Parameter Estimation Error in Tests of Predictive Performance under Discrete Loss Functions

Francisco Javier Eransus, Alfonso Novales
Universidad Complutense de Madrid
Departamento de Economía Cuantitativa

October, 2012

Abstract

We analyze the effect of parameter estimation error on the size of unconditional population level tests of predictive ability when they are implemented under a class of loss functions we refer to as 'discrete functions'. The analysis is restricted to linear models in stationary variables. We obtain analytical results for nonnested models guaranteeing asymptotic irrelevance of parameter estimation error under a plausible predictive environment and three subsets of discrete loss functions that seem quite appropriate for many economic applications. For nested models, we provide some Monte Carlo evidence suggesting that the asymptotic distribution of the Diebold and Mariano (1995) test is relatively robust to parameter estimation error in many cases if it is implemented under discrete loss functions, unlike what happens under the squared forecast error or the absolute value error loss functions.

1 Introduction

Evaluating forecasting performance requires the use of a statistical test under a specific loss function. Most often, the squared forecast error (SFE) or the absolute forecast error (AFE) loss functions are assumed, even when applying the popular test introduced in Diebold and Mariano (1995) (DM test) that allows for a general loss function to be defined on the (data,forecast)-pair. However, there are many situations in which these functions may not be a good choice or even a valid option. An obvious one is decision-making environments, where the economic benefits or losses produced by the use of forecasts are given by the nature of the deci sion problem. Then, an economic evaluation of forecasts is appropriate, and the loss function is explicitly determined so as to reflect these economic values.

Even when forecasts are made for an unspecified use and the loss function is viewed as a purely statistical measure of the quality of forecasts, SFE or AFE functions may be inadequate. In forecasting situations often encountered in economics and finance, the analyst is just interested in anticipating one among a finite set of categories for the data, even if that requires the prediction of a continuous underlying variable. Pesaran and Timmermann (2009) offer several examples: macroeconomic survey participants are often asked about the most likely range of values for GDP growth, changes in interest rates or inflation. Financial analysts often categorize stocks into "buy", "strong buy", "hold", "sell" or "strong sell" on the basis of a prediction of future price changes. In these directional forecasting situations, the sign of the forecast is essential, while it is unclear that the size of the forecast error should be assigned a continuous valuation like SFE or AFE which, in contrast, incorporates a high sensibility to outliers in either the data or the forecasts. Another situation would be qualitative forecasting, when the space of observed events cannot unambiguously be reduced to real numbers.

For such situations, a loss function can be defined on a finite partition of the domain of (data,forecast)-pairs into n^2 quadrants, with a specific cost or utility function assigned to each quadrant. A simple application with $n = 2$ is discussed in Granger and Pesaran (2000), with more general versions being mentioned in McCracken (2004) and Blaskowitz and Herwatz (2011), among others. Our work deals with this class of loss functions when the utility associated to each (data,forecast)-pair is a real number, which we will refer to as discrete loss functions. Such functions are the natural criterion to evaluate forecasts in some decision-making problems, and have advantages as a statistical criterion for forecast evaluation: *i*) they can be used in situations when the information provided by the (data,forecast)-pair cannot be summarized by their difference, like in qualitative forecasting, *ii*) asymmetries in the evaluation of forecasts can be easily incorporated; *iii*) they can place a bound on the potential distortion in forecast evaluation due to an occasional extreme prediction error or data outlier; *iv*) the sign of the forecast can be easily taken into account for its evaluation, as it is needed for directional forecasts. As an alternative to AFE loss function in many contexts, we specifically propose the SDAFE function, which is a discrete version of the former with the properties *iii* and *iv* mentioned above.

Considerable effort has been devoted to characterizing the effects of parameter error estimation (PEE) on tests of forecasting performance like DM, a major issue when assessing the performance of a forecasting model.[1] The main contribution of this paper is to show, firstly, that the properties of these tests are robust to the presence of PEE under a variety of classes of discrete loss functions, when certain plausible forecasting conditions are satisfied, and, secondly, that the distortion caused by PEE on the properties of the tests in other forecasting frameworks can be far less important under discrete loss functions than under standard continuous loss functions as SFE or AFE.

Our analysis is restricted to the following framework: we examine tests of forecasting performance following the unconditional population level approach, that tests the hypothesis $E(f(\beta^*)) = Ef$, where β^* denotes the stacked parameter vectors of the l models and $f = (f_1, ..., f_l)$ with f_i being the loss function used to evaluate the i-th model forecasts. We discard the alternative finite sample approach because results in Giacomini and White (2006), a key reference, do not hold for the recursive forecasting scheme, the most common in practice and the one we consider throughout the paper, as opposed to the fixed or rolling forecasting schemes. Furthermore, we focus on analyzing the size of the tests rather than their power. Finally, our results are restricted to linear models in stationary variables.

When comparing nonnested models in our framework, the basic references regarding the effect of PEE on the distribution of the test statistic are West (1996) and McCracken (2000). McCracken (2000) extends the results in West (1996) to loss functions that are non-differentiable in the model parameters, as it is the case of discrete loss functions. Both authors characterize a Normal limiting distribution for the sample mean $\overline{f}$ with a variance covariance matrix that involves an F matrix, defined as $F = E\left[\frac{\partial f(\beta)}{\partial \beta}\right]_{\beta=\beta^*}$, whose analytical expression depends on the specification of the loss function f. A central part of our paper consists precisely in characterizing this matrix for a discrete f. For nested models, results are much less general. They are specific to each particular test and most of them hold only for the SFE loss function, with non-standard limiting distributions for the test statistic. McCracken

[1] The properties of the DM test were originally derived in Diebold and Mariano (1995) without taking into account PEE. Later work has described how to incorporate PEE in the DM test (through its asymptotic distribution) in different forecasting frameworks. Some authors use different notation to refer to the original version of the test, as opposed to the PEE version. In contrast, we will regularly refer to the DM test, making explicit in each occasion whether PEE is being taking into account.

(2007) is a very relevant reference for the comparison of nested models using the DM test under an SFE loss. There are no results in the literature applying to nested forecasting models under discrete loss functions.

Unfortunately, even the results in West (1996) and McCracken (2000) may be difficult to incorporate to a forecasting test in practice, at a difference of what happens when PEE is not taken into account. First of all, because derivation of the analytical expression for the F matrix can be rather complicated when f is not differentiable (see McCracken (2004)); secondly, because estimating the variance covariance matrix may be too complex for not specialized practitioners; thirdly, because such estimation requires the knowledge of estimation details of the competing forecasting models, which is not always available.

It is therefore very fortunate that, in the context of West (1996) and McCracken (2000), there are situations in which PEE is irrelevant in terms of statistical inference, so that the tests can be implemented with the original asymptotic variance and critical values for the test statistic. We denote this circumstance by AIPEE (asymptotic irrelevance of parameter estimation error). Other than in setups characterized by the length of the sample and forecasting period, AIPEE arises for nonnested models when $F = 0$, a condition which can hold for specific loss functions in certain estimation and forecasting situations. For instance, it holds for SFE and AFE under quite plausible conditions. However, as McCracken (2004) and West (2006) state, it is not easy to find additional loss functions for which AIPEE holds.

The results of the paper can be summarized as follows:

i) in the case of nonnested models, we find three sets of sufficient conditions under which the F matrix in McCracken (2000) turns out to be a matrix of zeroes for discrete loss functions, so that the AIPEE property holds. The first two theorems require conditions related to the models which are quite plausible in practice and analogous to those imposed in McCracken (2000) for an AFE loss. Simpler versions of the SDAFE loss function verify our first theorem whereas many other interesting versions of SDAFE verify the second one. Our third theorem requires a relatively restrictive assumption on the explanatory variables of the models and it is valid for a subset of discrete loss functions which includes the function used in the decision environment example in section 2. These three theorems are the main results in the paper.

ii) we obtain simulation results on the distortions in the size of tests implemented under discrete loss functions when AIPEE does not hold, but that fact is ignored. We perform exercises for non-nested models in cases where $F \neq 0$, as well as for nested models. We focus on the DM test and a recursive forecasting scheme. Our results suggest that the size of the DM test is more robust to the presence of PEE in most predictive setups if it is implemented under a discrete loss function than under more standard loss functions. Given the inconvenience and complexity of incorporating PEE in the implementation of tests of forecasting ability, this is also a valuable conclusion.

The paper is organized as follows. Sections 2 and 3 introduce and motivate the use of discrete loss functions when evaluating forecasting ability. In Section 4 we review the literature. In Sections 5 and 6 we present the results. Section 5 contains the theorems on AIPEE. Section 6 presents the results of the simulation exercises mentioned above, as well as their sensitivity to the specific discrete loss function used. Additional analysis on this sensitivity is carried out in Section 7. The main conclusions are summarized in Section 8.

2 Discrete loss functions

There are many forecasting environments for which evaluating forecasts is naturally defined on a finite partition of the data-forecast domain with a utility function $U_{jk}(.)$ associated to each of the resulting (j, k)-quadrants. Granger and Pesaran (2000) introduce this type of loss functions for two-state, two-action decision problems and McCracken (2004) and Blaskowitz and Herwatz (2011) consider alternative versions of these functions. We focus on the particular case when U_{jk} are constants, which we will refer to as 'discrete loss functions'.

If y_{t+h} denotes the data at time $t+h$ and $\widehat{y}_{t+h}$ the forecast made at time t for y_{t+h}, a formal definition of a discrete loss function f is given by:

$(a.1)$ establishing a partition of the data domain into $n < \infty$ intervals: $\{l_0, l_1, l_2, ..., l_{n-2}, l_{n-1}, l_n\}$, where $l_i < l_{i+1}$ for $i = 0, ..., n-1$ with $l_0 = -\infty$ and $l_n = \infty$. Without loss of generality, we assume that the partition of the forecast domain is the same as that of the data domain.

$(a.2)$ assigning to each (j,k)-quadrant a real number a_{jk} :[2]

$$f(y_{t+h}, \widehat{y}_{t+h}) = a_{jk} < \infty \Leftrightarrow l_{j-1} < y_{t+h} \leq l_j,\ l_{k-1} < \widehat{y}_{t+h} \leq l_k. \tag{1}$$

The discrete loss function f can be summarized in matrix form:

		$\widehat{y}_{t+h}$			
		$(l_0, l_1]$	$(l_1, l_2]$	...	(l_{n-1}, l_n)
	$(l_0, l_1]$	a_{11}	a_{12}	...	a_{1n}
y_{t+h}	$(l_1, l_2]$	a_{21}	a_{22}	...	a_{2n}
	...	...	...	...	
	(l_{n-1}, l_n)	a_{n1}	a_{n2}	...	a_{nn}

The use of a discrete loss function as an alternative to SFE and AFE for forecast evaluation requires the choice of penalties into the quadrants defined by the partition chosen for the data space. In cases when the sign is relevant for evaluating forecasting performance, zero will be an l_s-point in the partition, and we propose the SDAFE ('Signed Discrete Absolute Forecast Error') function, a discrete version of AFE that takes into account whether the sign of forecast is correct while minimizing the impact of outliers. Starting from the partition $\{l_0, l_1, ..., l_{s-1}, l_s = 0, l_{s+1} ..., l_{n-1}, l_n\}$, $0 < s < n$, SDAFE penalties will be given by:

$$\tilde{l}_0 = l_1 - v_0;\ \tilde{l}_n = l_{n-1} + v_n;\ \tilde{l}_i = l_i,\ i = 1...n-1, \text{ for some } 0 < v_0 < \infty \text{ and } 0 < v_n < \infty,$$
$$\alpha_i = \theta \tilde{l}_{i-1} + (1-\theta)\tilde{l}_i,\ \theta \in (0,1),\ i = 1...n,$$
$$a_{jk} = |\alpha_j - \alpha_k| + \delta_j,\ j = 1...n,\ k = 1...n, \text{ with } \begin{array}{l} \delta_j = 0 \text{ if } sgn(\widehat{y}_{t+h}) = sgn(y_{t+h}) \\ \delta_j \geq 0 \text{ otherwise,} \end{array} \tag{2}$$

where $sgn(x)$ is a function taking the value 1 if x is positive, and -1 else.

Redefining the extreme points of the partition $\tilde{l}_0$ and $\tilde{l}_n$ prevents the subsequent loss function to take an infinite value, while α_z selects a representative point in the $[l_{z-1}, l_z)$ interval and δ_j reflects the extra penalty when the forecast has a different sign than the data. In the special case when forecasting the sign of the data has no value, the SDAFE function could be used with $\delta_j = 0$ for every j , thereby approximating a discrete version of the standard AFE function. More generally, the extra loss δ_j may be constant or change with the level of the data. A reasonable option is to choose δ_j to guarantee that for each data category, any forecast with the correct sign gets assigned a lower loss than any forecast with the wrong sign. One such possibility would be to define δ_j by:

[2] A natural convention would be $a_{jj} = 0$ for $j = 1, ..., n$, but this is not necessary for any of the results in the paper.

$$\delta_j = 0 \text{ if } sgn(\widehat{y}_{t+h}) = sgn(y_{t+h}) \text{ and } \delta_j = \max_k \{a_{jk} | \alpha_k \alpha_j > 0\}, \text{ otherwise.} \tag{3}$$

Along the paper we will use the SDAFE function with the particular specifications (3) and (4):

$$\theta = 1/2;\ v_0 = v_n. \tag{4}$$

In decision making environments, there are many situations in which a discrete loss function turns out to be the unique way to evaluate forecasts under a 'decision-based loss function' approach (Granger and Machina (2006)). That is the case when the state and decision variables can take just a finite number of values, and the payoff associated to each (state, decision)-pair is a real number that can be derived form the nature of the problem, as in the two-state, two-action decision problem in Granger and Pesaran (2000). We now present another decision problem, potentially representative of many other forecasting situations in practice. This example is specially interesting for our work because we will show later that AIPEE is guaranteed under a class of discrete loss functions with a structure like the loss function in the example.

3 A discrete decision problem

Consider a firm that produces an item which is a continuous variable, but it is sold in packs of r units of the item. For instance, each pack could be a bottle of r liters of milk. The potential weekly demand of the firm (measured in liters of milk) is an stationary variable x_t whose mean is rq_m and takes values in the range $(0, 2rq_m]$. At the beginning of each week $t+1$, the firm decides the number of bottles to produce, denoted by q_{t+1}, with $1 \leq q_{t+1} \leq 2q_m$.[3] Each bottle has a production cost of a and its price is p, generating a revenue of either p or 0 depending on whether the demand x_{t+1} is large enough to sell the bottle. We denote by $b = p - a$ the benefit per bottle, when sold. To produce a number of bottles above q_m, the firm would need to rent some additional equipment, whose cost is $A \geq 0$. If it turns out that $x_{t+1} > rq_m$ and so the equipment is used, there is a revenue for the firm of $A + B$, with $B \geq 0$.[4]

Since x_{t+1} is unknown, the firm takes its decisions on the basis of a point forecast $\widehat{x}_{t+1}$, so that the number of bottles to produce is given by $\widehat{q}_{t+1} = \lceil \widehat{x}_{t+1}/r \rceil$, where $\lceil z \rceil$ denotes the smallest integer larger or equal to z. This would be the 'optimal action function', in terms of Granger and Machina (2006). The states of the decision problem are the $((q-1)r, qr]$ intervals of values of x_{t+1}, for $q = ..., q_m - 2, q_m - 1, q_m, q_m + 1, q_m + 2, ...$

The weekly cost function of the firm is $C(\widehat{x}_{t+1}) = a\widehat{q}_{t+1} + A\widehat{s}_{t+1}$, where $\widehat{s}_{t+1} = 1\{\widehat{x}_{t+1} > rq_m\}$ is an indicator function which is equal to one if the firm demand forecast is larger than rq_m, so that the firm rents the new equipment to produce above q_m bottles. The weekly revenue function is $I(x_{t+1}, \widehat{x}_{t+1}) = \min(\widehat{q}_{t+1}, q^*_{t+1})p + (A + B)\widehat{s}_{t+1}s^*_{t+1}$, where $q^*_{t+1} = \lceil x_{t+1}/r \rceil$ and $s^*_{t+1} = 1\{x_{t+1} > rq_m\}$ would represent the optimal decisions if x_{t+1} was known. Finally, the utility function of the decision problem is the weekly profit of the firm $U(x_{t+1}, \widehat{x}_{t+1}) = I(x_{t+1}, \widehat{x}_{t+1}) - C(\widehat{x}_{t+1})$.

The problem can be rewritten in terms of the deviation of the demand around its mean, i.e., $y_t = x_t - rq_m$, whose mean is zero. Then the decision variable would be $w_{t+1} = q_{t+1} - q_m$, taking integer values between $1 - q_m$ and q_m, and consequently the optimal action function

[3] When $(q-1)r < x_{t+1} < qr$, we will assume that the firm could sell q packs. This arbitrary convention is not qualitatively relevant for the analysis.

[4] This can occur because of the positive effect in workers of managing a new equipment, because of marketing effects for the owner of the equipment, etc.

would be $\widehat{w}_{t+1} = \widehat{q}_{t+1} - q_m$. The states of the problem are now the intervals $((w-1)r, wr]$ of values of y_{t+1}, for $w = ..., -2, -1, 0, +1, +2, ...$It is straightforward to see that the weekly profit of the firm would be $U(y_{t+1}, \widehat{y}_{t+1}) = bq_m + \min(\widehat{w}_{t+1}, w^*_{t+1})p + \widehat{n}_{t+1}(n^*_{t+1}(A+B) - A) - a\widehat{w}_{t+1}$, being $w^*_t = q^*_t - q_m$, $\widehat{n}_{t+1} = 1\{\widehat{y}_{t+1} > 0\}$ and $n^*_{t+1} = 1\left\{y^*_{t+1} > 0\right\}$. For instance, if $q_m = 2$, the payoff matrix of the problem can be written as follows:

$\widehat{y}_{t+1}$

y_{t+1} States ↓	Actions →	$(-\infty, -r]$: $\widehat{w}_{t+1} = -1$, $\widehat{n}_{t+1} = 0$	$(-r, 0]$: $\widehat{w}_{t+1} = 0$, $\widehat{n}_{t+1} = 0$	$(0, +r]$: $\widehat{w}_{t+1} = +1$, $\widehat{n}_{t+1} = 1$	$(+r, +\infty)$: $\widehat{w}_{t+1} = +2$, $\widehat{n}_{t+1} = 1$
$(-\infty, -r]$	$w^*_{t+1} = -1$, $n^*_{t+1} = 0$	$bq_m - b$	$bq_m - p$	$bq_m - b$ $-2a - A$	$bq_m - b$ $-3a - A$
$(-r, 0]$	$w^*_{t+1} = 0$, $n^*_{t+1} = 0$	$bq_m - b$	bq_m	$bq_m - a$ $-A$	$bq_m - 2a$ $-A$
$(0, +r]$	$w^*_{t+1} = +1$, $n^*_{t+1} = 1$	$bq_m - b$	bq_m	$bq_m + b$ $+B$	$bq_m + b$ $-a + B$
$(+r, +\infty)$	$w^*_{t+1} = +2$, $n^*_{t+1} = 1$	$bq_m - b$	bq_m	$bq_m + b$ $+B$	$bq_m + 2b$ $+B$

(5)

Following Granger and Machina (2006), the 'decision-based loss function' is defined by the difference between the optimal utility level, i.e., the utility obtained if the forecast had been equal to the realized value, and the realized utility. In our context, the optimal utility U^* is obtained under $\widehat{y}_{t+1} = y_{t+1}$, so that $U^* = U(y_{t+1}, y_{t+1}) = bq_m + bw^*_{t+1} + n^*_{t+1}B$, and the discrete loss function $f(y_{t+1}, \widehat{y}_{t+1}) = U^*(y_{t+1}) - U(y_{t+1}, \widehat{y}_{t+1})$ corresponding to the payoff matrix (5) is:

$\widehat{y}_{t+1}$

y_{t+1}	$(-\infty, -r]$	$(-r, 0]$	$(0, +r]$	$(+r, +\infty)$
$(-\infty, -r]$	0	a	$2a + A$	$3a + A$
$(-r, 0]$	b	0	$a + A$	$2a + A$
$(0, +r]$	$B + 2b$	$B + b$	0	a
$(+r, +\infty)$	$B + 3b$	$B + 2b$	b	0

(6)

Because of the nature of the problem, the loss function is necessarily 'discrete', as defined by $(a.1) - (a.2)$. Moreover, the partition is symmetric around zero. We will later show that the AIPEE property holds for discrete loss functions based on a symmetric partition around zero for given patterns of the penalty values, provided some conditions on the forecasting model are satisfied. One of such patterns is (6). This is an interesting result because this pattern for the payoff matrix and the associated loss function appears in many other decision problems, maybe with $A = B = 0$. Indeed, for industrial production decisions where the item is sold in indivisible units, the demand for the next period is unknown, while prices are given and the cost of each aditional unit produced is constant.

In this example we assume that the decision is made on the basis of point forecasts, following Blaskowitz and Herwatz (2011), among many others. This assumption is maintained throughout the paper. Alternatively, the decision making environment suggested in Granger and Pesaran (2000) and Pesaran and Skouras (2002) considers an optimal action that maximizes expected utility computed under the probability distribution forecast on the set of states.

4 Review of theory

Let M_i be a forecasting model $y_{t+h} = \varphi_i(\beta_i^*, Z_{it}) + u_{it+h}$ that generates forecasts $\widehat{y}_{i,t+h}$ for y_{t+h} at time t , with β_i being the parameter vector in the model and β_i^* its true value, and Z_{it} a vector of variables known at time t. Variables in Z can either be exogenous regressors or lags of y. By u_{it+h} we denote the forecast error at time $t+h$ that would arise if β_i^* was known. We will assume we have a sample of size $T+h$ for y and Z, although just T data points will be used in estimation. The forecast $\widehat{y}_{i,t+h}$ is obtained as $\widehat{y}_{i,t+h} = \varphi_i(\widehat{\beta}_{it}, Z_{it})$, where $\widehat{\beta}_{it}$ denotes the estimate of β_i from a set of in-sample observations up to time t. There are three common forecasting schemes: fixed, rolling and recursive, and we focus on the last one: the first forecast is made at $t=R$ for y_{R+h}, estimating β_i with the sample up to $t=R$; each period we add a new data point in estimation and compute the h-step ahead forecast. We repeat the exercise up to period T, the last one used for forecasting. The total number of forecasts will be P, with $P+R=T+1$.

Given that we have l forecasting models, we would stack all the β_i parameter vectors in a single vector $\beta = (\beta_1', \beta_2', ..., \beta_l')'$ with dimension $K \times 1$, whose true value will be denoted by β^*. We would also stack the estimates $\widehat{\beta}_{it}$ and regressors Z_{it} into vectors $\widehat{\beta}_t$ and Z_t. We define a $l \times 1$ vector function $f_{t+h}(y_{t+h}, \beta, Z_t)$, whose i-th component is the value at time $t+h$ of the loss function associated to model M_i. The parametric estimation of $f_{t+h}(y_{t+h}, \beta^*, Z_t)$ will be $f_{t+h}(y_{t+h}, \widehat{\beta}_t, Z_t)$. To simplify notation, we denote $f_{t+h}(y_{t+h}, \beta, Z_t)$ by $f_{t+h}(\beta)$.

We focus on unconditional population-level tests of predictive ability, as described in Clark and McCracken (2010), that test for the unconditional value of $E(f_{t+h}(\beta^*))$ and only require the asymptotic distribution of the estimator $\overline{f} - E(f_{t+h}(\beta^*))$, with $\overline{f} = P^{-1} \sum_{t=R}^{T} f_{t+h}(\widehat{\beta}_t)$. Many tests for the adequacy of a single model or for comparing the forecast accuracy of several models belong to this category, the DM being the best known test in that class.

4.1 Nonnested models. Differentiable loss

For nonnested models, West (1996) obtains the analytical expression for the variance covariance matrix for the $\overline{f} - E(f_t(\beta^*))$ statistic under PEE. Assuming differentiability of $f_t(\beta)$ with respect to β, stationarity of the variables in the model,[5] convergence of $\pi = \lim_{T\to\infty} P/R$ and a specific assumption on the estimation method,[6] West (1996) shows:

$$\sqrt{P}(\overline{f} - E(f_{t+h}(\beta^*))) \overset{a}{\sim} N(0, \Omega), \tag{7a}$$

$$\Omega = S_{ff} + \Pi(FBS_{fh}' + S_{fh}B'F') + 2\Pi F V_\beta F', \tag{7b}$$

where $S_{ff} = \sum_{j=-\infty}^{+\infty} \Gamma_{ff}(j)$, $\Gamma_{ff}(j) = E\left[(f_t - E(f_t))(f_{t-j} - E(f_t))'\right]$ is the variance-covariance matrix when β^* is known, i.e., in the absence of PEE. Matrix $F = E\left[\frac{\partial f(\beta)}{\partial \beta}\right]_{\beta=\beta^*}$

[5] For integrated and possibly cointegrated variables, see Corradi, Swanson and Olivetti (2001) and Rossi (2005).

[6] For details on the specific assumptions see pages 1070-1071 in West (1996).

plays a very relevant role in our work. V_β is the asymptotic variance covariance matrix of $\sqrt{T}(\widehat{\beta}_T - \beta^*)$, while Π is an increasing function of π, bounded between 0 and 1. Hence, the higher is π, the larger is the divergence between the variance covariance matrix under PEE, Ω, and the component associated to sampling error, S_{ff}. Finally, matrices B and S_{fh} are determined by the estimation method, the characteristics of the forecasting model and the loss function.[7] The second and third terms in $\Omega, \Pi(FBS'_{fh} + S_{fh}B'F')$ and $2\Pi F V_\beta F'$, capture the covariance between f and the estimation error associated to $\widehat{\beta}_t$ and the variance due to estimation error in $\widehat{\beta}$, respectively.

A specially important situation arises when parameter estimation is irrelevant for statistical inference, i.e., when $\Omega = S_{ff}$, which we define as AIPEE. That is the case if $\pi = 0$ or $F = 0$.[8] The first condition holds when the number of available data points for estimation is arbitrarily large in relation to the number of forecasts, so that we can treat β^* as known. The second condition is more relevant, and it can be obtained for specific combinations of forecasting models and loss functions. As shown in West (1996), a situation of this type arises if the SFE function is used to evaluate forecasts of nonnested models when the probability distribution of forecast errors conditional on regressors has a zero mean.

4.2 Nonnested models. Non-differentiable loss

Theorem 2.3.1 in McCracken (2000) extends the theorem in West (1996) to non-differentiable loss functions, obtaining the same expression for the asymptotic distribution of the test statistic as in West (1996), with $F = E\left[\frac{\partial f(\beta)}{\partial \beta}\right]_{\beta=\beta^*}$ replaced by $F = \left[\frac{\partial Ef(\beta)}{\partial \beta}\right]_{\beta=\beta^*}$. That is the theorem that applies to discrete loss functions. McCracken (2000) derives the expression for F in nonnested linear models under a typical non-differentiable loss, the AFE function: $F = -E\left[sgn(u_{t+1})Z'_{t+1}\right]$. Consequently, $F = 0$ in this case if the probability distribution of errors u_{t+1} conditional on regressors Z_{t+1} has a zero median. When $F \neq 0$, estimation of F under an AFE loss function is simple, which is unfortunately not the case for most non-differentiable loss functions.

Application of theorems in West (1996) and McCracken (2000) to the DM test is straightforward. Then, $l = 2$, $f_t = (f_{1t}, f_{2t})'$ and the loss functions f_{1t} and f_{2t} are the same for both models. The test statistic is $\overline{d} = \overline{f}_1 - \overline{f}_2$, so that the version of the DM test under PEE is:

$$\sqrt{P}(\overline{d} - E(d_t)) \overset{a}{\sim} N(0, \Omega_d), \tag{8}$$

$$\Omega_d = \Omega_{11} + \Omega_{22} - 2\Omega_{12}, \tag{9}$$

with Ω_{ij} being the (i,j)-element in the 2×2 matrix Ω defined in (7b). Therefore, Ω_d is the correct asymptotic variance covariance matrix under PEE for the DM test, while $S_d = S_{ff\ 11} + S_{ff\ 22} - 2S_{ff\ 12}$ is the variance-covariance matrix presented in Diebold and Mariano (1995) under the assumption that β^* is known.

4.3 Nested models

When $l = 2$ and the models are nested, theorems in West (1996) and McCracken (2000) are no longer valid because Ω_d then becomes zero. The main result for unconditional population-level equal forecast accuracy tests under a recursive sampling scheme is obtained by McCracken (2007), who derives the asymptotic distribution for a restricted version of the DM

[7] See West (1996) for a detailed definition of these matrices.

[8] Actually, there is a third condition for AIPEE: $\Pi(FBS'_{fh} + S_{fh}B'F') = -2\Pi F V_\beta F'$, an infrequent situation, although West (1996), pgs. 1073-4, characterizes some forecasting setups in which such condition holds.

test (OOS-t test). The restrictions refers to a case in which forecast errors are serially uncorrelated and conditionally homoskedastic, and the objective function that characterizes the estimation method is the same as the loss function $f_t(\beta)$ used to evaluate the set of forecasts, with that loss function $f_t(\beta)$ being differentiable in the parameters β. Under this set of assumptions, McCracken (2007) obtains an asymptotically limiting distribution for the $\sqrt{P}\bar{d}\widehat{\Omega}_d^{-1/2}$ statistic free of nuisance parameters, where $\widehat{\Omega}_d = P^{-1}\sum_{t=R}^{T}(f_{1t}(\widehat{\beta}_{1t}) - f_{2t}(\widehat{\beta}_{2t}))^2$.

The resulting distribution has a representation as functionals of Brownian motions. McCracken (2007) tabulates critical values as a function of π and the difference k_2 in the number of parameters between the two models. The asymptotic distribution differs notably from a $N(0,1)$, having a negative mean. Numerical results in that paper suggest that the mathematical expectation of the limiting distribution decreases with π and k_2, while the variance increases with π. It is only when $\pi = 0$ that the DM test for nested models keeps its asymptotic Normality in a setup with PEE, provided the conditions in McCracken (2007) hold. The limitations of the result in McCracken (2007) refer to the fact that the resulting distribution is not standard, which is inconvenient for the practical implementation of the test, and also that the assumptions about the loss function are rather restrictive. Essentially, the test in McCracken (2007) can be applied in practice only under an SFE loss function, provided models are estimated by linear or nonlinear least squares, and forecasts are made one step ahead.

There are other relevant results related to the use of the SFE loss in this framework. Clark and McCracken (2001) establish results for tests of forecast encompassing when comparing two nested models. Clark and West (2007) propose testing for equal forecast accuracy through an 'adjusted-MSE' statistic. Even though the limiting distribution of this statistic is not Normal, the use of standard critical values yields little size distortion.

Unfortunately, there does not exist an analogous literature for nested models under non-differentiable loss functions and a recursive sampling scheme.

5 Analytical results

Let us consider l forecasting models for y_{t+h}. Model M_i is $y_{t+h} = \varphi_i(\beta_i^*, Z_{it}) + u_{it+h}$, where β_i^* denotes the true value of vector β_i ($K_i \times 1$) and u_{it+h} is the theoretical forecasting error for M_i. Vectors β_i and their true values are stacked into the $K \times 1$ vectors β and β^*, respectively.

We start by providing the analytical expression for matrix F in McCracken (2000) under a general discrete loss function, as defined in $(a.1) - (a.2)$. We will obtain the expression for the generic element in $F = \left[\frac{\partial Ef(\beta)}{\partial \beta}\right]_{\beta=\beta^*}$ for any model M_i and for each parameter in β. Obviously, if the parameter does not appear in the specification of M_i, the corresponding element of the derivative in F will be zero, so we will focus on $\left[\frac{\partial Ef_{it+h}(\beta)}{\partial \beta_{i,r}}\right]_{\beta=\beta^*}$, with $\beta_{i,r}$ being the r-th element in β_i. Let us write model M_i as: $y_{t+h} = \varphi_i(z_{irt}, \beta_{i,r}^*, Z'_{i(r)t}, \beta_{i,(r)}^*) + u_{it+h}$, where $Z'_{i(r)t}$ represents the $K_i - 1$ vector made up by the regressors in M_i except z_{irt}, while $\beta_{i,(r)}^*$ are the true values of the corresponding parameter vector. Let $g_x(x)$ denote the marginal density function of vector x, $g_{x_1|x_2}(x_1|x_2)$ the density of x_1 conditioned on x_2, $G_x(x)$ the distribution function for x and $G_{x_1|x_2}(x_1|x_2)$ the distribution function for x_1 conditioned on x_2.[9]

We will restrict our results to forecasting models M_i verifying the following assumptions:[10]

[9] To ease the lecture, we skip the time indices in the density and distribution functions for u_{it+h}, z_{irt+h} and $Z_{i(r)t+h}$.

[10] Except for Theorem 6, which explicitely requires linearity in model M_i, all the results on

($b.1$) The model is linear, i.e., $y_{t+h} = Z'_{it}\beta_i^* + u_{it+h}$.

($b.2$) The distribution of any explanatory variable, conditional on the other explanatory variables, has a finite variance.

($b.3$) The probability distribution of the theoretical forecasting error u_{it+h} , conditional on the explanatory variables, is continuous. The probability distribution of any explanatory variable, conditional on the rest, is also continuous.[11]

We start with a lemma that is valid under no assumptions on the discrete loss function.

Lemma 1 *Let M_i be a forecasting model satisfying assumptions* $(b.1) - (b.3)$*, and let f_i be a discrete loss function as defined in* $(a.1) - (a.2)$*. For each element $\beta_{i,r}$ in the parameter vector β_i we have:*

$$\left[\frac{\partial E f_{it+h}(\beta)}{\partial \beta_{i,r}}\right]_{\beta=\beta^*} = -\int_{Z_{i(r)}} \left(\sum_{k=1}^{n-1} H_k^* Q_k^*\right) g_{Z_{i(r)}}(Z_{i(r)}) dZ_{i(r)}, \tag{10}$$

with

$$Q_k^* = \sum_{j=1}^{n} B^*(j,k)(a_{jk} - a_{jk+1});\ H_k^* = g_{z_{ir}|Z_{i(r)}}(d_k^*|Z_{i(r)})\frac{d_k^*}{\beta_{i,r}^*} \tag{11a}$$

$$B^*(j,k) = G^*(j,k) - G^*(j-1,k);\ G^*(j,k) = G_{u_i|z_{ir},Z_{i(r)}}(l_j - l_k|d_k^*, Z_{i(r)}); \tag{11b}$$

$$d_k^* = \frac{l_k - Z'_{i(r)t}\beta^*_{i,(r)}}{\beta^*_{i,r}}. \tag{11c}$$

Proof. See Appendix A. ■

The general expression (10) could be used to particularize the theorem in McCracken (2000) to tests under a discrete loss function. But it would not be simple to implement it in practice, since it requires assumptions on the probability distribution of the error term conditional on the set of explanatory variables, as well as on the probability distribution of each regressor conditional on the rest.

5.1 AIPEE under discrete loss functions

To provide sufficient conditions for AIPEE we will need one of the following two assumptions on the distribution function $G_{u_i|z_{ir},Z_{i(r)}}$:

($b.4$) The probability distribution of the theoretical forecast error u_{it+h} conditional on the regressors $Z_{it} = (z_{irt}, Z'_{i(r)t})'$ has a zero median.

($b.4'$) The probability distribution of the theoretical forecast error u_{it+h} conditional on the regressors $Z_{it} = (z_{irt}, Z'_{i(r)t})'$ is symmetric around zero.

For forecasting models verifying conditions $(b.1)-(b.3)$ and either $(b.4)$ or $(b.4')$, we show that AIPEE holds for three types of discrete loss functions. The first type are those functions satisfying:

---

the paper can be shown analogously for non linear models as long as they verify the condition $\lim_{l_k \to +\infty,-\infty} g_{z_{ir}|Z_{i(r)}}(d_k^*|Z_{i(r)}) \left[\frac{\partial d_k}{\partial \beta_{i,r}}\right]_{\beta=\beta^*} = 0$, where $d_k = \varphi_{i,z_{ir}}^{-1}(l_k, \beta_{i,r}, Z'_{i(r)t}, \beta_{i,(r)})$, with $\varphi_{i,z_{ir}}^{-1}$ being the inverse function of φ_i with respect to z_{ir}, provided this inverse function exists, and $d_k^* = d_k(\beta^*)$. Assumptions $(b.1)$ and $(b.2)$ are sufficient conditions for it.

[11] This assumption is made just for analytical convenience. The results would still be true under discrete distributions, but the proof becomes notationally burdensome.

$$(a.3) \quad a_{jk} - a_{jk+1} = \begin{matrix} +c_k \text{ if } j \leq k \\ -c_k \text{ if } j > k. \end{matrix}$$

Assumption $(a.3)$ restricts the change in the loss function when we move from left to right to have the same absolute value in all rows in a given column.

Theorem 2 *Let M_i be a forecasting model satisfying assumptions $(b.1) - (b.4)$. Let f_i be a discrete loss function defined by $(a.1) - (a.2)$ and satisfying assumption $(a.3)$. Then, $\left[\frac{\partial E f_{it+h}(\beta)}{\partial \beta_{i,r}}\right]_{\beta=\beta^*} = 0$, $r = 1, ..., K_i$. Hence, the i-th row in matrix F in McCracken (2000) is $F_i = 0_{1\times K}$.*

Proof. See Appendix B. ■

Corollary 3 *Let us consider l nonnested forecasting models, each one verifying conditions $(b.1) - (b.4)$. Let f be a discrete loss function defined by $(a.1) - (a.2)$ satisfying assumption $(a.3)$. Then, the matrix F in McCracken (2000) is $F = 0_{l\times K}$. Consequently, under the conditions in Theorem 2.3.1 in McCracken (2000), AIPEE holds.*

It is straightforward to show that the use of SDAFE with $\delta_j = \overline{\delta} \geq 0$ for every j satisfies $(a.3)$ for any partition of the data space if there exists an element $l_s = 0$, $0 < s < n$.[12] The statement is shown in Appendix D. Hence, in this case, the use of SDAFE guarantees AIPEE.

While the result on AIPEE is clearly very relevant, the assumption $(a.3)$ that we have imposed on the structure of the loss function excludes a class of situations that may be important in practice. Let us consider a partition separating positive from negative values, with the same number of regions in the negative as in the positive zone, i.e., $l_{n/2} = 0$, and with the (data,forecast)-pairs classified into different regions according to size:

		$\widehat{y}_t$					
		L−	M−	S−	S+	M+	L+
y_t	L−	0	1	2	3	4	5
	M−	1	0	1	2	3	4
	S−	2	1	0	3	4	5
	S+	5	4	3	0	1	2
	M+	4	3	2	1	0	1
	L+	5	4	3	2	1	0

(12)

where L,M,S, refer to "large, "medium", or small" when referring to either data or forecast, while their sign is indicated next to the L,M,S symbols.

It would then look natural to desire that *"given some data, forecasts which have the right sign receive a lower (or at least not larger) penalty than any forecast with the wrong sign"*. The loss matrix (12) has been constructed with $c_k = -1$ for all k except for the $(n/2, n/2+1)$ and $(n/2+1, n/2)$ quadrants. The greatest loss in row 3 associated to forecasts with the right sign is 2, in quadrant (3,1). Therefore, for the previous principle to hold, we need losses in the quadrants to the right of the main diagonal in that row to be larger than 2, as it is the case in the matrix, even though $(a.3)$ would have imposed the (3,4), (3,5) and (3,6) quadrants to be equal to 1,2,3, respectively. This example suggests that it might be desirable

[12] This is not too restrictive, since it can be extended to a classification around a value $l_s = \overline{y}$, with data and forecasts then being transformed into $y_{t+h} - \overline{y}$ and $\widehat{y}_{i,t+h} - \overline{y}$, respectively.

to allow for changes in penalties from the $n/2$-th to the $n/2+1$-th columns to be specific for each row. For a discrete function satisfying $l_{n/2}=0$ these are the two columns in which the forecast moves from having the right sign to missing it.

To allow for that possibility, we consider a more general version of assumption $(a.3)$ for discrete loss functions with a partition verifying $l_{n/2}=0$:

$$(a.4)\quad a_{jk}-a_{jk+1}=\begin{array}{l} +c_{j,n/2} \text{ if } k=n/2 \text{ and } j\leq k \\ -c_{n-j+1,n/2} \text{ if } k=n/2 \text{ and } j>k \\ +c_k \text{ if } k\neq n/2 \text{ and } j\leq k \\ -c_k \text{ if } k\neq n/2 \text{ and } j>k. \end{array}$$

We now provide an alternative set of sufficient conditions, including $(a.4)$, under which we again show AIPEE. We also need to incorporate the assumption that the partition is symmetric around zero:

$$(a.5)\quad l_{n/2}=0 \text{ and } l_{n/2-i}=-l_{n/2+i}, \text{ for } i=1,...,n/2.$$

Theorem 4 *Let M_i be a forecasting model satisfying assumptions $(b.1)-(b.3)$ and $(b.4')$. Let f_i be a discrete loss function defined by $(a.1)-(a.2)$ and satisfying assumptions $(a.4)$ and $(a.5)$. Then, $\left[\frac{\partial Ef_{it+h}(\beta)}{\partial\beta_{i,r}}\right]_{\beta=\beta^*}=0$, $r=1,...,K_i$. Hence, the i-th row in matrix F in McCracken (2000) is $F_i=0_{1\times K}$.*

Proof. See Appendix B. ■

Corollary 5 *Let us consider l nonnested forecasting models, each one verifying conditions $(b.1)-(b.3)$ and $(b.4')$. Let f be a discrete loss function defined by $(a.1)-(a.2)$ and satisfying assumptions $(a.4)$ and $(a.5)$. Then, the matrix F in McCracken (2000) is $F=0_{l\times K}$. Consequently, under the conditions in Theorem 2.3.1 in McCracken (2000), AIPEE holds.*

Theorem 2 and Corollary 3 apply under no restriction on the partition used to define the discrete loss function, so long as the matrix of losses satisfies $(a.3)$. On the other hand, Theorem 4 and Corollary 5 apply to discrete functions defined on a symmetric partition of the data space around zero, but they can accommodate any loss function satisfying $(a.4)$, a more appropriate assumption when the sign of the forecast is relevant.

It is not difficult to show that, under a partition satisfying $(a.5)$, the SDAFE loss function with specifications (3)-(4) verifies $(a.4)$ and, consequently, Theorem 4 and Corollary 5 then hold. See Appendix D for a proof.

McCracken (2000) showed that the AFE loss function verifies the AIPEE property in the same forecasting context considered in Corollary 3. On the other hand, the SFE loss guarantees $F=0$ whenever $E(u_{t+h}|Z_t)=0$. However, it is hard to show that $F=0$ for any other statistical criterion for forecast evaluation, so that our results for discrete loss functions characterized by either $(a.3)$ or by $(a.4)-(a.5)$ are very relevant.

Besides their use as statistical measure of forecast accuracy, discrete loss functions may become the natural way to evaluate forecasts in the basis of their economic value in many applications, as it was explained in section 3. The example introduced there required a loss function characterized by a symmetric partition around zero while penalties would be under the following pattern, which may appear in many other problems similar to that in section 3:

$$(a.6)\quad a_{jk}-a_{jk+1}=\begin{array}{l} c_{n/2} \text{ if } k=n/2 \text{ and } j\leq k \\ c'_{n/2} \text{ if } k=n/2 \text{ and } j>k \\ c \text{ if } k\neq n/2 \text{ and } j\leq k \\ c' \text{ if } k\neq n/2 \text{ and } j>k. \end{array}$$

Assumption $(a.6)$ restricts the change in the loss function when we move from the (j,k) quadrant to the $(j,k{+}1)$ to take the same value in the cells above the main diagonal and again the same value in the cells below it. Besides, these two values must be the same for every column, except the $n/2$-th column. In contrast to the pattern in $(a.3)$ and $(a.4)$, assumption $(a.6)$ does not restrict the changes between two adjacent columns to have the same absolute value for all rows, so that $(a.6)$ is more general than $(a.3)$ and $(a.4)$ in that sense. However, while those changes can be specific to each column in the case of $(a.3)$ and $(a.4)$, they must be the same for every column except the $n/2$-th column in the case of $(a.6)$.

The loss function (6) in section 3 satisfies $(a.6)$,with $c=-a$, $c'=b$, $c_{n/2}=-a-A$ and $c'_{n/2}=B+b$. Moreover, many decision problems on how many units to produce along a period of an indivisible item will verify the assumption $(a.6)$ if prices can be considered fixed and the cost of each aditional unit produced is constant, often with $A=B=0$.

To obtain an alternative set of sufficient conditions for AIPEE to hold for the class of discrete loss functions given by $(a.5)$ and $(a.6)$, we need to impose a new condition on the forecasting model:

$(b.5)$ The probability distribution of any explanatory variable, conditional on the rest, is symmetric around zero. The probability distribution of any subset of explanatory variables is radially symmetric around zero.

Theorem 6 *Let M_i be a forecasting model satisfying assumptions $(b.1)-(b.5)$. Let f_i be a discrete loss function defined by $(a.1)-(a.2)$ and satisfying assumptions $(a.5)$ and $(a.6)$. Then, $\left[\frac{\partial E f_{it+h}(\beta)}{\partial \beta_{i,r}}\right]_{\beta=\beta^*}=0$, $r=1,...,K_i$. Hence, the i-th row in matrix F in McCracken (2000) is $F_i=0_{1\times K}$.*

Proof. See Appendix B. ■

Corollary 7 *Let us consider l nonnested forecasting models, each one verifying the conditions $(b.1)-(b.5)$. Let f be a discrete loss function defined by $(a.1)-(a.2)$ and satisfying assumptions $(a.5)$ and $(a.6)$. Then, the matrix F in McCracken (2000) is $F=0_{l\times K}$. Consequently, under the conditions in Theorem 2.3.1 in McCracken (2000), AIPEE holds.*

Notice that no assumption on the method to obtain forecasts is needed in the theorems above, so that when $F=0$, AIPEE is guaranteed not only for the recursive sampling scheme but for the rolling and fixed forecasting schemes too (see McCracken (2000)), although we only consider the recursive scheme.

6 Simulations

We lack analytical results on the asymptotic distribution of the test statistics under a discrete loss function f when conditions in the corollaries above do not hold. To gain some insight into this issue, we perform in this section some simulation exercises ignoring PEE when using the DM test, undoubtedly the most popular test for equal forecast accuracy. All simulations will use a recursive forecasting scheme. Our goal is to quantify the bias in the size of the tests as a consequence of the fact that the true probability distribution of the test statistics in the situations considered may differ from $N(0,S_{ff})$. We use a first forecasting scenario comparing nonnested models in which there is correlation between forecast errors and regressors, and a second scenario comparing nested models.

The experiments will consider nine loss functions: the standard SFE and AFE continuous functions, plus seven discrete loss functions L_1 to L_7, all of them defined on partitions that are symmetric around zero. These discrete loss functions will be of two types: on the one hand,

L_1 to L_4,which belong to the SDAFE class of loss functions, used as a statistical criterion for forecast accuracy. On the other hand, L_5 to L_7 are three alternative loss functions following the pattern (6) in the example of section 3, as if the experiments reflected a problem of economic evaluation of forecasts. The SDAFE losses will use four alternative partitions of decreasing fineness, allowing us to analyze the robustness of our results to the fineness of the partition. L_5 to L_7 will use the same partition but very different numerical values, so as to check the robustness of the size of the DM test to the payoffs of the problem.

The partition will be built with the same pattern for all seven discrete losses. The extreme values of the partition will be $l_1 = -m\sigma_y$ and $l_{n-1} = +m\sigma_y$, where σ_y denotes the standard deviation of y_t, and the interval $(-m\sigma_y, +m\sigma_y]$ will be partitioned in r equal-sized regions:

$$-(\infty, m\sigma_y] \quad \underbrace{\{-m\sigma_y \qquad\qquad\qquad 0\}}_{\text{partitioned in } r/2 \text{ equal-sized regions}} \quad \underbrace{\{0 \qquad\qquad\qquad +m\sigma_y\}}_{\text{partitioned in } r/2 \text{ equal-sized regions}} \quad +(m\sigma_y, \infty)$$

In the case of the SDAFE loss, r will take the values $r = 80$ (L_1), $r = 8$ (L_2), $r = 4$ (L_3) and $r = 2$ (L_4). The loss functions L_5 to L_7 use the same partition as L_4:

$$\begin{array}{c|c|c|c} R_1 & R_2 & R_3 & R_4 \\ \hline -(\infty, m\sigma_y] & -(m\sigma_y, 0] & +(0, m\sigma_y] & +(m\sigma_y, \infty) \end{array} \qquad (13)$$

Concerning numerical penalties, L_1 to L_4 will be directly defined by (2) with δ_j given by (3) and θ, v_0, v_n verifying (4), with $v_0 = l_2 - l_1$ and $v_n = l_{n-1} - l_{n-2}$. Loss functions L_5 to L_7 will follow the pattern (6) in the example of the section 2, with different values for a, b, A and B (see Table 1), reflecting how costly is an incorrect forecast of the sign and/or the magnitude of data.

TABLE 1. Specification of discrete loss functions used for simulation exercises

Loss Function	R-partition m: Nonnested models	R-partition m: Nested models	r	Size of R^2-partition	Penalties
L_1	1	0.8	80	82×82	SDAFE (2)-(3)-(4)
L_2	1	0.4	8	10×10	SDAFE (2)-(3)-(4)
L_3	1	0.4	4	6×6	SDAFE (2)-(3)-(4)
L_4	1	0.4	2	4×4	SDAFE (2)-(3)-(4)
L_5	1	0.4	2	4×4	(6) with $a = b = 1$, $A = B = 11$
L_6	1	0.4	2	4×4	(6) with $a = b = 1$, $A = 11$, $B = 21$
L_7	1	0.4	2	4×4	(6) with $a = 1$, $b = 2$, $A = 4$, $B = 6$

As it can be seen in Table 1, the central zone of the partition ($(-m\sigma_y, +m\sigma_y]$) is narrower in the experiments with nested models than in those with nonnested models. We reduce the width of the central zone of the partition reducing the value of m, while maintaining r unchanged. The reason is that forecasts produced by alternative nested models will be very similar to each other and hence, it will often be desirable to use a finer partition in order to capture differences in predictive accuracy between the competing models. In loss functions L_2 to L_7 the width of the central zone for nested-models is 60% lower than in exercises for nonnested models ($m = 0.4$ against $m = 1$). For L_1, the reduction is of only 20% ($m = 0.8$ against $m = 1$), because in that case the partition is already fine by construction.

The seven discrete loss functions verify assumption (a.5). Moreover, L_1 to L_4 also verify assumption (a.4) and hence, the conditions on the loss function required in Corollary 5 hold.

In contrast, L_5 to L_7 verify $(a.6)$, so that for them the conditions relative to the loss function required in Corollary 7 hold.

6.1 Nonnested models

We perform two simulation exercises for nonnested models. In the first one, all the conditions relative to the forecasting models required in the corollaries above hold so that AIPEE is guaranteed, and we evaluate the size of the bias in small samples. The design of our second exercise violates conditions $(b.4)$ and $(b.4')$ that are required in the corollaries, so that AIPEE cannot be shown to hold.

Experiment 1 uses:

$[PGD]$ $y_t = z_{1t} + z_{2t} + \varepsilon_t$, with $(z_{1t}, z_{2t}, \varepsilon_t)' \overset{iid}{\sim} N(0_{3\times1}, I_{3\times3})$.
[M$_1$] $y_t = \beta_{10} + \beta_{11} z_{1t} + u_{1t}$.
[M$_2$] $y_t = \beta_{20} + \beta_{22} z_{2t} + u_{2t}$.
and parameters are estimated by OLS.

Experiment 2 is the same as in section 5.2 in West (1996):

$[PGD]$ $y_t = w_{1t} + w_{2t} + \varepsilon_t$, with $w_{it} = z_{it} + \varepsilon_t$, $(z_{1t}, z_{2t}, \varepsilon_t)' \overset{iid}{\sim} N(0_{3\times1}, I_{3\times3})$.
[M$_1$] $y_t = \beta_{10} + \beta_{11} w_{1t} + u_{1t}$.
[M$_2$] $y_t = \beta_{20} + \beta_{22} w_{2t} + u_{2t}$.

In this latter case, $E(u_{it}|w_{it}) = w_{it}$ [see Appendix C], so that forecast errors u_{it} are correlated with the w_{it} regressors through ε_t and therefore, conditions $(b.4)$ and $(b.4')$ in our corollaries fail to hold. For the same reason, the conditions required to guarantee AIPEE under the SFE or the AFE loss functions also fail to hold in this experiment. West (1996) showed that PEE is very relevant in this setup regarding the properties of the DM test under a SFE loss function, so we can use Experiment 2 to check the effect of PEE on test size. Estimation of parameters β_{i0} and β_{ii} will be made by 2SLS using $A_{it} = (1, z_{it})'$ as instruments to avoid the inconsistency of OLS.

We perform P one-period ahead forecasts with models M$_1$ and M$_2$, starting at $t = R$, following the recursive scheme. In both experiments we will assume that the values of z_{it+1} and w_{it+1}, respectively, are known at time t. We use the same values for R and P as in West (1996), with the P/R ratio oscillating between 0.25 and 7, to perform 5000 repetitions of each forecasting exercise. In each repetition, we implement the DM test for the null hypothesis: $E(f_{1t}) = E(f_{2t})$, which is clearly true by construction in both experiments. The significance level is $\alpha = 5\%$ in all cases, and the test is applied using standard critical values, i.e., using the asymptotic variance in Diebold and Mariano (1995) rather than the variance in West (1996) and McCracken (2000). This procedure will not produce a significant bias in size in Experiment 1 as long as the asymptotic results in Corollary 5 and Corollary 7 are approximately true in finite samples. On the other hand, the result in West (1996) for the SFE loss showed that the bias in size of the DM test could be high in Experiment 2.

Simulation results are summarized in Tables 2 and 3. Table 2 illustrates the fact that the AIPEE property shown in Corollary 5 and Corollary 7 holds as a good approximation in finite samples. On the other hand, the results in Table 3 suggest that the effect on the DM test of ignoring PEE when it is relevant is smaller under a discrete loss than under the AFE or SFE loss functions, even though it remains sizeable in most cases. For $R = 100$, it might be considered admissible to implement the DM test under a discrete loss f ignoring PEE, as it is usually done, so long as we use a partition that is not too fine (like in L_4 to L_7), with only a small bias in size. This is a very convenient result, since the condition $R \geq 100$ is likely to hold in many practical forecasting exercises. In Section 7 we illustrate the reasons for the lower impact of PEE under a discrete loss function. Results in Table 3 suggest that the fineness of the partition plays a key role in the bias of the size of tests (the finer the partition, the bigger the bias), unlike the numeric values of penalties, whose effect on the size of the test is minor.

Our simulation results suggest that with small samples in forecasting environments where the conditions in our corollaries do not hold, the test should be implemented with the asymptotic variance in McCracken (2000), using expression (10) to estimate matrix F . The problem is that such expression requires some assumptions on the probability distributions of the forecast errors and the explanatory variables in the models. The alternative is to apply the numerical methods proposed in McCracken (2004) for estimating F.

TABLE 2. Size of DM Test, ignoring PEE. Experiment 1; Nominal Size: 5%

R	P	π	L_1	L_2	L_3	L_4	L_5	L_6	L_7	SFE	AFE
25	25	1	7.6	7.2	7.0	7.0	6.7	6.8	6.7	6.6	6.8
	50	2	6.8	6.8	6.4	6.4	6.3	6.2	6.5	6.5	6.2
	100	4	5.7	5.8	5.6	5.2	5.2	5.3	5.3	6.0	6.0
	150	6	5.9	6.0	5.5	5.5	5.5	5.2	5.2	5.6	5.9
	175	7	5.6	5.7	5.8	5.6	5.1	4.9	5.3	5.3	5.6
50	25	0.5	7.2	7.4	7.2	6.6	6.8	6.7	6.7	6.2	7.3
	50	1	6.2	6.3	6.3	6.0	5.9	6.2	6.1	5.4	6.1
	100	2	5.6	5.7	5.9	5.6	5.3	5.6	5.7	5.9	6.0
	150	3	5.5	5.4	5.5	5.5	5.5	5.6	5.6	5.4	5.7
100	25	0.25	6.7	6.8	6.8	6.9	6.9	6.7	6.9	6.0	7.0
	50	0.5	6.4	6.4	6.4	6.4	6.4	6.0	5.9	5.9	6.1
	100	1	5.9	6.0	6.0	6.1	6.0	5.8	5.9	5.4	5.6

P: number of forecasts; R: starting sample size for estimation; $\pi = P/R$.

The table shows relative frequencies over 5000 repetitions.

SFE: squared forecast error; AFE: absolute forecast error. See Table 1 for definition of losses L_1-L_7 .

TABLE 3. Size of DM Test, ignoring PEE. Experiment 2; Nominal Size: 5%

R	P	π	L_1	L_2	L_3	L_4	L_5	L_6	L_7	SFE	AFE
25	25	1	27.0	26.1	23.2	19.0	17.5	16.8	18.0	38.1	36.0
	50	2	28.4	27.7	24.9	18.8	16.4	15.2	16.1	44.0	41.2
	100	4	30.0	28.9	25.2	17.1	14.1	13.4	14.6	49.8	45.7
	150	6	29.8	28.6	24.8	15.5	11.4	11.3	12.4	51.5	46.9
	175	7	29.9	28.4	24.2	14.5	11.1	10.6	11.8	52.5	48.0
50	25	0.5	18.2	16.8	15.5	11.9	11.2	10.5	10.9	29.1	27.6
	50	1	20.7	19.6	16.8	11.6	10.2	9.8	10.3	36.8	33.5
	100	2	21.9	20.5	17.9	11.1	8.9	8.1	9.0	43.8	39.1
	150	3	23.4	22.0	19.2	10.8	8.3	8.3	9.3	46.9	41.9
100	25	0.25	12.3	11.6	10.4	8.0	7.5	7.1	7.4	20.6	19.1
	50	0.5	13.3	12.2	10.7	7.4	6.3	6.9	7.3	28.2	25.0
	100	1	15.2	14.6	12.1	7.3	6.1	6.2	6.4	36.0	31.6

See notes for Table 2.

6.2 Nested models

Under a recursive forecasting scheme, the asymptotic distribution of the DM test statistic is known only under certain conditions on the loss function[13] and the forecasting environment. The distribution is non standard except if $\pi = 0$ and the needed conditions can be summarized in practice as *i*) the loss is a SFE function, *ii*) the model is estimated by least squares and *iii*) forecasts are made one-period ahead. These results and the percentiles for the implementation of the test are obtained by McCracken (2007). There does not exist a parallel literature for discrete loss functions, and our goal in this Section is to produce some simulation evidence for such a case.

We specifically want to know how the asymptotic distribution of the DM test looks like under a discrete loss function in the presence of PEE for nested models. To that end, we use again the DM test with the standard critical values, under the same loss functions as in the previous Section. Experiment 3 considers the same design as an exercise used in Clark and McCracken (2001) to evaluate the properties of tests of equal forecast accuracy in this framework, one of them being precisely the DM test under an SFE loss. The DGP is:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0.3 & \delta^* \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{y,t} \\ u_{x,t} \end{pmatrix},$$

with $(u_{y,t}, u_{x,t})' \overset{iid}{\sim} N(0_{2\times 1}, I_{2\times 2})$.

We just want to forecast variable y_t using two models: an AR(1) for y_t (M_1, restricted model) and a VAR(1) for vector $(y_t, x_t)'$ (M_2, unrestricted model). The null hypothesis is $H_0 \equiv E(f_{1t}(\beta_1^*)) \leq E(f_{2t}(\beta_2^*))$, and the alternative, $H_1 \equiv E(f_{1t}(\beta_1^*)) > E(f_{2t}(\beta_2^*))$, with β_1 and β_2 being the parameter vectors to be estimated in M_1 and M_2 , respectively. They differ just in the presence of δ in β_2. These hypothesis can also be written as $H_0 \equiv \delta^* = 0$ and $H_1 \equiv \delta^* \neq 0$. We will restrict $\delta^* = 0$, so that the null hypothesis is true by construction.

Parameters are estimated by least squares, and P one-period ahead forecasts are produced using the recursive scheme, starting at $t = R$. The implementation of the DM test under a discrete loss function is subject to the following problem: since M_1 and M_2 make similar forecasts then, for small values of P we sometimes have realizations for which the P loss differentials $d_t = f_{2t}(\widehat{\beta}_2) - f_{1t}(\widehat{\beta}_1)$ turn out to be equal to zero. Then, both the numerator and denominator of the DM test statistic will be equal to zero. We will discard these realizations.[14]

We perform two types of exercises. On the one hand, the standard exercise of estimating the size of the DM test in finite samples, ignoring PEE. Sample sizes, P and R, will then be exactly equal to those in McCracken (2007), covering values of π usually found in practical applications, between 0.1 and 2.[15] The results of this experiment are shown in Table 4.[16] On the other hand, we want to produce some evidence on the asymptotic distribution of the DM test statistic applied to competing nested models under discrete loss functions when there is PEE. To that end, we repeat Experiment 3 using sufficiently large values of P and R so as to have a good approximation to the asymptotic distribution of the DM test. We will use the same values of π as before. In this exercise, we estimate the asymptotic size of the test under

[13] The functional form of the loss function must be the same as the one that defines the optimization criterion on which the estimator is based.

[14] An alternative way to deal with these realizations would be to not reject the null hypothesis in these cases. For very small forecast samples, then the empirical size would be lower than that obtained under discarding these realizations. Results obtained under this method are available from the authors upon request.

[15] Some of the π values used by Clark and McCracken (2001) are unusual in practice, so we prefer using those in McCracken (2007).

[16] We use now a significance level $\alpha = 10\%$, while in the experiments with nonnested models, we used $\alpha = 5\%$. This change is in coherence with the design of the simulation exercises in the literature in each case. Experiment 3 replicates the exercise of Clark and McCracken (2001), who use $\alpha = 10\%$, whereas Experiment 2 is the same as in West (1996), where $\alpha = 5\%$.

the seven discrete loss functions as well as under the SFE and AFE loss functions (Table 5). We also estimate the empirical density functions for the DM test statistic (Figure 1), to check for deviations with respect to a $N(0,1)$ distribution.

TABLE 4. Size of DM Test, ignoring PEE. Experiment 3; Nominal Size: 10%

R	P	π	L_1	L_2	L_3	L_4	L_5	L_6	L_7	SFE	AFE
50	5	0.1	13.8	14.1	12.3	9.2	7.2	7.0	9.9	13.2	12.6
	20	0.4	6.8	7.3	8.3	8.4	8.3	8.6	8.4	5.1	4.9
	50	1.0	5.0	5.1	5.8	6.5	7.2	7.8	7.2	2.3	2.7
	100	2.0	4.1	4.0	4.5	5.6	7.0	7.3	6.7	1.0	1.5
100	10	0.1	9.5	10.6	10.6	9.8	7.8	7.9	9.5	9.1	8.8
	40	0.4	6.5	7.2	8.1	8.8	8.8	8.9	8.7	4.5	4.5
	100	1.0	5.6	5.7	6.5	7.0	8.0	8.3	7.7	2.1	2.7
	200	2.0	5.1	5.2	5.6	6.9	7.4	7.7	7.4	1.3	2.0
150	15	0.1	8.3	9.7	10.0	9.7	7.6	7.7	8.8	8.1	8.0
	60	0.4	6.4	7.4	8.0	9.3	8.4	8.8	8.7	3.7	4.5
	150	1.0	5.4	5.8	6.6	7.6	8.1	8.3	7.9	1.9	2.5
	300	2.0	4.9	5.0	5.6	6.2	7.6	7.4	7.0	0.9	1.7
200	20	0.1	8.2	9.9	10.3	10.5	8.6	9.0	9.3	7.4	7.0
	80	0.4	7.2	7.4	7.9	9.0	9.6	9.7	9.2	3.5	4.3
	200	1.0	5.9	6.1	6.8	7.7	8.6	8.9	8.2	1.6	2.5
	400	2.0	5.9	5.6	6.3	7.7	8.4	8.0	8.2	0.7	1.7

See notes for Table 2.

TABLE 5. Size of DM Test, ignoring PEE. Experiment 3; Nominal Size: 10%

R	P	π	L_1	L_2	L_3	L_4	L_5	L_6	L_7	SFE	AFE
10000	1000	0.1	9.3	9.2	9.0	9.5	10.1	9.8	9.6	5.4	5.6
2500	1000	0.4	8.3	8.4	8.5	9.0	9.4	9.3	9.4	2.9	3.8
1000	1000	1.0	6.7	6.8	7.4	8.6	8.9	8.8	8.7	1.4	2.6
500	1000	2.0	5.8	5.4	6.6	7.0	7.6	7.7	7.3	0.9	1.6

See notes for Table 2.

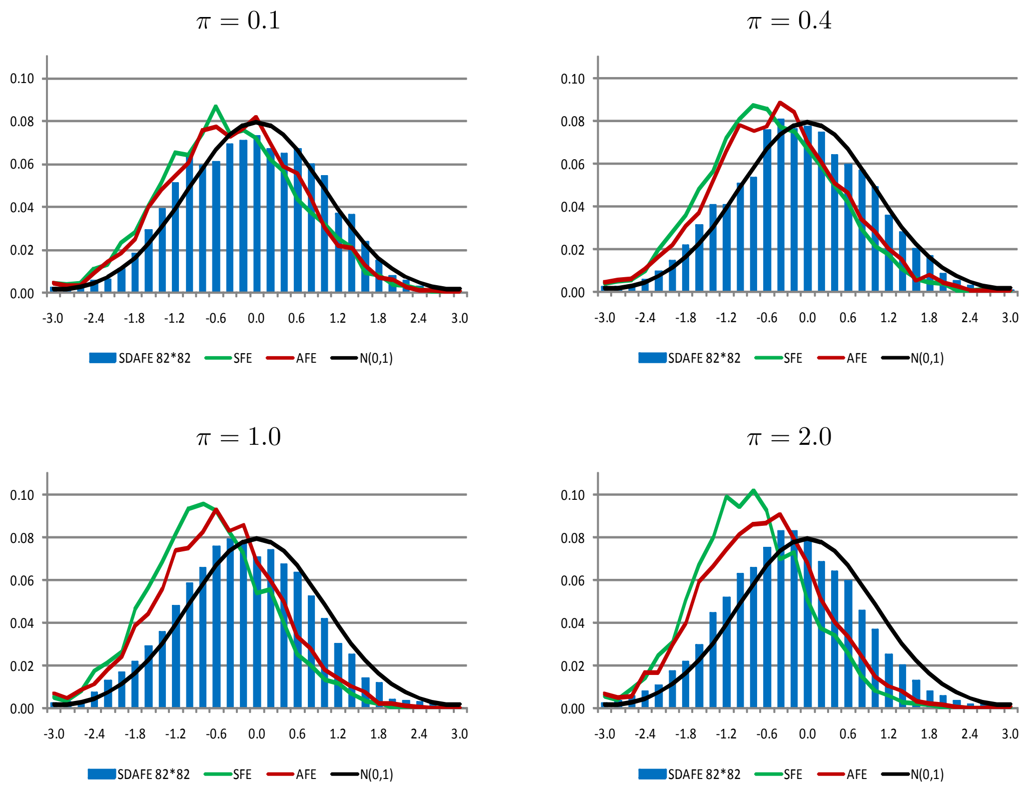
Results in Tables 4 and 5 for the SFE loss are in line with Clark and McCracken (2001) and McCracken (2007), respectively. McCracken (2007) had already shown that the asymptotic distribution of the DM statistic in this context is not Normal. However, for a discrete loss function, the effects on the size of the DM test of ignoring PEE in nested models are less important than if we ignore PEE under the SFE or AFE loss functions. Indeed, the bias in size of the DM test implemented with a discrete loss f is relatively small in cases which are most frequent in practice, those with R large and P small ($R \geq 100$, $\pi \leq 0.4$). This result is encouraging, since we lack analytical results on the asymptotic distribution of the DM test statistic under any loss function other than SFE, and using the standard critical values is the only simple procedure for the user to implement the test in those cases.

McCracken (2007) showed analytically that the DM test satisfies AIPEE under an SFE loss when $\pi = 0$, but the results in Table 5 for $\pi = 0.1$ do not confirm it. The observed pattern is consistent with the percentiles in McCracken (2007) and it is due to the fact that $\pi = 0.1$ is still too large for the theoretical results obtained under $\pi = 0$ for an SFE loss function to apply. On the contrary, results in Table 5 suggest that the AIPEE property is satisfied by the DM test under discrete loss functions too. In these cases, the result holds with sufficient approximation even for $\pi = 0.1$, in contrast to what happens for the SFE loss.

The empirical density functions obtained for large samples throw some light on the asymptotic distribution of the DM statistic for nested models under discrete loss functions. In order to make the picture as clear as possible, we just show the empirical densities for L_1 (82×82), AFE and SFE loss functions. As shown in Figure 1 for the L_1 loss, the empirical distribution for discrete loss functions resembles a $N(0, 1)$ in most cases,[17] at a difference of the empirical densities obtained under SFE and AFE losses. McCracken (2007) had already shown that the asymptotic distribution is not standard in the case of SFE, with its shifting to the left increasing with π, as shown by our empirical densities. Under a discrete loss, the distribution of the DM test in that context shifts to the left for $\pi \geq 1$, but the shifting is much lower than under SFE and AFE losses. This explains the acceptable results shown in Tables 4 and 5 for discrete losses, in spite of the fact that the test was implemented using the standard critical values.

[17] Graphs for other simulation cases are available from the authors upon request. The similarity to $N(0, 1)$ is still clearer for discrete losses with a coarser partition than L_1, but we prefer to show the least favourable of our results.

Figure 1. Empirical asymptotic density of DM statistic under PEE.
Nested Models. Experiment 3. Loss functions: SFE, AFE and 82×82 SDAFE.



In consistency with the conclusions obtained for nonnested models, distortions in the size of the DM test increase with the fineness of the partition defining the discrete loss, although not as much as in the nonnested framework, and they depend very little on the numerical values of penalties.

7 Why is the distribution of the test statistic more robust to PEE under a discrete loss function?

The results from Monte Carlo experiments presented in Section 6 suggest that the effect of PEE on the asymptotic distribution of the test statistic is lower under a discrete loss than under SFE and AFE loss functions. Intuitively, the reason is that under a discrete loss, changes in the numerical value of the estimate affect the value of f only if the implied changes in forecasts move the (data,forecast)-pair to a different quadrant in the partition of R^2. At a difference of what happens under the SFE or AFE loss functions, there will be just a few points in the space of β for which an infinitesimal change in the estimate will imply a change in f, although these changes will be larger than under standard continuous losses.

This argument can be seen analytically in the case of nonnested models, for which only the asymptotic variance (7b), obtained by West (1996) and McCracken (2000), is affected by

PEE. Essentially, PEE affects the variance of f through F, the gradient of $E(f(\beta))$ evaluated at β^*.[18] Under a discrete f, $f(\beta)$ will be a flat function in infinitely many points, with a zero gradient except in points of the space of β that imply a change in the (data,forecast)-quadrant. As a consequence, even when the sufficient conditions for AIPEE do not hold and F is not a zero matrix, F will easily be smaller under a discrete f than under the standard SFE and AFE loss functions and hence the discrepancy between Ω and S_{ff} will also be smaller.

Let us illustrate this argument with an example. The correct asymptotic variance under PEE can be decomposed in three terms: $\Omega = S_{ff} + T_1 + T_2$, with $T_1 = \Pi(FBS'_{fh} + S_{fh}B'F')$ and $T_2 = 2\Pi FV_\beta F'$ [see (7b)]. In the case of a test comparing the predictive ability of two models, Ω, S_{ff}, T_1 and T_2 will be scalars that we can denote as Ω_d, S_d, T_{1d} and T_{2d}. Ω_d was defined in (8)-(9) and the other three elements are defined accordingly, verifying $\Omega_d = S_d + T_{1d} + T_{2d}$. In this section we proceed to estimate them as well as matrix F using the same design and loss functions as in Experiment 2, a case when AIPEE does not hold. Since we want to estimate the asymptotic variance of the test statistic, we will use large values of P and R in the simulations. We will focus on the case $\pi = 1$, with $P = R = 500$.[19] Estimates for matrices B, S_{fh}, V_β, S_{ff} and for the parameter Π are the same as in the exercise in section 5.2 in West (1996) and do not require any special discussion. Furthermore, matrix F is estimated following the analytical expression corresponding to each of the loss functions considered. The expression for F under SFE and AFE loss functions is well known in linear models, while for discrete losses, we will use (10). Details of these estimates can be seen in Appendix C. Due to the chosen design, the matrix F resulting from any f-function in this experiment, can be written as: $F = k_F \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, which makes our exposition easier. Results are shown in Table 6.

The bias of the test size depends on the ratio of variances that arise depending on whether PEE is taken into account or it is ignored, which can be written: $\Omega_d/S_d = 1 + T_{1d}/S_d + T_{2d}/S_d$. The main influence of F on the bias arises through the T_{2d} term, a quadratic function of F, while the T_{1d} term is linear in F, and the effect of F on the bias through the T_{2d} term can be approximated by $|k_F|/S_d^{1/2}$. As an example, $|k_F|/S_d^{1/2}$ is 0.59 in the case of an AFE loss and then T_{2d}/S_d is 1.73. In contrast, for the SDAFE loss function, even when used with a very fine partition (L_1), the value of $|k_F|/S_d^{1/2}$ falls to 0.28 and then T_{2d}/S_d is only 0.38. Since the T_{1d} term is close to zero in all cases, the final result is that under an AFE loss, the parameter estimates add to the variance a term $T_{1d} + T_{2d}$ which is equal to 1.75 times the variance S_d, while the addition to the variance using L_1 is of only 0.39 times S_d. For the case of an SFE loss, the $T_{1d} + T_{2d}$ amounts to 2.26 times the variance S_d. As results in Table 6 show, as we move from L_1 to L_4, the coarser the partition of the discrete loss, the smaller is $|k_F|/S_d^{1/2}$ and so is the bias in test size, in consistency with results in Table 3.

[18] This definition of F is also valid for a differentiable loss f, since the continuity of f guarantees the identity: $F = E\left[\frac{\partial f(\beta)}{\partial\beta}\right]_{\beta=\beta^*} = \left[\frac{\partial Ef(\beta)}{\partial\beta}\right]_{\beta=\beta^*}$.

[19] The results of the analysis would be qualitatively similar for alternative values of π.

TABLE 6. Variance estimates for DM statistic under PEE.
Experiment 2. Case $\pi = 1, P = R = 500$.

Loss functions	k_F	$\|k_F\| \ / \ S_d^{1/2}$	$T_{1d} \ / \ S_d$	$T_{2d} \ / \ S_d$	$\Omega_d \ / \ S_d$
L_1	-0.65	0.28	0.01	0.38	1.39
L_2	-0.68	0.27	0.00	0.36	1.36
L_3	-0.70	0.24	0.00	0.29	1.29
L_4	-0.37	0.10	0.00	0.05	1.05
L_5	-0.12	0.02	0.00	0.00	1.00
L_6	-0.12	0.01	0.00	0.00	1.00
L_7	-0.17	0.04	0.00	0.01	1.01
SFE	-4.01	0.67	0.06	2.20	3.26
AFE	-0.71	0.59	0.02	1.73	2.75

S_d, Ω_d: asymptotic variance of DM statistic in the absence of PEE and under PEE, respectively.
$\Omega_d = S_d + T_{1d} + T_{2d}$; T_{1d} and T_{2d} are products of matrices involving F .
In this experiment, F can be written as a product of k_F and a 2×4 matrix of zeroes and ones.

8 Conclusions

We have analyzed the effect of parameter estimation error (PEE) on the size of tests of predictive ability of linear models with stationary variables under a recursive forecasting scheme. Our analysis has focused on discrete loss functions, which have a wide range of economic applications in decision making frameworks as well as for forecast evaluation under a statistical approach. In the latter case, we have suggested a particular class of discrete loss functions, the SDAFE ("Signed Discrete Absolute Forecast Error"), as an alternative to SFE and AFE loss functions.

For the setups considered by West (1996) and McCracken (2000) for non nested models, we have found three sets of sufficient conditions guaranteeing the asymptotic irrelevance of PEE (AIPEE) for discrete loss functions. Two of them only require quite plausible conditions on the forecasting models, like a zero median (for the first set of conditions) or symmetry around zero (for the second one) for the distribution of forecast errors conditional on regressors, and hold for different loss structures. Most interesting versions of the SDAFE loss function have a structure that leads to AIPEE. The third result is valid for a class of discrete loss functions that may arise in many decision making problems like the one presented in section 3 of this paper, but it requires an additional condition on the explanatory variables of the forecasting models.

Our result is of considerable practical value, because it is not known whether AIPEE holds for continuous loss functions other than the mean square forecast error (SFE) and the mean absolute forecast error (AFE), and it is rather difficult to estimate the asymptotic variance of the test statistics under PEE. Such estimation is rather complicated, specially if the loss function is not differentiable, and it may even be unfeasible if we do not know the details of the estimation methods used for the competing forecasting models. We also present Monte Carlo experiments suggesting that the AIPEE property holds approximately in finite samples under the conditions specified in our three analytical results.

Our second analysis has consisted on simulation exercises to check the effects of PEE in forecasting frameworks where AIPEE does not hold, with a focus on the DM test. First, under correlation between forecast errors and regressors in non nested models. Secondly, for the comparison of two nested forecasting models. Both cases produce similar evidence, suggesting that the distortion in the size of the test is sensibly lower under a discrete loss function than under SFE or AFE loss functions. The magnitude of the distortion depends very much on the fineness of the partition on which the discrete loss is based (the finer the partition, the bigger the size distortion), as well as on the values of P (number of forecasts) and R (the number of in-sample observations used to obtain the first forecast). In contrast, the influence on the size of the test of the numerical values chosen for the discrete loss function is minor. For some of the discrete loss functions we have examined, we have found simulation situations for which the DM test could be applied even in a standard fashion, ignoring PEE, with a negligible bias in size.

A Appendix: Proof of Lemma 1

For the ease of exposition, we eliminate the subindex i that indicates the forecasting model and substitute subindices r and (r) by 1 and 2. Hence, we use the equivalent expressions:

$$z_{1t} = z_{irt};\ Z_{2t} = Z_{i(r)t};\ u_{t+h} = u_{it+h};\ \beta_1 = \beta_{i,r};\ \beta_2 = \beta_{i,(r)}.$$

Then, under assumption $(b.1)$, the definition of a discrete loss f is:

$$\begin{aligned}
&f_{t+h}(\beta) = a_{jk}, \text{ if } l_{j-1} < y_{t+h} \le l_j, \text{ and } l_{k-1} < \widehat{y}_{t+h} \le l_k \Leftrightarrow \\
&\Leftrightarrow l_{j-1} < z_{1t}\beta_1^* + Z_{2t}'\beta_2^* + u_{t+h} \le l_j, \text{ and } l_{k-1} < z_{1t}\beta_1 + Z_{2t}'\beta_2 \le l_k \Leftrightarrow \\
&\Leftrightarrow b_{j-1} < u_{t+h} \le b_j \text{ and } d_{k-1} < z_{1t} \le d_k,
\end{aligned}$$

where

$$b_j = l_j - z_{1t}\beta_1^* - Z_{2t}'\beta_2^* \text{ and } d_k = \frac{l_k - Z_{2t}'\beta_2}{\beta_1}. \tag{14}$$

Before we move to the proof of Lemma 1, we must obtain the analytical expression for $E(f_{t+h}(\beta))$. For cdfs and pdfs we keep the notation introduced at the beginning of Section 5. Furthermore, in what follows, we eliminate the time subindices in $\widehat{y}_{t+h}, y_{t+h}, z_{1t},\ Z_{2t}$, u_{t+h}, f_{t+h}, to obtain:

$$\begin{aligned}
&E(f(\beta)) = \sum_{k=1}^{n}\sum_{j=1}^{n} a_{jk} P(l_{j-1} < y \le l_j, l_{k-1} < \widehat{y} \le l_k) = \\
&= \sum_{k=1}^{n}\sum_{j=1}^{n} a_{jk} \int_{Z_2} \left[\int\limits_{d_{k-1}(\beta_1)}^{d_k(\beta_1)} \left[\int\limits_{b_{j-1}}^{b_j} g_{u|z_1,Z_2}(u|z_1,Z_2)du \right] g_{z_1|Z_2}(z_1|Z_2)dz_1 \right] g_{Z_2}(Z_2)dZ_2 = \\
&= \sum_{k=1}^{n}\sum_{j=1}^{n} a_{jk} S(j,k),
\end{aligned}$$

where

$$S(j,k) = \int_{Z_2} \left[\int\limits_{d_{k-1}(\beta_1)}^{d_k(\beta_1)} \left(G_{u|z_1,Z_2}(b_j|z_1,Z_2) - G_{u|z_1,Z_2}(b_{j-1}|z_1,Z_2) \right) g_{z_1|Z_2}(z_1|Z_2)dz_1 \right] g_{Z_2}(Z_2)dZ_2.$$

Lemma 1 deals with the analytical expression for the generic element in matrix F in McCracken (2000) $\left[\frac{\partial Ef(\beta)}{\partial\beta_1}\right]_{\beta=\beta^*}$, whose analytical expresion holds in this context because of the differentiability of $E(f_{t+h}(\beta))$, condition guaranteed by assumption $(b.3)$.

Proof of Lemma 1. We first apply Leibniz's rule to the $S(j,k)$-terms:

$$\begin{aligned}
&\tfrac{\partial}{\partial\beta_1} S(j,k) = \int_{Z_2} [G_{u|z_1,Z_2}(l_j - d_k\beta_1^* - Z_2'\beta_2^*|d_k, Z_2) - G_{u|z_1,Z_2}(l_{j-1} - d_k\beta_1^* - Z_2'\beta_2^*|d_k, Z_2)] \\
&g_{z_1|Z_2}(d_k|Z_2)\tfrac{\partial d_k}{\partial\beta_1} g_{Z_2}(Z_2)dZ_2 \\
&- \int_{Z_2} [G_{u|z_1,Z_2}(l_j - d_{k-1}\beta_1^* - Z_2'\beta_2^*|d_{k-1}, Z_2) - G_{u|z_1,Z_2}(l_{j-1} - d_{k-1}\beta_1^* - Z_2'\beta_2^*|d_{k-1}, Z_2)] \\
&g_{z_1|Z_2}(d_{k-1}|Z_2)\tfrac{\partial d_{k-1}}{\partial\beta_1} g_{Z_2}(Z_2)dZ_2 = \\
&= -\int_{Z_2} [B(j,k)H_k - B(j,k-1)H_{k-1}]\, g_{Z_2}(Z_2)dZ_2,
\end{aligned}$$

where

$$H_k = g_{z_1|Z_2}(d_k|Z_2)\frac{d_k}{\beta_1},\ B(j,k) = G(j,k) - G(j-1,k) \tag{15}$$

$$\text{and } G(j,k) = G_{u|z_1,Z_2}(l_j - d_k\beta_1^* - Z_2'\beta_2^*|d_k, Z_2).$$

So the generic element $\frac{\partial Ef(\beta)}{\partial\beta_1}$ can be written as:

$$\frac{\partial Ef(\beta)}{\partial\beta_1} = -\int_{Z_2} \sum_{k=1}^{n-1} H_k \left(\sum_{j=1}^{n} B(j,k)(a_{jk} - a_{jk+1}) \right) + H_n \sum_{j=1}^{n} B(j,n)a_{jn} - H_0 \sum_{j=1}^{n} B(j,0)a_{j1} \quad g_{Z_2}(Z_2)dZ_2,$$

But, since $l_0 = -\infty$ and $l_n = +\infty$, we have that $d_0 = -\infty$ and $d_n = +\infty$ and, consequently, from assumption $(b.2)$, we have: $H_n = \lim_{x \to +\infty} H_x = g_{z_1|Z_2}(x|Z_2)\frac{x}{\beta_1} = 0$ and $H_0 = \lim_{x \to -\infty} H_x = g_{z_1|Z_2}(x|Z_2)\frac{x}{\beta_1} = 0$. [20] Hence, since $B(j,k)$ is a bounded function, we have:

$$\frac{\partial Ef(\beta)}{\partial \beta_1} = -\int_{Z_2} \left(\sum_{k=1}^{n-1} H_k Q_k \right) g_{Z_2}(Z_2) dZ_2, \text{ with } Q_k = \sum_{j=1}^{n} B(j,k)(a_{jk} - a_{jk+1}), \tag{16}$$

where $B(j,k)$, H_k , d_k are given by (15) and (14), respectively.

When we evaluate this derivative at $\beta = \beta^*$, we obtain:

$$\begin{aligned}
&d_k^* = d_k(\beta^*) = \tfrac{l_k - Z_2'\beta_2^*}{\beta_1^*},\\
&l_j - d_k^*\beta_1^* - Z_2'\beta_2^* = l_j - \tfrac{l_k - Z_2'\beta_2^*}{\beta_1^*}\beta_1^* - Z_2'\beta_2^* = l_j - l_k,\\
&G^*(j,k) = G_{u|z_1,Z_2}(l_j - l_k | d_k^*, Z_2),\\
&B^*(j,k) = G^*(j,k) - G^*(j-1,k),\\
&Q_k^* = \sum_{j=1}^{n} B^*(j,k)(a_{jk} - a_{jk+1}),\\
&H_k^* = H_k(d_k^*) = g_{z_1|Z_2}(d_k^*|Z_2)\tfrac{d_k^*}{\beta_1^*},
\end{aligned}$$

so that:

$$\left[\frac{\partial Ef(\beta)}{\partial \beta_1}\right]_{\beta=\beta^*} = -\int_{Z_2} \left(\sum_{k=1}^{n-1} H_k^* Q_k^* \right) g_{Z_2}(Z_2) dZ_2.$$

A model may be misspecified by including some variable that does not enter in the true data generating process, i.e., $\beta_1^* = 0$. Even in that case, the value of $\left[\frac{\partial Ef(\beta)}{\partial \beta_1}\right]_{\beta=\beta^*}$ is finite under the assumption $(b.2)$. This happens because $G^*(j,k) = G_{u|z_1,Z_2}(l_j - l_k | d_k^* = \pm\infty, Z_2)$ is a value in $[0,1]$, so that Q_k^* remains a bounded function, and the value of $\left[\frac{\partial Ef(\beta)}{\partial \beta_1}\right]_{\beta=\beta^*}$ will depend on the product $\left[g_{z_1|Z_2}\left(d_k(\beta_1)|Z_2\right)\frac{d_k(\beta_1)}{\beta_1}\right]_{\beta=\beta^*}$. The finite variance assumption on the conditional distribution of z_1 given Z_2 guarantees that: $\lim_{\beta_1 \to 0}\left[g_{z_1|Z_2}\left(d_k(\beta_1)|Z_2\right)\frac{d_k(\beta_1)}{\beta_1}\right] < \infty$, so it is fully justified to consider the partial derivative $\left[\frac{\partial Ef(\beta)}{\partial \beta_1}\right]_{\beta=\beta^*}$ even in this case. ■

B Appendix: Proof of Theorems 2, 4 and 6

To proceed with the following proofs, we use again the notation introduced at the beginning of Appendix A.

Proof of Theorem 2. Using the definition of Q_k^* and $B^*(j,k)$ above, together with assumption $(a.3)$ on the discrete loss function, we have:

$Q_k^* = \sum_{j=1}^{n} B^*(j,k)(a_{jk} - a_{jk+1}) = c_k \sum_{j=1}^{k} B^*(j,k) - c_k \sum_{j=k+1}^{n} B^*(j,k)$. Using the definition of $B^*(j,k)$ in (11b): $B^*(j,k) = G^*(j,k) - G^*(j-1,k)$, and the fact that $G_{u|z_1,Z_2}(-\infty|d_k^*, Z_2) = 0$ and $G_{u|z_1,Z_2}(+\infty|d_k^*, Z_2) = 1$, we have:

[20] A finite variance for the distribution of $z_1|Z_2$ guarantees that the convergence towards zero of that conditional density at both ends of the real line is faster than linear and hence, that the product $g_{z_1|Z_2}(x|Z_2)x$ converges to zero as x goes to $\pm\infty$.

$$Q_k^* = \sum_{j=1}^{n} B^*(j,k)(a_{jk} - a_{jk+1}) = c_k \sum_{j=1}^{k} B^*(j,k) - c_k \sum_{j=k+1}^{n} B^*(j,k) =$$
$$c_k \left[G^*(1,k) - G^*(0,k) + G^*(2,k) - G^*(1,k) + ... + G^*(k,k) - G^*(k-1,k)\right] -$$
$$c_k \left[G^*(k+1,k) - G^*(k,k) + G^*(k+2,k) - G^*(k+1,k) + ... + G^*(n,k) - G^*(n-1,k)\right] =$$
$$c_k \left(2G^*(k,k) - G^*(0,k) - G^*(n,k)\right) = c_k \left(2G^*(k,k) - 1\right).$$

Finally, assumption $(b.4)$ implies: $G^*(k,k) = 1/2$ and hence, $Q_k^* = 0$, for $0 < k < n$. Therefore, $\sum_{k=1}^{n-1} H_k^* Q_k^* = 0 \Rightarrow \left[\frac{\partial Ef_i(\beta)}{\partial \beta_{i,r}}\right]_{\beta=\beta^*} = 0$, for $r = 1,...,K_i$. Furthermore, if the parameter with respect to which we compute the derivative corresponds to a regressor that is not in model M_i, the derivative of $Ef_i(\beta)$ with respect to that parameter is zero, so that $F_i = 0_{1\times K}$. ■

Proof of Theorem 4. For $k \neq n/2$, the same argument above implies $Q_k^* = 0$ under assumption $(a.4)$. For $k = n/2$:

$$Q_{n/2}^* = \sum_{j=1}^{n} B^*(j,n/2)(a_{jn/2} - a_{jn/2+1}) = c_{1,n/2}[G^*(1,n/2) - G^*(0,n/2)] +$$
$$c_{2,n/2}[G^*(2,n/2) - G^*(1,n/2)] + ... + c_{n/2,n/2}[G^*(n/2,n/2) - G^*(n/2-1,n/2)] +$$
$$c_{n/2+1,n/2}[G^*(n/2+1,n/2) - G^*(n/2,n/2)] + c_{n/2+2,n/2}[G^*(n/2+2,n/2) - G^*(n/2+1,n/2)]$$
$$+ ... + c_{n,n/2}[G^*(n,n/2) - G^*(n-1,n/2)].$$

Applying assumption $(a.4)$ for $k = n/2$, the expression for $Q_{n/2}^*$ becomes:

$$Q_{n/2}^* = c_{1,n/2}[G^*(1,n/2) - G^*(0,n/2) - G^*(n,n/2) + G^*(n-1,n/2)] +$$
$$c_{2,n/2}[G^*(2,n/2) - G^*(1,n/2) - G^*(n-1,n/2) + G^*(n-2,n/2)]$$
$$+ ... + c_{n/2,n/2}[G^*(n/2,n/2) - G^*(n/2-1,n/2) - G^*(n/2+1,n/2) + G^*(n/2,n/2)].$$

We can apply assumption $(a.5)$ together with $(b.4')$ to conclude that the expression inside each square bracket is zero. Hence, we also have: $Q_{n/2}^* = 0$. Therefore, $\sum_{k=1}^{n-1} H_k^* Q_k^* = 0 \Rightarrow \left[\frac{\partial Ef_i(\beta)}{\partial \beta_{i,r}}\right]_{\beta=\beta^*} = 0$, for $r = 1,...,K_i$. Besides, if the parameter with respect to which we compute the derivative corresponds to a regressor that is not in model M_i, the derivative of $Ef_i(\beta)$ with respect to that parameter is zero, so that $F_i = 0_{1\times K}$. ■

Proof of Theorem 6. By assumption $(a.6)$, the expression of Q_k^* becomes:

$$Q_k^* = c_k \sum_{j=1}^{k} B^*(j,k) + c_k' \sum_{j=k+1}^{n} B^*(j,k) =$$
$$c_k \left[G^*(1,k) - G^*(0,k) + G^*(2,k) - G^*(1,k) + ... + G^*(k,k) - G^*(k-1,k)\right] +$$
$$c_k' \left[G^*(k+1,k) - G^*(k,k) + G^*(k+2,k) - G^*(k+1,k) + ... + G^*(n,k) - G^*(n-1,k)\right]$$
$$= c_k \left[G^*(k,k) - G^*(0,k)\right] + c_k' \left[G^*(n,k) - G^*(k,k)\right],$$

with $c_k = c$ and $c_k' = c'$ if $k \neq n/2$, while $c_k = c_{n/2}$ and $c_k' = c_{n/2}'$ if $k = n/2$, following the notation in $(a.6)$.

Finally, the properties of a probability distribution function imply $G^*(0,k) = 0$ and $G^*(n,k) = 1$, whereas $G^*(k,k) = 1/2$ by assumption $(b.4)$. So:

$$Q_k^* = \begin{array}{l} 1/2(c + c') \text{ if } k \neq n/2 \\ 1/2(c_{n/2} + c_{n/2}') \text{ if } k = n/2. \end{array}$$

Hence, the expression for $\left[\frac{\partial Ef(\beta)}{\partial \beta_1}\right]_{\beta=\beta^*}$ becomes:

$$\left[\frac{\partial Ef(\beta)}{\partial\beta_1}\right]_{\beta=\beta^*} = -1/2(c+c')\sum_{k=1,k\neq n/2}^{n-1}\int_{Z_2} H_k^* g_{Z_2}(Z_2)dZ_2 \,-\, 1/2(c_{n/2}+c'_{n/2})\int_{Z_2} H_{n/2}^* g_{Z_2}(Z_2)dZ_2. \tag{17}$$

Let us denote $E_k = \int_{Z_2} H_k^* g_{Z_2}(Z_2)dZ_2$, so that: $\left[\frac{\partial Ef(\beta)}{\partial\beta_1}\right]_{\beta=\beta^*} = -1/2(c+c')\sum_{k=1,k\neq n/2}^{n-1} E_k$ $- 1/2(c_{n/2}+c'_{n/2})E_{n/2}$. Applying the definition of H_k^*, $H_k^* = H_k(d_k^*) = g_{z_1|Z_2}(d_k^*|Z_2)\frac{d_k^*}{\beta_1^*}$, we get, for $k = 1, ..., n/2-1$:

$$\begin{aligned}
E_k &= \int_{-\infty}^{0}...\int_{-\infty}^{0} g_{z_1|Z_2}(\tfrac{l_k-Z_2'\beta_2^*}{\beta_1^*}|Z_2)\tfrac{l_k-Z_2'\beta_2^*}{(\beta_1^*)^2}g_{Z_2}(Z_2)dZ_2\\
&+\int_{0}^{\infty}...\int_{0}^{\infty} g_{z_1|Z_2}(\tfrac{l_k-Z_2'\beta_2^*}{\beta_1^*}|Z_2)\tfrac{l_k-Z_2'\beta_2^*}{(\beta_1^*)^2}g_{Z_2}(Z_2)dZ_2\\
&=\int_{0}^{\infty}...\int_{0}^{\infty} g_{z_1|Z_2}(\tfrac{l_k+Z_2'\beta_2^*}{\beta_1^*}|Z_2)\tfrac{l_k+Z_2'\beta_2^*}{(\beta_1^*)^2}g_{Z_2}(-Z_2)dZ_2 \qquad \to I_k^1\\
&+\int_{0}^{\infty}...\int_{0}^{\infty} g_{z_1|Z_2}(\tfrac{l_k-Z_2'\beta_2^*}{\beta_1^*}|Z_2)\tfrac{l_k-Z_2'\beta_2^*}{(\beta_1^*)^2}g_{Z_2}(Z_2)dZ_2 \qquad \to I_k^2\\
&= I_k^1 + I_k^2.
\end{aligned}$$

Taking into account that $l_{n-k} = -l_k$ for $k = 1, ..., n/2-1$ by assumption $(a.5)$, the expression for E_{n-k}, $k = 1, ..., n/2-1$ can be written as:

$$\begin{aligned}
E_{n-k} &= \int_{-\infty}^{0}...\int_{-\infty}^{0} g_{z_1|Z_2}(\tfrac{-l_k-Z_2'\beta_2^*}{\beta_1^*}|Z_2)\tfrac{-l_k-Z_2'\beta_2^*}{(\beta_1^*)^2}g_{Z_2}(Z_2)dZ_2\\
&+\int_{0}^{\infty}...\int_{0}^{\infty} g_{z_1|Z_2}(\tfrac{-l_k-Z_2'\beta_2^*}{\beta_1^*}|Z_2)\tfrac{-l_k-Z_2'\beta_2^*}{(\beta_1^*)^2}g_{Z_2}(Z_2)dZ_2\\
&=\int_{0}^{\infty}...\int_{0}^{\infty} g_{z_1|Z_2}(\tfrac{-l_k+Z_2'\beta_2^*}{\beta_1^*}|Z_2)\tfrac{-l_k+Z_2'\beta_2^*}{(\beta_1^*)^2}g_{Z_2}(-Z_2)dZ_2 \qquad \to I_{n-k}^1\\
&+\int_{0}^{\infty}...\int_{0}^{\infty} g_{z_1|Z_2}(\tfrac{-l_k-Z_2'\beta_2^*}{\beta_1^*}|Z_2)\tfrac{-l_k-Z_2'\beta_2^*}{(\beta_1^*)^2}g_{Z_2}(Z_2)dZ_2 \qquad \to I_{n-k}^2\\
&= I_{n-k}^1 + I_{n-k}^2.
\end{aligned}$$

Assumption $(b.5)$ guaranteeing symmetry around the zero vector for the distribution of vector Z_2 as well as symmetry around zero for the conditional distribution $z_1|Z_2$ implies that $I_k^1 = -I_{n-k}^2$ and $I_k^2 = -I_{n-k}^1$ and so $E_k + E_{n-k} = 0$ for each $k = 1, ..., n/2-1$.

Finally, the expression for $E_{n/2}$ can be written as:

$$\begin{aligned}
E_{n-k} &= \int_{-\infty}^{0}...\int_{-\infty}^{0} g_{z_1|Z_2}(\tfrac{-Z_2'\beta_2^*}{\beta_1^*}|Z_2)\tfrac{-Z_2'\beta_2^*}{(\beta_1^*)^2}g_{Z_2}(Z_2)dZ_2\\
&+\int_{0}^{\infty}...\int_{0}^{\infty} g_{z_1|Z_2}(\tfrac{-Z_2'\beta_2^*}{\beta_1^*}|Z_2)\tfrac{-Z_2'\beta_2^*}{(\beta_1^*)^2}g_{Z_2}(Z_2)dZ_2 =\\
&=\int_{0}^{\infty}...\int_{0}^{\infty} g_{z_1|Z_2}(\tfrac{Z_2'\beta_2^*}{\beta_1^*}|Z_2)\tfrac{Z_2'\beta_2^*}{(\beta_1^*)^2}g_{Z_2}(-Z_2)dZ_2 \qquad \to I_{n/2}^1\\
&+\int_{0}^{\infty}...\int_{0}^{\infty} g_{z_1|Z_2}(\tfrac{-Z_2'\beta_2^*}{\beta_1^*}|Z_2)\tfrac{-Z_2'\beta_2^*}{(\beta_1^*)^2}g_{Z_2}(Z_2)dZ_2 \qquad \to I_{n/2}^2\\
&= I_{n/2}^1 + I_{n/2}^2.
\end{aligned}$$

Again, under assumption $(b.5)$, $I_{n/2}^1 = -I_{n/2}^1$ and so $E_{n/2} = 0$. Consequently, expression (17) is equal to zero $\Rightarrow \left[\frac{\partial Ef_i(\beta)}{\partial\beta_{i,r}}\right]_{\beta=\beta^*} = 0$, for $r = 1, ..., K_i$. Besides, if the parameter with respect to which we compute the derivative corresponds to a regressor that is not in model M_i, the derivative of $Ef_i(\beta)$ with respect to that parameter is zero, so that $F_i = 0_{1\times K}$. ■

C Appendix: Estimation of matrix F for the example in Section 7

In the example described in Section 7 we used numerical estimates for matrix F under the SFE and AFE loss functions as well as under a discrete loss function. We now describe the details of the estimation of matrix F .

In that exercise, $\beta = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{22})'$ so that F is a 2×4 matrix with zeroes in all positions except for $F_{1,2} = \left[\frac{\partial Ef(\beta)}{\partial \beta_{11}}\right]_{\beta=\beta^*}$ and $F_{2,4} = \left[\frac{\partial Ef(\beta)}{\partial \beta_{22}}\right]_{\beta=\beta^*}$. As obtained by West (1996) for a SFE loss, the expressions for those derivatives are $F_{1,2} = -2E(u_{1t+1}w_{1t+1})$ and $F_{2,4} = -2E(u_{2t+1}w_{2t+1})$. On the other hand, McCracken (2000) shows that if f is the AFE loss function, then $F_{1,2} = -E\left[sgn(u_{1t+1})w_{1t+1}\right]$ and $F_{2,4} = -E\left[sgn(u_{2t+1})w_{2t+1}\right]$. Therefore, the estimation for these elements of matrix F in the case of SFE and AFE loss functions are, respectively, $\widehat{F}_{1,2} = -2P^{-1}\sum_{t=R}^{T}\widehat{u}_{1t+1}w_{1t+1}$ an $\widehat{F}_{1,2} = -P^{-1}\sum_{t=R}^{T}sgn(\widehat{u}_{1t+1})w_{1t+1}$. The estimation $\widehat{F}_{2,4}$ is obtained analogously.

On the other hand, the estimation of $F_{1,2}$ and $F_{2,4}$ in the case of a discrete loss f is given by (10). In this case, each model M_i employs a single regressor in addition to a constant term, so that the variables in that expression are $z_{ir} = w_i$ and $Z_{i(r)} = 1$. Therefore, $F_{1,2}$ reduces to $F_{1,2} = -\sum_{k=1}^{n}H_k^*Q_k^*$, with $H_k^* = g_{w_1}(d_k^*)\frac{d_k^*}{\beta_{11}^*}$, $d_k^* = \frac{l_k - \beta_{10}^*}{\beta_{11}^*}$, $Q_k^* = \sum_{j=1}^{n}B^*(j,k)(a_{jk} - a_{jk+1})$, $B^*(j,k) = G_{u_1|w_1}(l_j - l_k|d_k^*) - G_{u_1|w_1}(l_{j-1} - l_k|d_k^*)$ (see (11a)-(11c)). Given the design of the forecasting exercise, g_{w_1} corresponds to the density of a $N(0,2)$. On the other hand, since u_1 and w_1 follow Normal distributions, the distribution of $G_{u_1|w_1}$ will be $N(b(w_1), \sigma_{u_1}^2(1-\rho_{u_1w_1}^2))$, with $b(w_1) = \mu_{u_1} + (w_1 - \mu_{w_1})\rho_{u_1w_1}\frac{\sigma_{u_1}}{\sigma_{w_1}}$, where μ_x, σ_x, ρ_{xy} denote the population mean and standard deviation of x, and the correlation coefficient between x and y, respectively. Substituting for the parameter values in the exercise, we get: $\mu_{u_1} = \mu_{w_1} = 0$, $\sigma_{u_1} = \sqrt{5}$, $\sigma_{w_1} = \sqrt{2}$ and $\rho_{u_1w_1} = \sqrt{\frac{2}{5}}$, so that the distribution of $G_{u_1|w_1}$ is $N(w_1, 3)$. To compute $B^*(j,k)$ that distribution becomes $N(d_k^*, 3)$ since it must be evaluated at $w_1 = d_k^*$, as indicated by the expression for $B^*(j,k)$. To estimate $F_{1,2}$, we use the mentioned distributions, together with the expressions for H_k^* and Q_k^*, once we substitute β_{11}^* by its 2SLS estimation in $t = T$. The estimation of $F_{2,4}$ is obtained analogously.

D Appendix: SDAFE properties

We now show propositions that give us sufficient conditions for the discrete function SDAFE [defined by (2), in Section 2] to verify properties $(a.3)$ and $(a.4)$ introduced in Section 5. Such properties are conditions needed for Theorems 2 and 4, respectively, to apply.

Proposition 8 *If the SDAFE loss function verifies: (4) $\theta = 1/2$; $v_0 = v_n$, (3) $\delta_j = 0$ if $sgn(\widehat{y}_{t+h}) = sgn(y_{t+h})$ and $\delta_j = \max_k\{a_{jk}|\alpha_k\alpha_j > 0\}$, otherwise, and $(a.5)$ $l_{n/2} = 0$ and $l_{n/2-z} = -l_{n/2+z}$, for $z = 1, ..., n/2$, then, SDAFE satisfies $(a.4)$.*

We start by showing some auxiliary lemmas that we will used to show the proposition:

Lemma 8.1 $|\alpha_j - \alpha_k| - |\alpha_j - \alpha_{k+1}| = \begin{matrix} \alpha_k - \alpha_{k+1} \text{ if } j \leq k \\ \alpha_{k+1} - \alpha_k \text{ if } j > k. \end{matrix}$

Proof. Since the α_z function given in (2) [Section 2] is monotonically increasing, we have:

i) If $j \leq k$, then $\alpha_j \leq \alpha_k < \alpha_{k+1}$. Consequently, $|\alpha_j - \alpha_k| - |\alpha_j - \alpha_{k+1}| = -(\alpha_j - \alpha_k) - (-(\alpha_j - \alpha_{k+1})) = \alpha_k - \alpha_j + \alpha_j - \alpha_{k+1} = \alpha_k - \alpha_{k+1}$.

ii) If $j > k$, then $\alpha_j \geq \alpha_{k+1} > \alpha_k$. Consequently, $|\alpha_j - \alpha_k| - |\alpha_j - \alpha_{k+1}| = (\alpha_j - \alpha_k) - (\alpha_j - \alpha_{k+1}) = \alpha_{k+1} - \alpha_k$. ■

Lemma 8.2 Under conditions (4) and $(a.5)$, $\alpha_z = -\alpha_{n-z+1}$, $z = 1, ..., n/2$.

Proof. Given (4) and $(a.5)$, we have:

$$\alpha_z = \begin{array}{l} l_1 - 1/2v_0 = -(l_{n-1} + 1/2v_n),\ z = 1 \\ 1/2l_{z-1} + 1/2l_z = -(1/2l_{n-z} + 1/2l_{n-z+1}),\ 1 < z < n \\ l_{n-1} + 1/2v_n = -(l_1 - 1/2v_0),\ z = n \end{array} = -\alpha_{n-z+1}. \ \blacksquare$$

Proof of Proposition 8. Under definition (2) for the SDAFE loss function in Section 2, and given $l_s = 0$ (with $s = n/2$ or not), we have:

$$a_{jk} = \begin{array}{l} |\alpha_j - \alpha_k| \text{ if } (j \leq s, k \leq s) \text{ or } (j > s, k > s) \\ |\alpha_j - \alpha_k| + \delta_j \text{ else.} \end{array}$$

and

$$a_{jk+1} = \begin{array}{l} |\alpha_j - \alpha_{k+1}| \text{ if } (j \leq s, k+1 \leq s) \text{ or } (j > s, k+1 > s) \\ |\alpha_j - \alpha_{k+1}| + \delta_j \text{ else.} \end{array}$$

Therefore:

a) If $k \neq s$: whether a_{jk} and a_{jk+1} both include δ_j, or neither one of them does. Hence, applying Lemma 8.1, we have:

$$a_{jk} - a_{jk+1} = \begin{array}{l} \alpha_k - \alpha_{k+1} = -c_k \text{ if } j \leq k \\ \alpha_{k+1} - \alpha_k = +c_k \text{ if } j > k. \end{array}$$

b) If $k = s$: we have:

$$a_{js} - a_{js+1} = \begin{array}{l} \alpha_s - \alpha_{s+1} - \delta_j \text{ si } j \leq s \\ \alpha_{s+1} - \alpha_s + \delta_j \text{ si } j > s. \end{array}$$

If $s = n/2$, assumption $(a.4)$ holds if and only if $\delta_j = \delta_{n-j+1}$, for $j = 1...n/2$. We just need to show that SDAFE verifies $\delta_j = \delta_{n-j+1}$ ($j = 1...n/2$) under conditions (4), (3) and $(a.5)$, with $\delta_j = \max\{a_{j1}, ..., a_{jn/2}\}$ for row $j \in [1, n/2]$ and $\delta_{n-j+1} = \max\{a_{n-j+1,n/2+1}, ..., a_{n-j+1,n}\}$ for row $n - j + 1 \in [n/2 + 1, n]$. Given the definition (2) of SDAFE, we have $\delta_j = \max_k\{|\alpha_j - \alpha_k||k = 1, ..., n/2\}$ for $j \in [1, n/2]$. Let k^* be the value of k maximizing such function on the mentioned set.

On the other hand, $\delta_{n-j+1} = \max_k\{|\alpha_{n-j+1} - \alpha_k||k = n/2 + 1, ..., n\}$. We would like to find the value k^{**} maximizing the function in this other set. But, $\{|\alpha_{n-j+1} - \alpha_k||k = n/2 + 1, ..., n\} = \{|\alpha_{n-j+1} - \alpha_{n-k+1}||k = 1, ..., n/2\}$. From Lemma 8.2 we have for each $k = 1, ..., n/2$ that $\alpha_j - \alpha_k = -(\alpha_{n-j+1} - \alpha_{n-k+1})$. Therefore, $k^{**} = n - k^* + 1$ y $\delta_{n-j+1} = |\alpha_{n-j+1} - \alpha_{n-k^*+1}| = |\alpha_j - \alpha_{k^*}| = \delta_j$. ■

Proposition 9 *The SDAFE loss function verifies assumption* $(a.3)$ *if* $\delta_j = \overline{\delta} \geq 0$ *for all* j*, under a partition that verifies* $l_s = 0$ *for some* $0 < s < n$.

Proof. The same proof as for Proposition 8 applies here up to point b). Now, for the case $k = s$ we have: $a_{jk} - a_{jk+1} = \begin{array}{l} \alpha_k - \alpha_{k+1} - \overline{\delta} = -c_k \text{ si } j \leq s \\ \alpha_{k+1} - \alpha_k + \overline{\delta} = +c_k \text{ si } j > s. \end{array}$ ■

References

[1] Blaskowitz, O. and Herwartz, H. (2011). On Economic Evaluation of Directional Forecasts, *International Journal of Forecasting*, in press.

[2] Clark, T. and McCracken, M.W. (2001). Tests of Equal Forecast Accuracy and Encompassing for Nested Models, *Journal of Econometrics* 105, 85-110.

[3] Clark, T. and McCracken, M.W. (2010). Testing for Unconditional Predictive Ability, Working Paper 2010-031 A, Federal Reserve Bank of St. Louis.

[4] Clark, T. and West, K.D. (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models, *Journal of Econometrics* 138, 291-311.

[5] Corradi, V., Swanson, N. and Olivetti, C. (2001). Predictive Ability with Cointegrated Variables, *Journal of Econometrics* 104, 315-358.

[6] Diebold, F.X. and Mariano, R. (1995). Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 253-263.

[7] Giacomini, R. and White, H. (2006). Tests of Conditional Predictive Ability, *Econometrica*, 74, 1545-1578.

[8] Granger, C.W.J. and Machina, M.J. (2006), Forecasting and Decision Theory, in Handbook of Economic Forecasting, Elliot G., Granger C.W.J., Timmermann A. (eds), North Holland.

[9] Granger, C.W.J. and Pesaran, M.H. (2000). Economic and Statistical Measures of Forecast Accuracy, *Journal of Forecasting*, 19, 537-560.

[10] McCracken, M.W. (2000). Robust Out of Sample Inference, *Journal of Econometrics*, 99, 195-223.

[11] McCracken, M.W. (2004). Parameter Estimation and Tests of Equal Forecast Accuracy between nonnested Models, *International Journal of Forecasting*, 20, 503-514.

[12] McCracken, M.W. (2007). Asymptotics for Out of Sample Tests of Granger Causality, *Journal of Econometrics*, 140, 717-752.

[13] Pesaran, M.H. and Skouras, S. (2002). Decision-based Methods for Forecast Evaluation, in A companion to Economic Forecasting, Clements, M.P. and Hendry, D.F. (eds), Oxford, Blackwell Publishing.

[14] Pesaran, M.H. and Timmermann, A. (2009). Testing Dependence among Serially Correlated Multi-category Variables", *Journal of the American Statistical Association*, 104 (485), 325-337.

[15] Rossi, B. (2005). Testing Long-Horizon Predictive Ability with High Persistence, and the Meese-Rogoff Puzzle, *International Economic Review*, 46, 61-92.

[16] West, K.D. (1996). Asymptotic Inference about Predictive Ability, *Econometrica*, 64, 1067-1084.

[17] West, K.D. (2006), Forecast Evaluation, in Handbook of Economic Forecasting, Elliot G., Granger C.W.J., Timmermann A. (eds), North Holland.