JEL Classification: C22; C53; E31; E37; E47
Keywords: inflation forecasts, benchmark models, univariate time-series models, out-of-sample comparison

Forecasting Inflation with a Simple and Accurate Benchmark: The Case of the US and a Set of Inflation Targeting Countries*

Pablo M. PINCHEIRA—Central Bank of Chile and School of Business, Adolfo Ibáñez University, Chile (pablo.pincheira@uai.cl), *corresponding author*
Carlos A. MEDEL—Central Bank of Chile (lexcm6@nottingham.ac.uk)

Abstract
We evaluate the ability of several univariate models to predict inflation in the US and in a number of inflation targeting countries at different forecasting horizons. We focus on forecasts coming from a family of ten seasonal models that we call the Driftless Extended Seasonal ARIMA (DESARIMA) family. Using out-of-sample Root Mean Squared Prediction Errors (RMSPE) we compare the forecasting accuracy of the DESARIMA family with that of traditional univariate time-series benchmarks available in the literature. Our results show that DESARIMA-based forecasts display lower RMSPE at short horizons for every single country, with the exception of one case. We obtain mixed results at longer horizons. In particular, when the family-median forecast is considered, in more than half of the countries our DESARIMA-based forecasts outperform the benchmarks at long horizons. Remarkably, the forecasting accuracy of our DESARIMA family is high in stable-inflation countries, for which the RMSPE is around 100 basis points when a prediction is made 24 and even 36 months ahead.

1. Introduction

Forecasts of economic and financial variables are important inputs for policy-makers in the decision-making process. From time to time new forecasting techniques appear in the literature with the hope of providing a better understanding of the evolution of key economic variables or with the simpler goal of reducing some measure of forecasting error. When evaluating whether a novel forecasting approach is useful for prediction, at least three elements are necessary: a measure of accuracy or loss function, a good enough benchmark against which to compare the new predictions, and an adequate test of predictive ability. In this work, we focus exclusively on the second point by introducing a family of models for benchmarking inflation forecasts. This family, denominated Driftless Extended Seasonal ARIMA (DESARIMA), contains ten seasonal univariate time-series models sharing the common feature of a unit root.[1]

These models produce competitive forecasts at short and long horizons when compared to traditional univariate benchmarks used in the literature. Besides accuracy, our DESARIMA family contains all the desirable features of well-behaved univariate time-series models, but also some of their shortcomings. Some of the main

* We would like to thank Yan Carrière-Swallow, Roman Horvath, two anonymous referees, and seminar participants at the Central Bank of Chile and the Chilean Economics Society Annual Meeting 2012 for their comments. We would also like to thank Nicolás Rivera for his wonderful assistance. The views and ideas expressed in this paper do not necessarily represent those of the Central Bank of Chile or its authorities. Any errors or omissions are the responsibility of the authors.

advantages of univariate time-series models are their simplicity and tractability. Since they rely only on lagged values of the dependent variable, they can be updated easily and out-of-sample analyses can be carried out in a short time. Furthermore, as Aiolfi, Capistrán and Timmermann (2011) point out, forecasts from time-series models can be used in combination with other strategies, including publicly available surveys, to greatly enhance forecast accuracy. Pincheira (2012a, 2012b) shows that adjusted combinations between univariate time-series models and surveys can substantially reduce the Root Mean Squared Prediction Error (RMSPE) of surveys. Similarly, Pincheira (2009) shows that when some models from the DESARIMA family are fed with highly accurate one-step-ahead forecasts, multi-step-ahead forecasts resulting from this particular type of combination substantially improve the accuracy of medium-term forecasts. This is important, because many big general equilibrium macro models could benefit from the higher accuracy of exogenous short- and medium-term forecasts.[2] The shortcomings of univariate time-series models are well known. They include the omission of variables that may be relevant in the forecasting process and their unsuitability to provide economic explanations of the forecasts they produce.

We evaluate our DESARIMA family in terms of its ability to produce accurate forecasts of the Consumer Price Index (CPI) year-on-year (YoY) inflation rate for a set of eleven inflation targeting countries plus the US.[3] We evaluate the forecasting performance of the DESARIMA family by comparing its out-of-sample RMSPE against a set of thirteen benchmark models commonly used in the literature. We also analyze the statistical significance of the differences between the best model of each family using the Giacomini and White (2006) test.[4] The largest

[1] We mainly focus on a whole family of models and not on a particular model within this family because of the extensive literature pointing out the unstable predictive ability of traditional forecasting methods for output and inflation (see, for instance, Rossi, 2013, and Stock and Watson, 1996, 2003). Due to this instability, it is hard to think of a particular model performing well in predicting inflation in different countries at several forecasting horizons and at different moments in time. In our opinion, it would be less surprising to find some stability within a family of models sharing some common features. We do report, however, good predictive behavior also of the median forecast, which in our opinion could be considered as a final forecast when there is no clear guidance about which single model to use.

[2] For conducting monetary policy at the Central Bank of Chile, for instance, it is fairly usual to use exogenous short- and medium-term inflation forecasts as inputs for the big macro general equilibrium models. The different macroeconomic implications of a variety of monetary policy decisions are then analyzed in light of these models. Accurate short- and medium-term inflation forecasts are, therefore, critical for an adequate evaluation of different macroeconomic scenarios.

[3] The countries are Canada, Chile, Colombia, Israel, Mexico, Peru, South Africa, Sweden, Switzerland, Turkey, the United Kingdom and the US.

[4] The Giacomini and White (2006) (henceforth GW) and the Diebold and Mariano (1995)-West (1996) (henceforth DM-W) frameworks are very different in their fundamentals, though under some conditions their numerical calculation is exactly the same. In fact, GW claims that in some environments their test statistic "coincides with that proposed by Diebold and Mariano (1995)" (see page 1557 in GW). In the context of our paper, we use the *t*-type statistic suggested by GW, but it is exactly the same as the DM *t*-type statistic when comparing forecasts built for horizons greater than one period. Therefore, the only practical difference appears when comparing one-step-ahead forecasts. In this case, we rely on the stronger null hypothesis considered by GW, which exploits the simplifying feature of a martingale difference sequence in the time structure of the loss differential. This implies that the asymptotic variance can be consistently estimated by the sample variance and not only by a HAC estimator. As suggested by GW, we prefer to use the simpler calculation of the sample variance because it may increase the power of the test.

evaluation sample spans from February 1999 to December 2011 (155 observations) and includes forecasts made one, three, six, 12, 24, and 36 months ahead.

We find that DESARIMA-based forecasts display lower out-of-sample RMSPE than forecasts coming from traditional benchmarks at short horizons for every single country in our sample, except Colombia. We obtain mixed results at longer horizons. In particular, when the median forecast is considered, in more than half of the countries our DESARIMA-based forecasts outperform the benchmarks at long horizons. Remarkably, the forecasting accuracy of our DESARIMA family is surprisingly high in stable-inflation countries, for which the RMSPE is around 100 basis points when a prediction is made 24 and even 36 months ahead.

While the most frequent winner model within the DESARIMA family may be an adequate candidate for applied forecasters, in this paper we also report a good behavior of the median across the set of DESARIMA forecasts, which can be favorably used as a unique forecasting method.

The rest of the article is organized as follows: in Section 2 we describe our econometric setup. In Section 3 we describe the dataset. In Section 4 we present and discuss the results of our forecast evaluation. Finally, Section 5 concludes the paper.

2. Econometric Setup

We focus on monthly CPI YoY inflation defined as follows:

$$\pi_t = \frac{CPI_t}{CPI_{t-12}} \cdot 100 - 100$$

Our choice of YoY inflation as a target variable relies on the fact that, to our knowledge, every inflation targeting country in the world defines its target in YoY terms. For instance, considering a wider sample of inflation targeters, the Czech Republic has a target of 2% for the medium term. The United Kingdom has the same target, but it is supposed to be met at all times. In Chile, Thailand and Mexico the target is 3%. Some countries, such as Iceland, Norway, Poland, Romania and South Korea, have a target of 2.5%. The list is long, but all these countries express their targets in YoY terms. Given that the vast majority of the economies in our sample are inflation targeting countries, we think that forecasting YoY inflation is a reasonable way to proceed.[5]

We make use of models for this particular variable π_t, or for its first differences defined as:

$$\Delta\pi_t = \pi_t - \pi_{t-1}$$

We are interested in h-step-ahead forecasts, where h takes the following values:

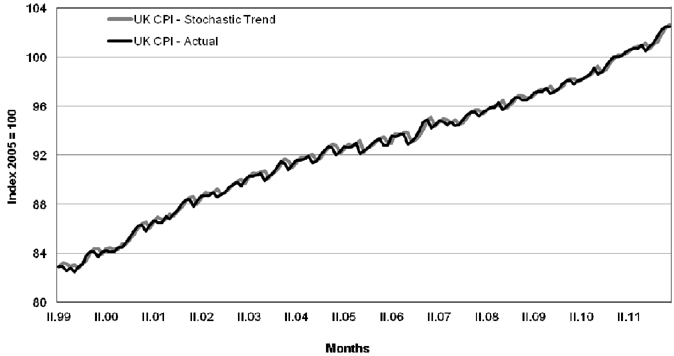
$$h \in \{1, 3, 6, 12, 24, 36\}$$

It is important to mention that for horizons longer than one month our forecasts are constructed using the iterated method rather than the direct method. In the following two subsections we introduce the family of candidate and benchmark

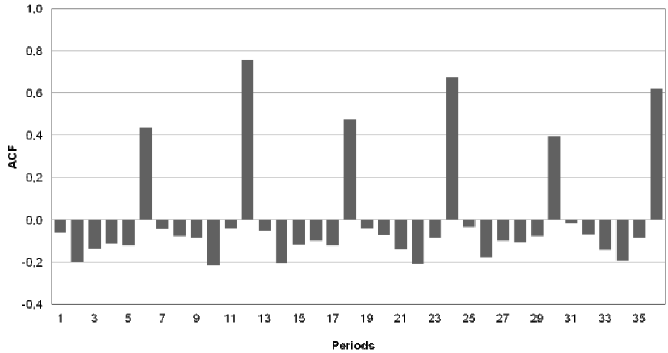
[5] To our knowledge, there are no articles comparing the predictive behavior of a variable based on YoY or month-to-month rates. This might be an interesting point to address in future research.

Figure 1 United Kingdom CPI – Actual, Trend, and Seasonality

Actual and Stochastic Trend Series



Autocorrelation Function of Detrended Series



Source: Bank of England and authors' computations.

models that we will use in our empirical application. The third subsection describes the framework we use to evaluate our forecasts.

2.1 Candidate Models: DESARIMA Family

For the construction of an alternative family of models we use as a starting point two stylized facts characterizing the CPI: stochastic trend and seasonality. These facts are depicted in *Figure 1* by taking the representative case of the United Kingdom.

A general SARIMA specification allowing for stochastic trends and seasonality in the natural logarithm of the CPI provides our basic forecasting framework:[6]

$$\Phi(L)\Phi_E\left(L^S\right)(1-L)^d\left(1-L^S\right)^D \ln\left(CPI_t\right) = \delta + \Theta_E\left(L^S\right)\Theta(L)\varepsilon_t \qquad (1)$$

where:

$$\Phi(L) = \left(1-\phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p\right)$$

[6] A recent survey of stationary ARMA models and their variations can be found in Holan, Lund and Davis (2010).

$$\Phi_E\left(L^S\right)=\left(1-\phi_{E1}L^S-\phi_{E2}L^{2S}-\ldots-\phi_{EP}L^{PS}\right)$$
$$\Theta(L)=\left(1-\theta_1L^S-\theta_2L^S-\ldots-\theta_qL^q\right)$$
$$\Theta_E\left(L^S\right)=\left(1-\theta_{E1}L^S-\theta_{E2}L^{2S}-\ldots-\phi_{EQ}L^{QS}\right)$$

are the autoregressive (AR) and moving average (MA) lag operators that are supposed to be stationary and invertible, respectively. Here, L is a lag operator $\left(L^j x_t = x_{t-j}\right)$, S represents the frequency of the series ($S = 12$ for monthly series), and δ, ϕ_j, ϕ_{Ej}, θ_j, and θ_{Ej} are parameters to be estimated. Finally, ε_t is a white noise process with variance σ^2.

This specification captures the two aforementioned salient features of the CPI with some flexibility by allowing both seasonal and nonseasonal trends, and also allowing both AR and MA components. Nevertheless, expression (1) represents a nonparsimonious specification (especially with long lag lengths) given the number of unknown parameters: one intercept, p non-seasonal autoregressive terms, q non-seasonal moving average terms, P seasonal autoregressive terms, and Q seasonal moving average terms. To alleviate the consequences of parameter uncertainty in our forecasts, we favor a parsimonious version of expression (1) by imposing the following restrictions on the lag operators:

$$p=P=q=Q=1$$

For simplicity, we also impose

$$d=1 \text{ and } D=0$$

Therefore, we now have the simpler expression:

$$(1-\rho L)\left(1-\phi L^{12}\right)(1-L)\ln\left(CPI_t\right)=\delta+(1-\theta L)\left(1-\theta_E L^{12}\right)\varepsilon_t \tag{2}$$

in which the number of parameters is only five. Following Box and Jenkins (1970), Brockwell and Davis (1991), and Harvey (1993), we impose $\phi=1$ to obtain an even more parsimonious expression as follows:[7]

$$(1-\rho L)\left(1-L^{12}\right)(1-L)\ln\left(CPI_t\right)=\delta+(1-\theta L)\left(1-\theta_E L^{12}\right)\varepsilon_t \tag{3}$$

This last expression is not only more parsimonious than expression (2), but also more convenient because by using the approximation that $\pi_t \approx \ln\left(CPI_t\right)-\ln\left(CPI_{t-12}\right)$, we can write equation (3) directly in terms of π_t:

$$(1-\rho L)\left(\pi_t-\pi_{t-1}\right)=\delta+(1-\theta L)\left(1-\theta_E L^{12}\right)\varepsilon_t$$

[7] Notice that (3) could also be derived directly from (1) by assuming:
$p=q=Q=D=d=1$ and $P=0$,
in the understanding that the seasonal AR operator is defined as follows:
$\Phi_E\left(L^S\right)=\left(1-\phi_{E1}L^S-\phi_{E2}L^{2S}-\ldots-\phi_{EP}L^{PS}\right)$ if $P>0$
$\Phi_E\left(L^S\right)=1$ if $P=0$
We thank an anonymous referee for pointing this out.

which is equivalent to:

$$\pi_t = \delta + (1-\rho L)\pi_{t-1} - \rho\pi_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} - \theta_E \varepsilon_{t-12} + \theta\theta_E \varepsilon_{t-13}$$

Notice that this expression can also be written as:

$$\pi_t - \pi_{t-1} = \delta + \rho\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon_t - \theta\varepsilon_{t-1} - \theta_E \varepsilon_{t-12} + \theta_3 \varepsilon_{t-13} \tag{4}$$

where $\theta_3 = \theta\theta_E$.

Expression (4) corresponds to an ARIMA(1,1,13) process for π_t in which several MA coefficients are set to zero. Following Box, Jenkins and Reinsel (2008), the eventual or explicit form of the forecast function for (4) is given by:

$$\hat{\pi}_{t+h|t} = c_t \rho^h + b_t + \left[\frac{\delta}{1-\rho}\right]h, \text{ for } h > 11 \tag{5}$$

where $\hat{\pi}_{t+h|t}$ denotes the best linear *h*-step-ahead forecast of the inflation $\pi_{t|t}$ given the information of the process available at time *t*. Furthermore, c_t and b_t represent adaptive coefficients, i.e. coefficients that are stochastic and functions of the process at time *t*.[8]

From expression (5) we can see that long-horizon forecasts of π_t will be divergent unless we impose the additional restriction of no intercept ($\delta = 0$).[9] This constraint leads us to the following specification:

$$\pi_t - \pi_{t-1} = \rho\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon_t - \theta\varepsilon_{t-1} - \theta_E \varepsilon_{t-12} + \theta_3 \varepsilon_{t-13} \tag{6}$$

It is interesting to point out that some models commonly used in the literature are nested in expression (6). By taking $\rho = 0$ and $\theta = \theta_E = 0$, we recover the random walk (RW) used by Groen, Kapetanios and Price (2009), which is also similar to the naive model used by Atkeson and Ohanian (2001) for the US. If we take only $\rho = 0$, expression (6) describes the *airline model* introduced by Box and Jenkins (1970), which is considered very useful for forecasting monthly time series with seasonal patterns according to Ghysels, Osborn and Rodrigues (2006). Also, by taking $\rho = 0$ and $\theta_E = 0$ we recover the IMA(1,1) model used, among others, by Box, Jenkins and Reinsel (2008) and more recently by Croushore (2010). Finally, by taking $\theta_E = 0$ we recover the ARIMA(1,1,1) model used in Proietti (2011) to compare the direct method versus the multistep iterated forecasting method.

Expression (6) depends only on three unknown parameters: ρ, θ, and θ_E. We could define a family of eight models by considering the eight different variations of (6) in which we either include or exclude the terms multiplied by ρ, θ, and θ_E. None of these eight specifications, however, is capable of including two MA terms of orders 1 and 12 and simultaneously excluding the MA component of order 13. This is because $\theta_3 = \theta\theta_E$, so setting either $\theta = 0$ or $\theta_E = 0$ leads necessarily to $\theta_3 = 0$. It might be relevant to add a couple of models allowing for two MA terms

[8] These adaptive terms are also function of the unknown parameters of the model.

[9] For a formal derivation and generalization of this result, see Pincheira and Medel (2012a).

Table 1 DESARIMA Family

1	$\pi_t - \pi_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1} - \theta_E \varepsilon_{t-12}$
2	$\pi_t - \pi_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1} - \theta_E \varepsilon_{t-12} + \theta\theta_E \varepsilon_{t-13}$
3	$\pi_t - \pi_{t-1} = \rho\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon_t - \theta\varepsilon_{t-1} - \theta_E \varepsilon_{t-12}$
4	$\pi_t - \pi_{t-1} = \rho\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon_t - \theta\varepsilon_{t-1} - \theta_E \varepsilon_{t-12} + \theta\theta_E \varepsilon_{t-13}$
5	$\pi_t - \pi_{t-1} = \varepsilon_t - \theta_E \varepsilon_{t-12}$
6	$\pi_t - \pi_{t-1} = \varepsilon_t$
7	$\pi_t - \pi_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1}$
8	$\pi_t - \pi_{t-1} = \rho\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon_t - \theta_E \varepsilon_{t-12}$
9	$\pi_t - \pi_{t-1} = \rho\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon_t$
10	$\pi_t - \pi_{t-1} = \rho\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon_t - \theta\varepsilon_{t-1}$

Source: Authors' elaboration.

of orders 1 and 12 without the inclusion of a MA term of order 13. This is so because in the particular case in which both coefficients θ or θ_E are small or close to zero, the parameter $\theta_3 = \theta\theta_E$ can be of negligible size. Its estimate could potentially be more harmful than useful because our forecasts might be substantially contaminated with a noisy estimation. For this reason, we propose the following specification:

$$\pi_t - \pi_{t-1} = \rho\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon_t - \theta\varepsilon_{t-1} - \theta_E \varepsilon_{t-12} + \theta\theta_E \Upsilon \varepsilon_{t-13} \tag{7}$$

in which Υ can take the value 1 or 0. When $\Upsilon = 1$ we obtain the same family of eight models coming from expression (6). When $\Upsilon = 0$, however, we allow for the inclusion of the following two models:

1. $\pi_t - \pi_{t-1} = \rho\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon_t - \theta\varepsilon_{t-1} - \theta_E \varepsilon_{t-12}$
2. $\pi_t - \pi_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1} - \theta_E \varepsilon_{t-12}$

Because these two models do not have a direct SARIMA representation, we call our family of ten candidate models a Driftless Extended SARIMA family (DESARIMA). *Table 1* provides the ten specifications belonging to our DESARIMA family.

Now, suppose that we disregard the seasonal behavior of the CPI. This would be an adequate assumption when working with seasonally adjusted data. If we focus mostly on modeling monthly variations of inflation rates we could start from the following non-seasonal expression:

$$(1-\beta L)(1-L)\ln(CPI_t) = \delta + (1-\theta L)\varepsilon_t \tag{8}$$

which is equivalent to:

$$(1-\beta L)\pi_t^1 = \delta + (1-\theta L)\varepsilon_t \tag{9}$$

where:

$$\pi_t^1 \approx \ln\left(CPI_t\right) - \ln\left(CPI_{t-1}\right)$$

Therefore we have:

$$\pi_t^1 = \delta + \beta\pi_{t-1}^1 + \varepsilon_t - \theta\varepsilon_{t-1} \tag{10}$$

This is a very simple expression characterizing monthly inflation rates, in which we have disregarded seasonal patterns. Now notice that:

$$\begin{aligned}\left(1-L^{12}\right)\pi_t^1 &= \pi_t^1 - \pi_{t-12}^1 = \ln\left(CPI_t\right) - \ln\left(CPI_{t-1}\right) - \ln\left(CPI_{t-12}\right) + \ln\left(CPI_{t-13}\right) = \\ &= \ln\left(CPI_t\right) - \ln\left(CPI_{t-12}\right) - \left[\ln\left(CPI_{t-1}\right) - \ln\left(CPI_{t-13}\right)\right] = \\ &= \pi_t - \pi_{t-1}\end{aligned}$$

therefore, using (10) we could derive the corresponding process for YoY inflation as follows:

$$\begin{aligned}\left(1-L^{12}\right)\pi_t^1 &= \left(1-L^{12}\right)\delta + \beta\left(1-L^{12}\right)\pi_{t-1}^1 + \left(1-L^{12}\right)\left(\varepsilon_t - \theta\varepsilon_{t-1}\right) \\ \pi_t - \pi_{t-1} &= \beta\left(\pi_{t-1} - \pi_{t-2}\right) + \varepsilon_t - \theta\varepsilon_{t-1} - \varepsilon_{t-12} + \theta\varepsilon_{t-13}\end{aligned} \tag{11}$$

We notice that the moving average component of this last expression has the following seasonal representation:

$$(1-\theta L)\left(1-\theta_{12}L^{12}\right)\varepsilon_t \text{ with } \theta_{12} \equiv 1$$

Notice also that (11) is the same model 4 in *Table 1* when setting:

$$\rho = \beta \text{ and } \theta_E = 1$$

Expression (11) shows that the YoY transformation of the data induces a seasonal pattern that otherwise would not be present in the original data.[10] In other words, either because annual differentiation is not able to eliminate non-additive seasonality in the original data or because annual differentiation induces a seasonal behavior that is not present in the original data, our family of models in *Table 1* offers reasonable candidates to provide a good fit to YoY inflation rates.

2.2 Benchmark Models

The use of different univariate time-series models to generate forecasts is fairly usual in the forecasting literature in general and in the inflation forecast literature in particular. For instance, Atkeson and Ohanian (2001) show that a simple variation of a RW model for YoY inflation in the US is very competitive when predicting inflation 12 months ahead. Giacomini and White (2006) present also for the US an empirical application in which several CPI forecasts are compared to those generated by a RW with drift and an autoregression in which the lag length is selected according to the Bayesian Information Criteria (BIC).

Another article using simple univariate benchmarks for the US is that of Ang, Bekaert and Wei (2007). Among the many methods the authors use, they include an ARMA(1,1) model, a RW, and an AR(p) model with lag length selection

[10] We would like to thank a referee for pointing this out. Furthermore, he/she correctly noted that if month-on-month inflation is generated by a simple data generating process like white noise, AR(1), MA(1), ARMA(1,1) or RW, then the corresponding model for the YoY inflation would be the DESARIMA 5, DESARIMA 8, DESARIMA 2, DESARIMA 4 and DESARIMA 8, respectively, with $\Theta_E = 1$ and, in the case of RW, with $\rho = 1$.

according to the BIC. Elliot and Timmermann (2008) also explore the ability of several simple univariate models to predict inflation in the US including a simple AR(p) model and a single exponential smoothing (ES), which generates the same forecasts as an IMA(1,1) model, in which some constraints are imposed over the parameters. More recently, Croushore (2010) also makes use of an IMA(1,1) model as a benchmark when evaluating survey-based inflation forecasts for the US. Finally, Marcellino, Stock and Watson (2006) make use of an AR(1) to perform a simulation exercise to compare the direct method against the multistep iterated method to forecast a number of macroeconomic series in the US.

Outside of the US the use of univariate time-series models has also become fairly usual. Groen, Kapetanios and Price (2009), for instance, evaluate the accuracy of the Bank of England inflation forecasts using several univariate models, including an AR(p) and the RW. Similarly, Andersson, Karlsson and Svensson (2007) make use of simple time-series models to compare inflation forecasts from the Riksbank. In addition, Pincheira and Alvarez (2009) and Pincheira (2010) also consider ARMA models to construct forecasts for Chilean inflation and GDP growth respectively.

Based on this selective review of the literature and our preliminary exploration, we define the family of benchmarks as that containing the following 13 univariate linear models for π_t: AR(1), AR(6), AR(12), ARMA(1,1), AR(p) with automatic lag selection based on the Akaike Information Criterion (AIC) and also on the BIC.[11] For this AR(p) process we consider different specifications, varying p from 1 to $p^{\max} = 12$. We also use ARMA(p,q) specifications with automatic lag selection according to the AIC and the BIC with $p^{\max} = 12$ and $p^{\max} = 6$. In addition, we include the models labeled 3 and 4 in Capistrán, Constandse and Ramos-Francia (2010) that include specific regressors for seasonality (henceforth CCR-F).[12] Finally, we also consider a single and double ES, and a Holt-Winters model with additive seasonal components.[13] These benchmark models are summarized in *Table 2*.

Models (9) and (10) from CCR-F contain the lag polynomial $\varphi(L)$. Its order is determined according to the BIC. In *Table 2*, D_{it} represents a seasonal dummy variable for the i-th month. For models (11) to (13) the initial value of inflation (π_0) is determined as the average of the first half of the estimation sample.

2.3 Forecast Evaluation Framework

We carry out an out-of-sample evaluation of our benchmark and DESARIMA family. To describe this evaluation, let us assume that for a given country we have

[11] The BIC is defined as $BIC = -2(\ell / T) + 2(k / T)$, whereas the AIC is defined as $AIC = -2(\ell / T) + +k \cdot (k / T) / T$, where ℓ is the log likelihood function, k the number of unknown parameters, and T the sample size. Therefore, the only difference in estimating the true order of ARMA model is the penalty term imposed on the number of unknown parameters. For more details, see Akaike (1974) and Schwarz (1978). An out-of-sample comparison between forecasts coming from models based on these criteria can be found in Granger and Jeon (2004).

[12] One of the models in Capistrán, Constandse and Ramos-Francia (2010) is defined in terms of π (model 3), while the other (model 4) is defined in terms of $\Delta\pi$.

[13] See Hyndman *et al.* (2008) for details.

Table 2 Benchmark Models

1. AR(1) with intercept	$\pi_t = \delta + \rho_1 \pi_{t-1} + \varepsilon_t$	11. Single ES	$\hat{\pi}_t = \alpha \pi_{t-1} + (1-\alpha)\hat{\pi}_{t-1},\ \ 0 < \alpha < 1$
2. AR(6) with intercept	$\pi_t = \delta + \sum_{i=1}^{6} \rho_i \pi_{t-i} + \varepsilon_t$	12. Double ES	$\hat{\pi}_{t+h} = \left(2 + \frac{\alpha h}{1-\alpha}\right) S_t - \left(1 + \frac{\alpha h}{1-\alpha}\right) D_t =$
3. AR(12) with intercept	$\pi_t = \delta + \sum_{i=1}^{12} \rho_i \pi_{t-i} + \varepsilon_t$		$= 2S_t - D_t + \frac{\alpha}{1-\alpha}\left(S_t - D_t\right) h$
4. ARMA(1,1) with intercept	$\pi_t = \delta + \rho_1 \pi_{t-1} - \theta_1 \varepsilon_{t-1} + \varepsilon_t$		$S_t = \alpha \pi_t + (1-\alpha) S_{t-1}$
5. AR(12) BIC-based	$\pi_t = \delta + \sum_{i=1}^{12} \rho_i \pi_{t-i} + \varepsilon_t$		$D_t = \beta S_t + (1-\beta) D_{t-1},\ \alpha = \beta$
6. AR(12) AIC-based	$\pi_t = \delta + \sum_{i=1}^{12} \rho_i \pi_{t-i} + \varepsilon_t$	13. Holt-Winters	$\hat{\pi}_{t+h} = a_t + b_t h + c_{t+h-s}$
7. ARMA(12,6) BIC-based	$\pi_t = \delta + \sum_{i=1}^{12} \rho_i \pi_{t-i} - \sum_{j=1}^{6} \theta_j \varepsilon_{t-j} + \varepsilon_t$		$a_t = \alpha\left(\pi_t - c_{t-s}\right) + (1-\alpha)\left(a_{t-1} + b_{t-1}\right)$
8. ARMA(12,6) AIC-based	$\pi_t = \delta + \sum_{i=1}^{12} \rho_i \pi_{t-i} - \sum_{j=1}^{6} \theta_j \varepsilon_{t-j} + \varepsilon_t$		$b_t = \beta\left(a_t - a_{t-1}\right) + 1 - \beta b_{t-1}$
9. CCR-F | Model 3	$\varphi(L)\left(\pi_t - \pi_{t-1}\right) = \sum_{i=1}^{12} \mu_i D_{it} + \varepsilon_t$		$c_t = \gamma\left(\pi_t - a_{t+1}\right) - \gamma c_{t-s}$
10. CCR-F | Model 4	$\varphi(L)\pi_t = \sum_{i=1}^{12} \mu_i D_{it} + \kappa t + \varepsilon_t$		$0 < \alpha, \beta, \lambda < 1$

Notes: The parameters δ, ρ_i, θ_j, μ_i, and κ are estimated. $\varphi(L)$ is a lag polynomial whose order is determined by the BIC. The parameters α, β, and γ, are chosen jointly by minimizing the in-sample mean squared error. h is the forecast horizon, and s is the seasonal frequency. $\varepsilon_t \sim iidN\left(0, \sigma_\varepsilon^2\right)$.

Source: Authors' elaboration.

T+1 observations of π_t. We generate a sequence of $P(h)$ h-step-ahead forecasts estimating the models in rolling windows of fixed size R. For instance, to generate the first h-step-ahead forecasts, we estimate our models with the first R observations of our sample. Then these forecasts are built with information available only at time R and are compared to the realization π_{R+h}. Next, we estimate our models with the second rolling window of size R that includes observations through R+1. These h-step-ahead forecasts are compared with the realization π_{R+1+h}. We iterate until the last forecasts are built using the last R available observations for estimation. These forecasts are compared with the realization π_{T+1}. We generate a total of $P(h)$ forecasts, with $P(h)$ satisfying $R+\left(P(h)-1\right)+h=T+1$. Thus:

$$P(h)=T+2-h-R$$

Forecast accuracy is measured in terms of RMSPE. Because this is a population moment, we estimate it using the following sample analog:

$$\widehat{RMSPE}_h=\left[\frac{1}{P(h)}\sum_{t=R}^{T+1-h}\left(\pi_{t+h}-\hat{\pi}_{t+h|t}\right)^2\right]^{\frac{1}{2}}$$

where $\hat{\pi}_{t+h|t}$ represents the forecast of π_{t+h} made with information known up until time t. We carry out inference about predictive ability by considering pairwise comparisons between the models with the best performance within each family. By doing this, we acknowledge that our inference approach does not control for a familywise false discovery rate.[14] Methods of correctly controlling for a familywise type-I error between two families of models are the subject of current research, and some unpublished papers are making progress in this direction (see for instance Calhoun, 2011, and Pincheira, 2013).

Inference is carried out within the framework developed by Giacomini and White (2006) (GW). We focus on the unconditional version of the t-type statistic proposed by GW. This test has the distinctive feature of allowing comparisons between two competing forecast methods instead of two competing models. This is desirable for our purpose, which is purely focused on the forecasts that different time-series models estimated with rolling windows of fixed size can provide.

Once the best forecasting models within each family are chosen, we test the following null hypothesis:

$$H_0=E\left(d_h\right)\leq 0$$

against the alternative:

$$H_A:\mathrm{E}\left(d_h\right)>0$$

where:

$$d_t(h)=\left(\pi_{t+h}-\hat{\pi}_{t+h|\mathrm{t}}^{Benchmark}\right)^2-\left(\pi_{t+h}-\hat{\pi}_{t+h|\mathrm{t}}^{DESARIMA}\right)^2$$

We test the null hypothesis of superior predictive ability in favor of the family of traditional benchmarks. Accordingly, we use a one-sided version of the t-type test statistic proposed by GW.

[14] For a formal definition and exposition of the problems that arise from the familywise error rate, see Corradi and Distaso (2011).

Table 3 Unit Root Testing—Full Sample

	Level (π_t)				First differences $(\pi_t - \pi_{t-1})$			
	ADF	DFGLS	PP	KPSS	ADF	DFGLS	PP	KPSS
Canada	-3.903**	-1.389	-3.680**	0.133*	-8.350***	-0.962	-14.337***	0.063
Chile	-2.328	-0.586	-2.104	0.405***	-5.069***	-2.102	-12.083***	0.400*
Colombia	-1.913	-2.002	-1.570	0.396***	-9.782***	-7.939***	-9.648***	0.134
Israel	-3.118	-2.916	-2.488	0.362***	-9.045***	-1.750	-8.783***	0.067
Mexico	-3.816**	-3.819***	-2.692	0.097	-4.451***	-3.910***	-5.095***	0.037
Peru	-1.930	-0.469	-2.156	0.340***	-6.368***	-1.609	-9.895***	0.295
South Africa	-2.139	-1.581	-2.894	0.231***	-7.030***	-4.762***	-10.974***	0.051
Sweden	-4.219***	-0.746	-2.760	0.292***	-7.985***	-4.084***	-15.384***	0.197
Switzerland	-2.127	-1.855	-2.522	0.311***	-9.701***	-1.465	-15.167***	0.058
Turkey	-2.392	-1.171	-2.764	0.241***	-7.678***	-0.724	-16.425***	0.106
U. Kingdom	-1.678	-0.908	-1.857	0.399***	-14.347***	-1.459	-14.402***	0.270
USA	-3.523**	-1.825	-3.683**	0.112	-9.672***	-1.703	-10.488***	0.041

Notes: * p<10%, ** p<5%, *** p<1%. ADF denotes the Augmented Dickey-Fuller test, DFGLS the GLS detrended Dickey-Fuller test (Elliot, Rothenberg and Stock), PP the Phillips-Perron test, and KPSS the Kwiatkowski, Phillips, Schmidt and Shin test. The null hypothesis for ADF, DFGLS and PP tests is the series has a unit root. For KPSS test the null hypothesis is the series is stationary.

Source: Authors' elaboration.

3. Our Dataset

We use monthly CPI inflation data for Canada, Chile, Colombia, Israel, Mexico, Peru, South Africa, Sweden, Switzerland, Turkey, the United Kingdom and the US, covering the period from October 1990 to December 2011.[15] This is a sub-sample of inflation targeting countries plus the US. We select this sample by making sure that they all have data at a monthly frequency and they all have sufficiently long series to carry out an out-of-sample analysis covering the same sample period. The sources of the dataset are national central banks. This implies that, despite changes to the CPI market basket implemented by certain countries during the sample period, we use the same official inflation observations used by policymakers. The target variable corresponds to YoY CPI. Descriptive statistics of all the series for different samples are shown in *Appendix A* of the working paper version of this article (Pincheira and Medel, 2012b).

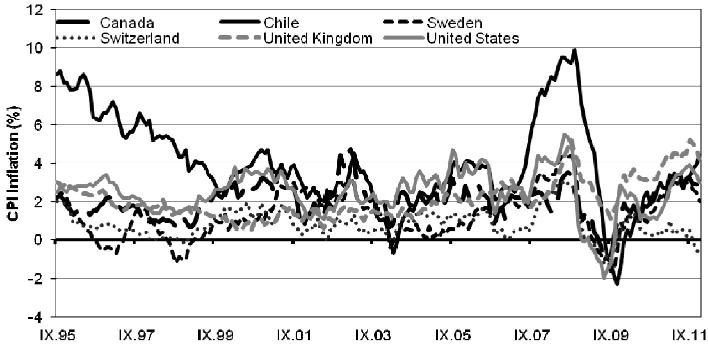
Table 3 shows the results of traditional unit root tests for our target variables for the sample period from October 1990 to December 2011. For most of the countries, the null hypothesis of a unit root cannot be rejected at the usual significance levels. Some exceptions are Canada, Mexico, Sweden and the US. The evidence for the first difference in inflation is, however, quite robust. In this case, the null hypothesis of a unit root is always rejected.

Note that some of the countries in our sample have converged to a seemingly stationary inflation process after experiencing a relatively long period of declining inflation. This mixture of regimes represents a special challenge to any method aimed

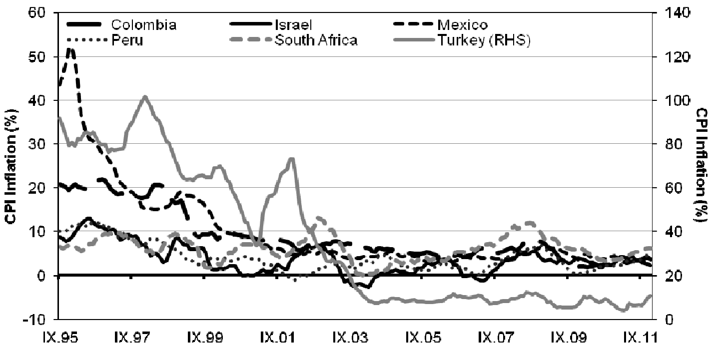
[15] We emphasize that we make use of non-seasonally adjusted data.

Figure 2 Inflation of CPI

A: Low-Inflation Countries



B: High-Inflation Countries



Notes: Vertical line = Evaluation sample start point (February 1999).
Source: National central banks.

at forecasting inflation during this sample period. Besides that, we have considered a relatively heterogeneous set of countries. This is clearly shown in *Figure 2*, in which we plot countries with low and high inflation separately. We will see in future sections that our DESARIMA family works well despite this heterogeneity.

We estimate the models with a rolling window of fixed length R. Because the exact choice of the rolling window size may play an important role in the accuracy of our forecasts, we generate predictions for two different values of R: 40 and 100 observations.[16] It is important to point out that our inference approach considers the size of the rolling window as given and not as a choice variable, as proposed by Hansen and Timmermann (2012). Our first estimation window of 40 observations covers the period from October 1995 to January 1999. When $R = 100$, the first

[16] We emphasize that, when selecting $R = 40$ and $R = 100$, our focus is on exploring the possible differences in the predictive behavior of our models when they are estimated with a small and moderate number of observations. Our analysis is mainly descriptive. The implementation of formal tests that are robust to windows size choice, such as those proposed by Rossi and Inoue (2011), may be an interesting extension for future research.

Table 4 Multi-Horizon RMSPE Estimates of the Best Model, Across *R*

		h = 1	*h* = 3	*h* = 6	*h* = 12	*h* = 24	*h* = 36
Canada	Best benchmark	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	AR(6)	AR[BIC]	AR(12)
	RMSPE	0.482	0.836	1.044	1.140	1.087	1.047
	Best DESARIMA	[5]	[5]	[5]	[5]	[1]	[2]
	RMSPE	0.366***	0.658**	0.890	1.105	1.068	0.957**
Chile	Best benchmark	AR[AIC]	AR(6)	AR(6)	AR(6)	AR(6)	AR(6)
	RMSPE	0.466	1.083	1.800	2.731	2.608	2.589
	Best DESARIMA	**[4]**	**[8]**	**[8]**	**[8]**	[4]	**[2]**
	RMSPE	0.392***	0.931**	1.695**	2.730	2.931	2.942
Colombia	Best benchmark	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	AR(6)	AR[AIC]
	RMSPE	0.404	1.018	1.407	1.930	2.360	2.669
	Best DESARIMA	**[8]**	**[8]**	**[8]**	**[7]**	**[7]**	**[9]**
	RMSPE	0.333**	0.860*	1.417	1.988	2.559	2.715
Israel	Best benchmark	AR(12)	AR(12)	**ARMA(1,1)**	**ARMA(1,1)**	**AR[BIC]**	AR(12)
	RMSPE	0.552	1.376	2.230	3.043	3.159	2.329
	Best DESARIMA	[8]	[8]	**[8]**	[1]	[1]	[1]
	RMSPE	0.436***	1.133***	1.891**	3.007	2.979	2.312
Mexico	Best benchmark	**ARMA(1,1)**	**ARMA(1,1)**	**ARMA(1,1)**	**CCR-F | M3**	ES[Sgl]-RW	ES[Sgl]-RW
	RMSPE	0.348	0.872	1.461	2.115	3.640	4.305
	Best DESARIMA	[1]	[1]	[9]	**[9]**	[9]	[9]
	RMSPE	0.315**	0.818*	1.357	2.380	3.115*	3.813*
Peru	Best benchmark	AR(6)	AR(6)	AR(6)	AR[AIC]	AR[AIC]	**AR[AIC]**
	RMSPE	0.415	0.921	1.503	2.267	2.619	2.339
	Best DESARIMA	[4]	**[8]**	**[8]**	**[8]**	**[3]**	[6]
	RMSPE	0.325***	0.791**	1.329*	2.276	2.698	2.417
South Africa	Best benchmark	AR(6)	AR(12)	AR(12)	AR[BIC]	AR[AIC]	AR[AIC]
	RMSPE	0.605	1.379	2.378	3.625	3.334	3.380
	Best DESARIMA	[4]	[4]	[4]	[4]	[8]	[5]
	RMSPE	0.434***	0.985***	1.768**	3.170*	3.508	3.674
Sweden	Best benchmark	AR(1)	ARMA(1,1)	AR(6)	AR(6)	AR(12)	AR(6)
	RMSPE	0.398	0.754	1.110	1.499	1.342	1.265
	Best DESARIMA	[5]	[3]	[3]	[1]	[5]	[5]
	RMSPE	0.318***	0.569***	0.868***	1.292	1.347	1.315
Switzerland	Best benchmark	ARMA[AIC]	ARMA[AIC]	ARMA[AIC]	ARMA(1,1)	AR[AIC]	AR[BIC]
	RMSPE	0.330	0.612	0.840	0.959	0.866	0.892
	Best DESARIMA	[3]	[1]	[4]	[4]	**[3]**	[1]
	RMSPE	0.295**	0.531***	0.700**	0.878*	0.846	0.927

Turkey	Best benchmark	**AR[AIC]**	CCR-F | M3	**ES[Single]**	ES[Sgl]-RW	**ES[Single]**	ES[Sgl]-RW
	RMSPE	1.945	5.590	9.442	14.319	18.092	24.270
	Best DESARIMA	**[8]**	**[8]**	**[8]**	**[7]**	[9]	[9]
	RMSPE	1.449***	4.237**	7.919**	14.271	17.867	21.149
U. Kingdom	Best benchmark	**ES[Sgl]-RW**	**ES[Single]**	**AR[AIC]**	**AR[AIC]**	**AR[AIC]**	**ES[Sgl]-RW**
	RMSPE	0.303	0.576	0.822	0.942	0.978	0.995
	Best DESARIMA	**[2]**	**[2]**	**[2]**	**[1]**	**[1]**	[6]
	RMSPE	0.259***	0.485**	0.711**	0.944	1.030	0.995
USA	Best benchmark	ARMA[AIC]	ARMA[AIC]	ARMA(1,1)	ARMA(1,1)	AR[BIC]	ARMA(1,1)
	RMSPE	0.442	1.003	1.350	1.490	1.438	1.492
	Best DESARIMA	[4]	[4]	[4]	[5]	[4]	[3]
	RMSPE	0.331***	0.751**	1.099**	1.495	1.471	1.487

Notes: We acknowledge that when making inference between the best individual forecasts from each family, we are not controlling for the familywise false discovery rate. Thus, these test results represent only auxiliary information which might not be precise enough. See *Tables 1* and *2* for DESARIMA and benchmark and specifications. GW test: * p<10%, ** p<5%, *** p<1%.

Source: Authors' elaboration.

estimation window covers the period from October 1990 to January 1999. The rest of the sample is used to compute forecast errors. We focus on one-, three-, six-, 12-, 24- and 36-months-ahead forecasts. Accordingly, we have a total of 155 observations for one-step-ahead forecast errors, 153 observations for three-steps-ahead forecast errors, 150 for six-steps-ahead forecast errors, 144 for 12-steps-ahead forecast errors, 132 for 24-steps-ahead forecast errors, and 120 observations for 36-steps-ahead forecasts errors.

4. Empirical Results

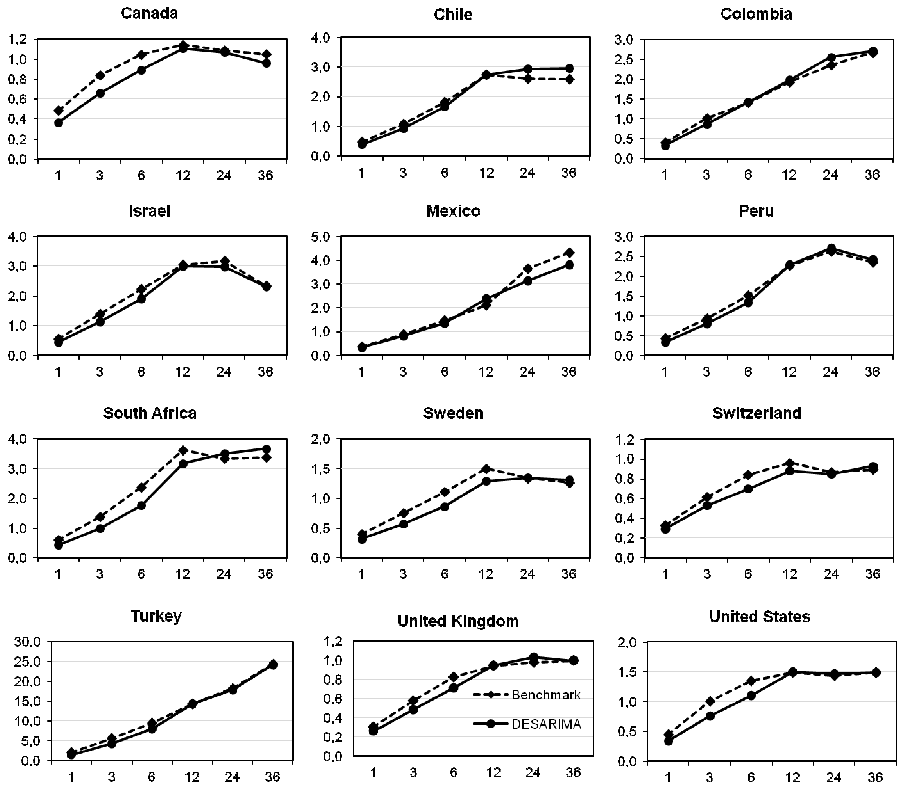
In a huge empirical exercise like the one we have carried out in this paper, there are a number of interesting findings that deserve mentioning. We will organize our discussion around four major topics: short-horizon accuracy, long-horizon accuracy, robustness check, and the role of the estimation window size. Most of our remarks focus on *Table 4* in which we report the RMSPE of the best-performing models in each family and the GW core statistic and its respective t-statistic when comparing the best models within each family.[17] Besides that, in *Appendix B* of the working paper version (Pincheira and Medel, 2012b), we show the RMSPE estimates for all the models and countries under evaluation.

4.1 Short-Horizon Accuracy

In *Table 4* we show sample RMSPE results for the best models within each family. These RMSPE are also depicted in *Figure 3* displayed below. Notice that for the construction of this figure and table, the DESARIMA family includes forecasts coming from 20 forecasting methods. These forecasting methods correspond to each

[17] We acknowledge that when making inference between the best individual forecasts from each family, we are not controlling for the familywise false discovery rate. Thus, these test results represent only auxiliary information which might not be precise enough.

Figure 3 Multi-horizon RMSPE Estimates of the Best Model, Across *R*



Source: Authors' elaboration.

DESARIMA model estimated with either $R = 40$ or $R = 100$ observations. Similarly, the benchmark family contains 26 forecasting methods coming from the 13 benchmark models estimated with either $R = 40$ or $R = 100$ observations. From *Table 4* we see that the best DESARIMA method always outperforms the best benchmark method when forecasting one and three months ahead. A similar result holds true six months ahead, with Colombia being the only exception. In summary, the best DESARIMA forecasts are almost always more accurate than the best benchmark forecast at short horizons. It is also interesting to note that the most frequent winners at short horizons (one, three and six months ahead) within the DESARIMA family are the models labeled DESARIMA 4 and 8, which are different from the more traditional IMA(1,1), RW and airline models, which correspond to those labeled as DESARIMA 7, 6 and 2, respectively.

4.2 Long-Horizon Accuracy

Table 4 depicts a different scenario when forecasting at the longer horizons of 12, 24, and 36 months ahead. Results are mixed and no clear winner between the two families under consideration arises from this table. In fact, we see that in seven countries the best one-year-ahead forecasting methods belong to the DESARIMA

Table 5 Average Ranking within the DESARIMA Family

	R = 40					
	***h* = 1**	***h* = 3**	***h* = 6**	***h* = 12**	***h* = 24**	***h* = 36**
Average RW rank	10	10	9	7	7	6
Average rank of models with MA(12)	4	4	4	5	5	5
Average rank of models without MA(12)	9	8	8	8	8	7
	***R* = 100**					
	***h* = 1**	***h* = 3**	***h* = 6**	***h* = 12**	***h* = 24**	***h* = 36**
Average RW rank	10	9	9	6	6	4
Average rank of models with MA(12)	4	4	5	6	6	6
Average rank of models without MA(12)	8	8	8	6	6	4

Source: Authors' elaboration.

family. Regarding two- and three-years-ahead forecasts, *Table 4* shows that only in four countries the lowest RMSPE is achieved by the DESARIMA family. It is important to mention that in some cases the RMSPE achieved at long horizons is remarkably low. For instance, for the cases of Canada, Switzerland and the United Kingdom, the RMSPE of the best models is lower than 100 basis points when forecasting three years ahead. Finally, we see that the most frequent winners at long horizons within the DESARIMA family are the models labeled DESARIMA 1 and 9.

It is important to emphasize that the good predictive performance of the DESARIMA family is closely related to the presence of seasonal MA components. We can see this in *Table 5*. In the first row of each panel of *Table 5* we display the average ranking across different countries of the RW within the ten different DESARIMA models. A ranking of 10 indicates that a given model X is the worst forecasting model within the DESARIMA family. A ranking of 1 would indicate that this model X is the best performing model. According to this table, when forecasting at short horizons (one, three and six months ahead), the RW is always the worst or the second worst performing model. At longer horizons the relative performance of the RW is better, with a best ranking of 4 and a worst ranking of 7. Each panel of the table also has two more rows indicating the average ranking (across countries and models) of the subfamily of models including a MA term of order 12, and the subfamily of models excluding this term. In general, models including a MA(12) term fare much better than models without this term. This is especially noticeable in the first panel and also in the second panel at short forecasting horizons. At longer horizons, however, the second panel shows either ties or an advantage for models without this MA(12) term. In summary, the RW does not explain the good performance of the DESARIMA family in the short run. This good performance is seemingly driven by models with an explicit seasonal MA(12) term. In particular, the most frequent winners within the DESARIMA family are the seasonal models 4 and 8 described in *Table 1*. As far as the family of benchmark models is concerned, and in line with the results reported by Makridakis *et al*. (1982), the ARMA(1,1), AR(6) and single ES models are the most frequent winners within the benchmark family.

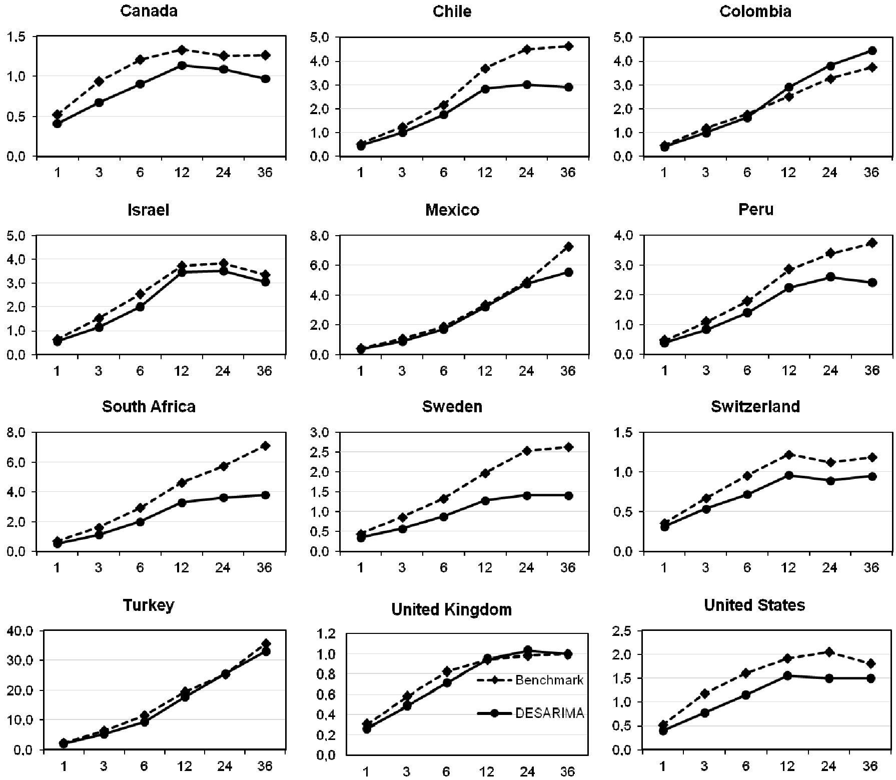
Table 6 Multi-Horizon RMSPE Estimates of the Median Forecast, Across *R*

	h = 1	*h* = 3	*h* = 6	*h* = 12	*h* = 24	*h* = 36
Canada						
RMSPE benchmark	0.493	0.852	1.088	1.307	1.185	1.107
RMSPE DESARIMA	0.399***	0.716**	0.973	1.209**	1.144	1.002**
Chile						
RMSPE benchmark	0.469	1.099	1.951	3.393	3.944	3.757
RMSPE DESARIMA	0.408***	1.003*	1.760*	2.828*	3.014**	2.924**
Colombia						
RMSPE benchmark	0.410	1.046	1.506	2.105	2.575	2.717
RMSPE DESARIMA	0.356***	0.942*	1.474	2.524	3.412	3.881
Israel						
RMSPE benchmark	0.553	1.379	2.242	3.396	3.386	2.418
RMSPE DESARIMA	0.463***	1.154***	1.976***	3.313	3.269	2.646
Mexico						
RMSPE benchmark	0.344	0.864	1.478	2.584	3.675	4.427
RMSPE DESARIMA	0.305***	0.825*	1.510	2.720	3.899	4.609
Peru						
RMSPE benchmark	0.424	0.973	1.601	2.617	3.079	2.757
RMSPE DESARIMA	0.352***	0.825***	1.393***	2.240**	2.604*	2.414
South Africa						
RMSPE benchmark	0.613	1.411	2.476	3.910	4.613	5.297
RMSPE DESARIMA	0.496***	1.163***	2.099***	3.421**	3.721***	4.161***
Sweden						
RMSPE benchmark	0.406	0.787	1.203	1.786	2.089	1.863
RMSPE DESARIMA	0.334***	0.608***	0.932**	1.392**	1.572***	1.489***
Switzerland						
RMSPE benchmark	0.340	0.634	0.914	1.166	1.028	1.083
RMSPE DESARIMA	0.302***	0.550***	0.748***	1.017*	0.895**	0.977*
Turkey						
RMSPE benchmark	1.999	5.570	9.699	15.869	20.024	27.982
RMSPE DESARIMA	1.718***	5.016**	8.979**	16.396	22.263	28.969
United Kingdom						
RMSPE benchmark	0.308	0.604	0.881	1.142	1.132	1.118
RMSPE DESARIMA	0.259***	0.497***	0.735***	1.013	1.071	1.118
United States						
RMSPE benchmark	0.471	1.055	1.454	1.771	1.757	1.722
RMSPE DESARIMA	0.380***	0.844***	1.212**	1.682**	1.570*	1.574**

Notes: See *Tables 1* and *2* for DESARIMA and benchmark and specifications. GW test: * *p*<10%, ** *p*<5%, *** *p*<1%.

Source: Authors' elaboration.

Figure 4 Multi-horizon RMSPE Estimates of the Median Forecast, Across *R*



Source: Authors' elaboration.

4.3 Robustness Check

Our previous analysis is focused on finding the best forecasting model within a set of forecasting methods. This task may be hard or, even worse, not very useful, as nothing ensures that the best model in a given sample will still be the best model in the future. Furthermore, our previous results may be questioned on the grounds of the method we have used to carry out inference about predictive ability. As we have not implemented methods to adequately control for the familywise type-I error rate, our inference may not be precise enough. To overcome these shortcomings, we also analyze the behavior of the median forecast coming from the 20 DESARIMA methods and from the 26 benchmark methods.[18] In *Table 6* and *Figure 4* we report the RMSPE of these forecasts. From *Table 6* we see that the median forecasts from the DESARIMA family always outperform the median benchmark forecasts at one and three months ahead. A similar result holds true six months ahead, with Mexico being the only exception. At longer horizons, *Table 6* shows that in only five countries there are cases in favor of the benchmark methods: Colombia, Israel, Turkey, Mexico and the United Kingdom. Therefore, at these horizons results are again

[18] Stock and Watson (2007) point out that the median acts as a good point estimator of a forecasting pool when errors are non-Gaussian.

Table 7 RMSPE of the Best-Performing Model of the Benchmark Family

		Pre and post-crisis analysis					
		h **= 1**	***h*** **= 3**	***h*** **= 6**	***h*** **= 12**	***h*** **= 24**	***h*** **= 36**
Canada	*Pre-crisis*	0.465	0.769	0.934	1.035	0.900	0.914
	Post-crisis	0.499	0.967	1.269	1.377	1.368	1.226
Chile	*Pre-crisis*	0.401	0.928	1.511	2.301	2.290	2.546
	Post-crisis	0.543	1.338	2.322	3.644	3.069	2.674
Colombia	*Pre-crisis*	0.425	1.065	1.359	1.685	2.098	2.580
	Post-crisis	0.310	0.850	1.538	2.422	2.743	2.200
Israel	*Pre-crisis*	0.599	1.503	2.468	3.485	3.574	2.675
	Post-crisis	0.369	0.808	1.002	1.223	1.478	1.306
Mexico	*Pre-crisis*	0.348	0.908	1.561	2.182	4.010	5.087
	Post-crisis	0.319	0.640	0.792	1.182	1.177	1.040
Peru	*Pre-crisis*	0.438	0.970	1.526	2.284	2.476	2.128
	Post-crisis	0.333	0.737	1.250	1.918	1.724	1.717
South Africa	*Pre-crisis*	0.663	1.507	2.634	3.974	3.672	3.753
	Post-crisis	0.369	0.857	1.336	2.013	2.277	2.232
Sweden	*Pre-crisis*	0.323	0.557	0.762	1.059	1.246	1.285
	Post-crisis	0.538	1.104	1.737	2.101	1.437	1.215
Switzerland	*Pre-crisis*	0.281	0.487	0.562	0.786	0.762	0.750
	Post-crisis	0.395	0.806	1.222	1.287	1.042	1.088
Turkey	*Pre-crisis*	2.170	6.326	10.77	16.59	21.00	28.51
	Post-crisis	0.952	1.745	2.231	2.348	2.641	2.641
U. Kingdom	*Pre-crisis*	0.260	0.463	0.648	0.764	0.786	0.944
	Post-crisis	0.402	0.802	1.135	1.278	1.221	0.923
US	*Pre-crisis*	0.357	0.665	0.895	1.127	1.023	1.132
	Post-crisis	0.610	1.559	2.154	2.032	2.121	2.016

Source: Authors' elaboration.

mixed but lean in favor of the DESARIMA forecasts, as in seven out of 12 countries the median DESARIMA forecast outperforms the median benchmark forecast at every single horizon. We see then that at short horizons the median forecast seems to be a very good alternative in terms of accuracy. The same happens at longer horizons for more than half of our sample of countries.

4.3.1 Pre- and Post-Crisis Analysis

The forecasts for the sample from February 1999 to December 2011 are evaluated as a whole. This sample includes a part of the Great Moderation and also the recent financial crisis and the Great Recession. So it is not impossible that this period includes structural breaks that affect the relative performance of the forecasting models. To check this, we carry out a separate analysis for the period before and after the Lehman default (September 2008). *Tables 7* and *8* show the RMSPE for both the pre- and post-crisis period. In *Table 7* we show the lowest RMSPE achieved

Table 8 RMSPE of the Best-Performing Model of the DESARIMA Family

		Pre and post-crisis analysis					
		***h* = 1**	***h* = 3**	***h* = 6**	***h* = 12**	***h* = 24**	***h* = 36**
Canada	*Pre-crisis*	**0.361**	**0.635**	**0.818**	**0.953**	**0.899**	**0.819**
	Post-crisis	**0.377**	**0.707**	**1.061**	1.429	1.389	**1.195**
Chile	*Pre-crisis*	**0.349**	**0.810**	**1.370**	**2.279**	2.290	2.773
	Post-crisis	**0.462**	**1.030**	**1.905**	**3.049**	3.112	**2.654**
Colombia	*Pre-crisis*	**0.355**	**0.922**	**1.353**	**1.649**	2.183	**2.517**
	Post-crisis	**0.255**	**0.643**	**1.126**	**1.964**	**2.467**	**2.444**
Israel	*Pre-crisis*	**0.468**	**1.226**	**2.078**	**3.449**	**3.375**	**2.559**
	Post-crisis	**0.310**	**0.646**	1.037	**1.144**	1.529	1.332
Mexico	*Pre-crisis*	**0.328**	**0.868**	**1.477**	2.625	**3.523**	**4.455**
	Post-crisis	**0.263**	**0.540**	**0.704**	**0.975**	**1.124**	1.245
Peru	*Pre-crisis*	**0.349**	**0.827**	**1.337**	**2.283**	2.476	2.454
	Post-crisis	**0.245**	**0.598**	**1.152**	2.012	2.072	**1.547**
South Africa	*Pre-crisis*	**0.457**	**1.060**	**1.922**	**3.516**	3.816	4.098
	Post-crisis	**0.322**	**0.588**	**1.020**	**1.922**	2.460	2.382
Sweden	*Pre-crisis*	**0.291**	**0.502**	**0.723**	1.133	1.285	1.360
	Post-crisis	**0.386**	**0.730**	**1.184**	**1.618**	1.483	**1.181**
Switzerland	*Pre-crisis*	0.282	**0.466**	**0.539**	**0.766**	**0.733**	**0.722**
	Post-crisis	**0.320**	**0.634**	**0.989**	**1.125**	**1.042**	1.139
Turkey	*Pre-crisis*	**1.598**	**4.809**	**9.105**	**16.497**	21.209	29.25
	Post-crisis	**0.852**	**1.651**	2.238	2.955	**2.441**	**3.824**
U. Kingdom	*Pre-crisis*	**0.224**	**0.389**	**0.592**	**0.749**	**0.785**	**0.941**
	Post-crisis	**0.302**	**0.627**	**0.969**	1.287	1.321	1.092
US	*Pre-crisis*	**0.301**	**0.567**	**0.708**	**1.051**	1.024	**1.082**
	Post-crisis	**0.403**	**1.113**	**1.785**	2.295	2.142	2.041

Source: Authors' elaboration.

by the benchmark models. In *Table 8*, we show the lowest RMSPE achieved by the DESARIMA family. In this last table we have colored the cells in which the DESARIMA family outperforms the benchmark family.

The results from *Tables 7 and 8* are robust in the short run. When one-step-ahead forecasts are considered, the DESARIMA family outperforms the benchmark family in all our countries with the only exception of Switzerland during the pre-crisis period. A similar situation occurs when considering three-months-ahead forecasts. In this case, Turkey is the only exception to the superiority of the DESARIMA family during the post-crisis period. Six months ahead, we have Turkey and Israel as exceptions during the post-crisis period. At longer horizons, the results are mixed when considering the best model in each family. We also constructed tables analyzing the median and average RMSPE of both families during the pre- and post-crisis. These tables, available upon request, are even more consistent in providing evidence in favor of the DESARIMA family when splitting the sample into the pre- and post-crisis periods. In summary, our results during the pre- and post-crisis periods are consistent with our overall results in terms of favoring the behavior of the DESARIMA family when forecasting at short horizons, i.e. up to six months. Results are again mixed when considering longer forecasting horizons.

4.3.2 The Average Forecast

As a robustness check, we also explore the behavior of the average forecast of both families of models. We first consider a weighted average using AIC-based weights according to Wagenmakers and Farrel (2004) and, second, the simple average of the forecasts, which imposes equal weights. Notice that when considering AIC-based weights, some shortcomings arise: first, the calculation of an AIC weight is not straightforward for all our benchmark methods. In the case of the subfamily of ES methods, there is no formal econometric model producing the forecasts. Therefore, the definition of the AIC weights in these cases is not obvious. Second, the benchmark family contains a model (CCR-F | Model 4) with a dependent variable that is transformed with a different operator than in the rest of the benchmark family. This difference may lead to an unfair comparison. Accordingly, when using AIC weights we consider only the average of the benchmark methods excluding the ES, double ES and Holt-Winters methods, for which an AIC weight is not obvious. For the sake of brevity, we do not show tables with our results (but they are available upon request). In general, when taking averages we obtain results that are very consistent with the previous results using the median forecast and the best model from each family. In the short run (one, three and six-month horizons) there is overwhelming superiority of the forecast combinations coming from the DESARIMA family, as DESARIMA-based combinations display a lower RMSPE than benchmark based combinations in all countries and for both estimation windows sizes; the only exception is Colombia when forecasting six months ahead. At longer horizons, the evidence is mixed, sometimes favoring the DESARIMA family and at other times the benchmark family. Therefore, the main conclusions remain the same when considering these types of forecast combinations.

4.3.3 Another Loss Function: The Hit Rate

Differing from traditional RMSPE comparisons, a researcher may also be interested in the ability that different forecasting methods may have to correctly predict if inflation rates are going up or down. We evaluate this dimension of our forecasting methods by computing the hit rate, i.e. the rate of correctly forecasting the direction of change in inflation rates. *Table 9* displays the average hit rate across models and estimation window sizes (R) within the two competing families: DESARIMA and benchmark. Shaded or colored cells indicate superiority of the DESARIMA family. The results are consistent with those obtained in terms of RMSPE: at short horizons the DESARIMA family is more effective when forecasting the direction of change of inflation rates. At longer horizons, however, results are again mixed.

Table 9 Average Hit Rate Across Models and *R* (in %)

		h = 1	*h* = 3	*h* = 6	*h* = 12	*h* = 24	*h* = 36
Canada	*Benchmark*	48	52	56	57	58	52
	DESARIMA	**60**	**62**	**63**	**61**	**53**	**56**
Chile	*Benchmark*	57	57	57	59	56	52
	DESARIMA	**66**	**63**	**62**	**62**	55	**55**
Colombia	*Benchmark*	48	50	51	52	61	64
	DESARIMA	**57**	**53**	**52**	45	43	39
Israel	*Benchmark*	50	51	54	57	53	57
	DESARIMA	**60**	**60**	**64**	**62**	**55**	56
Mexico	*Benchmark*	50	47	49	48	51	52
	DESARIMA	**56**	**53**	**54**	48	44	43
Peru	*Benchmark*	50	51	50	54	61	58
	DESARIMA	**60**	**60**	**60**	**59**	58	50
South Africa	*Benchmark*	53	56	56	55	52	55
	DESARIMA	**61**	**64**	**66**	**61**	**63**	**65**
Sweden	*Benchmark*	42	44	46	49	52	52
	DESARIMA	**51**	**53**	**55**	**56**	**67**	**71**
Switzerland	*Benchmark*	43	46	49	57	54	55
	DESARIMA	**52**	**55**	**59**	**64**	**58**	**59**
Turkey	*Benchmark*	56	54	55	52	47	49
	DESARIMA	**60**	**57**	**56**	46	44	38
U. Kingdom	*Benchmark*	45	45	45	48	49	46
	DESARIMA	**54**	**54**	**54**	**52**	**53**	**47**
US	*Benchmark*	47	51	51	58	56	55
	DESARIMA	**64**	**63**	**62**	**64**	**62**	**59**

Source: Authors' elaboration.

We also compare the highest hit rates achieved by models within the competing families. The results of this exercise, not reported for the sake of brevity, confirm those in the previous table: the superiority of the DESARIMA family in the short run and mixed results in the rest of the horizons. The only exception in this exercise is Colombia, for which the DESARIMA family fares better than the benchmark family only at the first forecasting horizon.

4.4 The Role of the Estimation Window Size (*R*)

We investigate the role that the size of the estimation window R may have in the accuracy of our forecasts. In stationary environments, we should expect a higher predictive performance of the methods estimated with longer samples. Nevertheless, an environment in which parameters are time-varying might be better handled by shorter estimation windows. *Table 10* shows how frequently each model produces

Table 10 Percentage of Times that a Model is Better when *R* = 40 vs *R* = 100 (in %)

	h = 1	*h* = 3	*h* = 6	*h* = 12	*h* = 24	*h* = 36
1. AR(1)	25	25	25	17	17	17
2. AR(6)	17	8	25	25	42	33
3. AR(12)	17	17	8	25	25	17
4. ARMA(1,1)	17	25	25	25	25	17
5. AR(12) AIC	33	17	17	25	33	42
6. AR(12) BIC	33	25	17	25	33	33
7. ARMA(12,6) AIC	8	17	8	25	33	17
8. ARMA(12,6) BIC	0	17	8	33	17	33
9. CCR-F | Model 3	0	0	8	8	17	25
10. CCR-F | Model 4	8	8	8	8	8	8
11. Single ES	17	25	42	42	42	58
12. Double ES	17	67	50	42	42	42
13. Holt-Winters	0	0	8	17	17	17
1. DESARIMA 1	8	33	33	33	50	33
2. DESARIMA 2	50	58	58	58	67	67
3. DESARIMA 3	50	58	58	58	67	67
4. DESARIMA 4	33	50	58	50	42	33
5. DESARIMA 5	58	58	58	58	67	67
6. DESARIMA 6	-	-	-	-	-	-
7. DESARIMA 7	33	50	67	58	58	67
8. DESARIMA 8	50	58	58	58	67	67
9. DESARIMA 9	50	67	67	17	17	33
10. DESARIMA 10	33	58	25	17	8	25

Notes: See *Table 1* and *2* for DESARIMA and benchmark and specifications.
Source: Authors' elaboration.

better forecasts when estimated with rolling windows of 40 observations. For most of the benchmark models these rates are lower than 50%, and sometimes much lower, indicating that in general they produce better forecasts when they are estimated with rolling windows of 100 observations. The only exceptions are the single and double ES, for which at a few forecasting horizons we find a better performance when estimating with only 40 observations.

The results from the DESARIMA family are different. In fact, the models labeled DESARIMA 2, 3, 5 and 8 show frequency rates greater than or equal to 50% at every horizon. For the rest of the models within the DESARIMA family, rates are not that high but in general are higher than in the benchmark family. In fact, the average rate of the DESARIMA family is 49%. This is in sharp contrast to the average rate of the benchmark family, which is only 22%.

Frequency rates offer only a partial view of the relative performance of the models estimated with rolling windows of 40 and 100 observations. This is because they are invariant to the size of RMSPE gains. For instance, a 1% RMSPE

Table 11 MSPE Ratios of the Model Estimated with R = 40 vs R = 100 Observations

	Average across families and horizons			
	Benchmark family	**DESARIMA family**	**Median benchmark**	**Median DESARIMA**
Canada	1.125	1.057	1.082	1.086
Chile	1.249	0.954	1.178	1.023
Colombia	2.248	0.858	1.174	0.880
Israel	1.286	0.930	1.026	0.954
Mexico	0.822	0.879	0.867	0.900
Peru	1.158	0.951	1.128	1.014
South Africa	1.578	1.282	1.320	1.099
Sweden	1.822	2.199	1.285	1.118
Switzerland	1.291	1.494	1.167	1.029
Turkey	1.097	0.892	0.880	0.863
United Kingdom	1.434	0.962	0.946	0.928
US	1.234	1.057	1.124	1.091
Overall	1.362	1.126	1.098	0.999

Notes: A number greater than 1 indicates that the MSPE of the models estimated with R=40 is higher than when estimated with R = 100.

Source: Authors' elaboration.

reduction is equivalent to a 50% RMSPE reduction when we only pay attention to frequency rates. Consequently, we need a more detailed analysis in terms of RMSPE gains. *Table 11* is useful for this purpose. The first two columns of *Table 11* show the average of the RMSPE ratio between the forecasts generated with 40 and 100 observations across horizons and models within a given family. The last two columns in the same table indicate the average of the RMSPE ratio between the median forecasts generated with 40 and 100 observations across horizons. A number greater than 1 indicates that the RMSPE of the forecasts generated with R = 40 is higher than when generated with R = 100.

According to the first column of *Table 11*, forecasts coming from the benchmark family display an average reduction of more than 25% in terms of RMSPE when these models are estimated with 100 observations. A reduction in RMSPE is achieved in every country except Mexico. The results are different for the DESARIMA family. While the RMSPE of the forecasts generated with 100 observations is lower on average, this result is not uniform across countries. Actually, in seven out of 12 countries it is better to generate the forecasts in the DESARIMA family using only 40 observations. With some variations, columns 3 and 4 in *Table 11* provide the same general picture when the median forecast within each family is considered.

A final word regarding one additional interesting feature of the median forecast is worth mentioning. From the first two columns in *Table 11*, we see that for some countries the choice of the estimation window size is very relevant for the construction of forecasts. In the case of Sweden, for instance, forecasts coming from the DESARIMA family estimated with only 40 observations display a RMSPE that more than doubles the RMSPE of the forecasts generated with 100 observations. Nevertheless, when we look at the fourth column in *Table 11*, we see, on average, only a minor edge in favor of the forecasts generated with 100 observations. According to *Table 11*, this is not an exception, as in most cases big ratios in the first two columns are associated with modest ratios in the third and fourth columns, suggesting that the median forecast seems to be a strategy that is relatively robust to the choice of the estimation window size.

5. Concluding Remarks

In this paper we introduce a family of univariate forecasting models that is shown to produce competitive inflation forecasts both at short and long horizons. This family of models is called Driftless Extended Seasonal ARIMA (DESARIMA) and contains ten seasonal driftless univariate time-series models sharing the common feature of a unit root.

Using out-of-sample Root Mean Squared Prediction Errors (RMSPE) we compare the forecasting accuracy of the DESARIMA family with that of traditional univariate time-series benchmarks for a sample of eleven inflation targeting countries plus the US. Our results show that DESARIMA-based forecasts display lower RMSPE at short horizons for every single country, with the exception of one case. We obtain mixed results at longer horizons. In particular, when the median forecast is considered, in more than half of the countries our DESARIMA-based forecasts outperform the benchmarks at long horizons and they are always superior at short horizons. This indicates that the median forecast of the DESARIMA family is an interesting and accurate forecast that should be seriously considered by applied forecasters.

Finally, we analyze the impact of the estimation window size on the accuracy of our forecasts. We do this by estimating all the models with two different sample sizes of 40 and 100 observations. While the traditional benchmarks tend to benefit from an increasing number of observations, this is less clear cut in the case of our DESARIMA-based forecasts. Interestingly, the median forecast seems to be a strategy that is relatively robust to the choice of the estimation window size.

A horse race between our DESARIMA family and more complex benchmarks seems to be an interesting issue to explore in subsequent research. For instance, we could either look at the class of Self-Exciting Threshold AR (SETAR) and Smooth Transition AR (STAR) non-linear models or exploit the cross-country angle of the data set by estimating a panel. Likewise, the construction of a multivariate DESARIMA family might also be of interest for future investigation.

REFERENCES

Aiolfi M, Capistrán C, Timmermann A (2011): Forecast Combinations. In: Clements MP, Hendry DF (Eds.): The *Oxford Handbook of Economic Forecasting.* Oxford University Press, USA.

Akaike H (1974): A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Andersson M, Karlsson G, Svensson J (2007): The Riksbank Forecasting Performance. *Economic Review*, 3:59–75.

Ang A, Bekaert G, Wei M (2007): Do Macro Variables, Assets Markets or Surveys Forecast Inflation Better? *Journal of Monetary Economics*, 54(4):1163–1212.

Atkeson A, Ohanian LE (2001): Are Phillips Curves Useful for Forecasting Inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1):2–11.

Box GEP, Jenkins GM (1970): *Time Series Analysis: Forecasting and Control*. 1st Edition. Holden-Day, San Francisco, USA.

Box GEP, Jenkins GM, Reinsel GC (2008): Time Series Analysis: Forecasting and Control. 4th Edition. Wiley, USA.

Brockwell PJ, Davis RA (1991): *Time Series: Theory and Methods*. 2nd Edition. Springer.

Calhoun G (2011): *An Asymptotically Normal Out-of-Sample Test Based on Mixed Estimation Windows*. Iowa State University, mimeo,

available at: http://www.econ.iastate.edu/~gcalhoun/dl/calhoun-mixed-window.pdf

Capistrán C, Constandse C, Ramos-Francia M (2010): Multi-Horizon Inflation Forecasts Using Disaggregated Data. *Economic Modelling*, 27:666–677.

Corradi V, Distaso W (2011): Multiple Forecast Model Evaluation. In: Clements MP, Hendry DF (Eds.): The *Oxford Handbook of Economic Forecasting*, Oxford University Press, USA.

Croushore D (2010): An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data. The *B.E. Journal of Macroeconomics*, 10(1)Article 10.

Diebold F, Mariano R (1995): Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13:253–265.

Elliot G, Timmermann A (2008): Economic Forecasting. *Journal of Economic Literature*, 46(1):3–56.

Ghysels E, Osborn D, Rodrigues PM (2006): Forecasting Seasonal Time Series. In: Elliot G, Granger CWJ, Timmermann A (Eds.): *Handbook of Economic Forecasting*, Volume 1, Elsevier, North Holland.

Giacomini R, White H (2006): Test of Conditional Predictive Ability. *Econometrica*, 74:1545–1578.

Granger CWJ, Jeon Y (2004): Forecasting Performance of Information Criteria with Many Macro Series. *Journal of Applied Statistics*, 31(10):1227–1240.

Groen J, Kapetanios G, Price S (2009): A Real Time Evaluation of Bank of England Forecasts of Inflation and Growth. *International Journal of Forecasting*, 25:74–80.

Hansen PR, Timmermann A (2012): Choice of Sample Split in Out-of-Sample Forecast Evaluation. *European University Institute, Economics Working Papers ECO*2012/10.

Harvey AC (1993): *Time Series Models*. 2nd Edition. Harvester-Wheatsheaf, New York.

Holan SH, Lund R, Davis G (2010): The ARMA Alphabet Soup: A Tour of ARMA Model Variants. *Statistics Survey*, 4:232–274.

Hyndman RJ, Koehler AB, Ord JK, Snyder RD (2008): Forecasting with Exponential Smoothing—The State Space Approach. *Springer Series on Statistics*, Berlin.

Makridakis S, Anderson A, Carbonne R, Fildes R, Hibon M, Lewandowski R, Newton J, Parzen E, Winkler R (1982): The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition. *Journal of Forecasting*, 1:111–153.

Marcellino M, Stock JH, Watson MW (2006): A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series. *Journal of Econometrics*, 135:499–526.

Pincheira P (2009): *Concatenation of Forecasts* [in Spanish]. Central Bank of Chile, mimeo, available upon request.

Pincheira P (2010): A Real Time Evaluation of the Central Bank of Chile GDP Growth Forecasts. *Money Affairs*, XXIII(1):37–73.

Pincheira P (2012a): Are Forecast Combinations Efficient? *Central Bank of Chile, Working Paper*, no. 661.

Pincheira P (2012b): *Are Forecast Combinations Efficient?* Available at Social Science Research Network, http://ssrn.com/abstract=2039670.

Pincheira P (2013): A Bunch of Models, a Bunch of Nulls and Inference About Predictive Ability. *Romanian Journal of Economic Forecasting*, XVI(3):26–43.

Pincheira P, Alvarez R (2009): Evaluation of Short Run Inflation Forecasts and Forecasters in Chile. *Money Affairs*, XXII(2):159–180.

Pincheira P, Medel CA (2012a): Forecasting Inflation with a Random Walk. *Central Bank of Chile, Working Paper*, no. 669.

Pincheira P, Medel CA (2012b): Forecasting Inflation with a Simple and Accurate Benchmark: A Cross-Country Analysis. *Central Bank of Chile, Working Paper*, no. 677.

Proietti T (2011): Direct and Iterated Multistep AR Methods for Difference Stationary Processes. *International Journal of Forecasting*, 27(2):266–280.

Rossi B, Inoue A (2011): Out-of-Sample Forecast Tests Robust to Window Size Choice. *Duke University, Dpt of Economics, Working Papers*, no. 11-04.

Rossi B (2013): Advances in Forecasting under Instabilities. In: Elliot G, Timmermann A (Eds.): *Handbook of Economic Forecasting*. Vol. 2. Elsevier, North Holland.

Schwarz G (1978): Estimating the Dimension of a Model. *Annals of Statistics*, 6(2):461–464.

Stock JH, Watson MW (1996): Evidence on Structural Stability in Macroeconomic Time Series Relations. *Journal of Business and Economic Statistics*, 14:11–30.

Stock JH, Watson MW (2003): Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature*, XLI:788–829.

Stock JH, Watson MW (2007): A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. In: Engle RF, White H (Eds.): *Cointegration, Causality, and Forecasting—A Festschrift in Honour of Clive W. J. Granger.* Oxford University Press, New York.

Wagenmakers E-J, Farrell S (2004): AIC Model Selection Using Akaike Weights. *Psychonomic Bulletin and Review*, 11:192–196.

West K (1996): Asymptotic Inference about Predictive Ability. *Econometrica*, 64:1067–1084.