Working Paper 2008 | 35

Forecasting Inflation in Argentina: Individual models or Forecast Pooling?

Laura D´Amato / Lorena Garegnani / Emilio Blanco
BCRA

July, 2008

ie | BCRA
INVESTIGACIONES ECONÓMICAS

Banco Central de la República Argentina
ie | Investigaciones Económicas

July, 2008
ISSN 1850-3977
Electronic Edition

Reconquista 266, C1003ABF
C.A. de Buenos Aires, Argentina
Phone: (5411) 4348-3719/21
Fax: (5411) 4000-1257
Email: investig@bcra.gov.ar
Web Page: www.bcra.gov.ar

Forecasting Inflation in Argentina: Individual Models or Forecast Pooling?*

Laura D'Amato
BCRA

Lorena Garegnani
BCRA

Emilio Blanco
BCRA

July 2008

Abstract

Inflation forecasting plays a central role in monetary policy formulation. At the same time, recent international empirical evidence suggests that with the decline in inflation of recent years, the joint dynamics of this variable and its potential predictors has changed and inflation has become more unpredictable. Using a univariate model as a benchmark, we evaluate the predictive capacity of certain causal models linked to different inflation theories, such as the Phillips Curve and a monetary VAR. We also analyze the predictive power of models that use factors that combine the overall variability of a large number of business cycle time series as predictors. We compare their relative performance using a set of parametric and non-parametric tests proposed by Diebold and Mariano (1995). Although the univariate model performs best, as the forecast horizon lengthens, multivariate models performance improves. In particular, a monetary VAR performs better than the univariate ARMA model in the case of a one-year horizon. Nevertheless, when tests are calculated to evaluate the statistical significance of differences in the predictive capacity of models, taking a univariate ARMA model as a benchmark, differences are not statistically significant. Finally, estimated models are pooled to forecast inflation. Some of the forecast combinations outperform the best individual forecast over a one-year horizon. Taking into account that a one year-horizon is relevant for economic policy decisions, the possibility of combining both univariate and multivariate models for forecasting purpose is interesting, because it it can also be helpful to answer specific economic policy questions.

JEL Classification: C32, E31, E37

*The opinions expressed in this work are those of the authors, and do not necessarily reflect the opinions of the Central Bank of Argentina or its authorities. Correspondence: ldamato@bcra.gov.ar

1 Introduction

Inflation forecasting plays a central role in monetary policy formulation. Recent international empirical evidence suggests that with the decline in inflation of recent years, a fairly widespread phenomenon, the combined dynamics of this variable and its potential predictors, such as money or different measures of the output gap, has changed, and inflation has become more unpredictable. Univariate models tend to show a better forecasting capacity than those based on various inflation theories, such as the Phillips curve. Traditionally, in industrialized countries the Phillips curve has played a predominant role in inflation forecasting, and according to Stock and Watson (1999), Atkenson and Ohanian (2001) and Canova, (2002), it would seem to perform better in terms of forecasting error than other alternative models. In recent years there have been indications, in the United States in particular, that the Phillips curve became unstable as from the eighties, and that perhaps for this reason, its forecasting ability has weakened, in general being overcome by univariate models.

Clements and Hendry (2006) suggest that this difficulty with causal models can be linked to the presence of changes in regime that mainly affect the deterministic components of the models, and propose a modeling strategy based on a battery of models to help overcome this difficulty. Another strategy in the forecasting literature is to work with combinations or the pooling of forecasts, a theory developed initially by Bates and Granger (1969). Hendry and Clements (2002) studied pooling or forecast combinations for non-stationary models because of breaks in the intercept or in a deterministic trend, and found that a simple average of forecasts can counteract the instability of individual forecasts for plausible parameter values, acting as a correction for the intercept. An alternative strategy is to use models that incorporate a very large number of predictors, as proposed by Stock and Watson (1999, 2002, 2006) among others, by means of the use of statistical techniques that help constructing summarized measures (factors) of their joint variability.

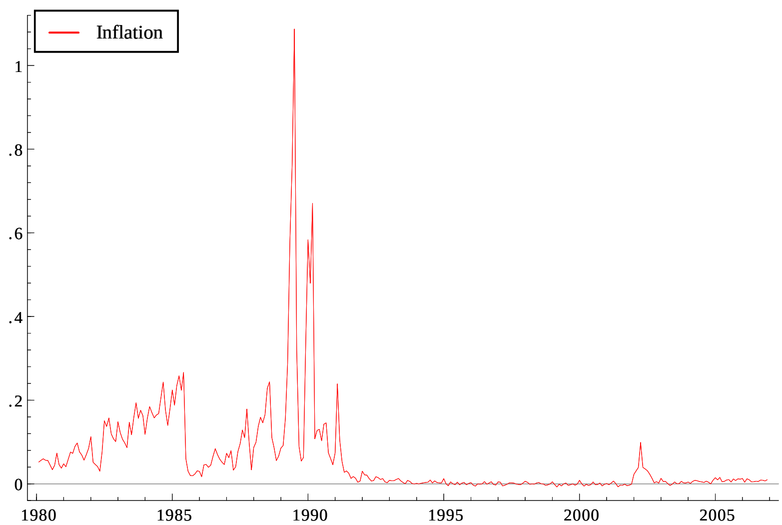
We evaluate the performance of various inflation-forecasting models for Argentina, some of which make use of these forecasting techniques. We also compare their relative performance using a series of tests proposed by Diebold and Mariano (1995). We are not aware of any previous work on inflation forecasting in Argentina, comparing the performance of alternative models. Using a univariate model as a benchmark, we evaluate the predictive power of certain causal models associated with different inflation theories, such as the Phillips curve and a monetary VAR. We also study the predictive capacity of factor models. On the basis of the break points identified in D'Amato, Garegnani and Sotes (2007), we focuse on the search for models with a good forecasting performance, restricting the analysis to the 1993-2006 period.

The paper is organized as follows: the following section briefly describes the dynamic of inflation in Argentina; section 3 reviews recent developments in forecasting literature; section 4 presents results in relation to the predictive power of the models. In section 5 we conclude.

2 Inflation dynamics in Argentina

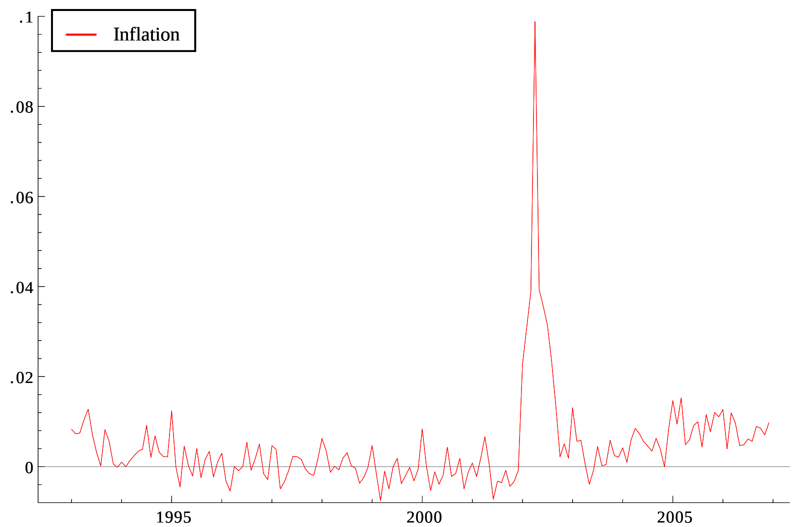
The dynamics of inflation in Argentina has had its own particular characteristics, although sharing some aspects in common with the rest of the region. During the seventies and eighties, inflation was a fairly widespread phenomenon in Latin America. Monetary financing of fiscal imbalances was a common feature of all these inflationary experiences. In Argentina inflation remained at very high levels during the eighties, despite successive attempts at stabilization. Towards the end of the decade this dynamic evolved into a hyperinflationary episode. In 1991 a currency board system was adopted (known as Convertibility) that was successful in causing a permanent reduction in the inflation rate. Inflation began to stabilize at a low level by 1993. Although the change in the exchange system was perceived to be permanent and inflation remained at very low levels, the inflationary tax was replaced by the issue of debt on international markets, and fiscal reform remained pending. As a result of both fiscal and external imbalances and a persistent real appreciation of the currency, an acute financial and external account crisis broke out in 2001 that led to the abandoning of the Convertibility regime in January 2002. Devaluation caused an abrupt change in relative prices and a consequent increase in the inflation rate, which peaked in April 2002, to then return to levels close to those of the Convertibility period. Despite remaining at low levels, inflation began to accelerate slightly towards the end of 2004, when the economy embarked on a period of strong growth.

Figure 1



Simple observation of the evolution of inflation over time suggests the presence of structural break points that hinder the finding of a model with stable parameters for forecasting purposes when considering the whole sample. In this regard, D'Amato, Garegnani and Sotes (2007) provide evidence of significant changes in both the mean and the autoregressive component of inflation between periods of high and low inflation. Although by restricting the sample to the period of low inflation it is possible to determine a break point in 2002, after the abandoning of Convertibility, the size of the jump in inflation and its return to levels closer to those of the previous period allow the crisis period to be treated as a temporary change in the mean that is partially reversed, with inflation showing a mean slightly above that of Convertibility in the 2003-2006 period. As a result, the low-inflation period between 1993 and 2006 can be modeled for forecasting purposes, controlling for that break point.

Figure 2



We evaluate models for different forecasting horizons with relevance for monetary policy: one year is the usual horizon for the setting of monetary policy targets, whether for the inflation rate, or as currently in Argentina, for expansion of a relevant monetary aggregate. Model forecasting performance is also examined for shorter periods: 6 and 3 months, also relevant for monitoring the cyclical position of the economy and for monetary policy decisions.

3 Forecasting models in the literature

Causal econometric models often provide a satisfactory representation of the data-generating process (DGP) in terms of the behavior suggested by economic theory. These models do however tend to perform poorly when forecasting relevant time series, compared with autoregressive models. One reason for this is that the latter tend to respond better to unanticipated changes in the data-generating process, given their intrinsically adaptive nature.

In recent years, forecasting literature has made progress in several directions in order to deal with these difficulties. On the one hand, authors such as Clements and Hendry (2006) propose a battery of forecasting models that take into account break points in the mean and changes in deterministic trend. In recent years models employing a large number of predictors for forecasts are widely used, following two: (i) forecast pooling, which combines a considerable number of models using different weighting criteria (ii) use of factor models, which make it possible to find summarized measures of the variability of a large number of relevant economic cycle series. In the first case, the path chosen aims to preserve the causal models and eventually achieve better forecasts by expanding the group of predictors. In the second case, a large set of business economic cycle indicators is considered, and by means of multivariate statistical techniques, a reduced number of factors underlying those series is extracted that explain a significant portion of their variability. Empirical evidence indicates that these variables add relevant information.

3.1 Pooling of forecasts

The pooling or combination of forecasts implies combining two or more forecasts derived from models that use different predictors to produce a forecast. This technique was originally developed by Bates and Granger (1969), and the basic idea is as follows:[1]

Let $\left\{Y^h_{i,t+h}, i = 1, ..., n\right\}$ be a panel of n forecasts. The combined forecast or forecasting pool will be given by the linear combination

$$Y^h_{t+h/t} = w_0 + \sum_{i=1}^{n} w_{it} Y^h_{i,t+h/t}$$

where w_{it} is the weight of the i^{th} forecast in period t.

Bates and Granger (1969) show that the weights that minimize the means squared forecast error (MSFE) are given by the projection to the population of $Y^h_{t+h/t}$ in a constant and the individual forecasts. Frequently the constant is omitted, and by imposing $\sum_{i=1}^{n} w_{it} = 1$ it is determined that if each of the forecasts is unbiased, so is $Y^h_{t+h/t}$. As long as none of the forecasts is generated by the real model, the optimal combination of forecasts spreads the weight over a multiple combination of forecasts. The minimum $RMSE$ combining those forecasts will be variable over time if the variance and covariance matrixes for $\left(Y^h_{t+h/t}, \left\{Y^h_{i,t+h/t}\right\}\right)$ change over time.

In practice, optimal weightings are not viable because the variance and covariance matrixes are unknown. Granger and Ramanathan (1984) propose estimating weights using minimum least squares or restricted least squares, if $w_0 = 0$ and $\sum_{i=1}^{n} w_{it} = 1$ is imposed, although if n is large it is expected that

[1] A detailed description of forecast pooling techniques and the principal developments contained in this literature can be found in Stock and Watson (2006), and in even greater detail in Timmerman (2006).

estimates will perform poorly, simply because by estimating a large number of parameters, uncertainty is introduced into the sample. If n is proportionate to the size of the sample, the OLS estimator is not consistent, and the combinations that use it are not asymptotically optimum. For this reason, research into the combination or pooling of forecasts has concentrated on imposing greater structure on the combination of forecasts. Possible techniques used include:

(i) *Simple combination of forecasts*, which provides a measure of the distribution center of the panel of forecasts. Weights are distributed equally, that is to say, $w_{it} = 1/n$. The mean combination or truncated mean are simple combinations that are less sensitive to the presence of extreme observations.

(ii) *Weights based on the root mean squared error (RMSE)*; in this case the combined forecast uses weightings that assign weight to forecasts inversely dependent on their discounted RMSE

$$w_{it} = m_{it}^{-1} / \sum_{j=1}^{n} m_{jt}^{-1}, \quad where \;\; m_{it} = \sum_{s=T_0}^{t-h} \rho^{t-h-s} \left(Y_{s+h}^{h} - Y_{s+h/s}^{h} \right)^2$$

where ρ is a discount factor.

Here we use a variant of the weights based on the RMSE proposed by Marcellino (2002).

(iii) *Forecast shrinkage*, a technique involving a convergence of weighting factors on an a priori value that usually tend to have equal weight.
Diebold and Pauly (1990) propose using $w_{it} = \lambda w_{it} + (1 - \lambda)(1/n)$where w_{it} is the estimated coefficient in an OLS regression of Y_{s+h}^{h} in ,......., $Y_{n,s+h/s}^{h}$ for $s = T_0, ...t - h$, where T_0 is the first date of the combined forecast and λ controls by the degree of shrinkage towards equal weights.

3.2 Dynamic Factor Models

The development of *Dynamic Factor Models* is based on factor analysis and principal components, longstanding techniques in multivariate statistical analysis. The idea underlying these techniques is that covariance between a large number of n economic time series with their leads and lags can be represented by a reduced number of unobserved q factors, with $n > q$ Disturbances in such factors could in this context represent shocks to aggregate supply or demand.

Therefore, the vector for n observables in the cycle can be explained by the distributed lags of q common factors plus n idiosyncratic disturbances which could eventually be serially correlated, as well as being correlated among i

$$X_{it} = \lambda_i(L)' f_t' + u_{it} \tag{1}$$

Where f_t is a vector $q\times$ 1 of unobserved factors, λ is a $q\times$ 1 vector lag polynomials of *dynamic factor loadings* and the u_{it} are the idiosyncratic disturbances that are assumed to be uncorrelated with the factors in all leads and lags, that is to say $E(f_t u_{it}) = 0$ for all i, s.

The objective is therefore to estimate $E(y_{t+1}/X_t)$ modeling y_{t+1} according to

$$y_{t+1} = \beta_t' F_t + \varepsilon_{t+1} \tag{2}$$

If the lag polynomials $\lambda_{it}(L)$ in (1)and $\beta(L)$ in(2) are of finite order p, Stock and Watson (2002a) show that the factors F can be estimated by principal components and y_{t+1} can be modeled as

$$Y_{t+1} = \beta F_t + \varepsilon_{t+1} \tag{3a}$$

where $F_t = \left[f_t', f_{t-1.}' ... f_{t-p}'\right]$ is a vector of dimension $r = (p+1)\overline{r}$ and p is the maximum number of lags and $\overline{r}$ the predefined number of factors to be extracted from the data. A brief description of the principal component multivariate statistical technique is presented in *Appendix A.*

4 Empirical results

Given the major changes experienced by trend inflation rate in Argentina and the volatility associated with this phenomenon, we consider a sample contained within 1993:1 and 2006:12 for the forecast. Although this period includes two sub-periods that are fairly different both in terms of the monetary regime in force and the inflation dynamics, we consider that such changes were not large enough to significantly hinder the inflation estimate and forecast.

We evaluate the predictive power of a set of forecasting models that includes: an $ARMA$ model and various multivariate models, a bivariate VAR monetary model, a hybrid Phillips curve, and two models that project inflation on the basis of factors obtained using the principal components method.

In general the lag structure was chosen following the conventional criteria of Akaike and Schwarz.

To evaluate the forecasting performance of the models we calculated the root mean squared errors ($RMSE$), the absolute mean error (MAE), the mean absolute percentage error ($MAPE$) and also the $RMSE$ ratio for each model regarding the $ARMA$ model selected as a benchmark.

We find that an $ARMA(1, 12)$ is an adequate representation of inflation in the 1993 -2006 period. Dummy variables for outliers.

We estimate two causal models that incorporate variables reflecting alternative inflation theories. On the one hand, a Hybrid New-Keynsian Philips Curve[2] specified for a small open economy, incorporating as inflation predictors nominal depreciation and a measure of international inflation, in addition to the output gap, as well as expected inflation and a backward looking term.

Furthermore, we estimate a monetary VAR model that includes inflation and the change in $M2$ monetary aggregate, as a measure of transactional money. In this case, the model reflects the notion that money should be a determinant of the rate of inflation in the long term.

In addition to these models we estimate factor models, following the methodology described in Section 2.1. The factors obtained by means of the principal components method[3] summarize the combined variability of a large number of business cycle indicators grouped according to whether they involved time series concerning aggregate demand in demand factors, those connected with supply in supply factors, a series of nominal indicators that include monetary aggregates, prices, interest rates and tax revenue, among others, and a summary indicator that includes all of these business cycle time series, which we name as the total factor. Only models estimated on the basis of total and nominal factors were satisfactory in explaining the inflation dynamics.[4] In general no use was made of series beyond the fourth principal component, as suggested by literature on the subject. The type of model estimated is

$$Y_{t+1} = \beta F_t + \epsilon_{t+1} \tag{3b}$$

where $F_t = \left[f'_t, f'_{t-1.}...f'_{t-p}\right]$ is a vector of dimension $r = (p+1)\overline{r}$ where p is the maximum number of lags, and $\overline{r}$ the previously-defined number of factors to be extracted from the data, which in our case is 4.

A detail of the models estimated is shown in *Appendix C*.

[2] Galí and Gertler (1999) propose a Hybrid New-Keynsian Philips curve, incorporating a backward looking term. The model used here for forecasting was estimated by D'Amato and Garegnani (2006) and is in line with that of Galí and Getler, although extended to cover the case of a small open economy.

[3] See *Appendix A* for a detailed description of the principal components method.

[4] See *Appendix B* for a description of the series that were considered to obtain the different factors.

4.1 Evaluating estimated models predictive performance

To compare the performance of the different models we report four types of statistics: root mean squared errors ($RMSE$), absolute mean errors (MAE), mean absolute percentage error ($MAPE$) and the $U-Theil$ ratio, which compares the $RMSE$ for each model against the best univariate model chosen as a benchmark. The results are shown on *Table 1*.

Table 1: Predictive performance of individual models

		RMSE			
Forecast period	ARMA(1,1)	Phillips-GMM Curve	VAR	Total factors	Nominal factors
2006:10-2006:12	0.0027	0.0036	0.0031	0.0034	0.0046
2006:7-2006:12	0.0026	0.0039	0.0028	0.0034	0.0043
2006:1-2006:12	0.0033	0.0037	0.0032	0.0036	0.0039
		MAE			
Forecast period	ARMA(1,1)	Phillips Curve	VAR	Total factors	Nominal factors
2006:10-2006:12	0.0024	0.0035	0.0031	0.0029	0.0043
2006:7-2006:12	0.0022	0.0033	0.0027	0.0028	0.0038
2006:1-2006:12	0.0027	0.0032	0.0025	0.0027	0.0031
		MAPE			
Forecast period	ARMA(1,1)	Phillips Curve	VAR	Total factors	Nominal factors
2006:10-2006:12	27.33	43.69	36.67	32.96	49.61
2006:7-2006:12	26.93	40.49	35.32	34.95	47.74
2006:1-2006:12	36.15	45.08	31.39	31.51	36.67
		U-Theil statistic			
Forecast period	ARMA(1,1)	Phillips Curve	VAR	Total factors	Nominal factors
2006:10-2006:12	1	1.33	1.15	1.23	1.67
2006:7-2006:12	1	1.53	1.09	1.30	1.66
2006:1-2006:12	1	1.12	0.97	1.08	1.18

One immediate question that arises is whether the univariate model systematically outperforms causal and multivariate models. The answer is somewhat mixed. Over a very short-term horizon, such as a quarter or a six-month period, the $ARMA$ model systematically outperforms the rest of the models ($U-Theil$ statistic exceeds 1). When the horizon is extended to one year, monetary VAR performs better than the $ARMA$ ($U-Theil$ 0.97). In the case of a one-year horizon the monetary VAR is very close to the best univariate model, and is better than the $ARMA$ model.

In short, results indicate that there is complementarity between forecasting models that can be exploited. Univariate models perform very well at very short horizons. As the forecasting horizon is extended, multivariate models, both theorethically based and atheorethical become closer to the univariate model performance and eventually outperform it.

4.2 Comparing predictive performance

Evaluation of the predictive capacity of forecasting models is important not only because in general forecasts are a fundamental element for policy-making decisions, but also because they imply a choice between alternative economic hypotheses.

As indicated by Diebold and Mariano (1995), a review of the empirical literature on forecasting reveals that evaluation of the forecasting performance of alternative models is usually based on comparison of specific estimates, without any evaluation of the uncertainty of the sample. The use of statistical tests is difficult, as there are usually problems of serial or contemporary correlation between forecasts and forecast errors.

Diebold and Mariano propose a series of tests to evaluate the null hypothesis of equal forecast accuracy of two alternative forecast methods. These tests are based on the evaluation of the presence of significant differences between the models and the data. The tests proposed by Diebold and Mariano in some cases allow non-normal forecast errors, serially correlated errors as well as contemporaneously correlated errors between models.

We will evaluate here the models described in the previous sub-section using some of the tests proposed by Diebold and Mariano. Because the forecast exercise carried out considers horizons not extending beyond one year, evaluation for all horizons is based principally on non-parametric tests. In the case of one-year forecasts evaluation has also been made of the hypothesis that there is no difference in predictive power, using parametric tests.

In many applications in which comparison is made of the predictive performance of different models, the loss function of each model can be represented by the forecasting error or as a direct function of it. In this study the forecasting error itself is considered as a loss function of the respective model. The null hypothesis of equal forecast accuracy for two forecasts will be the differential between the two forecast errors. For the three forecast horizons considered, tests are used adapted to the case of a small number of forecast-error observations. Two of these tests are based on observed loss differentials (the sign test) or in their ranks (Wilcoxon signed-rank test). In the one-year case the mentioned tests are complemented by parametric tests provided in the literature: a simple F test and the Morgan-Granger-Newbold (MGN) test. In the case of these two latter tests the loss function is defined as the square of the distance between the two forecasting errors.[5]

Non-parametric sign and Wilcoxon signed-rank tests make it possible to work with all horizons. *Table 2* shows the number of positive differences observed for forecasts of all the models compared with the $ARMA$ model. For example, in the case of the Phillips curve for a one-year horizon, the number 5 in the sign test column indicates that of the 12 forecasting error differences recorded by the Phillips curve compared with the $ARMA$, 5 are positive. According to a binomial distribution with parameters $T = 12$ and $1/2$, under the null hypothesis, the two models do not differ as regards their predictive performance. This result is repeated for the rest of the models and forecast horizons. In the case of the Wilcoxon signed-rank test, the studentized version enables use of the normal standard distribution, and at 5% of significance the hypothesis of equal predictive capacity is not rejected, with the exception of the nominal factors model for the one-year horizon. Rejection implies in this case that the outperforms the predictive capability of nominal factors.[6]

Parametric tests require the fulfillment of certain assumptions in relation to forecast errors: (i) the loss function needs to be quadratic; (ii) forecast errors must be zero mean; (iii) their distribution should be normal; (iv) they should not be serially correlated and (v) they should not be contemporaneously correlated among themselves. The MGN test permits contemporaneous correlation between forecast errors. In the case of the F-test, when the values of the observed test statistic are compared with the critical values for an $F(12, 12)$, it can be seen that the null hypothesis at 5% is not rejected. The results of the MGN test are not very different when the null hypothesis is evaluated considering the Student-t distribution with $T - 1$ degrees of freedom. This means that for both tests the predictive capacity of the models is not statistically different from the $ARMA$ model.

[5] *Appendix D* contains a description of the tests implemented.

[6] The test was calculated for horizons of 6 and 12 months, as it requires symmetry in the loss function.

Table 2

Comparison of the predictive capacity of the ARMA model

Forecast model	Sign Test (1)	Wilcoxon signed-test rank (2)	F-Test	Morgan- Granger Newbold test
Phillips curve				
3 months	2			
6 months	4	-1.3628		
1 year	5	-0.3922	1.3845	1.0635
Monetary VAR				
3 months	2			
6 months	3	-1.5724		
1 year	5	-0.4707	0.9362	0.1390
Total factors				
3 months	2			
6 months	4	-1.3628		
1 year	7	0.4707	1.4020	0.4186
Nominal factors				
3 months	3			
6 months	6	-0.9435		
1 year	10	2.3534	1.7828	0.4844

(1) Parameters of binomial distribution T (number of periods to be forecasted) and 1/2
(2) Test statistic based on Wilcoxon in its normal asymptotic studentized version

In short, tests indicate that no model outperforms the remainder for all horizons. These results suggest the possibility of working with a pooling of forecasts, taking into account that each of the models considered could contain relevant information for the forecast. It should also be considered that these models could complement each other when making the forecast, in the sense that they make it possible to answer different questions and guide different policy decisions. For example, monetary VAR enables a reply to be provided to a relevant question such as the lag with which monetary impulses are transmitted to prices, and the Phillips curve is informative in relation to the impact of changes in the output gap on the inflation rate. Timmermann (2003) points out that unless one can identify ex-ante a model with a better predictive power than its competitors, a combination of forecasts provides diversification gains that make it more attractive than forecasts derived from an individual model.[7]

Considering a combination of these forecasts provides advantages at various levels: (i) forecast combinations provide diversification. Intuitively, when there is a quadratic loss function, even if one of the models outperforms another in predictive power, by generating a lower loss a linear combination could be preferable; (ii) in the case of economies subject to structural changes, forecast combinations offer better prediction than individual models. In general, the

[7] For a detailed view of the advantages of combining forecasts, see Hendry y Clements (2002). Marcellino (2002) and Timmermann (2006).

speed at which models adapt to structural changes tends to differ. In such an instance, combination of models with differing adaptability to change could improve on individual models; (iii) forecast combination could be seen as a way of making forecasts more robust in the face of specification bias and variable measurement errors in individual forecasts. For example, if two forecasts have different biases, in opposing directions, it is easy to imagine that combination could generate an improvement in the forecast.[8]

On the basis of the suggestion by Marcellino (2002), we constructed lineal forecast combinations according to weighted forecasts, with weights calculated as follows:

$$\widehat{y}_{t+h} = \sum_{m=1}^{M} k_{m,h,t}\widehat{y}_{t+h,m}, \text{ with } k_{m,h,t} = (\frac{1}{RMSE_{m,h,t}})^{w} / \sum_{j=1}^{M} \left(\frac{1}{RMSE_{j,h,t}}\right)^{w} \tag{4}$$

where m indexes the models, $k_{m,h,t}$ indicates the weighting factors, and $RMSE$ is the root mean squared error.

Weights for each model are chosen inversely proportional to their three predictive power statistics in the case of $w = 1$ in the (4) equation. We also considered the case of $w = 5$, in which greater weight is assigned to the models with better predictive performance.

In our case, combinations were made for all models, using for the weights the previously-mentioned three measurements of predictive capacity: $RMSE$, MAE and $MAPE$.

Table 3 presents: the $RMSE$, the MAE, the $MAPE$ and the $U - Theil$ for combinations weighted by $w = 1$ and $w = 5$ using as weighting factors the respective $RMSE$, MAE and $MAPE$ for each model. The statistic is determined in relation to the $ARMA$.

[8]For a detailed view of the advantages of combining forecasts, see Hendry y Clements (2002). Marcellino (2002) and Timmermann (2006).

Table 3

	RMSE					
	w=1			w=5		
Forecast Period	RMSE	MAE	MAPE	RMSE	MAE	MAPE
2006:10-2006:12	0.0032	0.0032	0.0031	0.0028	0.0028	0.0028
2006:7-2006:12	0.0030	0.0030	0.0030	0.0027	0.0027	0.0027
2006:1-2006:12	0.0034	0.0034	0.0034	0.0032	0.0033	0.0034
	MAE					
	w=1			w=5		
Forecast Period	RMSE	MAE	MAPE	RMSE	MAE	MAPE
2006:10-2006:12	0.0030	0.0030	0.0030	0.0024	0.0024	0.0023
2006:7-2006:12	0.0027	0.0027	0.0027	0.0023	0.0023	0.0023
2006:1-2006:12	0.0027	0.0027	0.0027	0.0026	0.0026	0.0026
	MAPE					
	w=1			w=5		
Forecast Period	RMSE	MAE	MAPE	RMSE	MAE	MAPE
2006:10-2006:12	34.61	34.47	34.16	27.15	26.20	24.99
2006:7-2006:12	55.79	55.69	55.81	46.37	45.40	45.50
2006:1-2006:12	32.53	32.45	32.44	31.28	30.92	30.95
	U-Theil statistic					
	w=1			w=6		
Forecast Period	RMSE	MAE	MAPE	RMSE	MAE	MAPE
2006:10-2006:12	1.160	1.156	1.154	1.031	1.025	1.024
2006:7-2006:12	1.180	1.179	1.181	1.051	1.047	1.050
2006:1-2006:12	1.010	1.012	1.019	0.973	0.984	1.020

Results indicate that forecast combinations outperform the best individual forecast for a one-year horizon, with weights more than inversely proportionate for the best individual model. This result is useful for practical purposes because the one-year horizon is relevant for economic policy decision.

5 Conclusions

We estimate a variety of inflation models for forecasting purposes that ranged from univariate models, causal models based on alternative inflation theories, to models based on the use of factors or summarized measures for the combined variability of a large number of economic datasets. We found that although the univariate model generally performs best, as the forecasting horizon lengthens multivariate model performance more closely approximates that of the univariate models. Nevertheless, when tests are performed to evaluate the statistical significance in the predictive power of the models, taking a univariate $ARMA$ as a benchmark, differences are not significant. Lastly, the estimated models are combined by means of a pooling of forecasts using the inverse of the $RMSE$, $MAPE$ and MAE of the respective models as weights. Results indicate that some of the combinations of forecasts outperform the best individual forecast for a one-year horizon. Bearing in mind that the one-year horizon is relevant for economic policy decision-making, the possibility of combining both univariate and multivariate models is of interest, because it also enables answers to specific economic policy questions.

References

[1] Ahumada, H. and L. Garegnani, (2007), "Forecasting a monetary aggregate in a changing environment: Argentina after 2002", mimeo.

[2] Atkenson, A. and L. Ohanian, (2001), "Are Phillips Curve Useful for Forecasting Inflation". Federal Reserve Bank of Minneapolis, Quarterly Review, 25 (1), 2-11.

[3] Canova, F. (2002), "G-7 Inflation Forecasts", European Central Bank, WP No. 151

[4] Bates, J.M. and C.W.I. Granger, (1969), "The combination of forecasts", Operations Research Quarterly 20, 451-468.

[5] Clements, M. and D. Hendry, (2006) "Forecasting with breaks" in Handbook of Economic Forecasting, Elliot, G., C. Granger and A. Timmermann, eds., Chapter 12, Vol. 1, North-Holland.

[6] D'Amato, L., L. Garegnani and J.Sotes, (2007), "Inflation persistence and changes in the monetary regime: The Argentine case", BCRA Working Paper No. Sep. 2007.

[7] D'Amato, L. and L. Garegnani, (2006), The Short -run Dynamics of Inflation:Estimating a Hybrid New Keynesian Phillips Curve for Argentina, in collaboration with Lorena Garegnani. Monetaria, Volume X, CEMLA.

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

[9] Diebold F.X. and P. Pauly, (1990), "Structural change and the combination of forecasts", Journal of Forecasting, 6, 503-508.

[10] Galí J.and M. Gertler, (1999), "Inflation Dynamics: A structural econometric analysis", Journal of Monetary Economics, vol. pp. 195-222.

[11] Granger and Ramanathan, (1984), "Improved methods of forecasting", Journal of Forecasting, 3, 197-204.

[12] Elliot, G., C. Granger and A. Timmermann, eds., (2006), Handbook of Economic Forecasting, Vol. 1, North-Holland.

[13] Kendall, M. (1975), Multivariate Analysis, Charles Griffin & Company Ltd.

[14] Hendry, D.F. y Clements, M.P., (2002), "Pooling of forecasts", Econometrics Journal 5, 1-26.

[15] Marcellino, M., (2002), "Forecasting pooling for short time series of macroeconomic variables", Oxford Bulletin of Economics and Statistics 66, 91-112.

[16] Morgan, W.A. (1939-1940), "A test for the Significance of the Difference Between Two Variances in a Sample from a Normal Bivariate Population", Biometrika, 31, 13-19.

[17] Press, S.J., (1972), Applied Multivariate Analysis, Holt, Rinehart and Witson, Inc.

[18] Stock, J. and M. Watson, (2006), "Forecasting with many predictors", in Handbook of Economic Forecasting, Elliot, G., C. Granger and A. Timmermann, eds., Chapter 10, Vol. 1, North-Holland (2005),

[19] Stock, J. and M. Watson (2005), "Why has U.S. Inflation become harder to forecast"?, WP N°12324,*NBER*.

[20] Stock, J. and M. Watson (2002a), "Macroeconomic Forecasting using diffusion indexes", *Journal of Business and Economic Statistics 20*, 147-162.

[21] Stock, J. and M. Watson (1999), "Forecasting Inflation", Journal of Monetary Economics,44:293-3

[22] Timmermann, A., (2006), "Forecast combinations", in Handbook of Economic Forecasting, Elliot, G., C. Granger and A. Timmermann, eds., Chapter 10, Vol. 1, North-Holland.

A The principal components method

Principal components methodology enables the obtaining of a synthetic measure of the joint variability of a collection of random variables from calculation of those linear combinations of such variables with a maximum variance.[9]

Given a set of n random variables that are known to be related, it is expected that a relatively reduced number of linear combinations - the principal components- could explain a large proportion of their total variability. This method enables the summarizing in a reduced set of $q < n$ principal components and their variables the information contained in a vector $X : nx1$of random variables and its variance and covariance matrixes Σ

Given a random variable vector $X : nx1$ with $E(X) = \Phi, var(X) = \Sigma$, it is possible to define $\alpha \equiv (\alpha i)$ as a vector of unknown weightings for components of X and z as a vector so that

$$z = X'\alpha \; with \; z_i = \sum_{i=1}^{n} \alpha_i X_i$$

If the elements of X are measured in the same units, it is possible to impose:

$$\alpha'\alpha = \sum_{i=1}^{n} \alpha_i = 1$$

As a result, it is possible to encounter a vector of weightings α that maximizes $var(z) = \alpha'\Sigma\alpha$, subject to $\alpha'\alpha = 1$

Then the problem of:

$$\max_{\alpha}(\alpha'\Sigma\alpha)$$

$$subject.to \; \alpha'\alpha = 1 \tag{1}$$

can be written as

$$L = \alpha'\Sigma\alpha - \lambda(\alpha'\alpha - 1)$$

Differentiating L with regard to α and equaling to it 0 is determined that

$$\frac{\partial L}{\partial \alpha} = 2\Sigma\alpha - 2\lambda\alpha = 0$$

[9] For a detailed description of the multivariate analysis technique for principal components, see Press (1972) and Kendall (1975).

and

$$(\Sigma - \lambda I)\alpha = 0 \tag{2}$$

And given that

$$\alpha \neq 0$$

There is a solution if

$$|\Sigma - \lambda I| = 0 \tag{3}$$

This implies that λ is a root characteristic of the variances and covariances matrix of X, Σ and α is a characteristic vector of that matrix. Given that Σ is a $n \times n$ matrix, there are n characteristic roots λ that satisfy (3) and n orthogonal linear combinations associated with characteristic vectors α. Taking into account (2)

$$\Sigma\alpha = \lambda\alpha$$

and pre-multiplying by α'

$$\alpha'\Sigma\alpha = \lambda\alpha'\alpha = \lambda$$

The greater value of λ is that which maximizes the variance of z. The solution to the problem (1) is given by (α_1, λ_1); and z_1 is known as the first principal component of this set of random variables.

B Detail of the cycle series used to calculate factors

All the series used have been seasonally adjusted using the X-12 ARIMA program, and were subsequently standardized

Demand factor	
Series	**Factor**
Synthetic energy indicator - seasonally adjusted	Demand/consumption
Supermarket sales at constant prices - seasonally adjusted	Demand/consumption
Malls sales at constant prices - seasonally adjusted	Demand/consumption
Public services statistics - synthetic general index	Demand/consumption
Imports - seasonally adjusted	Demand/consumption
Domestic market automobile sales - units	Demand/consumption
Domestic market domestically-manufactured auto sales - units	Demand/consumption
Automobile imports - units	Demand/consumption
Sugar sales - thousands of tons	Demand/consumption
Beer sales - thousands of hectoliters	Demand/consumption
Wine sales - thousands of hectoliters	Demand/consumption
Soft drink sales - thousands of hectoliters	Demand/consumption
Cigarette sales - millions of packs	Demand/consumption
Pharmaceutical products sales - millions of units	Demand/consumption
Gasoline sales - thousands of m3	Demand/consumption
Cement sale shipments to domestic market - thousands of tons	Demand/consumption
Asphalt sales - thousands of tons	Demand/consumption
Car sales - units	Demand/consumption
Utility vehicle sales - units	Demand/consumption
Passenger and goods vehicle sales - units	Demand/consumption
Energy demand sales - GWh	Demand/consumption

Supply factor	
Capacity Utilization Index (CUI) - manufacturing industry	Production/supply
CUI - non-durable consumer goods	Production/supply
CUI - durable consumer goods	Production/supply
CUI - capital goods	Production/supply
CUI - intermediate goods	Production/supply
FIEL Survey - general situation manufacturing industry	Production/supply
FIEL Survey - general situation non-durable consumer goods	Production/supply
FIEL Survey - general situation consumer durables	Production/supply
FIEL Survey - general situation capital gods	Production/supply
FIEL Survey - general situation intermediate goods	Production/supply
FIEL Survey -outlook manufacturing industry	Production/supply
FIEL Survey -outlook non-durable cons. goods	Production/supply
FIEL Survey -outlook consumer durables	Production/supply
FIEL Survey -outlook capital gods	Production/supply
FIEL Survey -outlook intermediate goods	Production/supply
FIEL Survey - manufacturing industry demand trend	Production/supply
FIEL Survey - non-durable cons. goods demand trend	Production/supply
FIEL Survey - consumer durables demand trend	Production/supply
FIEL Survey - capital gods demand trend	Production/supply
FIEL Survey - intermediate goods demand trend	Production/supply
FIEL Survey - industry stock levels manufacturing	Production/supply
FIEL Survey - non-durable cons. goods stock levels	Production/supply
FIEL Survey - consumer durables stock levels	Production/supply
FIEL Survey - capital gods stock levels	Production/supply
FIEL Survey - intermediate goods stock levels	Production/supply
Autobile exports - units	Production/supply
Monthly industrial estimator - seasonally adjusted	Production/supply
Monthly economic activity estimator - seasonally adjusted	Production/supply
Industrial production index (IPI) - general level	Production/supply
IPI - non-durable consumer goods	Production/supply
IPI - durable consumer goods	Production/supply
IPI - intermediate goods	Production/supply
IPI - capital goods	Production/supply
IPI - food and beverages	Production/supply
IPI - cigarettes	Production/supply
IPI - textiles input	Production/supply
IPI - pulp and paper	Production/supply
IPI - fuels	Production/supply
IPI - chemicals and plastics	Production/supply
IPI - non-metallic minerals	Production/supply

Supply factor (cont.)	
IPI - steel	Production/supply
IPI - metalworking	Production/supply
IPI - automobiles	Production/supply
Total vehicle production - units	Production/supply
Car production - units	Production/supply
Total cement dispatches	Production/supply
Primary iron production - thousands of tons	Production/supply
Raw steel production - thousands of tons	Production/supply
Crude oil production -thousands of m3	Production/supply
Processed petroleum production - thousands of m3	Production/supply
Natural gas production - millions of m3	Production/supply
Wheat flour production - thousands of tons	Production/supply
Vegetable oil production - thousands of tons	Production/supply
Oilseed by-products production- thousands of tons	Production/supply
Biscuit and cracker production - thousands of tons	Production/supply
Cattle slaughter - thousands of head	Production/supply
Poultry slaughter - thousands of birds	Production/supply
Spirit beverages - thousands of liters	Production/supply
Cellulose thread woven goods - tons	Production/supply
Paper pulp production - thousands of tons	Production/supply
Newsprint production - thousands of tons	Production/supply
Toilet and washing soap production - tons	Production/supply
Vehicle tire production - thousands of units	Production/supply
Urea production - thousands of tons	Production/supply
Caustic soda production - thousands of tons	Production/supply
PVC production - thousands of tons	Production/supply
Ethylene production - thousands of tons	Production/supply
Polyethylene production - thousands of tons	Production/supply
Polypropylene production - thousands of tons	Production/supply
Sec butanol production - tons	Production/supply
Isopropanol production - tons	Production/supply
Sulfuric acid production - thousands of tons	Production/supply
Chlorine production - thousands of tons	Production/supply
Gasoline production - thousands of m3	Production/supply
Diesel fuel production - thousands of m3	Production/supply
Fuel oil production - thousands of tons	Production/supply
Synthetic rubber production - tons	Production/supply
Carbon black production - tons	Production/supply
Construction paint production - tons	Production/supply
Portland cement production - thousands of tons	Production/supply

Supply factor (cont.)	
Steel rods for concrete production - tons	Production/supply
Utility vehicle production - units	Production/supply
Passenger and freight vehicles - units	Production/supply
Cold rolled steel production - thousands of tons	Production/supply
Hot rolled non-flat steel - thousands of tons	Production/supply
Flat hot rolled steel - thousands of tons	Production/supply
Electrolytic zinc production - tons	Production/supply
Tractor production - units	Production/supply

Nominal factor (cont.)	
Rates on sight deposits	Nominal/Rates/Prices
Rates on 30-59 day time deposits	Nominal/Rates/Prices
Rates for 60 days or more deposits	Nominal/Rates/Prices
Rates for sight deposits in dollars	Nominal/Rates/Prices
Rates on 30-59 day time deposits	Nominal/Rates/Prices
Rates for 60 days or more	Nominal/Rates/Prices
Real Multilateral Exchange Rate index - Dec 2001=100	Nominal/Rates/Prices
Consumer Price Index (CPI) - general level	Nominal/Rates/Prices
CPI - food and beverages	Nominal/Rates/Prices
CPI - clothing	Nominal/Rates/Prices
CPI - housing and basic services	Nominal/Rates/Prices
CPI - household equipment and maintenance	Nominal/Rates/Prices
CPI - medical and healthcare expenses	Nominal/Rates/Prices
CPI - transport and communications	Nominal/Rates/Prices
CPI - leisure	Nominal/Rates/Prices
CPI - education	Nominal/Rates/Prices
CPI - other sundry goods and services	Nominal/Rates/Prices
wholesale price index (IPIM) - general level	Nominal/Rates/Prices
IPIM - domestic goods total	Nominal/Rates/Prices
IPIM - domestic primary goods	Nominal/Rates/Prices
IPIM - domestic manufactured goods and electricity	Nominal/Rates/Prices
IPIM - imported goods	Nominal/Rates/Prices
Domestic internal wholesale prices (IPIB) - general level	Nominal/Rates/Prices
IPIB - domestic goods total	Nominal/Rates/Prices
IPIB - domestic primary goods	Nominal/Rates/Prices
IPIB - domestic manufactured goods and electricity	Nominal/Rates/Prices
IPIB - imported goods	Nominal/Rates/Prices
Producer price index (PPI) - general level	Nominal/Rates/Prices
PPI - domestic primary goods	Nominal/Rates/Prices
PPI - domestic manufactured goods and electricity	Nominal/Rates/Prices
VAT revenue -DGI	Nominal/Rates/Prices
Import duty revenue	Nominal/Rates/Prices
Private cash and banks at month-end	Nominal/Rates/Prices
Private sector broad M1 at month-end	Nominal/Rates/Prices
Private sector broad M2 at month-end	Nominal/Rates/Prices
Private sector broad M3 at month-end	Nominal/Rates/Prices
Merval index at month-end	Nominal/Rates/Prices
Total average cash and banks at month-end	Nominal/Rates/Prices
Total average broad M1 for month	Nominal/Rates/Prices
Total average broad M2 for month	Nominal/Rates/Prices
Total average broad M3 for month	Nominal/Rates/Prices
VAT revenue - DGA	Nominal/Rates/Prices

C Forecast models

Table C presents estimated forecast models

Table C

ARMA(1,1)		Phillips curve		Forecast Mdels Monetary VAR			Total Factors		Nominal Factors	
dep. var. : inflation		dep. var. : inflation		dep. var.	D(LM2)	D(LIPC)	dep. var. : inflation		dep. var. : inflation	
C	0.0012	INFLA(-1)	0.57714	D(LM2(-1))	0.070025	-0.000231	C	0.0050	C	0.00491
	0.0005		*0.06337*		*0.09122*	*0.01724*		*0.0004*		*0.00034*
INFLA(-1)	0.4605	INFLA(1)	0.18159	D(LM2(-2))	0.065873	0.042099	FT1	0.0027	FN1	0.00294
	0.0587		*0.04029*		*0.07114*	*0.01345*		*0.0002*		*0.00015*
DUM0219	0.0149	GAP(-1)	0.00972	D(LM2(-3))	-0.087637	0.035181	FT1(-1)	-0.0006	FN1(-1)	-0.00066
	0.0020		*0.00299*		*0.07819*	*0.01478*		*0.0002*		*0.00017*
DUM024	0.0654	DEVNOM(-1)	0.02346	D(LM2(-4))	0.096694	0.000191	FT2	0.0004	FN1(-3)	0.00032
	0.0037		*0.00801*		*0.07795*	*0.01473*		*0.0001*		*0.00012*
DUM025	-0.0182	DIPPUSA(-1)	0.21346	D(LIPC(-1))	-0.07621	0.303032	FT2(-3)	0.0006	FN2	0.00058
	0.0055		*0.05667*		*0.35616*	*0.06731*		*0.0001*		*0.00013*
MA(12)	0.4854	D021	0.01983	D(LIPC(-2))	-0.417894	0.149748	FT3(-3)	-0.0006	FN2(-3)	-0.00026
	0.0768		*0.00578*		*0.28985*	*0.05478*		*0.0001*		*0.00016*
		D024	0.06581	D(LIPC(-3))	0.677121	0.030016	FT4	0.0007	FN4	-0.00106
			0.00254		*0.27385*	*0.05176*		*0.0001*		*0.00016*
		D025	-0.02662	D(LIPC(-4))	0.491274	0.04276	FT4(-3)	-0.0005		
			0.00605		*0.22779*	*0.04305*		*0.0002*		
		D031	0.01320	C	0.083111	0.001235				
			0.00214		*0.00541*	*0.00102*				
				D0219	-0.054569	0.005724				
					0.01618	*0.00306*				
				D0112	0.271904	0.002375				
					0.02075	*0.00392*				
				D017	-0.07793	-0.004301				
					0.01915	*0.00362*				
				D022	0.121785	0.007495				
					0.03605	*0.00681*				
				D024	0.05552	0.061302				
					0.03247	*0.00614*				
				D013	-0.05333	0.004544				
					0.01891	*0.00357*				
				D0110	-0.048168	0.002288				
					0.02016	*0.00381*				
				D021	0.302341	0.012345				
					0.0375	*0.00709*				
				D018	-0.084922	0.001322				
					0.02032	*0.00384*				
Sample (adj.): 1993M01 2005M12		Sample (adj.): 1994M02 2005M12		Sample (adj.): 1993M06 2005M12			Sample (adj.): 1997M05 2005M12		Sample (adj.): 1994M02 2005M12	
Inc. obs.: 156 after adj.		Inc.obs.: 143 after adj.		Inc. Obs.: 151 after adjustments			Inc. obs: 104 after adjustments		Inc.obs.: 143 after adj.	
R-sq	0.8757	R-sq	0.8896	R-sq	0.888107	0.92313	R-sq	0.9274	R-sq.	0.9348
S.E. of reg	0.0039	S.E. of reg	0.0039	Sum sq. resids	0.039174	0.001399	Sum sq. resids	0.0012	Sum sq. resid.	0.0011
SSR	0.0023	J-statistic	0.11391							

D Tests to compare predictive power

D.1 Tests for finite samples

1. **Sign test**

   Given the loss differential d_t, between two models i and j, defined as

$$d_t = [g(e_{it}) - g(e_{jt})]$$

   the null hypothesis of the test is that the median loss differential is 0

$$med(g(e_{it}) - g(e_{jt})) = 0$$

   Assuming that the loss differential is an *iid* variable , the number of positive differentials in a sample of size T follows a binomial distribution with parameters T, $\frac{1}{2}$, under the null hypothesis. The test statistic is therefore

$$S_1 = \sum_{t=1}^{T} I_+(d_t)$$

where

$$\begin{aligned} I_+(d_t) &= 1 \quad \text{if } d_t > 0 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

The significance of the statistic can be confirmed on the table for the accumulated binomial distribution.

2. **Wilcoxon signed-rank test**

   This non-parametric test requires symmetry in the loss differential. Again, it is assumed that the loss differential is *iid*.

   The test statistic is the sum of the ranks of the absolute value of the positive differences.

$$S_2 = \sum_{t=1}^{T} I_+(d_t)\ rank\ (|d_t|)$$

   The studentized distribution of the S_2 statistic is asymptotically distributed as a standard normal

$$S_{2std} = \frac{S_2 - \frac{T(T+1)}{4}}{\sqrt{\frac{T(T+1)(2T+1)}{24}}} \overset{a}{\sim} N(0,1)$$

3. **F Test**

   If (i) the loss function is quadratic ;(ii) the forecast mean error is 0 and (iii) the distribution is normal; (iv) the errors are not serially correlated, and (v) they are not contemporaneously correlated among themselves, the ratio of the sample variances follows the F distribution under the null hypothesis of no differences in predictive power. The statistic to be evaluated is

$$F = \frac{\frac{e_i' e_i}{T}}{\frac{e_j' e_j}{T}} = \frac{e_i' e_i}{e_j' e_j}$$

   and is distributed as $F(T,T)$.

4. **Test de Morgan-Granger-Newbold**

Granger and Newbold (1977) seek to resolve the problem of contemporaneous correlation between forecast errors on the basis of an orthogonal transformation proposed by Morgan (1939). Defining $x_t = (e_{it} + e_{jt})$ and $z_t = (e_{it} - e_{jt})$, maintaining assumptions (i) to (iv), the null hypothesis of equal predictive capability between the models i and j is equivalent to a 0 correlation between x and z, that is to say $\rho_{zx} = 0$. Statistic

$$MNG = \frac{\widehat{\rho_{zx}}}{\sqrt{\frac{1-\widehat{\rho_{zx}}}{T-1}}}$$

is distributed as a Student's t with T-1 degrees of freedom.