DOCUMENTOS DE TRABAJO

Forecasting Inflation with a Random Walk

Pablo Pincheira
Carlos Medel

N.º 669 Julio 2012

BANCO CENTRAL DE CHILE

DOCUMENTOS DE TRABAJO

Forecasting Inflation with a Random Walk

Pablo Pincheira
Carlos Medel

N.º 669 Julio 2012

BANCO CENTRAL DE CHILE

BANCO CENTRAL DE CHILE

CENTRAL BANK OF CHILE

Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Agustinas 1180, Santiago, Chile
Teléfono: (56-2) 3882475; Fax: (56-2) 3882231

Documento de Trabajo
N° 669

Working Paper
N° 669

FORECASTING INFLATION WITH A RANDOM WALK*

Pablo Pincheira
Banco Central de Chile

Carlos Medel
Banco Central de Chile

Abstract

The use of different time-series models to generate forecasts is fairly usual in the forecasting literature in general, and in the inflation forecast literature in particular. When the predicted 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 driftless unit-root-based forecasts generates bounded Mean Squared Prediction Errors errors at every single horizon. We also show via simulations that persistent stationary processes may be better predicted by unit-root-based forecasts than by forecasts coming from a model that is correctly specified but that is subject to a higher degree of parameter uncertainty. Finally we provide an empirical illustration in the context of CPI inflation forecasts for three industrialized countries.

Resumen

El uso de diferentes modelos de series de tiempo para generar pronósticos es usual en la literatura predictiva en general, y en la literatura de predicción de inflación en particular. Cuando la variable a predecir es estacionaria, el uso de procesos con raíz unitaria puede parecer contraintuitivo. No obstante, en este artículo nosotros demostramos que la predicción de variables estacionarias con pronósticos basados en procesos con raíz unitaria sin intercepto, genera Errores de Predicción Cuadrático Medios acotados en todo horizonte. También mostramos con simulaciones, que procesos estacionarios persistentes pueden ser pronosticados de mejor manera por predicciones generadas a partir de un modelo con raíz unitaria que por predicciones generadas a partir de un modelo correctamente especificado pero sujeto a un grado mayor de incertidumbre paramétrica. Finalmente mostramos una ilustración empírica en el contexto de proyecciones de inflación del IPC para tres países industrializados.

---

* The views expressed in this paper do not necessarily represent those of the Central Bank of Chile or its authorities. Correspondence: Agustinas 1180, Santiago, Chile. Tel: +56-2-3882874. Fax: +56-2-3882836. E-mails: ppinchei@bcentral.cl (corresponding author); cmedel@bcentral.cl.

1 Introduction

Some of the univariate models used to predict macroeconomic time-series, as inflation, 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 overdifferenced. Second, unit-root-based forecasts may have a deterministic trend approaching to infinity (or minus infinity) as the forecasting horizon lengthens, which is in clear 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 to infinity. This may lead 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), Capistrán, Constandse and Ramos-Francia (2010), Groen, Kapetanios, and Price (2009), Elliot and Timmermann (2008), Stock and Watson (2008), among others, show that unit-root-based forecasts perform well when forecasting inflation or GDP growth, variables that may be considered stationary in a number of countries. It is in the context of these findings that we pose the two following questions: (i) when predicting a stationary variable with unit-root-based forecasts, does the MSPE diverge as the forecasting horizon lengthens? and (ii) is it possible that unit-root-based forecasts for a persistent stationary process perform better than forecasts generated from a correctly specified model in finite samples? In this paper we aim at answering these two 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. 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, and the United States is presented in section 4. Finally, section 5 concludes.

2 Forecasting inflation with a unit root process

To set some preliminary ideas, let us consider that the true model of a 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 a 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:

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

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

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

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,

$e_t^{RW}(h)$:

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

The MSPE thus is given by:[1]

$$\begin{aligned} \mathsf{MSPE}(h) &= \mathbb{E}(Y_{t+h} - Y_t^{RW}(h))^2 \\ \mathsf{MSPE}(h) &= \mathbb{E}\left[\alpha \left[\frac{1-\rho^h}{1-\rho}\right] - (1-\rho^h)Y_t\right]^2 + \mathbb{E}(\sum_{i=0}^{h-1} \rho^i \varepsilon_{t+h-i})^2 \\ \mathsf{MSPE}(h) &= (1-\rho^h)^2 \mathbb{V}(Y_t) + \sigma_\varepsilon^2 (\frac{1-\rho^{2h}}{1-\rho^2}) \end{aligned}$$

then:

$$\lim_{h \longrightarrow \infty} \mathsf{MSPE}(h) = 2\mathbb{V}(Y_t)$$

Thus, because Y_t is stationary, forecast errors coming from a RW-based forecast do not display an explosive behavior as the forecasting horizon lengthens. Note that for this implication, the "no-drift" assumption (denoted as $\delta = 0$) plays a key role. In fact, if we had assumed that the true DGP is a RW with drift, $Y_{t+1} = \delta + Y_t + \varepsilon_{t+1}$, we would have ended with forecasts $Y_t^{RWD}(h)$ and corresponding forecast errors $e_t^{RWD}(h)$ according to:

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

In this case, the MSPE is given by:

$$\begin{aligned} \mathsf{MSPE}_D(h) &= \mathbb{E}(Y_{t+h} - Y_t^{RWD}(h))^2 \\ \mathsf{MSPE}_D(h) &= \mathbb{E}\left[\alpha \left[\frac{1-\rho^h}{1-\rho}\right] - \delta h - (1-\rho^h)Y_t\right]^2 + \mathbb{E}(\sum_{i=0}^{h-1} \rho^i \varepsilon_{t+h-i})^2 \\ \mathsf{MSPE}_D(h) &= (1-\rho^h)^2 \mathbb{V}(Y_t) + (\delta h)^2 + \sigma_\varepsilon^2 \sum_{i=0}^{h-1} \rho^{2i} \\ \mathsf{MSPE}_D(h) &= \mathsf{MSPE}(h) + (\delta h)^2 \end{aligned}$$

then,

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

and it is clear that the drift will generate a divergent MSPE.

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

$$Y_{t+1} = \alpha + Y_t + \varepsilon_{t+1}$$

[1] Through this paper we denote the expected value, variance, covariance, and autocovariance with $\mathbb{E}(\cdot)$, $\mathbb{V}(\cdot)$, $\mathbb{C}(\cdot,\cdot)$, and γ_t, respectively.

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:

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

and the optimal MSPE diverges as the horizon lengthens:

$$\begin{aligned} \mathsf{MSPE}(h) &= \mathbb{E}(Y_{t+h} - Y_t^f(h))^2 \\ \mathsf{MSPE}(h) &= \mathbb{E}\left[\sum_{i=0}^{h-1} \varepsilon_{t+h-i}\right]^2 = \sigma_\varepsilon^2 h \underset{h\to\infty}{\longrightarrow} \infty \end{aligned}$$

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 unit-root-based forecasts display a bounded MSPE(h) sequence as the forecasting horizons 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 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 $e_t^{RW}(h)$:

$$\begin{aligned} Y_t^{RW}(h) &= Y_t \\ e_t^{RW}(h) &= Y_{t+h} - Y_t \end{aligned}$$

The MSPE is given by:

$$\begin{aligned} \mathsf{MSPE}(h) &= \mathbb{E}(Y_{t+h} - Y_t^{RW}(h))^2 = \mathbb{E}(Y_{t+h} - Y_t)^2 \\ \mathsf{MSPE}(h) &= \mathbb{V}(Y_{t+h}) + \mathbb{V}(Y_t) - 2\mathbb{C}(Y_{t+h}, Y_t) \\ \mathsf{MSPE}(h) &= 2\mathbb{V}(Y_t) - 2\gamma_h \end{aligned}$$

So,

$$\begin{aligned} \mathsf{MSPE}(h) &= |2\mathbb{V}(Y_t) - 2\gamma_h| \leq 2\mathbb{V}(Y_t) + 2|\gamma_h| \\ \mathsf{MSPE}(h) &\leq 2\mathbb{V}(Y_t) + 2\sqrt{\mathbb{V}(Y_{t+h})\mathbb{V}(Y_t)} = 2\mathbb{V}(Y_t) + 2\sqrt{\mathbb{V}(Y_t)\mathbb{V}(Y_t)} \\ \mathsf{MSPE}(h) &\leq 4\mathbb{V}(Y_t) \end{aligned}$$

and

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

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, that is to say,*

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

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

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

therefore

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

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 United States is very competitive when predicting 12-months ahead. Giacomini and White (2006), also for the United States, present an empirical application in which several CPI forecasts are compared to those generated by a RW with drift and an autoregression (AR) whose lag length is selected according to the Bayesian Information Criteria (BIC). Another article using simple univariate benchmarks for the United States is Ang, Bekaert, and Wei (2007). Among the many methods the authors use, they include a RW. In addition, Croushore (2010) makes use of an integrated moving average IMA(1,1) model as a benchmark when evaluating survey-based inflation forecast for the United States.

In other countries the use of unit-root-based forecasts is also fairly usual. 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 García (2012) also consider seasonal ARMA models with unit root to construct forecasts for Chilean CPI inflation. Finally, Pincheira and Medel (2012) also make use of unit-root-based forecasts to predict YoY CPI inflation for twelve countries both at short and long horizons.

We proceed next to show another proposition that goes in the same line as Proposition 1. The novelty 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 also consider a white noise process $\{\varepsilon_{t+1}\}_{t=-\infty}^{\infty}$ with variance σ_ε^2, such that the moments $\mathbb{C}(Y_{t+j}, \varepsilon_{t+i})$ for all $i, j \in \mathbb{Z}$ are well defined. Then, forecasts coming from a driftless ARIMA(p,1,q) process with $0 \leq p,\ q < \infty$ will display a bounded MSPE sequence as the forecasting horizon approaches to infinity.*

Proof. In appendix A we provide a proof for the particular case in which forecasts comes from an ARIMA(*0,1,*q) model. In appendix B, however, we consider forecasts coming from an ARIMA(p,*1,0*) model. At last, in appendix C we give a more general proof for the case ARIMA(p,*1*,q). A way to proceed is to proof only the last case that obviously encompasses the others. Nevertheless, as we have used different approaches in the proofs, we think it is worth showing them all. ■

So far we have shown that the construction of unit-root-based forecasts for stationary variables does not imply an explosive behavior of the MSPE as the forecast horizon lengthens. In the next section we will show with simulations that parameter uncertainty in combination with persistence may generate a large noise in the ordinary least squares (OLS) estimates of simple stationary processes. Under this scenario, we will provide evidence that unit-root-based forecasts may offer more accuracy than correctly specified forecasts in small and moderate samples due to their relative parsimony. This is particularly relevant at long horizons.

3 Monte Carlo simulations

We generate 10,000 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 equals to $\sigma_\varepsilon^2 = 0.25$. Thus, the models look as follows:

$$\begin{aligned} \mathsf{AR(1)} &: \quad Y_{t+1} = \alpha + \rho Y_t + \varepsilon_{t+1} \\ \sigma_\varepsilon^2 &= \quad 0.25 \text{ and } 0 < \rho < 1 \end{aligned}$$

$$\begin{aligned} \mathsf{AR(2)} &: \quad Y_{t+1} = \alpha + \phi_1 Y_t + \phi_2 Y_{t-1} + \varepsilon_{t+1} \\ \sigma_\varepsilon^2 &= \quad 0.25 \text{ and } 0 < \phi_1 + \phi_2 < 1 \end{aligned}$$

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 varies between 1 and 2 depending on the process we are considering, an 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 1-step ahead predictions we construct. In all our simulations we set $P = 500$. We are not only interested in 1-step ahead forecasts, so we engage in an out-of-sample h-step ahead evaluation, with $h = \{1, 12, 24, 36\}$, where the parameters of the processes are estimated with rolling OLS.

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 and must be estimated with rolling OLS.[2] 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 the 10,000 replications we compute the average across all the 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

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

We report these ratios for the three values of $R = \{50, 100, 200\}$, and several choices of the parameters that defines the AR(1) and AR(2) models. In particular, we consider AR(1) specifications

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

with the following parameter values:

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

For the AR(2) model we consider the following parameter values[3]

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

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

Results in table 1 show three salient features that are worth mentioning:

1. First, as the sample size gets larger, all the ratios become larger as well. This is easy to understand, because larger estimation samples implies more precise parameter estimation. With small estimation noise, we should expect a better performance of correctly over incorrectly specified forecasts.

2. Second, as the persistence of the processes increases, all the ratios show a tendency to decrease. 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 drives behind these results. Firstly, as ρ gets larger the process approaches to a RW. Therefore, the RW becomes closer and closer to be the correct specification. It also happens that as ρ gets larger, the small sample bias of the OLS estimates of ρ gets worse and worse. These two forces point to the same direction and helps 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 the RW is not nested in the AR(2) process. 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 (2007) 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 as those in Clark and West (2006, 2007) are based on the problems that not vanishing parameter uncertainty may generate when carrying out out-of-sample inference.

3. Third, table 1 shows an interesting interaction between persistence, sample size and forecasting horizon. We can see that given a sample size of R, there is a persistence threshold after which the MSPE-ratios are decreasing 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 decreasing pattern. Of course as R gets larger this decreasing pattern is smoother. All this means that the problem of noisy 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 profitable strategy to use in the long run. To our knowledge, this third salient feature has not been covered by the literature thus far.

[3] These values belong to the stationarity region for an AR(2) process, which is characterized by the following expressions

$$\begin{aligned} \phi_1 + \phi_2 &< 1, \\ \phi_2 - \phi_1 &< 1, \text{ and} \\ -1 &< \phi_2 < 1. \end{aligned}$$

Results in table 2 reinforces the argument 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 higher ratios the smaller the sample size is. Similarly, higher ratios are obtained when the persistence of the processes is higher. Interestingly, higher ratios are also achieved when forecasting at longer horizons. This pattern is more striking with higher levels of persistence.

Results from tables 1-2 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 in the need of long run forecasts.

Table 1: MSPE-ratio estimates

	AR(1)						AR(2)				
ρ	0.500	0.900	0.950	0.975	0.990	ϕ_1	0.400	0.500	0.500	0.500	0.500
	-	-	-	-	-	ϕ_2	0.100	0.400	0.450	0.475	0.490
						$\alpha = 0$					
						R=50					
h=1	1.277	0.996	0.908	0.900	0.897		1.311	1.208	1.209	1.218	1.228
h=12	1.884	1.213	0.676	0.611	0.586		1.877	0.904	0.907	0.832	0.804
h=24	1.885	1.300	0.023	0.013	0.034		1.880	0.787	0.799	0.669	0.630
h=36	1.884	1.256	0.000	0.000	0.000		1.880	0.476	0.523	0.255	0.342
						R=100					
h=1	1.306	1.027	0.997	0.982	0.973		1.356	1.262	1.262	1.269	1.274
h=12	1.941	1.374	1.094	0.946	0.867		1.936	1.034	1.032	0.928	0.876
h=24	1.943	1.573	1.205	0.960	0.825		1.940	1.077	1.073	0.899	0.812
h=36	1.941	1.643	1.271	0.955	0.777		1.938	1.099	1.091	0.863	0.750
						R=200					
h=1	1.320	1.041	1.013	0.998	0.989		1.378	1.276	1.290	1.296	1.300
h=12	1.970	1.463	1.190	1.031	0.932		1.969	1.343	1.123	1.006	0.938
h=24	1.971	1.708	1.364	1.095	0.920		1.969	1.556	1.224	1.018	0.897
h=36	1.971	1.792	1.488	1.158	0.918		1.965	1.674	1.313	1.041	0.877
						$\alpha = 1$					
						R=50					
h=1	1.278	0.996	0.965	0.951	0.946		1.312	1.200	1.209	1.219	1.227
h=12	1.885	1.216	0.954	0.837	0.788		1.878	1.099	0.906	0.831	0.802
h=24	1.889	1.283	0.906	0.723	0.633		1.882	1.089	0.778	0.674	0.627
h=36	1.884	0.696	0.660	0.480	0.330		1.881	0.724	0.326	0.388	0.301
						R=100					
h=1	1.306	1.027	0.997	0.981	0.973		1.355	1.251	1.262	1.269	1.276
h=12	1.939	1.372	1.095	0.944	0.867		1.935	1.254	1.033	0.925	0.880
h=24	1.941	1.573	1.206	0.959	0.827		1.938	1.412	1.075	0.895	0.818
h=36	1.940	1.642	1.270	0.956	0.778		1.938	1.493	1.096	0.860	0.755
						R=200					
h=1	1.320	1.041	1.013	0.998	0.989		1.378	1.277	1.290	1.297	1.300
h=12	1.969	1.458	1.190	1.032	0.931		1.967	1.347	1.122	1.006	0.939
h=24	1.970	1.701	1.366	1.096	0.918		1.969	1.558	1.221	1.016	0.898
h=36	1.974	1.790	1.488	1.158	0.914		1.968	1.673	1.308	1.038	0.878

Source: Authors' elaboration.

Table 2: The impact of noisy estimation on MSPE

	AR(1)						AR(2)				
ρ	0.500	0.900	0.950	0.975	0.990	ϕ_1	0.400	0.500	0.500	0.500	0.500
	-	-	-	-	-	ϕ_2	0.100	0.400	0.450	0.475	0.490
						$\alpha = 1$					
						R=50					
h=1	1.044	1.056	1.064	1.065	1.062		1.066	1.082	1.086	1.086	1.082
h=12	1.060	1.280	1.364	1.374	1.347		1.063	1.317	1.367	1.361	1.332
h=24	1.060	1.438	1.711	1.711	1.769		1.062	1.577	1.818	1.807	1.753
h=36	1.060	2.796	2.624	2.624	3.568		1.062	2.574	4.813	3.368	3.772
						R=100					
h=1	1.020	1.026	1.029	1.033	1.032		1.034	1.038	1.040	1.042	1.042
h=12	1.030	1.136	1.184	1.226	1.220		1.035	1.156	1.194	1.223	1.215
h=24	1.030	1.178	1.285	1.361	1.352		1.035	1.222	1.311	1.362	1.342
h=36	1.030	1.191	1.366	1.506	1.513		1.036	1.256	1.426	1.524	1.504
						R=200					
h=1	1.010	1.013	1.012	1.015	1.014		1.016	1.017	1.018	1.020	1.019
h=12	1.015	1.072	1.094	1.117	1.140		1.014	1.073	1.103	1.128	1.133
h=24	1.015	1.092	1.137	1.182	1.224		1.014	1.102	1.161	1.201	1.210
h=36	1.015	1.098	1.163	1.233	1.294		1.014	1.116	1.204	1.265	1.277

Source: Authors' elaboration.

4 Empirical evidence

In this section we illustrate the benefits of unit-root-based forecasts from the point of view of a practitioner in which different models are used to generate inflation forecasts. We first describe the used dataset, and then the models. Finally, we evaluate the relative accuracy between the 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, and the United States. The source of the dataset are country-specific central banks and the Federal Reserve Bank of St. Louis. This implies that despite some CPI basket changes experimented in some countries during the sample period, we only use official inflation data for each country. Our models work with YoY CPI inflation rates, defined as

$$\pi_t = (CPI_t/CPI_{t-12}) \cdot 100 - 100$$

Table 3 shows results of traditional unit root tests for these three series for the sample period covering from September 1995 to December 2011. At standard significance values, these traditional tests rejects the presence of a unit root in the data for Canada and the United States. For Sweden the evidence is mixed: the Augmented Dickey-Fuller tests rejects the null of a unit root, but the Phillips-Perron test is not able to reject it at 10% significance level.

Table 3: Unit root testing –full sample

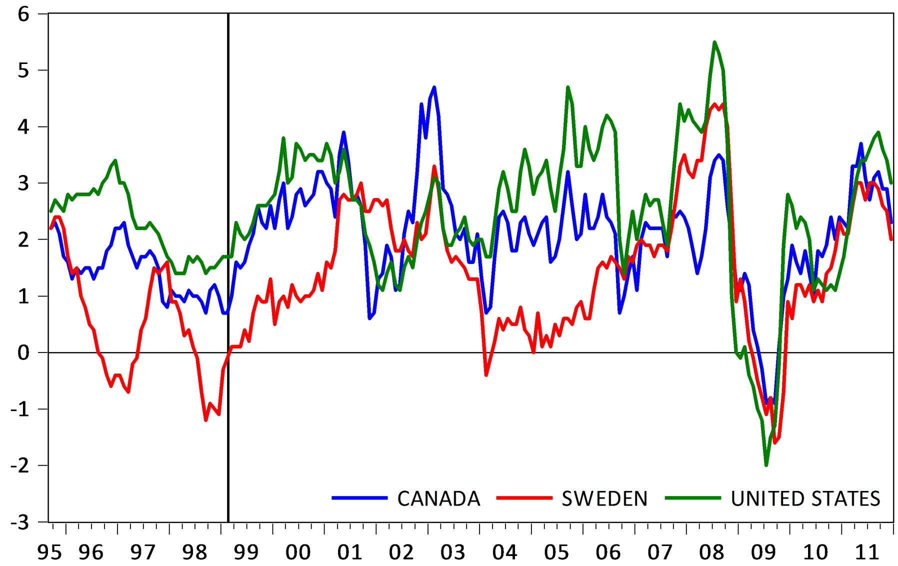
	Level (π_t)	
	Aug. Dickey-Fuller	Phillips-Perron
Canada	-3.903	-3.680
(p-value)	(0.013)	(0.025)
Sweden	-4.219	-2.760
(p-value)	(0.004)	(0.213)
United States	-3.523	-3.683
(p-value)	(0.039)	(0.025)

The null hypothesis is that the series have a unit root.

Source: Authors' computations.

Figure 1 shows the series of YoY CPI inflation for these three countries. Some descriptive statistics for different subsamples are available in appendix D.

Figure 1: CPI Inflation of Canada, Sweden, and the United States



Vertical line: Evaluation sample startpoint (Feb-1999). Source: Country-specific central banks.

Under the plausible assumption of stationarity we will use the following models to generate out-of-sample forecasts for these three series:

1. AR(1): $\pi_t = \alpha + \rho_1 \pi_{t-1} + \varepsilon_t$
2. AR(6): $\pi_t = \alpha + \sum_{i=1}^{6} \rho_i \pi_{t-i} + \varepsilon_t$
3. AR(12): $\pi_t = \alpha + \sum_{i=1}^{12} \rho_i \pi_{t-i} + \varepsilon_t$

We will also construct unit-root-based forecasts based upon two commonly used models in the forecasting literature: the driftless RW and the *airline model* introduced by Box and Jenkins (1970):[4]

[4] As Ghysels, Osborn, and Rodrigues (2006) point out, this specification has proved to be very useful to forecast monthly time series with seasonal patterns.

1. Random walk: $\pi_t - \pi_{t-1} = \varepsilon_t$
2. Airline model: $\pi_t - \pi_{t-1} = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_{12}\varepsilon_{t-12} + \theta_1\theta_{12}\varepsilon_{t-13}$

We estimate the models with a fixed-size rolling window of length R'. We consider two values for R': 40 and 100 observations. The first estimation sample with $R' = 40$ covers the period from September 1995 to January 1999, while the first estimation sample with $R' = 100$ covers the period from September 1990 to January1999. The remaining sample is used for the evaluation of the forecasts, covering from February 1999 to December 2011. This means that we consider 155 one-month ahead forecasts, 144 twelve-, 132 two-, and 120 three-years ahead forecasts.

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 4 we report estimates of the sample-RMSPE for all the forecasts under consideration. This sample estimate is calculated as follows

$$\widehat{\mathsf{RMSPE}}_h = \left[\frac{1}{T(h)}\sum_{t=1}^{T(h)}(\pi_{t+h} - \pi_t(h))^2\right]^{\frac{1}{2}}$$

where $\pi_t(h)$ is the h-step ahead forecast of π_t, and $T(h)$ represents the total number of out-of-sample forecast errors available for a given methodology and forecasting horizon.

Table 4: Multi-horizon RMSPE estimates

	Rolling-window size: 40				Rolling-window size: 100			
	h=1	h=12	h=24	h=36	h=1	h=12	h=24	h=36
				Canada				
1. AR(1)	0.493	1.275	1.191	1.060	0.487	1.231	1.106	1.086
2. AR(6)	0.522	1.336	1.213	1.251	0.497	1.140	1.098	1.064
3. AR(12)	0.552	1.703	2.092	1.307	0.518	1.307	1.271	1.047
1. RW	0.493	1.560	1.476	1.130	0.493	1.560	1.476	1.130
2. *Airline model*	0.403	1.208	1.143	1.045	0.368	1.114	1.075	0.957
				Sweden				
1. AR(1)	0.417	2.182	2.900	2.841	0.398	1.601	1.660	1.357
2. AR(6)	0.462	2.392	3.466	5.124	0.414	1.499	1.448	1.265
3. AR(12)	0.522	4.019	4.977	18.620	0.430	1.511	1.342	1.338
1. RW	0.400	1.742	2.088	1.863	0.400	1.742	2.088	1.863
2. *Airline model*	0.338	1.481	1.591	1.438	0.322	1.299	1.351	1.319
				United States				
1. AR(1)	0.534	1.855	2.119	1.604	0.519	1.634	1.619	1.519
2. AR(6)	0.508	1.837	1.956	1.715	0.474	1.614	1.516	1.512
3. AR(12)	0.550	3.499	6.828	2.585	0.483	2.040	2.259	1.646
1. RW	0.521	2.125	2.050	1.842	0.521	2.125	2.050	1.842
2. *Airline model*	0.373	1.652	1.569	1.578	0.331	1.501	1.474	1.499

Source: Authors' computations.

Results in table 4 indicate that the airline model outperforms the rest of the forecasting methods at every single forecasting horizon and for every country with only two exceptions: the case of Sweden when forecasting two and three years ahead, when estimation is carried out with R=100 observations. If we assume that YoY inflation for Canada, Sweden, and the United States are stationary, then our empirical findings are consistent with our simulation results. The two exceptions mentioned above are also consistent with the evidence shown in our simulations because the airline model outperforms all of the benchmarks when 40 observations are used for estimation. It is only when we increase the number of observations that the airline model is outperformed at long horizons for the case of Sweden.

Table 5 shows the p-values of the Giacomini and White (2006) test of superior predictive ability between each AR(p) model and the models with unit roots we are considering in this application (RW and airline model). The null hypothesis is that of superior predictive ability of the AR(p) models, while the alternative is that our unit root models 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 in favor of the airline model in a number of occasions. This happens both at short and long horizons. In particular, the airline model beats all of the AR(p) specifications for Canada, at the longest forecasting horizon, when the parameters are estimated with 100 observations.

Table 5: Giacomini-White test results –p-value

		Rolling-window size: 40			Rolling-window size: 100		
		AR(1)	AR(6)	AR(12)	AR(1)	AR(6)	AR(12)
Canada							
Random walk	h=1	0.519	0.085	0.039	0.711	0.416	0.174
	h=12	0.998	0.959	0.267	0.999	0.998	0.956
	h=24	0.999	0.984	0.077	0.979	0.970	0.948
	h=36	0.777	0.181	0.175	0.627	0.697	0.750
Airline model	h=1	0.000	0.000	0.000	0.000	0.000	0.000
	h=12	0.163	0.091	0.025	0.005	0.257	0.065
	h=24	0.287	0.236	0.044	0.257	0.238	0.038
	h=36	0.369	0.050	0.059	0.021	0.017	0.025
Sweden							
Random walk	h=1	0.006	0.013	0.002	0.706	0.179	0.075
	h=12	0.007	0.023	0.029	0.998	0.999	0.972
	h=24	0.013	0.047	0.026	1.000	1.000	1.000
	h=36	0.066	0.048	0.132	1.000	1.000	1.000
Airline model	h=1	0.007	0.002	0.000	0.014	0.001	0.000
	h=12	0.019	0.027	0.025	0.062	0.147	0.079
	h=24	0.005	0.024	0.011	0.007	0.133	0.586
	h=36	0.025	0.038	0.131	0.263	0.869	0.402
United States							
Random walk	h=1	0.061	0.653	0.314	0.593	0.925	0.786
	h=12	0.978	0.985	0.095	0.999	1.000	0.589
	h=24	0.344	0.764	0.068	0.992	0.987	0.243
	h=36	0.977	0.898	0.010	0.989	0.990	0.915
Airline model	h=1	0.001	0.000	0.005	0.001	0.000	0.000
	h=12	0.043	0.015	0.058	0.002	0.039	0.080
	h=24	0.062	0.073	0.065	0.069	0.053	0.023
	h=36	0.266	0.072	0.003	0.187	0.323	0.020

Source: Authors' computations.

5 Concluding remarks

The use of different time-series models to generate forecasts is usual in the forecasting literature. When the target variable is stationary, the construction of forecasts coming from processes with unit roots may seem counterintuitive. Nevertheless, in this paper we demonstrate that forecasting a stationary variable with driftless unit-root-based forecasts generates bounded Mean Squared Prediction Errors at every single horizon.

We also show via simulations that persistent stationary processes may be better predicted by unit-root-based forecasts than by forecasts coming from a model that is correctly specified but that is subject to a higher degree of parameter uncertainty.

Our simulations also provide evidence indicating that the benefits of using unit-root-based forecasts is not only confined within the boundaries of short horizons. In fact, the benefits may be sizable at long horizons as well. Future research might explore if the results we have found here may still hold true in more general environments. A natural extension would be the analysis with vector autoregressions and nonlinear processes.

References

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

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

3. Box, G.E.P. and G.M. Jenkins, 1970, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, USA.

4. Box, G.E.P, G.M. Jenkins and G. Reinsel, 2008, *Time Series Analysis: Forecasting and Control*, fourth edition, Wiley.

5. Capistrán, C., C. Constandse, and M. Ramos-Francia, 2010, "Multi-Horizon Inflation Forecasts using Disaggregated Data," *Economic Modelling* **27**: 666–677.

6. Clark, T. and K. West, 2006, "Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis," *Journal of Econometrics* **135**: 155-186.

7. Clark, T. and K. West, 2007, "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics* **138:** 291–311.

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

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

10. Giacomini, R. and H. White, 2006, "Test of Conditional Predictive Ability," *Econometrica* **74**: 1545–1578.

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

12. Ghysels E., D. Osborn, and P.M. Rodrigues, 2006, *Forecasting Seasonal Time Series*, in Elliot, G., C.W.J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 1, Elsevier, North Holland.

13. Hamilton J., 1994, *Time Series Analysis*, Princeton University Press.

14. Pincheira, P. and A. García, 2012, "En Busca de un Buen Marco de Referencia Predictivo para la Inflación en Chile," *El Trimestre Económico* **LXXIX**(313): 85–123.

15. Pincheira P. and C. Medel, 2012, "Forecasting Inflation with a Simple and Accurate Benchmark: A Cross-Country Analysis," *manuscript*, Central Bank of Chile.

16. Stock J. and M. Watson, 2007, *Introduction to Econometrics*, Second Edition, Pearson, Addison-Wesley.

17. Stock J. and M. Watson, 2008, "Phillips Curve Inflation Forecasts" in *Understanding Inflation and the Implications for Monetary Policy, A Phillips Curve Retrospective*, Federal Reserve Bank of Boston, 2009.

A Proof of Proposition 2: ARIMA(*0,1,q*)

Let us consider $\{X_{t+1}\}_{t=-\infty}^{\infty}$ following an ARIMA(p,d,q) process, where $W_{t+1} = (1-B)^d X_{t+1}$ represents a stationary and invertible ARMA(p,q) process, with B a backshift operator ($B^j Z_t = Z_{t-j}$). Thus,

$$\begin{aligned} W_t - \sum_{j=1}^{p} \phi_j W_{t-j} &= \delta + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \\ \phi(B)(1-B)^d X_{t+1} &= \delta + \theta(B)\varepsilon_t \end{aligned}$$

We are concerned with the particular case in which $d = 1$ and $p = 0$. Therefore, we have that $\{X_{t+1}\}_{t=\infty}^{\infty}$ satisfies $W_{t+1} = (1-B)X_{t+1}$, implying that:

$$W_{t+1} = \delta + \varepsilon_{t+1} - \sum_{j=1}^{q} \theta_j \varepsilon_{t+1-j}$$

We notice that:

$$\begin{aligned} X_{t+h} - X_t &= W_{t+h} + W_{t+h-1} + ... + W_{t+1} = \sum_{i=1}^{h} W_{t+i} \\ X_{t+h} - X_t &= \delta h + \sum_{i=1}^{h} \varepsilon_{t+i} - \sum_{i=1}^{h}\sum_{j=1}^{q} \theta_j \varepsilon_{t+i-j} \end{aligned}$$

so

$$\begin{aligned} h = 1: \quad & X_{t+1} = \delta + X_t + \varepsilon_{t+1} - \textstyle\sum_{j=1}^{q} \theta_j \varepsilon_{t+1-j} \\ h = q: \quad & X_{t+q} = q\delta + X_t + \textstyle\sum_{j=1}^{q} \varepsilon_{t+j} - \sum_{j=1}^{q} \theta_j \varepsilon_{t+1-j} - \sum_{j=1}^{q} \theta_j \varepsilon_{t+2-j} - \sum_{j=1}^{q} \theta_j \varepsilon_{t+3-j} - \\ & -... - \textstyle\sum_{j=1}^{q} \theta_j \varepsilon_{t+q-j} \\ h = q+l: \quad & X_{t+q+l} = (q+l)\,\delta + X_t + \textstyle\sum_{j=1}^{q+l} \varepsilon_{t+j} - \sum_{j=1}^{q} \theta_j \varepsilon_{t+1-j} - \sum_{j=1}^{q} \theta_j \varepsilon_{t+2-j} - \sum_{j=1}^{q} \theta_j \varepsilon_{t+3-j} - \\ & -... - \textstyle\sum_{j=1}^{q} \theta_j \varepsilon_{t+q-j} - ... - \sum_{j=1}^{q} \theta_j \varepsilon_{t+q+l-j} \end{aligned}$$

For $l > 0$ we have that the best linear forecast based upon information available until time t is given by:

$$X_t^f(q+l) = (q+l)\,\delta + X_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t+1-j} - \sum_{j=2}^{q} \theta_j \varepsilon_{t+2-j} - \sum_{j=3}^{q} \theta_j \varepsilon_{t+3-j} - \ldots - \sum_{j=q}^{q} \theta_j \varepsilon_{t+q-j}$$

Let us use the following notation:

$$\varkappa_t(q) \equiv \sum_{j=1}^{q} \theta_j \varepsilon_{t+1-j} + \sum_{j=2}^{q} \theta_j \varepsilon_{t+2-j} + \sum_{j=3}^{q} \theta_j \varepsilon_{t+3-j} + \ldots + \sum_{j=q}^{q} \theta_j \varepsilon_{t+q-j}$$

Note that $\varkappa_t(q)$ is a zero-mean stationary process that does not depend on h or l. Thus, the optimal forecast $X_t^f(q+l)$, and the corresponding forecast error $e_t^f(q+l)$ are given by:

$$\begin{aligned} X_t^f(q+l) &= (q+l)\,\delta + X_t + \varkappa_t(q) \\ e_t^f(q+l) &= \sum_{j=1}^{q+l} \varepsilon_{t+j} - \sum_{j=1}^{1} \theta_j \varepsilon_{t+2-j} - \sum_{j=1}^{2} \theta_j \varepsilon_{t+3-j} - \ldots - \\ &\quad - \sum_{j=1}^{q-1} \theta_j \varepsilon_{t+q-j} - \sum_{j=1}^{q} \theta_j \varepsilon_{t+q+1-j} - \ldots - \sum_{j=1}^{q} \theta_j \varepsilon_{t+q+l-j} \end{aligned}$$

If the econometrician mistakenly assumes that the stationary process Y_{t+1} follows an ARIMA(*0,1,q*) process with $0 \leq q < \infty$, innovations given by $\varepsilon_t \sim iid\mathcal{N}(0, \sigma_\varepsilon^2)$, and moving average parameters θ_j, then he or she would construct the following optimal forecast for Y_{t+q+l}:

$$Y_t^f(q+l) = (q+l)\,\delta + Y_t + \varkappa_t(q)$$

In a driftless manner ($\delta = 0$) the MSPE is given by:

$$\begin{aligned} \mathsf{MSPE}(q+l) &= \mathbb{E}(Y_{t+q+l} - Y_t^f(q+l))^2 \\ \mathsf{MSPE}(q+l) &= \mathbb{E}(Y_{t+h} - Y_t - \varkappa_t(q))^2 \end{aligned}$$

Notice also that under stationarity assumptions for Y_t we will have that:

$$\mathbb{E}(Y_{t+h} - Y_t) = 0$$

implying that:

$$\mathbb{E}(Y_{t+h} - Y_t - \varkappa_t(q))^2 = \mathbb{V}(Y_{t+h} - Y_t - \varkappa_t(q))$$

With this in mind, we have that:

$$\begin{aligned} \mathsf{MSPE}(q+l) &= \mathbb{E}(Y_{t+q+l} - Y_t^f(q+l))^2 \\ \mathsf{MSPE}(q+l) &= \mathbb{E}(Y_{t+h} - Y_t - \varkappa_t(q))^2 \\ \mathsf{MSPE}(q+l) &= \mathbb{V}(Y_{t+q+l} - Y_t - \varkappa_t(q)) \\ \mathsf{MSPE}(q+l) &= \mathbb{V}(Y_{t+q+l} - Y_t) + \mathbb{V}(\varkappa_t(q)) - 2\mathbb{C}(Y_{t+q+l} - Y_t, \varkappa_t(q)) \\ \mathsf{MSPE}(q+l) &= \mathbb{V}(Y_{t+q+l}) + \mathbb{V}(Y_t) - 2\gamma_{q+l} + 2\mathbb{C}(Y_t, \varkappa_t(q)) - 2\mathbb{C}(Y_{t+q+l}, \varkappa_t(q)) \\ &= 2\mathbb{V}(Y_t) - 2\gamma_{q+l} + 2\mathbb{C}(Y_t, \varkappa_t(q)) - 2\mathbb{C}(Y_{t+q+l}, \varkappa_t(q)) \end{aligned}$$

Therefore:

$$
\begin{aligned}
|\mathsf{MSPE}(q+l)| &\leq 2\left[|\mathbb{V}(Y_t)| + |\gamma_{q+l}| + |\mathbb{C}(Y_t, \varkappa_t(q))| + |\mathbb{C}(Y_{t+q+l}, \varkappa_t(q))|\right] \\
&\leq 2\left[|\mathbb{V}(Y_t)| + |\gamma_{q+l}| + |\mathbb{C}(Y_t, \varkappa_t(q))| + \sqrt{\mathbb{V}(Y_{t+q+l})}\sqrt{\mathbb{V}(\varkappa_t(q))}\right] \\
&\leq 2\left[2|\mathbb{V}(Y_t)| + |\mathbb{C}(Y_t, \varkappa_t(q))| + \sqrt{\mathbb{V}(Y_t)}\sqrt{\mathbb{V}(\varkappa_t(q))}\right]
\end{aligned}
$$

hence,

$$
\begin{aligned}
\lim_{l\to\infty} \mathsf{MSPE}(q+l) &\leq \lim_{l\to\infty} 2\left[2|\mathbb{V}(Y_t)| + |\mathbb{C}(Y_t, \varkappa_t(q))| + \sqrt{\mathbb{V}(Y_t)}\sqrt{\mathbb{V}(\varkappa_t(q))}\right] \\
&= 2\left[2\mathbb{V}(Y_t) + |\mathbb{C}(Y_t, \varkappa_t(q))| + \sqrt{\mathbb{V}(Y_t)}\sqrt{\mathbb{V}(\varkappa_t(q))}\right] \\
&\leq 2\left[2\mathbb{V}(Y_t) + 2\sqrt{\mathbb{V}(Y_t)}\sqrt{\mathbb{V}(\varkappa_t(q))}\right]
\end{aligned}
$$

implying that the $\mathsf{MSPE}(q+l)$ is a bounded sequence.

B Proof of Proposition 2: ARIMA(p,1,0)

Let us consider $\{X_{t+1}\}_{t=\infty}^{\infty}$ following an ARIMA(p,d,q) process, where $W_{t+1} = (1-B)^d X_{t+1}$ represents a stationary and invertible ARMA(p,q) process, with B a backshift operator ($B^j Z_t = Z_{t-j}$). Thus,

$$
\begin{aligned}
W_t - \sum_{j=1}^{p} \phi_j W_{t-j} &= \delta + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \\
\phi(B)(1-B)^d X_{t+1} &= \delta + \theta(B)\varepsilon_{t+1}
\end{aligned}
$$

We are concerned with the particular case in which $d = 1$ and $q = 0$. Therefore, we have that $\{X_{t+1}\}_{t=\infty}^{\infty}$ satisfies $W_{t+1} = (1-B)X_{t+1}$, implying that:

$$
\begin{aligned}
\phi(B)(1-B)X_{t+1} &= \delta + \varepsilon_{t+1} \\
\phi(B)W_{t+1} &= \delta + \varepsilon_{t+1} \\
W_t - \sum_{j=1}^{p} \phi_j W_{t-j} &= \delta + \varepsilon_{t+1}
\end{aligned}
$$

Let us define

$$
Z_t = W_t - \mu; \quad \mu \equiv \frac{\delta}{1 - \sum\limits_{j=1}^{p} \phi_j}
$$

then

$$
\begin{aligned}
Z_t - \sum_{j=1}^{p} \phi_j Z_{t-j} &= W_t - \mu - \sum_{j=1}^{p} \phi_j \left(W_{t-j} - \mu\right) \\
Z_t - \sum_{j=1}^{p} \phi_j Z_{t-j} &= W_t - \sum_{j=1}^{p} \phi_j W_{t-j} - \mu\left(1 - \sum_{j=1}^{p} \phi_j\right) \\
Z_t - \sum_{j=1}^{p} \phi_j Z_{t-j} &= W_t - \sum_{j=1}^{p} \phi_j W_{t-j} - \delta = \varepsilon_{t+1}
\end{aligned}
$$

Consequently, we could write

$$W_t - \sum_{j=1}^{p} \phi_j W_{t-j} = \delta + \varepsilon_{t+1}$$

or equivalently as

$$Z_t - \sum_{j=1}^{p} \phi_j Z_{t-j} = \varepsilon_{t+1}$$

Let us write down the previous process as the following VAR(1):

$$\xi_t = F\xi_{t-1} + v_t$$

where

$$\xi_t = \begin{pmatrix} Z_t \\ Z_{t-1} \\ . \\ . \\ . \\ Z_{t-p+1} \end{pmatrix}_{p\times 1} \quad \text{and } v_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}_{p\times 1}$$

$$F = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & ... & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & ... & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}_{p\times p}$$

with

$$\mathbb{E} v_t v_\tau^T = Q \mathbb{I}_{\{t=\tau\}}$$

$$Q = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & 0 & ... & 0 & 0 \\ 0 & 0 & 0 & ... & 0 & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}_{p\times p}$$

We notice that the first equation in

$$\xi_t = F\xi_{t-1} + v_t$$

will capture the initial process

$$\phi(B) W_{t+1} = \delta + \varepsilon_{t+1}$$

perfectly well. We notice that:

$$\begin{aligned} \xi_{t+1} &= F\xi_t + v_{t+1} \\ \xi_{t+2} &= F\left(F\xi_t + v_{t+1}\right) + v_{t+2} = F^2\xi_t + Fv_{t+1} + v_{t+2} \\ \xi_{t+3} &= F^2\xi_{t+1} + Fv_{t+2} + v_{t+3} = F^2\left(F\xi_t + v_{t+1}\right) + Fv_{t+2} + v_{t+3} = F^3\xi_t + F^3 v_{t+1} + Fv_{t+2} + v_{t+3} \end{aligned}$$

So, in general we have

$$\xi_{t+h} = F^h\xi_t + F^{h-1}v_{t+1} + F^{h-2}v_{t+2} + F^{h-3}v_{t+3} + ... + Fv_{t+h-1} + v_{t+h}$$

Let us consider the first component of the vector $\sum_{i=1}^{h} \xi_{t+i}$. This first component is equal to

$$\left[\sum_{i=1}^{h} \xi_{t+i}\right]_1 = \sum_{i=1}^{h} Z_{t+i} = \sum_{i=1}^{h} (W_{t+i} - \mu) = \sum_{i=1}^{h} (Y_{t+i} - Y_{t+i-1}) - \mu h = Y_{t+h} - Y_t - \mu h$$

therefore

$$Y_{t+h} = \mu h + Y_t + \left[\sum_{i=1}^{h} \xi_{t+i}\right]_1$$

In other words

$$Y_{t+h} = \mu h + Y_t + \left[\sum_{i=1}^{h} F^i \xi_t + F^{i-1} v_{t+1} + F^{i-2} v_{t+2} + F^{i-3} v_{t+3} + ... + F v_{t+i-1} + v_{t+i}\right]_1$$

or

$$\begin{aligned} Y_{t+h} &= \mu h + Y_t + [F\xi_t + v_{t+1}]_1 + \left[F^2 \xi_t + F v_{t+1} + v_{t+2}\right]_1 + ... \\ &... + \left[F^h \xi_t + F^{h-1} v_{t+1} + F^{h-2} v_{t+2} + F^{h-3} v_{t+3} + ... + F v_{t+h-1} + v_{t+h}\right]_1 \end{aligned}$$

The best linear forecast of Y_{t+h} with information available at time t would be:

$$Y_t^f(h) = \mu h + Y_t + [F\xi_t]_1 + \left[F^2 \xi_t\right]_1 + ... + \left[F^h \xi_t\right]_1$$

but

$$\begin{aligned} [F\xi_t]_1 + \left[F^2\xi_t\right]_1 + ... + \left[F^h \xi_t\right]_1 &= \left[\left(F + F^2 + ... + F^h\right) \xi_t\right]_1 = \left[\left(I + F + F^2 + ... + F^h - I\right) \xi_t\right]_1 \\ &= \left[\left(I - F^{h+1}\right)(I - F)^{-1} \xi_t - \xi_t\right]_1 \\ &= \left[\left(I - F^{h+1}\right)(I - F)^{-1} \xi_t\right]_1 - [\xi_t]_1 \\ &= \left[\left(I - F^{h+1}\right)(I - F)^{-1} \xi_t\right]_1 - Z_t \\ &= \left[\left(I - F^{h+1}\right)(I - F)^{-1} \xi_t\right]_1 - W_t + \mu \\ &= \left[\left(I - F^{h+1}\right)(I - F)^{-1} \xi_t\right]_1 - Y_t + Y_{t-1} + \mu \end{aligned}$$

hence

$$\begin{aligned} Y_t^f(h) &= \mu h + Y_t + [F\xi_t]_1 + \left[F^2 \xi_t\right]_1 + ... + \left[F^h \xi_t\right]_1 \\ Y_t^f(h) &= \mu h + Y_t + \left[\left(I - F^{h+1}\right)(I - F)^{-1} \xi_t\right]_1 - Y_t + Y_{t-1} + \mu \\ Y_t^f(h) &= \mu (h+1) + Y_{t-1} + \left[\left(I - F^{h+1}\right)(I - F)^{-1} \xi_t\right]_1 \\ Y_t^f(h) &= \mu (h+1) + Y_{t-1} + \left[(I - F)^{-1} \xi_t\right]_1 - \left[F^{h+1}(I - F)^{-1} \xi_t\right]_1 \end{aligned}$$

But

$$\begin{aligned} \left[(I - F)^{-1} \xi_t\right]_1 &= \sum_{j=0}^{p-1} a_{j+1} Z_{t-j} = \sum_{j=0}^{p-1} a_{j+1} W_{t-j} - \mu \sum_{j=0}^{p-1} a_{j+1} \\ &= \sum_{j=0}^{p-1} a_{j+1} Y_{t-j} - \sum_{j=0}^{p-1} a_{j+1} Y_{t-j-1} - \mu \sum_{j=0}^{p-1} a_{j+1} \end{aligned}$$

where $a_1, ..., a_p$ represents the components of the first row of the invertible $p \times p$ matrix $(I - F)^{-1}$. We are interested in the case in which the drift is zero, thus, $\delta = \mu = 0$. In this case, the MSPE is given by

$$
\begin{aligned}
\mathsf{MSPE}(h) &= \mathbb{E}(Y_{t+h} - Y_t^f(h))^2 \\
\mathsf{MSPE}(h) &= \mathbb{E}(Y_{t+h} - Y_{t-1} - \left[(I-F)^{-1}\xi_t\right]_1 + \left[F^{h+1}(I-F)^{-1}\xi_t\right]_1)^2 \\
\mathsf{MSPE}(h) &= \mathbb{V}(Y_{t+h} - Y_{t-1} - \left[(I-F)^{-1}\xi_t\right]_1 + \left[F^{h+1}(I-F)^{-1}\xi_t\right]_1) \\
\mathsf{MSPE}(h) &= \mathbb{V}(Y_{t+h} - Y_t) + \mathbb{V}(\left[(I-F)^{-1}\xi_t\right]_1) + \mathbb{V}(\left[F^{h+1}(I-F)^{-1}\xi_t\right]_1) - 2\mathbb{C}(Y_{t+h} - Y_t, \left[(I-F)^{-1}\xi_t\right]_1) \\
&\quad +2\mathbb{C}(Y_{t+h} - Y_t, \left[F^{h+1}(I-F)^{-1}\xi_t\right]_1) - 2\mathbb{C}(\left[(I-F)^{-1}\xi_t\right]_1, \left[F^{h+1}(I-F)^{-1}\xi_t\right]_1) \\
\mathsf{MSPE}(h) &= \mathbb{V}(Y_{t+h}) + \mathbb{V}(Y_t) - 2\gamma_h + \mathbb{V}(\left[(I-F)^{-1}\xi_t\right]_1) + \mathbb{V}(\left[F^{h+1}(I-F)^{-1}\xi_t\right]_1) \\
&\quad -2\mathbb{C}(Y_{t+h} - Y_t, \left[(I-F)^{-1}\xi_t\right]_1) + 2\mathbb{C}(Y_{t+h} - Y_t, \left[F^{h+1}(I-F)^{-1}\xi_t\right]_1) \\
&\quad -2\mathbb{C}(\left[(I-F)^{-1}\xi_t\right]_1, \left[F^{h+1}(I-F)^{-1}\xi_t\right]_1)
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
|\mathsf{MSPE}(h)| \leq\ & 2\mathbb{V}(Y_t) + 2|\gamma_h| + \mathbb{V}(\left[(I-F)^{-1}\xi_t\right]_1) + \mathbb{V}(\left[F^{h+1}(I-F)^{-1}\xi_t\right]_1) + \\
& +2\left[\sqrt{\mathbb{V}(Y_{t+h} - Y_t)}\sqrt{\mathbb{V}([(I-F)^{-1}\xi_t]_1)}\right] + 2\left[\sqrt{\mathbb{V}(Y_{t+h} - Y_t)}\sqrt{\mathbb{V}([F^{h+1}(I-F)^{-1}\xi_t]_1)}\right] + \\
& +2\left[\sqrt{\mathbb{V}([(I-F)^{-1}\xi_t]_1)}\sqrt{\mathbb{V}([F^{h+1}(I-F)^{-1}\xi_t]_1)}\right]
\end{aligned}
$$

hence,

$$
\begin{aligned}
|\mathsf{MSPE}(h)| \leq\ & 2\mathbb{V}(Y_t) + 2\sqrt{\mathbb{V}(Y_{t+h})\mathbb{V}(Y_t)} + \mathbb{V}(\left[(I-F)^{-1}\xi_t\right]_1) + \mathbb{V}(\left[F^{h+1}(I-F)^{-1}\xi_t\right]_1) + \\
& +2\left[\sqrt{2\mathbb{V}(Y_t) + 2\sqrt{\mathbb{V}(Y_{t+h})\mathbb{V}(Y_t)}}\sqrt{\mathbb{V}([(I-F)^{-1}\xi_t]_1)}\right] + \\
& +2\left[\sqrt{2\mathbb{V}(Y_t) + 2\sqrt{\mathbb{V}(Y_{t+h})\mathbb{V}(Y_t)}}\sqrt{\mathbb{V}([F^{h+1}(I-F)^{-1}\xi_t]_1)}\right] + \\
& +2\left[\sqrt{\mathbb{V}([(I-F)^{-1}\xi_t]_1)}\sqrt{\mathbb{V}([F^{h+1}(I-F)^{-1}\xi_t]_1)}\right]
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\mathsf{MSPE}(h) \leq\ & 4\mathbb{V}(Y_t) + \mathbb{V}(\left[(I-F)^{-1}\xi_t\right]_1) + \mathbb{V}(\left[F^{h+1}(I-F)^{-1}\xi_t\right]_1) + \\
& +4\left[\sqrt{\mathbb{V}(Y_t)}\sqrt{\mathbb{V}([(I-F)^{-1}\xi_t]_1)}\right] + 4\left[\sqrt{\mathbb{V}(Y_t)}\sqrt{\mathbb{V}([F^{h+1}(I-F)^{-1}\xi_t]_1)}\right] \\
& +2\left[\sqrt{\mathbb{V}([(I-F)^{-1}\xi_t]_1)}\sqrt{\mathbb{V}([F^{h+1}(I-F)^{-1}\xi_t]_1)}\right]
\end{aligned}
$$

and given that the AR(p) operator is stationary, all the eigenvalues of the F matrix will have absolute value less than one. Furthermore, Hamilton (1994) uses the following Jordan decomposition for F:

$$F = MJM^{-1}$$

Here M is a $p \times p$ matrix and J has the following Jordan structure:

$$J = \begin{pmatrix} J_1 & 0 & 0 & ... & 0 \\ 0 & J_2 & 0 & ... & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & 0 & J_s \end{pmatrix}_{s \times s}$$

where

$$J_i = \begin{pmatrix} \lambda_i & 1 & 0 & ... & 0 \\ 0 & \lambda_i & 1 & ... & 0 \\ . & . & . & . & . \\ . & . & . & . & 1 \\ 0 & 0 & 0 & 0 & \lambda_i \end{pmatrix}_{s \times s}$$

and λ_i corresponds to one of the s different eigenvalues of F. From the Jordan decomposition follows that F^h satisfies

$$F^h = MJ^hM^{-1}$$

where

$$J^h = \begin{pmatrix} J_1^h & 0 & 0 & ... & 0 \\ 0 & J_2^h & 0 & ... & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & 0 & J_s^h \end{pmatrix}_{s \times s}$$

where $s < p$ denotes the number of linearly independent eigenvectors. Each one of the terms J_i^h has the following shape

$$J_i^h = \begin{pmatrix} \lambda_i^h & \begin{pmatrix} h \\ 1 \end{pmatrix} \lambda_i^{h-1} & \begin{pmatrix} h \\ 2 \end{pmatrix} \lambda_i^{h-2} & ... & \begin{pmatrix} h \\ n_i - 1 \end{pmatrix} \lambda_i^{h-n_i+1} \\ 0 & \lambda_i^h & \begin{pmatrix} h \\ 1 \end{pmatrix} \lambda_i^{h-1} & ... & \begin{pmatrix} h \\ n_i - 2 \end{pmatrix} \lambda_i^{h-n_i+2} \\ . & . & \lambda_i^h & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & 0 & \lambda_i^h \end{pmatrix}_{s \times s}$$

where

$$\begin{pmatrix} h \\ n \end{pmatrix} = \frac{h(h-1)(h-2)...(h-n+1)}{n(n-1)(n-2)...3 \cdot 2 \cdot 1} \tag{2}$$

and n_i is the multiplicity of the eigenvalue λ_i. We notice that the terms in (2) are polynomial in h of order n. None of these polynomial have the divergence speed required to beat the exponential speed of the the terms λ_i^{h-j}.Therefore, as the forecasting horizon lengthens:

$$\begin{aligned} \lim_{h\to\infty} J_i^h &= 0 \\ \lim_{h\to\infty} J^h &= \lim_{h\to\infty} F^h = 0 \end{aligned}$$

implying that

$$\lim_{h\to\infty} \mathbb{V}(\left[F^{h+1}(I-F)^{-1}\xi_t\right]_1) = 0$$

therefore

$$\lim_{h\to\infty} \mathsf{MSPE}(h) \leq 4\mathbb{V}(Y_t) + \mathbb{V}(\left[(I-F)^{-1}\xi_t\right]_1) + 4\left[\sqrt{\mathbb{V}(Y_t)}\sqrt{\mathbb{V}(\left[(I-F)^{-1}\xi_t\right]_1)}\right]$$

and the $\mathsf{MSPE}(h)$ is a bounded sequence.

C Proof of Proposition 2: ARIMA(p,1,q)

Let us consider $\{X_{t+1}\}_{t=\infty}^{\infty}$ following an ARIMA(p,d,q) process, where $W_{t+1} = (1-B)^d X_{t+1}$ represents a stationary and invertible ARMA(p,q) process, with B a backshift operator ($B^j Z_t = Z_{t-j}$). Thus,

$$\begin{aligned} W_t - \sum_{j=1}^{p} \phi_j W_{t-j} &= \delta + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \\ \phi(B)(1-B)^d X_{t+1} &= \delta + \theta(B)\varepsilon_{t+1} \end{aligned}$$

Following Box, Jenkins, and Reinsel (2008), let us define the generalized autoregressive operator as

$$\begin{aligned} Q(B) &\equiv \phi(B)(1-B)^d = (1-\phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1-B)^d \\ &= (1-\varphi_1 B - \varphi_2 B^2 - \dots - \varphi_{p+d} B^{p+d}) \end{aligned}$$

Hence, we could write the process X_{t+1} as follows:

$$Q(B)X_t = \delta + \theta(B)\varepsilon_t$$

For every single forecasting horizon h, the optimal forecast satisfies

$$X_t^f(h) = \left\{ \begin{array}{c} \sum_{i=1}^{p+d} \varphi_i X_t^f(h-i) + \delta - \sum_{i=l}^{q} \theta_i \varepsilon_{t+h-i} \quad if \quad h \leq q \\ \sum_{i=1}^{p+d} \varphi_i X_t^f(h-i) + \delta \quad if \quad h > q \end{array} \right\} \tag{3}$$

The general solution for the homogeneous difference equation (3) when $h > q$ is given by

$$X_t^f(h) = \sum_{i=1}^{p} c_i(t) m_i^h + (b_0(t) + b_1(t)h + b_2(t)h^2 + \dots + b_{d-1}(t)h^{d-1})$$

where $c(t)$ and $b(t)$ represents adaptative coefficients, that is, coefficients that are stochastic and functions of the process at time t, and the terms m_i corresponds to the roots of the following expression

$$m^p - \phi_1 m^{p-1} - \phi_2 m^{p-2} - \dots - \phi_p = 0$$

Expression (3) is not homogeneous, so we need to add a particular solution, which is given by

$$b_d l^d$$

where b_d is a deterministic coefficient.

Thus, the eventual or explicit forecast function is given by

$$X_t^f(h) = \sum_{i=1}^{p} c_i(t) m_i^h + (b_0(t) + b_1(t)h + b_2(t)h^2 + ... + b_{d-1}(t)h^{d-1}) + \mathbb{I}_{(\delta \neq 0)} b_d h^d; \quad h > q - p - d \quad (4)$$

The previous expression characterizes long forecast horizons from ARIMA(p,d,q) models. It is interesting that the moving average terms only play a role in the determination of the adaptative coefficients. Besides, stationary roots of the autoregressive operator will vanish while the forecasting horizon lengthens as they have an absolute value less than one. Finally, the influence of unit roots determines the presence of a polynomial of order d in the forecast horizon, in which some of the coefficients are adaptative. If the econometrician mistakenly considers that the Y_t process follows an ARIMA(p,1,d) then he or she will compute the forecasts according to (4). When $\delta = 0$, and for large values of h we will have that the MSPE is given by

$$\begin{aligned}
\mathsf{MSPE}(h) &= \mathbb{E}(Y_{t+h} - Y_t^f(h))^2 \\
\mathsf{MSPE}(h) &= \mathbb{E}(Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h - b_0(t))^2 \\
\mathsf{MSPE}(h) &= \mathbb{V}(Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h - b_0(t)) + \left[\mathbb{E}(Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h - b_0(t))\right]^2 \\
\mathsf{MSPE}(h) &= \mathbb{V}(Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h) + \mathbb{V}(b_0(t)) + 2\mathbb{C}\left(Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h, b_0(t)\right) + \\
&\quad + \left[\mathbb{E}\left[Y_t\right] - \sum_{i=1}^{p} m_i^h \mathbb{E}\left[c_i(t)\right] - \mathbb{E}\left[b_0(t)\right]\right]^2
\end{aligned}$$

We notice that

$$\lim_{h \to \infty} \left[\mathbb{E}\left[Y_t\right] - \sum_{i=1}^{p} m_i^h \mathbb{E}\left[c_i(t)\right] - \mathbb{E}\left[b_0(t)\right]\right]^2 + \mathbb{V}(b_0(t)) = \left[\mathbb{E}\left[Y_t\right] - \mathbb{E}\left[b_0(t)\right]\right]^2 + \mathbb{V}(b_0(t))$$

Therefore we will place attention on the following terms

$$\mathbb{V}(Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h) + 2\mathbb{C}\left(Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h, b_0(t)\right)$$

First notice that

$$\begin{aligned}
\mathbb{V}(Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h) &= \mathbb{V}(Y_{t+h}) + \mathbb{V}\left[\sum_{i=1}^{p} c_i(t) m_i^h\right] - 2\mathbb{C}\left(Y_{t+h}, \sum_{i=1}^{p} c_i(t) m_i^h\right) \\
&= \mathbb{V}(Y_t) + \mathbb{V}\left[\sum_{i=1}^{p} c_i(t) m_i^h\right] - 2\sum_{i=1}^{p} \mathbb{C}\left(Y_{t+h}, c_i(t) m_i^h\right) \\
&\leq \mathbb{V}(Y_t) + \sum_{i=1}^{p} m_i^{2h} \mathbb{V}\left[c_i(t)\right] + 2\sum_{i=1}^{p} \sum_{j \neq i}^{p} \mathbb{C}(c_i(t) m_i^h, c_j(t) m_j^h) - 2\sum_{i=1}^{p} \mathbb{C}\left(Y_{t+h}, c_i(t) m_i^h\right)
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbb{V}(Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h) &\leq \mathbb{V}(Y_t) + \sum_{i=1}^{p} |m_i^{2h}| \mathbb{V}\left[c_i(t)\right] + \sum_{i=1}^{p} \sum_{j<i}^{p} \sqrt{\mathbb{V}(c_i(t) m_i^h) \mathbb{V}(c_j(t) m_j^h)} + \\
&\quad + 2 \sum_{i=1}^{p} |\mathbb{C}\left(Y_{t+h}, c_i(t) m_i^h\right)| \\
&\leq \mathbb{V}(Y_t) + \sum_{i=1}^{p} |m_i^{2h}| \mathbb{V}\left[c_i(t)\right] + \sum_{i=1}^{p} \sum_{j<i}^{p} \sqrt{\mathbb{V}(c_i(t) m_i^h) \mathbb{V}(c_j(t) m_j^h)} + \\
&\quad + 2 \sum_{i=1}^{p} \sqrt{\mathbb{V}(c_i(t) m_i^h) \mathbb{V}(Y_{t+h})} \\
&= \mathbb{V}(Y_t) + \sum_{i=1}^{p} |m_i^{2h}| \mathbb{V}\left[c_i(t)\right] + \sum_{i=1}^{p} \sum_{j<i}^{p} |m_i^h| |m_j^h| \sqrt{\mathbb{V}(c_i(t)) \mathbb{V}(c_j(t))} + \\
&\quad + 2 \sum_{i=1}^{p} |m_i^h| \sqrt{\mathbb{V}(c_i(t)) \mathbb{V}(Y_t)}
\end{aligned}$$

so

$$\lim_{h \to \infty} \mathbb{V}(Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h) \leq \mathbb{V}(Y_t)$$

provided that

$$\lim_{h \to \infty} |m_i^h| = 0$$

Now,

$$\begin{aligned}
|\mathbb{C}\left(Y_{t+h} - \sum_{i=1}^{p} c_i(t) m_i^h, b_0(t)\right)| &= |\mathbb{C}\left(Y_{t+h}, b_0(t)\right) - \mathbb{C}\left(\sum_{i=1}^{p} c_i(t) m_i^h, b_0(t)\right)| \\
&= |\mathbb{C}\left(Y_{t+h}, b_0(t)\right) - \sum_{i=1}^{p} \mathbb{C}\left(c_i(t) m_i^h, b_0(t)\right)| \\
&\leq |\mathbb{C}\left(Y_{t+h}, b_0(t)\right)| + \sum_{i=1}^{p} |\mathbb{C}\left(c_i(t) m_i^h, b_0(t)\right)| \\
&\leq \sqrt{\mathbb{V}\left[Y_{t+h}\right] \mathbb{V}\left[b_0(t)\right]} + \sum_{i=1}^{p} \sqrt{\mathbb{V}\left[c_i(t) m_i^h\right] \mathbb{V}\left[b_0(t)\right]} \\
&= \sqrt{\mathbb{V}\left[Y_t\right] \mathbb{V}\left[b_0(t)\right]} + \sum_{i=1}^{p} |m_i^h| \sqrt{\mathbb{V}\left[c_i(t)\right] \mathbb{V}\left[b_0(t)\right]} \underset{h \to \infty}{\to} \sqrt{\mathbb{V}\left[Y_t\right] \mathbb{V}\left[b_0(t)\right]}
\end{aligned}$$

Finally,

$$\lim_{h \to \infty} \mathsf{MSPE}(h) \leq \mathbb{V}(Y_t) + \sqrt{\mathbb{V}\left[Y_t\right] \mathbb{V}\left[b_0(t)\right]} + \left[\mathbb{E}\left[Y_t\right] - \mathbb{E}\left[b_0(t)\right]\right]^2 + \mathbb{V}(b_0(t))$$

implying that the $\mathsf{MSPE}(h)$ is a bounded sequence.

D Descriptive statistics of the series

Descriptive statistics –three samples

	Larger estimation sample				Evaluation sample				Full sample			
	Sep-1990 – Jan-1999				Feb-1999 – Dec-2011				Sep-1990 – Dec-2011			
	Mean	St. Dev.	Max.	Min.	Mean	St. Dev.	Max.	Min.	Mean	St. Dev.	Max.	Min.
Canada	2.00	1.70	6.90	-0.20	2.10	0.90	4.70	-0.90	2.10	1.30	6.90	-0.90
Sweden	3.00	3.40	12.60	-1.20	1.50	1.20	4.40	-1.60	2.10	2.40	12.60	-1.60
United States	2.90	1.10	6.40	1.40	2.50	1.30	5.50	-2.00	2.70	1.20	6.40	-2.00

Source: Authors' computations.

Documentos de Trabajo **Banco Central de Chile**	**Working Papers** **Central Bank of Chile**
NÚMEROS ANTERIORES	PAST ISSUES

La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: www.bcentral.cl/esp/estpub/estudios/dtbc. Existe la posibilidad de solicitar una copia impresa con un costo de $500 si es dentro de Chile y US$12 si es para fuera de Chile. Las solicitudes se pueden hacer por fax: (56-2) 6702231 o a través de correo electrónico: bcch@bcentral.cl.

Working Papers in PDF format can be downloaded free of charge from: www.bcentral.cl/eng/stdpub/studies/workingpaper. Printed versions can be ordered individually for US$12 per copy (for orders inside Chile the charge is Ch$500.) Orders can be placed by fax: (56-2) 6702231 or e-mail: bcch@bcentral.cl.

DTBC – 668 Junio 2012
On the International Transmission of Shocks: Micro Evidence From Mutual Fund Portfolios
Claudio Raddatz y Sergio L. Schmukler

DTBC – 667 Mayo 2012
Heterogeneous Inflation Expectations, Learning and Market Outcomes
Carlos Madeira y Basit Zafar

DTBC – 666 Marzo 2012
Financial Development, Exporting and Firm Heterogeneity in Chile
Roberto Alvarez y Ricardo López

DTBC – 665 Marzo 2012
Determinantes e Impacto de Episodios de Reversión Abrupta de Flujos de Capitales: ¿Es Distinto un *Sudden Stop* de un *Sudden Flight*?
Gabriela Contreras, Alfredo Pistelli y Mariel Siravegna

DTBC – 664 Febrero 2012
Rational Inattention, Multi-Product Firms and the Neutrality of Money
Ernesto Pastén

DTBC – 663 Febrero 2012
Non – Ricardian Aspects of the Fiscal Policy in Chile
Luis Felipe Céspedes, Jorge Fornero y Jordi Galí

DTBC – 662 Febrero 2012
Cubrir o no Cubrir: ¿Ese es el Dilema?
Rodrigo Alfaro y Natán Goldberger

DTBC – 661 Febrero 2012
Are Forecast Combinations Efficient?
Pablo Pincheira

DTBC – 660 Enero 2012
Combinación de Proyecciones para el Precio del Petróleo: Aplicación y Evaluación de Metodologías.
Ercio Muñoz, Miguel Ricaurte y Mariel Siravegna

DTBC – 659 Enero 2012
Dinámica Laboral en Chile
Macarena García y Alberto Naudon

DTBC – 658 Enero 2012
¿Akaike o Schwarz? ¿Cuál elegir para predecir el PIB Chileno?
Carlos Medel

DTBC – 657 Enero 2012
How Informative are In – Sample Information Criteria to Forecasting? The Case of Chilean GDP
Carlos Medel

DTBC – 656 Enero 2012
Indicadores Sintéticos para la Proyección de Imacec en Chile
Gonzalo Calvo y Miguel Ricaurte

DTBC – 655 Enero 2012
Government Spending and the Real Exchange Rate: a Cross – Country Perspective
Rodrigo Caputo y Miguel Fuentes

BANCO CENTRAL
DE CHILE