JEL classification: C22, C53, E31, E37
Keywords: inflation forecasts, unit root, univariate time-series models, out-of-sample comparison, random walk

Forecasting with a Random Walk*

Pablo M. PINCHEIRA—School of Business, Adolfo Ibáñez University, Chile (pablo.pincheira@uai.cl), *corresponding author*

Carlos A. MEDEL—School of Economics, University of Nottingham, United Kingdom (carlos_medel@yahoo.com)

Abstract

The use of different time-series models to generate forecasts is fairly usual in the fields of macroeconomics and financial economics. When the target variable is stationary, the use of processes with unit roots may seem counterintuitive. Nevertheless, in this paper we demonstrate that forecasting a stationary variable with forecasts based on driftless unit-root processes generates bounded mean squared prediction errors at every single horizon. We also show that these forecasts are unbiased. In addition, we show via simulations that persistent stationary processes may be better predicted by driftless unit-root-based forecasts than by forecasts coming from a model that is correctly specified but is subject to a higher degree of parameter uncertainty. Finally, we provide an empirical illustration of our findings in the context of CPI inflation forecasts for a sample of industrialized economies.

1. Introduction

Some of the univariate models used to predict macroeconomic time series, such as inflation or GDP growth, involve the explicit presence of a unit root. Under the assumption of stationarity of the target variable, this may seem counterintuitive. In principle, one could think that unit-root-based forecasts may not be appropriate to predict a stationary process. This is so for a number of reasons. First, unit-root-based forecasts would have been generated from a model that is misspecified and over-differenced. Second, unit-root-based forecasts may have a deterministic trend approaching infinity (or minus infinity) as the forecasting horizon lengthens, which is clearly in conflict with a stationary process. Third, the optimal forecasts of a unit-root process display a divergent mean squared prediction error (MSPE) as the forecasting horizon approaches infinity. This may lead one to think that a similar property might hold true when forecasting a stationary process with unit-root-based forecasts.

Despite these arguments, results in Atkeson and Ohanian (2001), Giacomini and White (2006), Elliot and Timmermann (2008), Stock and Watson (2008), Groen, Kapetanios and Price (2009) and Capistrán, Constandse and Ramos-Francia (2010), among others, show that unit-root-based forecasts perform well when forecasting inflation or GDP growth, which are variables that may be considered stationary in a number of countries. It is in the context of these findings that we pose the three following questions: When predicting a stationary variable with unit-root-based forecasts, does the MSPE diverge as the forecasting horizon lengthens? Are these unit-root-based forecasts biased? Is it possible that unit-root-based forecasts for a per-

* The authors would like to thank Martin Fukač (Editor), four anonymous referees, seminar participants at the Economics Society of Chile and the Central Bank of Chile, and the 11th Prague International Academic Conference for their comments. Any errors or omissions are the responsibility of the authors.

sistent stationary process perform better than forecasts generated from a correctly specified model in finite samples? In this paper, we aim to answer these three questions.

In order to do so, in Section 2 we analyze the behavior of the MSPE of unit-root-based forecasts for stationary variables as the forecast horizon lengthens. We also analyze the bias of unit-root-based forecasts. In Section 3 we report Monte Carlo simulations evaluating the ability of unit-root-based forecasts to predict a stationary process. An empirical illustration based on year-on-year (YoY) Consumer Price Index (CPI) inflation for Canada, Sweden, Switzerland, the US and the UK is presented in Section 4. Finally, Section 5 concludes the paper.[1]

2. Forecasting with a Unit-Root Process

To set forth some preliminary ideas, let us consider that the true model of variable Y_t is the following Gaussian stationary AR(1) process, $Y_{t+1} = \alpha + \rho Y_t + \varepsilon_{t+1}$, where ε_{t+1} is white noise with variance σ_ε^2, $\alpha \neq 0$, and $0 < \rho < 1$. By iterating forward, it is possible to show that for an arbitrary horizon $h \in \mathbb{N}$ we have

$$Y_{t+1} = \alpha\left[\frac{1-\rho^h}{1-\rho}\right] + \rho^h Y_t + \sum_{i=0}^{h-1} \rho^i \varepsilon_{t+h-i}$$

The best linear h-step-ahead forecast $Y_t^f(h)$ and its corresponding error $e_t^f(h)$ are given by

$$Y_t^f(h) = \alpha\left[\frac{1-\rho^h}{1-\rho}\right] + \rho^h Y_t$$

$$e_t^f(h) = Y_{t+h} - Y_t^f(h) = \varepsilon_{t+h} + \sum_{i=1}^{h-1} \rho^i \varepsilon_{t+h-i}$$

Suppose now that we forecast Y_{t+h}, assuming that the true data generating process (DGP) is a driftless random walk (RW) that delivers the following forecast, $Y_t^{RW}(h)$, and forecast error, $e_t^{RW}(h)$:

$$Y_t^{RW}(h) = Y_t$$

$$e_t^{RW}(h) = Y_{t+h} - Y_t = \alpha\left[\frac{1-\rho^h}{1-\rho}\right] - \left(1-\rho^h\right) Y_t + \sum_{i=0}^{h-1} \rho^i \varepsilon_{t+h-i}$$

The MSPE thus is given by

$$MSPE(h) = \mathbb{E}\left[Y_{t+h} - Y_t^{RW}(h)\right]^2$$

$$MSPE(h) = \mathbb{E}\left[\alpha\left[\frac{1-\rho^h}{1-\rho}\right] - \left(1-\rho^h\right) Y_t\right]^2 + \mathbb{E}\left[\sum_{i=0}^{h-1} \rho^i \varepsilon_{t+h-i}\right]^2$$

[1] We would like to point out here that the focus of our paper is on global measures of forecasts accuracy. In other words, we are concerned with the average relative performance between two forecasting strategies. Recent literature has placed attention on the stability of different forecasting methods. See, for instance, Giacomini and Rossi (2009, 2010). While the question about stability is very important, we will leave it as an extension for further research.

$$MSPE(h) = \left(1-\rho^h\right)^2 \mathbb{V}\left[Y_t\right] + \sigma_\varepsilon^2 \left(\frac{1-\rho^{2h}}{1-\rho^2}\right)$$

then

$$\lim_{h\to\infty} MSPE(h) = 2\mathbb{V}\left(Y_t\right)$$

Thus, because Y_t is stationary, forecast errors coming from an RW-based forecast do not display explosive behavior as the forecasting horizon lengthens. Furthermore, RW-based forecasts display another interesting property: they are unbiased. To see this, let us consider the expected value of the forecast error:

$$\begin{aligned} Bias(h) &\equiv \mathbb{E}\left[e_t^{RW}(h)\right] = \mathbb{E}\left[Y_{t+h} - Y_t^{RW}(h)\right] \\ &= \mathbb{E}\left[\alpha\left[\frac{1-\rho^h}{1-\rho}\right] - \left(1-\rho^h\right)Y_t + \sum_{i=0}^{h-1}\rho^i\varepsilon_{t+h-i}\right] \\ &= \alpha\left[\frac{1-\rho^h}{1-\rho}\right] - \left(1-\rho^h\right)\mathbb{E}\left[Y_t\right] \\ &= \alpha\left[\frac{1-\rho^h}{1-\rho}\right] - \left(1-\rho^h\right)\frac{\alpha}{1-\rho} \\ &= 0 \end{aligned}$$

Note that for these two implications, the "no-drift" assumption (denoted as $\delta = 0$) plays a key role. In fact, if we had assumed that the true DGP is an RW with drift, $Y_{t+1} = \delta + Y_t + \varepsilon_t$, we would have ended with forecasts $Y_t^{RWD}(h)$ and corresponding forecast errors $\varepsilon_t^{RWD}(h)$ according to

$$Y_t^{RWD}(h) = \delta h + Y_t$$

$$e_t^{RWD}(h) = \alpha\left[\frac{1-\rho^h}{1-\rho}\right] - \delta h - \left(1-\rho^h\right)Y_t + \sum_{i=0}^{h-1}\rho^i\varepsilon_{t+h-i}$$

In this case, showing that the MSPE satisfies is straightforward:

$$\lim_{h\to\infty} MSPE_D(h) = \lim_{h\to\infty}\left[MSPE(h) + (\delta h)^2\right] = 2\mathbb{V}\left[Y_t\right] + \lim_{h\to\infty}(\delta h)^2 = +\infty$$

and the drift will generate a divergent MSPE. Similarly, with a little algebra it is possible to show that the drift will generate biased forecasts:

$$Bias_D(h) \equiv \mathbb{E}\left[e_t^{RWD}(h)\right] = \mathbb{E}\left[Y_{t+h} - Y_t^{RWD}(h)\right] = -\delta h$$

So, again, the "no-drift" assumption is key for keeping the forecasts unbiased.

Now, let us assume that the true DGP is the same AR(1) but with $\rho = 1$. Accordingly:

$$Y_{t+1} = \alpha h + Y_t + \sum_{i=0}^{h-1} \varepsilon_{t+h-i}$$

By iterating forward, it is possible to show that for an arbitrary horizon $h \in \mathbb{N}$ we have

$$Y_{t+h} = \alpha h + Y_t + \sum_{i=0}^{h-1} \varepsilon_{t+h-i}$$

The best linear *h*-step ahead forecast $Y_t^f(h)$ and its corresponding error $e_t^f(h)$ are given by

$$Y_t^f(h) = \alpha h + Y_t$$

$$e_t^f = Y_{t+h} - Y_t^f(h) = \varepsilon_{t+h} + \sum_{i=1}^{h-1} \varepsilon_{t+h-i}$$

and the optimal MSPE diverges as the horizon lengthens:

$$MSPE(h) = \mathbb{E}\left[Y_{t+h} - Y_t^f(h)\right]^2$$

$$MSPE(h) = \mathbb{E}\left[\sum_{i=0}^{h-1} \varepsilon_{t+h-i}\right]^2 = \sigma_\varepsilon^2 h \rightarrow_{h\rightarrow\infty} +\infty$$

This last result is general to unit-root processes. Their optimal forecasts have increasing confidence bands (see Box, Jenkins and Reinsel, 2008). Nevertheless, when used to predict stationary variables, driftless RW-based forecasts display a bounded $MSPE(h)$ sequence as the forecasting horizon goes to infinity. The next proposition generalizes the previous AR(1) example to a broader class of stationary processes.

Proposition 1: *Let Y_t be a stationary process, then driftless RW-based forecasts are unbiased and display a bounded MSPE as the forecasting horizon goes to infinity.*

Proof: Suppose that we forecast Y_{t+h}, assuming that the true DGP is a driftless RW that delivers the following forecast $Y_t^{RW}(h)$ and forecasting errors $\varepsilon_t^{RW}(h)$:

$$Y_t^{RW}(h) = Y_t$$

$$e_t^{RW}(h) = Y_{t+h} - Y_t$$

Because of the stationarity of Y_{t+h} we have

$$\mathbb{E}\left[Y_{t+h} - Y_t^{RW}(h)\right] = \mathbb{E}\left[Y_{t+h} - Y_t\right] = \mathbb{E}\left[Y_{t+h}\right] - \mathbb{E}\left[Y_t\right] = 0$$

and therefore, driftless RW-based forecasts are unbiased. The MSPE is given by

$$MSPE(h) = \mathbb{E}\left[Y_{t+h} - Y_t^{RW}(h)\right]^2 = \mathbb{E}\left[Y_{t+h} - Y_t\right]^2$$

$$MSPE(h) = \mathbb{V}\left[Y_{t+h}\right]^2 + \mathbb{V}\left[Y_t\right]^2 - 2\mathbb{C}\left[Y_{t+h}, Y_t\right]$$

$$MSPE(h) = 2\mathbb{V}\left[Y_t\right] - 2\gamma_h$$

So,

$$MSPE(h) = \left|2\mathbb{V}[Y_t] - 2\gamma_h\right| \leq 2\mathbb{V}[Y_t] + 2\left|\gamma_h\right|$$

$$MSPE(h) \leq \mathbb{V}[Y_t] + 2\sqrt{\mathbb{V}[Y_{t+h}]\mathbb{V}[Y_t]} = 2\mathbb{V}[Y_t] + 2\sqrt{\mathbb{V}[Y_t]\mathbb{V}[Y_t]}$$

$$MSPE(h) \leq 4\mathbb{V}[Y_t]$$

and

$$\lim_{h\to\infty} MSPE(h) \leq 4\mathbb{V}[Y_t] < +\infty$$

and then $MSPE(h)$ is a bounded sequence.

Remark: If, in addition to stationarity, we assume that the process Y_t has absolutely summable autocovariances, which is to say

$$\sum_{i=0}^{\infty} \left|\gamma_i\right| < +\infty \tag{1}$$

where $\gamma_i = \mathbb{E}\left[Y_t - \mathbb{E}[Y_t]\right]\left[Y_{t-1} - \mathbb{E}[Y_{t-1}]\right]$, and $\mu = \mathbb{E}[Y_t] = \mathbb{E}[Y_{t-i}]$, for all $i \in \mathbb{Z}$, then we can achieve tighter bounds, because (1) implies

$$\lim_{h\to\infty} \left|\gamma_h\right| \leq \lim_{h\to\infty} \sum_{i=h}^{\infty} \left|\gamma_i\right| = 0$$

Therefore,

$$\lim_{h\to\infty} MSPE(h) = \lim_{h\to\infty} 2\mathbb{V}[Y_t] - 2\lim_{h\to\infty} \gamma_h = 2\mathbb{V}[Y_t] < +\infty$$

and the sequence $MSPE(h)$ is not only bounded but also convergent.

Unit-root-based forecasts are commonly used in the literature. For instance, Atkeson and Ohanian (2001) show that a simple RW model for inflation in the US is very competitive when predicting 12 months ahead. Giacomini and White (2006), also for the US, present an empirical application in which several CPI forecasts are compared to those generated by an RW with drift and an autoregression (AR) whose lag length is selected according to the Bayesian information criterion (BIC). Another article using simple univariate benchmarks for the US is Ang, Bekaert and Wei (2007). Among the many methods the authors use, they include an RW. Stock and Watson (1999, 2008) also consider integrated processes of order one when evaluating inflation forecasts for the US. They mention that modeling inflation as I(1) is usual in the literature. In addition, Croushore (2010) makes use of an integrated moving average IMA(1,1) model as a benchmark when evaluating survey-based inflation forecasts for the US. Similarly, Stock and Watson (2007a) make use of a time-varying IMA(1,1) process when analyzing inflation in the US, concluding that inflation in that country has become both easier and harder to forecast depending on one's point of view.[2]

[2] On the one hand, it has become harder to forecast in the sense that the value added of traditional Phillips curve models has declined relative to simple univariate benchmarks. On the other hand, it has become easier to forecast because the MSPE of the forecast has fallen.

The use of unit-root-based forecasts for macro variables is fairly usual in other countries as well. Groen, Kapetanios and Price (2009), for instance, evaluate the accuracy of the Bank of England inflation and GDP growth forecasts using several univariate models, including an $AR(p)$ and the RW as benchmarks. Capistrán, Constandse and Ramos-Francia (2010) make use of seasonal unit-root models to forecast inflation in Mexico. Similarly, Pincheira and Medel (2015) also make use of unit-root-based forecasts to predict YoY CPI inflation for twelve countries both at short and long horizons.

Next, we show another proposition that goes in the same line as Proposition 1. The new aspect is that now we allow for more general types of unit-root-based forecasts coming from the $ARIMA(p,1,q)$ family.

Proposition 2: *Let Y_t be a stationary process as in Proposition 1. Let us also consider a white noise process $\{\varepsilon_{t+1}\}_{t=-\infty}^{+\infty}$ with variance σ_ε^2, so that the moments $\mathbb{C}\left[Y_{t+j},\varepsilon_{t+i}\right]$ for all $i, j\epsilon\mathbb{Z}$ are well defined. Then optimal linear forecasts coming from a driftless $ARIMA(p,1,q)$ process using $\{\varepsilon_{t+1}\}_{t=-\infty}^{+\infty}$ as innovations and MA terms, with $0 \leq p,q \leq \infty$, will display a bounded $MSPE$ sequence as the forecasting horizon approaches infinity.*

Proof: See *Appendix A* (on the website of this journal).

So far we have shown that the construction of driftless unit-root-based forecasts for stationary variables does not imply explosive behavior of the MSPE as the forecast horizon lengthens. Nevertheless, to some extent this is a weak result. It is saying that driftless unit-root-based forecasts for stationary variables are not extremely bad in the long run. The next proposition shows a more powerful result. It shows that, in general, unit-root-based forecasts are unbiased.

Proposition 3: *Let Y_t be a stationary process as in Proposition 1. Let us also consider a white noise process $\{\varepsilon_{t+1}\}_{t=-\infty}^{+\infty}$ with variance σ_ε^2. Then optimal linear forecasts coming from a driftless $ARIMA(p,1,q)$ process using $\{\varepsilon_{t+1}\}_{t=-\infty}^{+\infty}$ as innovations and MA terms, with $0 \leq p,q \leq \infty$, are unbiased at every forecasting horizon.*

Proof: See *Appendix B* (on the website of this journal).

In the next section, we will show with simulations that parameter uncertainty in combination with persistence may generate a high degree of noise in the ordinary least squares (OLS) estimates of simple stationary processes. Under this scenario, we will provide evidence that driftless unit-root-based forecasts may offer more accuracy than correctly specified forecasts in small and moderate samples. This may be particularly relevant at long horizons.

3. Monte Carlo Simulations

3.1 AR(1) and AR(2) Models

We generate 5,000 replications of two stationary processes: an AR(1) and an AR(2) model, first setting the drift to zero and later setting it to one. We generate these four processes from independent zero-mean homoskedastic Gaussian shocks with variance equal to $\sigma_\varepsilon^2 = 1$. Thus, the models appear as follows:

$$AR(1): Y_{t+1} = \alpha + \rho Y_t + \varepsilon_{t+1}$$

$$\sigma_\varepsilon^2 = 1 \, and \, 0 < \rho < 1$$

$$AR(2): Y_{t+1} = \alpha + \phi_1 Y_t + \phi_2 Y_{t-1} + \varepsilon_{t+1}$$

$$\sigma_\varepsilon^2 = 1 \, and \, 0 < \phi_1 + \phi_2 < 1$$

We will be interested in the persistence of the processes. We will use ρ and $\phi_1 + \phi_2$ as measures of persistence in the AR(1) and AR(2) models respectively.

In each replication, we generate a total of $R+P+l$ observations, where *R* represents the estimation sample size used in our simulations. We consider different exercises with *R* taking three different values: 50, 100 and 200. The parameter *l* takes the values 1 or 2 depending on the process we are considering—AR(1) or AR(2). We do this because we drop one observation to estimate an AR(1) model and we drop two observations when estimating the AR(2) model. *P* represents the number of one-step-ahead predictions. In all our simulations, we set $P = 500$. We are not interested only in one-step-ahead forecasts, so we engage in an out-of-sample *h*-step ahead evaluation, with $h \in \{1;12;24;36;48;96;120\}$, where the parameters of the processes are estimated with rolling OLS.

The construction of multistep-ahead forecasts is done using the iterative method. This means that *h*-step-ahead forecasts are built as a recursive function of forecasts built for shorter horizons. An alternative way to construct multistep-ahead forecasts is the so-called "direct method". This approach differs from the iterative method in that a direct relationship between the target variable Y_{t+h} and the set of predictors known at time *t* is established and there is no need to construct short horizon forecasts in order to build long horizon forecasts.[3] Although in the tables we present next we make use of the iterative method for multistep forecasts, we have also constructed tables using the direct method. For the sake of brevity, we do not report these tables in the paper, but they are available upon request. Generally speaking, the same point that we will see in *Table 1* holds for the direct method: driftless unit-root-based forecasts may outperform forecasts coming from the true DGP in small and moderate samples when the process is stationary but persistent.

For each of the processes we construct forecasts using two different methodologies. First, we generate optimal forecasts assuming that we know the specification of the models, but also assuming that the parameters of these models are unknown

[3] See Jordà (2005) and Marcellino, Stock and Watson (2006) for a discussion of the advantages of the direct method and a comparison between direct and iterative multistep forecasts.

Table 1 MSPE Ratios between Driftless RW-Based Forecasts and DGP-Based Forecasts

	AR(1)					AR(2)					
ρ	**0.50**	**0.90**	**0.95**	**0.98**	**0.99**	**ϕ_1**	**0.40**	**0.50**	**0.50**	**0.50**	**0.50**
	-	**-**	**-**	**-**	**-**	**ϕ_2**	**0.10**	**0.40**	**0.45**	**0.48**	**0.49**
α = 1											
R = 50											
h = 1	1.278	0.996	0.965	0.951	0.946		1.312	1.200	1.209	1.218	1.227
h = 12	1.886	1.211	0.956	0.836	0.790		1.877	1.098	0.905	0.830	0.803
h = 24	1.885	1.297	0.913	0.720	0.637		1.881	1.094	0.769	0.672	0.631
h = 36	1.888	1.219	0.715	0.455	0.320		1.881	0.801	0.142	0.345	0.329
h = 48	1.887	0.686	0.307	0.102	0.035		1.881	0.124	0.001	0.038	0.053
h = 96	1.886	0.000	0.000	0.000	0.000		1.882	0.000	0.000	0.000	0.000
h = 120	1.887	0.000	0.000	0.000	0.000		1.880	0.000	0.000	0.000	0.000
R = 100											
h = 1	1.306	1.027	0.997	0.981	0.973		1.355	1.252	1.261	1.268	1.274
h = 12	1.942	1.373	1.097	0.944	0.870		1.937	1.255	1.032	0.927	0.878
h = 24	1.942	1.573	1.213	0.959	0.831		1.939	1.413	1.074	0.897	0.816
h = 36	1.945	1.642	1.279	0.955	0.785		1.938	1.494	1.095	0.864	0.755
h = 48	1.941	1.665	1.309	0.923	0.718		1.938	1.528	1.085	0.806	0.674
h = 96	1.942	1.677	1.141	0.272	0.166		1.939	1.446	0.486	0.170	0.112
h = 120	1.941	1.671	0.687	0.025	0.018		1.937	1.000	0.070	0.014	0.010
R = 200											
h = 1	1.320	1.041	1.013	0.998	0.989		1.378	1.276	1.290	1.296	1.301
h = 12	1.970	1.459	1.189	1.031	0.933		1.968	1.344	1.123	1.005	0.939
h = 24	1.970	1.703	1.363	1.095	0.921		1.970	1.557	1.222	1.015	0.898
h = 36	1.971	1.789	1.482	1.154	0.919		1.970	1.673	1.311	1.038	0.879
h = 48	1.970	1.815	1.555	1.204	0.916		1.969	1.728	1.379	1.060	0.862
h = 96	1.972	1.831	1.655	1.297	0.857		1.968	1.772	1.497	1.079	0.759
h = 120	1.970	1.832	1.662	1.293	0.789		1.968	1.771	1.511	1.039	0.671

Note: Figures below unity favor RW based forecasts.

Source: Authors' elaboration.

and must be estimated with rolling OLS.[4] Second, we generate optimal forecasts under the assumption that the processes are driftless RW. In each replication, we compute the sample MSPE of the forecasts. Then, using 5,000 replications, we compute the average across the entire sample MSPE to get a good estimate of the population MSPE. In *Table 1*, under the columns "AR(1)" and "AR(2)", we report the MSPE-ratio defined as

$$MSPE\ Ratio = \frac{MSPE^{RW}(h)}{MSPE^{AR}(h)}$$

[4] We always estimate the processes including a constant in our regressions.

We report these ratios for the three values of $R \in \{50;100;200\}$ and several choices of the parameters that define the AR(1) and AR(2) models. In particular, we consider AR(1) specifications with the following parameter values:

$$\rho \epsilon \{0.5; 0.9; 0.95; 0.975; 0.99\}$$

For the AR(2) model we consider the following parameter values:[5]

$$(\phi_1, \phi_2) \in \{(0.4, 0.1); (0.5, 0.4); (0.50, 0.45); (0.500, 0.475); (0.50, 0.49)\} \quad (2)$$

An MSPE ratio below unity implies that RW-based forecasts outperform those coming from the correctly specified model.

For the sake of brevity, in *Table 1* we only present results for the case in which the drift α is set to one. Results when the drift is set to zero are qualitatively similar and are available upon request. *Table 1* shows three salient features that are worth mentioning:

1. As the estimation sample size gets larger, all the ratios become larger as well. This is easy to understand, because a larger estimation sample implies more precise parameter estimation. With little estimation noise, we should expect better performance of correctly over incorrectly specified forecasts.
2. As the persistence of the processes increases, all the ratios show a tendency to decrease (with only a few exceptions). In fact, most of the ratios are below one when persistence equals 0.99 in *Table 1*. In the case of the AR(1) process, we detect two major drivers behind these results. First, as ρ gets larger, the process approaches an RW. Therefore, the RW becomes closer to the correct specification. Second, as ρ gets larger, the small sample bias of the OLS estimates of ρ gets worse. These two forces point in the same direction and help to explain the good behavior of RW-based forecasts over correctly specified forecasts when ρ is close to one and sample sizes are not large. For the AR(2) process, the first reason stated above might not be very compelling, as when $\phi_1 + \phi_2$ is close to 1 with our choice of parameters in (2), the AR(2) process is still an AR(2) and does not approach an RW. Nevertheless, the second reason holds perfectly well in this scenario. These two salient features are relatively well known in the literature. Actually, Stock and Watson (2007a) and Hamilton (1994) provide interesting discussions regarding OLS estimation of parameters from persistent process. Furthermore, the development of out-of-sample tests of Granger causality such as those in Clark and West (2006, 2007) are based on the problems that non-vanishing parameter uncertainty may generate when performing out of sample inference.

Table 1 shows an interesting interaction between persistence, sample size and forecasting horizon. We see that given a sample size of R, there is a persistence threshold beyond which the MSPE ratios do not increase with the forecasting horizon. For instance, for $R = 50$ and the AR(1), when ρ is greater or equal to 0.95, we get these non-increasing patterns. Of course, as R gets larger, these non-increasing patterns are smoother. All this means that the problem of noisy

[5] These values belong to the stationarity region for an AR(2) process, which is characterized by the following expressions: $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$, and $|\phi_2| < 1$.

Table 2 MSPE Ratios between OLS-Based Forecasts and Forecasts Made with the True Parameters

	AR(1)					AR(2)					
ρ	**0.50**	**0.90**	**0.95**	**0.98**	**0.99**	**ϕ_1**	**0.40**	**0.50**	**0.50**	**0.50**	**0.50**
	-	**-**	**-**	**-**	**-**	**ϕ_2**	**0.10**	**0.40**	**0.45**	**0.48**	**0.49**
$\alpha = 1$											
R = 50											
h = 1	1.044	1.057	1.063	1.067	1.062		1.067	1.082	1.086	1.087	1.084
h = 12	1.061	1.288	1.356	1.379	1.342		1.065	1.317	1.372	1.366	1.327
h = 24	1.061	1.432	1.688	1.806	1.757		1.064	1.572	1.85	1.818	1.732
h = 36	1.061	1.610	2.409	3.154	3.683		1.064	2.337	11.131	3.799	3.429
h = 48	1.061	2.910	5.983	15.284	35.252		1.064	15.634	1326	36.83	22.123
h = 96	1.061	175258	285031	11M	102M		1.063	896773	1337M	202M	32M
h = 120	1.061	81M	149M	158M	246M		1.063	532M	442M	6531MM	761M
R = 100											
h = 1	1.020	1.027	1.028	1.033	1.033		1.031	1.037	1.040	1.042	1.041
h = 12	1.029	1.133	1.182	1.219	1.224		1.032	1.154	1.201	1.221	1.216
h = 24	1.029	1.175	1.277	1.352	1.350		1.032	1.221	1.317	1.359	1.342
h = 36	1.029	1.187	1.350	1.496	1.500		1.032	1.258	1.431	1.518	1.499
h = 48	1.029	1.190	1.406	1.674	1.717		1.032	1.276	1.557	1.736	1.734
h = 96	1.030	1.190	1.734	6.763	8.666		1.032	1.385	3.969	9.865	11.801
h = 120	1.030	1.191	2.891	76.00	85.39		1.032	2.006	27.85	128.8	141.5
R = 200											
h = 1	1.010	1.011	1.013	1.015	1.016		1.014	1.017	1.018	1.021	1.024
h = 12	1.016	1.070	1.090	1.118	1.135		1.014	1.079	1.102	1.128	1.142
h = 24	1.015	1.090	1.131	1.182	1.216		1.014	1.107	1.160	1.205	1.225
h = 36	1.016	1.096	1.156	1.235	1.283		1.014	1.121	1.202	1.268	1.294
h = 48	1.016	1.098	1.171	1.283	1.348		1.014	1.128	1.234	1.327	1.363
h = 96	1.016	1.098	1.191	1.426	1.680		1.014	1.130	1.298	1.566	1.748
h = 120	1.015	1.096	1.190	1.481	1.934		1.014	1.129	1.311	1.720	2.084

Note: Figures greater than unity indicate that OLS-based forecasts inflate the MSPE of correct specified models.

Source: Authors' elaboration.

estimates may be much more serious when forecasting persistent series at long horizons than at short horizons. Under these circumstances, a parsimonious RW-based forecast may be a much more accurate strategy to use in the long run.[6]

The results in *Table 2* reinforce the arguments given above. In this table, we show the ratio between the MSPE of the optimal forecasts constructed with estimated parameters and the MSPE of the optimal forecasts constructed with the true parameters. All figures in *Table 2* are above unity, indicating that estimation noise inflates the MSPE. As expected, *Table 2* shows the smaller the sample size, the higher

[5] Clements and Hendry (2001) show a similar result using a somewhat different environment.

the ratios. Similarly, there is a tendency to observe higher ratios when the persistence of the processes is higher, although there are some departures from this tendency when the sample size is either small or moderate. Interestingly, there is an increasing pattern in the ratios across the diagonals on each block in *Table 2*. In other words, higher ratios are achieved when both the forecasting horizon and the persistence of the processes grow. This pattern is sharper when the sample size is low.

The results in *Tables 1* and *2* are important. They provide evidence in favor of using unit-root-based forecasts to predict stationary variables when the parameters of the correctly specified models are not properly estimated due, for instance, to data restrictions. Interestingly, this recommendation may also be convenient if long-run forecasts are needed.

Tables 1 and *2* were constructed using rolling OLS. This means that in each replication a sequence of h-step-ahead forecasts is built estimating the parameters of the models using rolling windows of fixed size R. A common alternative approach uses windows of expanding size and considers more information as time goes by. For stationary processes, this strategy provides more accurate forecasts. We have also used an expanding-window approach to build tables similar to *Tables 1* and *2*.[7] While we do not report these tables in order to save space, there are two main features to point out from these expanding-window exercises:

1. Reinforcing our argument, persistent stationary processes are better predicted using driftless unit-root-based forecasts in small samples.
2. The level of persistence required for the previous result to hold true is higher than in the exercise with rolling windows.

The intuition is simple: when we compare the exercise using rolling windows of fixed size R with the exercise using expanding windows in which the sample size of the first window is R, then we are comparing exercises in which the noise in the estimation of the parameters is totally different, being much lower in the expanding-window approach. As estimation noise offers an opportunity for forecasts from misspecified models to outperform forecasts generated using the true DGP, it is easier to beat the forecasts from the true model in strategies with higher parameter uncertainty, such as the rolling-window strategy.

Table 2 shows the pervasive effect of estimating the parameters of a persistent stationary process in small samples using rolling OLS. This effect is even more noticeable when forecasting at longer horizons. It is natural to look for the existence of other estimators with better predictive performance than OLS. After all, there is a vast literature reporting the good predictive behavior of models estimated either with shrinkage estimators or with methods aimed at reducing the small sample bias of OLS.

One such method is the median-unbiased estimator proposed by Andrews (1993) for the AR(1) process. By means of simulations, we also explore the out-of-sample predictive behavior of driftless unit-root-based forecasts against those coming from the true AR(1) model in which parameters are estimated using the median-unbiased estimator proposed by Andrews (see Medel and Pincheira, 2016). As expected, the performance of forecasts constructed with the median-unbiased esti-

[7] These tables are available upon request.

mator is generally much better than that of OLS-based forecasts. This is especially noticeable at longer horizons. In fact, the ratio of MSPE between median-unbiased forecasts and OLS-based forecasts is always lower than or equal to one, indicating that median-unbiased-based forecasts are never outperformed by OLS-based forecasts. We also constructed the ratio of MSPE between driftless RW-based forecasts and median-unbiased-based forecasts when the true DGP is an AR(1) model. Despite the relatively good performance of median-unbiased-based forecasts, they are outperformed at short horizons by driftless RW-based forecasts when persistence is high ($\rho \geq 0.975$) and samples are either small or moderate. Reductions of up to 10% in MSPE are obtained using the driftless RW. These reductions are generally much smaller than those in *Table 1*. In another related article, Kim and Durmaz (2012) show via simulations that some other bias-correction methods can outperform OLS-based forecasts in out-of-sample exercises when the true DGP is persistent but still stationary. A thorough comparison between unit-root-based forecasts and forecasts generated using the vast range of alternative estimators is beyond the scope of this paper, but it certainly should be the subject of further research.

Tables 1 and *2* are useful for showing the relationship between parameter uncertainty, estimation sample size and forecasting horizon in the case of the simple AR(1) and AR(2) models. In the next subsection we explore whether the same general pattern found for the simple AR(1) and AR(2) models holds true when working with more general and realistic processes.

3.2 AR(12) and SARIMA Models

We estimate four simple models for monthly YoY CPI inflation, which is a persistent time series. We focus on the sample period October 1990–December 2011 for Canada, Sweden, Switzerland, the UK and the US. We consider the following four models:

1. $AR(1): \pi_t^{12} = \alpha + \rho_1 \pi_{t-1}^{12} + \varepsilon_t$
2. $AR(2): \pi_t^{12} = \alpha + \sum_{i=1}^{2} \phi_i \pi_{t-i}^{12} + \varepsilon_t$
3. $AR(12): \pi_t^{12} = \alpha + \sum_{i=1}^{12} \phi_i \pi_{t-i}^{12} + \varepsilon_t$
4. $SARIMA: \pi_t^{12} = \alpha + \rho \pi_{t-1}^{12} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_{12} \varepsilon_{t-12} + \theta_1 \theta_{12} \varepsilon_{t-13}$

We have used the abbreviation SARIMA to denote the Seasonal ARIMA model with the following parameters $(p,d,q) \times (P,D,Q) = (1,0,1) \times (0,0,1)$

Variations of this model have received a lot of attention in the forecasting literature. In particular, Ghysels, Osborn and Rodrigues (2006) point out the good forecasting performance of such models when applied to seasonal time series.

As mentioned before, our models are estimated using YoY CPI inflation rates, defined as:

$$\pi_t^{12} = \left(\frac{CPI_t}{CPI_{t-12}} \right) * 100 - 100$$

Table 3 In-Sample Diagnostic Statistics for Inflation Models

October 1990–December 2011					
	Canada	**Sweden**	**Switzerland**	**UK**	**US**
			AIC		
AR(1)	1.305	1.625	0.559	0.591	1.106
AR(2)	1.289	1.630	0.562	0.588	0.924
AR(12)	1.201	1.405	0.508	0.625	0.912
SARIMA	0.724	0.986	0.285	0.331	0.156
			BIC		
AR(1)	1.333	1.653	0.581	0.619	1.134
AR(2)	1.331	1.672	0.604	0.630	0.965
AR(12)	1.385	1.589	0.692	0.809	1.095
SARIMA	0.780	1.042	0.341	0.387	0.211
			Adjusted R^2		
AR(1)	0.953	0.961	0.883	0.945	0.866
AR(2)	0.952	0.959	0.900	0.942	0.867
AR(12)	0.931	0.940	0.889	0.914	0.810
SARIMA	0.966	0.971	0.954	0.974	0.925
			Durbin-Watson statistic		
AR(1)	1.932	1.820	1.229	1.773	1.679
AR(2)	2.019	2.027	1.903	1.996	1.983
AR(12)	1.966	1.969	2.053	2.212	1.942
SARIMA	1.994	1.981	1.910	1.978	1.994

Source: Authors' elaboration.

In *Table 3* we show the Akaike information criterion (AIC) and Bayesian information criterion (BIC), the adjusted R^2 and the Durbin-Watson statistic for all the models and countries in our exercise.

According to *Table 3*, if a researcher decided to pick the best model using either the AIC, BIC or the adjusted R^2 statistic, he or she would pick the SARIMA model for all five countries in our sample. The second best model depends both on the country and the specific information criterion. For instance, the second best model for Sweden is the AR(12), for the US it is the AR(12) when AIC is used, but the AR(2) when BIC is considered instead. All in all, we detect a slight edge in favor of the AR(12) model as the second best model. Accordingly, it seems reasonable to explore via simulations the forecasting behavior of driftless unit-root-based forecasts when the DGP corresponds to either a SARIMA or an AR(12) model.

Differing from the results in *Table 1*, where we use the driftless RW model to generate unit-root-based forecasts, we now compare the predictive behavior of the true models with estimated parameters against forecasts coming from the very same models in which two restrictions are imposed: the drift is set to zero and a unit root is imposed on the autoregressive component of the model. We do this to generate a fair comparison between the forecasts: the only difference between them will come from the drift and the imposition of a unit root. Everything else is kept the same.

3.2.1 SARIMA Model

As the first step, we estimate a SARIMA model for the US with the following results:[8]

$$\pi_t^{US} = \hat{\alpha} + \hat{\rho}\pi_{t-1}^{US} + \varepsilon_t - \widehat{\theta_1}\varepsilon_{t-1} - \widehat{\theta_{12}}\varepsilon_{t-12} + \widehat{\theta_1}\widehat{\theta_{12}}\varepsilon_{t-13} \qquad (3)$$

where

$$\hat{\alpha} = 0.16; \hat{\rho} = 0.94; \widehat{\theta_1} = 0.47; \widehat{\theta_{12}} = 0.92; \widehat{\sigma_\varepsilon} = 0.27$$

We notice also that during our sample period US inflation displays an average of $\bar{\pi}^{US} = 2.44$

For the design of the Monte Carlo experiments, we generate 5,000 replications of the following SARIMA specification:

$$y_t = \bar{\pi}^{US}(1-\rho) + \rho y_{t-1} + u_t - \widehat{\theta_1}u_{t-1} - \widehat{\theta_{12}}u_{t-12} + \widehat{\theta_1}\widehat{\theta_{12}}u_{t-13} \qquad (4)$$

$u_t \sim N\left(0, \hat{\sigma}_\varepsilon^2\right)$.

We notice that the only free parameter in (4) is the persistence parameter ρ. Consequently, as this parameter changes so does the drift:

$$Drift = \bar{\pi}^{US}(1-\rho)$$

We use this strategy to generate processes with the same average of US inflation. We consider the following values for ρ:

$$\rho \in \{0.50; 0.90; 0.95; 0.975; 0.99\}$$

We then estimate each of the processes and generate 500 h-step-ahead forecasts using a rolling non-linear LS method. We consider $h \in \{1;12;24;36\}$ and three different rolling windows of size $R \in \{50;100;200\}$. In *Table 4* we report the following ratio:

$$MSPE\,Ratio = \frac{MSPE^{UR}(h)}{MSPE^{SARIMA}(h)}$$

where $MSPE^{SARIMA}(h)$ stands for the out-of-sample MSPE of the SARIMA model in (4) and $MSPE^{UR}(h)$ for the out-of-sample MSPE when forecasting the same process but assuming that ρ is equal to one. In other words, we generate the forecasts estimating the following misspecified model:

$$y_t = y_{t-1} + \varepsilon_t - \gamma_1\varepsilon_{t-1} - \gamma_{12}\varepsilon_{t-12} + \gamma_1\gamma_{12}\varepsilon_{t-13} \qquad (5)$$

From *Table 4* we see that the three main features that we point out for the AR(1) and AR(2) processes hold true. First, as the sample size gets larger, all the ratios become larger as well. Second, as the persistence of the process increases, all the ratios show a tendency to decrease (with only a few exceptions). Finally, we see that given a sample size R, there is a persistence threshold beyond which the MSPE ratios decrease with the forecasting horizon. For instance, for $R = 100$ when ρ is greater or equal to 0.95, we get this decreasing pattern. In other words, we

[8] We picked the US inflation process simply because of the size and relevance of the US economy.

Table 4 Ratio of MSPE between an Unrestricted SARIMA Specification and a Restricted One Imposing a Unit Root and No Drift

ρ:	**0.50**	**0.90**	**0.95**	**0.98**	**0.99**
			R = 50		
h = 1	1.131	0.898	0.870	0.860	0.860
h = 12	1.692	0.746	0.603	0.560	0.554
h = 24	0.967	0.338	0.166	0.126	0.086
h = 36	0.074	0.005	0.000	0.000	0.000
			R = 100		
h = 1	1.223	0.994	0.965	0.953	0.947
h = 12	2.021	1.153	0.906	0.810	0.776
h = 24	1.039	1.042	0.940	0.835	0.783
h = 36	1.038	1.020	0.924	0.786	0.663
			R = 200		
h = 1	1.257	1.030	1.003	0.988	0.980
h = 12	2.102	1.371	1.093	0.950	0.879
h = 24	1.014	1.057	1.053	0.989	0.923
h = 36	1.014	1.021	1.025	0.991	0.930

Note: A figure below unity favors driftless unit-root-based forecasts.
Source: Authors' elaboration.

can also see that in this environment, forecasts generated by a driftless unit-root process that is nested in the true DGP can have greater accuracy in small and moderate samples when the persistence of the process is high.

3.2.2 AR(12) Model

First we estimate an AR(12) model for the US with the following results:

$$\pi_t^{US} = \hat{\delta} + \widehat{\phi_1}\pi_{t-1}^{US} + \widehat{\phi_2}\pi_{t-2}^{US} + \ldots + \widehat{\phi_{12}}\pi_{t-12}^{US} + w_t \qquad (6)$$

where:

$$\hat{\delta} = 0.34; \widehat{\phi_1} = 1.39; \widehat{\phi_2} = -0.66; \widehat{\phi_3} = 0.20; \widehat{\phi_4} = 0.10 \qquad (7)$$

$$\widehat{\phi_5} = -0.28; \widehat{\phi_6} = 0.25; \widehat{\phi_7} = -0.06; \widehat{\phi_8} = -0.09; \widehat{\phi_9} = 0.08$$

$$\widehat{\phi_{10}} = 0.06; \widehat{\phi_{11}} = -0.13; \widehat{\phi_{12}} = 0.00; \widehat{\sigma_w} = 0.63$$

To explain the design of the Monte Carlo experiment, it is useful to recall that the $AR(p)$ model

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t$$

can be alternatively represented as

$$Y_t = \delta + \rho Y_{t-1} + \xi_1 \Delta Y_{t-1} + \xi_2 \Delta Y_{t-2} + \ldots + \xi_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

where

$$\rho \equiv \sum_{i=1}^{p} \phi_i$$

$$\zeta_j \equiv -\left[\phi_{j+1} + \phi_{j+2} + \ldots + \phi_p\right]; j = 1, \ldots, p-1$$

(see Hamilton, 1994). In a stationary $AR(p)$ we have

$$\rho \equiv \sum_{i=1}^{p} \phi_i < 1$$

In the particular case in which $\rho = 1$ the process will have a unit-root and could be written as

$$\Delta Y_t = \delta + \xi_1 \Delta Y_{t-1} + \xi_2 \Delta Y_{t-2} + \xi_3 \Delta Y_{t-3} + \ldots + \xi_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

For our simulations, we generate 1,000 replications of the following $AR(12)$ specification:

$$\begin{gathered} y_t = \bar{\pi}^{US}(1-\rho) + \phi_1 y_{t-1} + \sum_{i=2}^{12} \widehat{\phi}_i y_{t-i} + v_t \\ v_t \sim N\left(0, \hat{\sigma}_v^2\right) \\ \phi_1 = \rho - \sum_{i=2}^{12} \widehat{\phi}_i \end{gathered} \tag{8}$$

We notice that the only free parameter in (8) is the persistence parameter ρ Consequently, changes in ρ will also affect the drift:

$$Drift = \bar{\pi}^{US}(1-\rho)$$

and the ϕ_1 coefficient:

$$\phi_1 = \rho - \sum_{i=2}^{12} \widehat{\phi}_i$$

We use this strategy to generate processes with the same average of US inflation. We consider the following values for ρ:

$$\rho \in \{0.50; 0.90; 0.95; 0.975; 0.99\}$$

We then estimate each of the processes and generate 500 h-step-ahead forecasts using a rolling OLS method. We consider $h \in \{1;12;24;36\}$ and three different rolling windows of size $R \in \{50;100;200\}$. In *Table 5* we report the following ratio:

$$MSPE\,Ratio = \frac{MSPE^{UR}(h)}{MSPE^{AR12}(h)}$$

where $MSPE^{AR12}(h)$ stands for the out-of-sample MSPE of the $AR(12)$ model in (8) and $MSPE^{UR}(h)$ for the same process in (8) but assuming that ρ is equal to 1. In other words, we generate these forecasts estimating the following misspecified model:

$$\Delta Y_t = \xi_1 \Delta Y_{t-1} + \xi_2 \Delta Y_{t-2} + \xi_3 \Delta Y_{t-3} + \ldots + \xi_{11} \Delta Y_{t-11} + \varepsilon_t \tag{9}$$

Table 5 Ratio of MSPE between an Unrestricted AR(12) Specification and a Restricted One Imposing a Unit Root and No Drift

ρ:	**0.50**	**0.90**	**0.95**	**0.98**	**0.99**
			R = 50		
h = 1	1.035	0.996	0.963	0.931	0.914
h = 12	1.279	1.241	0.906	0.680	0.573
h = 24	1.080	1.466	0.827	0.540	0.386
h = 36	0.594	1.089	0.026	0.149	0.113
			R = 100		
h = 1	1.070	1.053	1.035	1.013	0.979
h = 12	1.374	1.790	1.495	1.204	0.880
h = 24	1.245	2.608	2.386	1.564	0.875
h = 36	1.275	1.925	3.016	2.203	0.932
			R = 200		
h = 1	1.084	1.070	1.056	1.038	1.015
h = 12	1.406	2.031	1.756	1.471	1.178
h = 24	1.254	2.888	2.940	2.107	1.377
h = 36	1.277	2.091	3.633	3.107	1.680

Note: A figure below unity favors driftless unit-root-based forecasts.
Source: Authors' elaboration.

From *Table 5,* we see that the three main features that we pointed out previously also hold true with minor differences. First, as the sample size gets larger, all the ratios become larger as well. Second, as the persistence of the process increases, all the ratios show a tendency to decrease (with only a few exceptions). Third, we see that given a sample size of $R \leq 100$, there is a persistence threshold beyond which the out-of-sample MSPE ratios decrease with the forecasting horizon. In other words, we see that in this environment driftless unit-root-based forecasts can display higher accuracy than forecasts coming from the correct model when samples are small and the process is persistent.

Despite these similarities, *Tables 4* and *5* indicate that the use of driftless unit-root-based forecasts is more advantageous for our SARIMA specifications than for our AR(12) models. This is so because percentage reductions in RMSPE, due to the imposition of a unit root, are higher in SARIMA relative to AR(12) models.

Table 6 is also interesting, as it reports the share of the MSPE displayed in *Tables 4* and *5* that is explained by the variance of the forecasts errors.[9] *Table 6* reports this share for driftless unit-root-based forecasts. As we can see, most of the MSPE in our previous simulations is due to variance and only a minor part of it corresponds to bias. These figures are consistent with our theoretical results indicating that driftless unit-root-based forecasts are unbiased.

Before moving to the empirical section, it is important to mention that *Tables 1, 4* and *5* show a linkage between sample size, persistence, forecasting horizon and the relative accuracy of driftless unit-root-based forecasts. These tables suggest that given a DGP, a sample size *R* and a forecasting horizon *h*, there is an invertible

[9] Let us recall the identity: $MSPE = Variance + Bias^2$.

Table 6 Share of the MSPE Explained by the Variance of Forecast Errors of Driftless Unit-Root-Based Forecasts

	SARIMA					AR(12)				
ρ	**0.50**	**0.90**	**0.95**	**0.98**	**0.99**	**0.50**	**0.90**	**0.95**	**0.98**	**0.99**
					R = 50					
h = 1	0.9999	0.9997	0.9997	0.9996	0.9995	0.9996	0.9993	0.9991	0.9988	0.9986
h = 12	0.9996	0.9988	0.9984	0.9980	0.9978	0.9990	0.9938	0.9921	0.9910	0.9902
h = 24	0.9998	0.9982	0.9973	0.9967	0.9962	0.9989	0.9930	0.9879	0.9856	0.9851
h = 36	0.9997	0.9978	0.9968	0.9961	0.9956	0.9985	0.9932	0.9898	0.9845	0.9833
					R = 100					
h = 1	0.9999	0.9998	0.9997	0.9995	0.9994	0.9996	0.9995	0.9994	0.9992	0.9988
h = 12	0.9994	0.9986	0.9980	0.9971	0.9964	0.9990	0.9952	0.9944	0.9930	0.9906
h = 24	0.9998	0.9991	0.9983	0.9970	0.9959	0.9991	0.9947	0.9908	0.9874	0.9843
h = 36	0.9998	0.9991	0.9981	0.9965	0.9949	0.9989	0.9958	0.9919	0.9835	0.9787
					R = 200					
h = 1	0.9998	0.9997	0.9996	0.9994	0.9991	0.9994	0.9994	0.9994	0.9992	0.9991
h = 12	0.9990	0.9974	0.9965	0.9952	0.9928	0.9982	0.9933	0.9933	0.9924	0.9911
h = 24	0.9998	0.9991	0.9981	0.9964	0.9936	0.9985	0.9919	0.9876	0.9861	0.9838
h = 36	0.9998	0.9992	0.9983	0.9966	0.9935	0.9982	0.9944	0.9880	0.9808	0.9765

Source: Authors' elaboration.

function between the persistence of the process ρ and the relative performance of driftless unit-root-based forecasts against forecasts coming from the true DGP with estimated parameters. It could be interesting to exploit this relationship to develop a decision rule aimed at determining which type of forecast we should use: one based upon a stationary process or one based upon a process with a unit root. In *Appendix D* (on the website of this journal), we shed some light on how to proceed in this respect. We focus on the empirical illustration of our findings next.

4. Empirical Evidence

In this section, we illustrate, from the point of view of a practitioner, the benefits of driftless unit-root-based forecasts in which different models are used to generate inflation forecasts.[10] We use this kind of macroeconomic data for two reasons. On the one hand, year-on-year inflation rates in many countries exhibit near-unity behavior. On the other hand, it is important to forecast inflation, especially in the current conditions where a number of developed countries face the threat of deflation. We first describe the dataset and then the models. Finally, we evaluate the relative accuracy between forecasts using out-of-sample MSPE pairwise comparisons and a superior predictive ability test based on Giacomini and White (2006).

4.1. Data

We use monthly CPI inflation data for Canada, Sweden, Switzerland, the UK and the US. The sources of the dataset are country-specific central banks and the Federal Reserve Bank of St. Louis. The inflation data comprises the same data that we used in the construction of *Table 3*. Our models work with YoY CPI inflation rates, defined as[11]

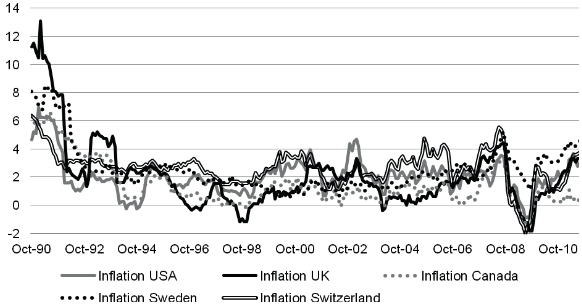
$$\pi_t^{12} = \left(\frac{CPI_t}{CPI_{t-12}} \right) * 100 - 100$$

Table 7 shows traditional unit-root tests for the five inflation series for the sample period from October 1990 to December 2011. At the 10% significance level, these traditional tests reject the presence of a unit root for all the series in our sample. *Figure 1* shows all the series in their stationary transformation. Some descriptive statistics for different subsamples are available in *Appendix C* (on the website of this journal).

[10] We also explored the behavior of unit-root-based forecasts with unemployment series for the US, Canada and the euro zone. The results are qualitatively similar to those obtained with inflation and are available upon request.

[11] We have a very simple justification for the choice of year-on-year inflation as a target variable: to our knowledge, most of the inflation-targeting countries in the world define their targets in year-on-year terms. For instance, the Czech Republic has a target of 2% for the medium term. The UK 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 have a target of 2.5%, such as Iceland, Norway, Poland, Romania and North Korea. The list is long but all of these countries express their targets in year-on-year terms. Coming from an inflation-targeting country, it is absolutely natural for us to focus on the year-on-year inflation rate. Furthermore, the choice of year-on-year inflation is not uncommon in the literature. For instance, the works of Ciccarelli and Mojon (2010), Clark (2001), Atkeson and Ohanian (2001), Pincheira and Medel (2015) and Pincheira and West (2016) consider year-on-year inflation as the predicted variable or as an independent variable in a forecasting equation.

Figure 1 YoY CPI inflation Rate for Selected Economies



Source: Country-specific central banks and the Federal Reserve Bank of St. Louis.

Table 7 Unit-Root Test for YoY CPI Inflation

	Levels		First differences	
	ADF	**PP**	**ADF**	**PP**
Canada	-4.149	-3.582	-7.994	-14.003
(*p*-value) {lag} [exog.]	(0.001) {12} [*C*]	(0.007) {3} [*C*]	(0.000) {11} [*N*]	(0.000) {1} [*N*]
Sweden	-4.950	-3.952	-7.908	-14.317
(*p*-value) {lag} [exog.]	(0.000) {12} [*C*]	(0.002) {5} [*C*]	(0.000) {11} [*N*]	(0.000) {5} [*N*]
Switzerland	-2.704	-2.819	-6.536	-15.442
(*p*-value) {lag} [exog.]	(0.075) {0} [*C*]	(0.057) {6} [*C*]	(0.000) {11} [*N*]	(0.000) {6} [*N*]
United Kingdom	-3.159	-3.169	-14.50	-14.494
(*p*-value) {lag} [exog.]	(0.024) {0} [*C*]	(0.023) {3} [*C*]	(0.000) {0} [*N*]	(0.000) {1} [*N*]
United States	-3.107	-4.032	-9.399	-10.452
(*p*-value) {lag} [exog.]	(0.027) {12} [*C*]	(0.002) {4} [*C*]	(0.000) {11} [*N*]	(0.000) {3} [*N*]

Notes: Null Hypothesis: The series has a unit root. *p*-value in (.). Regression lag order or bandwidth in {.}. Lag length chosen according to BIC; maximum lag length = 18. Bandwidth chosen according to Newey-West (1994) using the Bartlett kernel. Exogenous regressors: *C* = Constant, *N* = None. Inflation series sample: October 1990–December 2011.

Source: Authors' elaboration.

We are aware that the results in *Table 7* provide no mathematical proof of stationarity in the series. These results come from statistical tests that suffer from both type-I and type-II errors. In particular, unit-root tests display low power against persistent stationary alternatives, as mentioned by Andrews and Chen (1994) and Cochrane (1991). Low power means that it is hard for the tests to reject the null hypothesis of a unit root.

The fact that the tests reject the null hypothesis in spite of this low power is somewhat reassuring. Nevertheless, we prefer to work with the assumption of stationarity rather than with certainty about it. Our tests in *Table 7* provide support to this assumption for all the series. Bearing this in mind, we will use the following models to generate out-of-sample forecasts for these five series:

1. $AR(12): \pi_t^{12} = \delta + \sum_{i=1}^{12} \phi_i \pi_{t-i}^{12} + \varepsilon_t$

2. $AR(12) - UR: \Delta\pi_t^{12} = \xi_1 \Delta\pi_{t-1}^{12} + \xi_2 \Delta\pi_{t-2}^{12} + \ldots + \xi_{11} \Delta\pi_{t-11}^{12} + \varepsilon_t$

with: $\zeta_j \equiv -\left[\phi_{j+1} + \phi_{j+2} + \ldots + \phi_{12}\right]; j = 1, \ldots, 11$

3. $SARIMA: \pi_t^{12} = \alpha + \rho\pi_{t-1}^{12} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_{12}\varepsilon_{t-12} + \theta_1\theta_{12}\varepsilon_{t-13}$

4. $SARIMA - UR: \pi_t^{12} = \pi_{t-1}^{12} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_{12}\varepsilon_{t-12} + \theta_1\theta_{12}\varepsilon_{t-13}$

The first model is a simple AR(12); the second model is the same AR(12) without drift and with the restriction that all the autoregressive coefficients must sum to one. Model 3 is the SARIMA model introduced in the section on simulations and Model 4 is the same SARIMA but with no drift and the restriction of a unit root. This model is labeled as the *airline model* by Box and Jenkins (1970).[12] Differing from the direct autoregressions of Stock and Watson, our models have π_t^{12} both on the right-hand side and on the left-hand side.

We estimate the models with a fixed-size rolling window of length *R*. We consider two values for *R*: 30 and 100 observations. The first estimation sample with $R = 30$ covers the period from August 1996 to January 1999, whereas with $R = 100$ it covers the period from October 1990 to January 1999. The remaining sample is used for evaluation of the forecasts, covering the period from February 1999 to December 2011.

4.2 Forecast Evaluation

We focus on comparing the predictive performance of unit-root-based forecasts versus forecasts coming from models in which no unit roots are imposed a *priori*. In *Table 8* we report estimates of the sample RMSPE for all the forecasts under consideration. This sample estimate is calculated as follows:

$$\widehat{RMSPE_h} = \left[\frac{1}{P(h)} \sum_{t=R}^{R+P(h)-1} \left(y_{t+h} - y_t^f(h)\right)^2\right]^{\frac{1}{2}}$$

where $y_t^f(h)$ is the *h*-step-ahead forecast of the generic variable y_t, and $P(h)$ represents the total number of out-of-sample forecast errors available for a given methodology and forecasting horizon.

The superior performance of driftless unit-root-based forecasts reported in *Table 8* is outstanding. The lowest MSPE is almost always reached by one of the models restricted to having a unit root—either AR(12)-UR or SARIMA-UR. Actually, there are only three exceptions, all of them occurring when rolling windows of size 100 are used: the case of Sweden when forecasting inflation 24 months ahead and the cases of US inflation when forecasting two and three years ahead. In all the remaining cases, driftless unit-root-based forecasts take the lead.

[12] As Ghysels, Osborn and Rodrigues (2006) point out, this specification has proven to be very useful for forecasting monthly time series with seasonal patterns.

Table 8 Multi-Horizon RMSPE of Forecasts for Selected Economies

	Rolling-window size: 30				Rolling-window size: 100			
	***h* = 1**	***h* = 12**	***h* = 24**	***h* = 36**	***h* = 1**	***h* = 12**	***h* = 24**	***h* = 36**
				Inflation: Canada				
1. AR(12)	0.576	1.717	2.128	2.806	0.518	1.307	1.271	1.047
2. AR(12)-UR	0.571	1.747	1.686	1.617	0.533	1.517	1.384	1.126
3. SARIMA	0.465	1.576	1.778	1.564	0.375	1.221	1.102	1.110
4. SARIMA-UR	0.426	1.260	1.217	1.008	0.368	1.114	1.075	0.957
Lowest RMSPE	0.426	1.260	1.217	1.008	0.368	1.114	1.075	0.957
				Inflation: Sweden				
1. AR(12)	0.581	4.171	6.669	31.150	0.430	1.511	1.342	1.338
2. AR(12)-UR	0.520	2.472	3.608	5.392	0.449	2.003	2.449	1.977
3. SARIMA	0.410	3.098	31.420	435.400	0.337	1.632	1.620	1.591
4. SARIMA-UR	0.353	1.549	1.710	1.539	0.322	1.299	1.351	1.319
Lowest RMSPE	0.353	1.549	1.710	1.539	0.322	1.299	1.342	1.319
				Inflation: Switzerland				
1. AR(12)	0.405	1.538	3.129	6.059	0.344	1.001	1.198	1.045
2. AR(12)-UR	0.396	1.743	2.722	5.753	0.356	1.531	1.319	1.406
3. SARIMA	0.349	1.670	3.613	16.637	0.317	0.964	0.945	0.961
4. SARIMA-UR	0.307	1.038	0.896	0.970	0.304	0.929	0.885	0.951
Lowest RMSPE	0.307	1.038	0.896	0.970	0.304	0.929	0.885	0.951
				Inflation: United Kingdom				
1. AR(12)	0.397	1.741	6.546	32.223	0.329	1.337	1.352	1.365
2. AR(12)-UR	0.372	1.233	1.455	1.861	0.321	1.123	1.062	1.018
3. SARIMA	0.288	1.866	6.892	38.640	0.270	1.240	1.383	1.858
4. SARIMA-UR	0.270	0.952	1.022	1.118	0.262	1.139	1.215	1.320
Lowest RMSPE	0.270	0.952	1.022	1.118	0.262	1.123	1.062	1.018
				Inflation: United States				
1. AR(12)	0.574	3.535	7.159	3.099	0.483	2.040	2.259	1.646
2. AR(12)-UR	0.535	2.937	3.707	5.357	0.496	2.675	2.971	2.735
3. SARIMA	0.464	2.189	2.749	4.726	0.335	1.569	1.439	1.455
4. SARIMA-UR	0.403	1.772	1.674	1.604	0.331	1.501	1.474	1.499
Lowest RMSPE	0.403	1.772	1.674	1.604	0.331	1.501	1.439	1.455

Source: Authors' elaboration.

We can also look at pairwise comparisons in *Table 8*. In other words, we could compare the performance of the unrestricted AR(12) model with its driftless unit-root version. We can engage in the same comparison using the SARIMA models. *Table 8* shows that when using rolling windows of 30 observations, the driftless unit-root versions of both the AR(12) and SARIMA models outperform their unrestricted versions in 92.5% of the pairwise comparisons. The exceptions are Canada and Switzerland when forecasting one year ahead and the US when forecasting three years ahead.

Table 9 One-Sided *p*-Values of the Giacomini and White (2006) Test of Superior Predictive Ability of Driftless Unit-Root-Based Forecasts

	Rolling-window size: 30				Rolling-window size: 100			
	***h* = 1**	***h* = 12**	***h* = 24**	***h* = 36**	***h* = 1**	***h* = 12**	***h* = 24**	***h* = 36**
				Inflation: Canada				
1. AR(12)	0.408	0.564	0.057*	0.107	0.844	0.941	0.865	0.776
2. SARIMA	0.003***	0.021**	0.047**	0.011**	0.144	0.125	0.280	0.005***
				Inflation: Sweden				
1. AR(12)	0.035***	0.045**	0.010**	0.129	0.977	0.991	0.999	1.000
2. SARIMA	0.005***	0.058*	0.149	0.153	0.045**	0.002***	0.051**	0.149
				Inflation: Switzerland				
1. AR(12)	0.352	0.719	0.346	0.445	0.719	0.955	0.685	0.989
2. SARIMA	0.002***	0.008***	0.110	0.151	0.059*	0.322	0.212	0.452
				Inflation: United Kingdom				
1. AR(12)	0.048**	0.067*	0.053*	0.095*	0.042*	0.002***	0.009***	0.027***
2. SARIMA	0.063*	0.088*	0.069*	0.088*	0.165	0.313	0.277	0.122
				Inflation: United States				
1. AR(12)	0.255	0.235	0.097*	0.852	0.684	0.944	0.894	0.984
2. SARIMA	0.000***	0.002***	0.035**	0.133	0.174	0.258	0.831	0.964

Notes: Figures below X/100 indicate rejection at the X% significance level of the null hypothesis of superior predictive ability models of forecasts coming from the correctly specified models in favor of driftless unit-root-based forecasts. (*) *p*-value<10%, (**) *p*-value<5%, (***) *p*-value<1%.

Source: Authors' elaboration.

When using rolling windows of 100 observations, *Table 8* shows mixed results. Actually, in 55% of the pairwise comparisons, driftless unit root-based-forecasts outperform forecasts coming from unrestricted models. If we assume that our series are stationary, then our empirical findings are consistent with our simulation results: driftless unit-root-based forecasts perform well, especially when using rolling windows of small size.[13]

Table 9 shows the *p*-values of the Giacomini and White (2006) test of superior predictive ability between forecasts coming from the unrestricted versions of the models and their driftless unit-root versions. The null hypothesis is that of superior predictive ability of forecasts coming from the unrestricted versions, while the alternative is that our driftless unit-root-based forecasts perform better. Therefore, we carry out a one-sided test. As usual, low *p*-values are associated with the rejection of the null hypothesis in favor of the alternative. We see that the null hypothesis is rejected at the 10% significance level in favor of driftless unit-root-based forecasts on a number of occasions. This happens both at short and long horizons.

[13] In an additional exercise, which is not reported for the sake of brevity, we compare the predictive performance of an AR(1) and an RW model in monthly unemployment series for Canada, the euro zone and the US. These results are fairly consistent with those shown for YoY inflation in two main aspects: First, in most cases the driftless RW model outperforms the AR(1). Second, the driftless RW model fares better relative to the AR(1) in rolling windows of size $R = 30$, i.e. when the AR(1) is estimated in short samples.

5. Conclusion

The use of different time-series models to generate forecasts is fairly usual in the fields of macroeconomics and financial economics. When the target variable is stationary, the use of processes with unit-roots may seem counterintuitive. Nevertheless, in this paper we demonstrate that forecasting a stationary variable with forecasts based on driftless unit-root processes generates bounded mean squared prediction errors at every single horizon. We also show that these forecasts are unbiased. In addition, we show via simulations that persistent stationary processes may be better predicted by driftless unit-root-based forecasts than by forecasts coming from a model that is correctly specified but is subject to a higher degree of parameter uncertainty. In addition, we provide an empirical illustration of our findings in the context of inflation forecasts for five industrialized economies.

Our simulations also provide evidence indicating that the benefits of using unit-root-based forecasts are not confined within the boundaries of short horizons. In fact, these benefits may be sizable at long horizons as well.

It is important to point out that in our simulations we have assumed that we know the true data generating process, which is a little unrealistic and unfair to the performance of some of the driftless unit-root-based forecasts. A natural extension of this paper could be an exploration of a more realistic environment in which the data is known to be stationary but there is uncertainty about the true parametric form of the data generating process. In this scenario, both unit-root-based and stationary forecasts will probably come from misspecified models and therefore a fairer comparison can be made. Moreover, one could explore whether forecasts from persistent stationary models can be better approximated by forecasts coming from the set of stationary models or from the set of models with the presence of a unit root. More generally, future research might explore if the results we have reached here may still hold true in broader environments. In particular, it would be very interesting to evaluate the local relative performance between unit-root-based forecasts and their stationary counterparts using the fluctuation test of Giacomini and Rossi (2010).

REFERENCES

Andrews D (1993): Exactly Median-Unbiased Estimation of First Order Autoregressive / Unit Root Models. *Econometrica*, 61(1):139–165.

Andrews D, Chen H-H (1994): Approximately Median-Unbiased Estimation of Autoregressive Models. *Journal of Business & Economic Statistics*, 12(2):187–204.

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*. Holden-Day, San Francisco.

Box GEP, Jenkins GM, Reinsel G (2008): *Time Series Analysis: Forecasting and Control*. Fourth edition, Wiley.

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

Ciccarelli M, Mojon B (2010): Global Inflation. *Review of Economics and Statistics*, 92(3):524–535.

Clark T (2001): Comparing Measures of Core Inflation. *Federal Reserve Bank of Kansas City Economic Review*, (Second Quarter):5–31.

Clark T, West KD (2006): Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis. *Journal of Econometrics*, 135:155–186.

Clark T, West KD (2007): Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics*, 138:291–311.

Clements MP, Hendry DF (2001): Forecasting with Difference Stationary and Trend Stationary Models. *Econometrics Journal*, 4:S1–S19.

Cochrane JH (1991): A Critique of the Application of Unit Root Tests. *Journal of Economic Dynamics and Control*, 15:275–284.

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

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, Rossi B (2009): Detecting and Predicting Forecast Breakdowns. *Review of Economic Studies*, 76(2):669–705.

Giacomini R, Rossi B (2010): Forecast Comparisons in Unstable Environments. *Journal of Applied Econometrics*, 25(4):595–620.

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

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.

Hamilton J (1994): *Time Series Analysis*. Princeton University Press.

Jordà Ò (2005): Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, 95:161–182.

Kim H, Durmaz N (2012): Bias Correction and Out-of-sample Forecast Accuracy. *International Journal of Forecasting*, 28:575–586.

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.

Medel CA, Pincheira PM (2016): The Out-of-sample Performance of an Exact Median-Unbiased Estimator for the Near-Unity AR(1) Model. *Applied Economics Letters*, 23(2):126–131.

Newey W, West K (1994): Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, 61:631–654.

Pincheira PM, Medel CA (2015): Forecasting Inflation with a Simple and Accurate Benchmark: The Case of the US and a Set of Inflation Targeting Countries. *Finance a úvěr-Czech Journal of Economics and Finance*, 65(1):2–29.

Pincheira PM, West KD (2016): A Comparison of Some Out-of-sample Tests of Predictability in Iterated Multi-step-ahead Forecasts. *Research in Economics*, 70(2):304–319.

Stock JH, Watson MW (1999): Forecasting Inflation. *Journal of Monetary Economics*, 44(2): 293–335.

Stock JH, Watson MW (2007a): *Introduction to Econometrics*. Second Edition. Pearson, Addison-Wesley.

Stock JH, Watson MW (2007b): Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit, and Banking*, 39(1):3–33.

Stock JH, Watson MW (2008): Phillips Curve Inflation Forecasts. *NBER Working Paper,* no. 14322.

Turner J (2004): Local to Unity, Long Horizon Forecasting Thresholds for Model Selection in the AR(1). *Journal of Forecasting*, 23(7):513–539.