out of Sample Test for Granger Causality. *Macroeconomic Dynamics* **5**: 598-620.
- Cheung Y, Chinn M, Garcia Pascual A. 2005. Empirical Exchange Rate Models of the Nineties : Are any Fit to Survive?. *Journal of International Money and Finance* **24** : 1150-1175
- Clark T, McCracken. 2001. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, Volume 105, Number 1, pp. 85-110(26).
- Clark T, McCracken M. 2005. Evaluating Direct Multistep Forecasts. *Econometric Reviews* **24**: 329-404.
- Clark T, West K. 2006. Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis. *Journal of Econometrics* **135**: 155-186.
- Clark T, West K. 2007. Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics* **138**: 291-311.

- Clemen R. 1989. Combining Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting* 5 559-583.
- Clements M, Hendry 2004. Pooling of Forecasts. *The Econometrics Journal*. Vol 7, Issue 1, 1-31.
- Diebold F, Mariano R. 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* **13:** 253-263.
- Giacomini R, White H. 2006. Test of Conditional Predictive Ability. *Econometrica* **74**: 1545-1578.
- Hamilton J. 1994. Time Series Analysis. Princeton University Press: Princeton, NJ.
- Inoue A, Kilian L. 2004. In-Sample or Out-of-Sample Tests of Predictability: Which one Should We Use?. *Econometric Reviews* **23**: 371-402.
- McCracken M. 2007. Asymptotics for Out of Sample Tests of Granger Causality. *Journal of Econometrics* **140**: 719-752.
- McCracken M, West. 2002. Inference About Predictive Ability.299-321 in M. Clements and D. Hendry (eds) Companion to Economic Forecasting, Oxford: Blackwell.
- Meese R, Rogoff K. 1983. Empirical Exchange Rate Models of the Seventies. Do They Fit Out-of-Sample?. *Journal of International Economics* **14**: 3-24.
- Obstfeld M, Rogoff 2002. Foundations of International Macroeconomics. The MIT Press.
- Pesaran H, Timmermannn. 1992. A Simple Nonparametric Test of Predictive Performance..*Journal of Business and Economic Statistics.* Vol.10 N°4, 461-465.
- Pincheira, P. 2006. Shrinkage Based Tests of the Martingale Difference Hypothesis. Working Paper N°376, Central Bank of Chile.
- Reineke D, Baggett J, Elfessi A. 2003. A Note on the Effect of Skewness, Kurtosis, and Shifting on One-Sample t and Sign Tests. *Journal of Statistics Education* **11** (3)
- Smith J, Wallis. 2005. Combining Point Forecasts: The Simple Average Rules:OK? manuscript Department of Economics University of Warwick.
- Timmermannn A. 2006. Forecast Combinations. in In *Handbook of Economic Forecasting*, Elliott G, Granger CWJ, Timmermannn A (eds): 135-194.
- Tsay R. 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. 2006. Forecast Evaluation. In *Handbook of Economic Forecasting*, Elliott G, Granger CWJ, Timmermannn A (eds): 99-134.
- White H. 2001. Asymptotic Theory for Econometricians. In *Economic Theory, Econometrics and Mathematical Economics* series. 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	**Working Papers**
Banco Central de Chile	**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-458 Diciembre 2007
Copper Price, Fiscal Policy and Business Cycle in Chile
Juan Pablo Medina y Claudio Soto

DTBC-457 Diciembre 2007
The Chilean Business Cycles Through the Lens of a Stochastic General Equilibrium Model
Juan Pablo Medina y Claudio Soto

DTBC-456 Diciembre 2007
Is Ownership Structure a Determinant of Bank Efficiency?
Rodrigo Fuentes y Marcos Vergara

DTBC-455 Diciembre 2007
Estimating the Output Gap for Chile
Rodrigo Fuentes, Fabián Gredig y Mauricio Larraín

DTBC-454 Diciembre 2007
Un Nuevo Marco Para la Elaboración de los Programas de Impresión y Acuñación
Rómulo Chumacero, Claudio Pardo y David Valdés

DTBC-453 Diciembre 2007
Development Paths and Dynamic Comparative Advantages: When Leamer Met Solow
Rodrigo Fuentes y Verónica Mies

DTBC-452 Diciembre 2007
Experiences With Current Account Deficits in Southeast Asia
Ramon Moreno

DTBC-451 Diciembre 2007
Asymmetric Monetary Policy Rules and the Achievement of the Inflation Target: The Case of Chile
Fabián Gredig

DTBC-450 Diciembre 2007
Current Account Deficits: The Australian Debate
Rochelle Belkar, Lynne Cockerell y Christopher Kent

DTBC-449 Diciembre 2007
International Reserves Management and the Current Account
Joshua Aizenman

DTBC-448 Diciembre 2007
Estimating the Chilean Natural Rate of Interest
Rodrigo Fuentes y Fabián Gredig

DTBC-447 Diciembre 2007
Valuation Effects and External Adjustment: A Review
Pierre-Oliver Gourinchas

DTBC-446 Diciembre 2007
What drives the Current Account in Commodity Exporting Countries? The cases of Chile and New Zealand
Juan Pablo Medina, Anella Munro y Claudio Soto

DTBC-445 Diciembre 2007
The Role of Interest Rates and Productivity Shocks in Emerging Market Fluctuations
Mark Aguiar y Guita Gopinath

DTBC-444 Diciembre 2007
Financial Frictions and Business Cycles in Middle Income Countries
Jaime Guajardo

DTBC-443 Diciembre 2007
Stocks, Flows and Valuation Effects of Foreign Assets and Liabilities: Do They Matter?
Alfredo Pistelli, Jorge Selaive y Rodrigo Valdés