

Penalised regressions vs. autoregressive moving average models for forecasting inflation

Regresiones penalizadas vs. modelos autorregresivos de media móvil para pronosticar la inflación



DOI: <https://doi.org/10.17981/econcuc.41.1.2020.Econ.3>

Artículo de investigación.
Fecha de recepción: 07/10/2019.
Fecha de aceptación: 10/11/2019.
Fecha de publicación: 15/11/2019

Javier Humberto Ospina-Holguín 
Universidad del Valle. Cali (Colombia)
javier.ospina@correounivalle.edu.co

Ana Milena Padilla-Ospina 
Universidad del Valle. Cali (Colombia)
ana.milena.padilla@correounivalle.edu.co

Para citar este artículo:

Ospina-Holguín, J. y Padilla-Ospina, A. (2020). Penalised regressions vs. autoregressive moving average models for forecasting inflation. *Económicas CUC*, 41(1). 65-80. DOI: <https://doi.org/10.17981/econcuc.41.1.2020.Econ.3>

Abstract

This article relates the Seasonal Autoregressive Moving Average Models (SARMA) to linear regression. Based on this relationship, the paper shows that penalized linear models can outperform the out-of-sample forecast accuracy of the best SARMA models in forecasting inflation as a function of past values, due to penalization and cross-validation. The paper constructs a minimal functional example using edge regression to compare both competing approaches to forecasting monthly inflation in 35 selected countries of the Organization for Economic Cooperation and Development and in three groups of countries. The results empirically test the hypothesis that penalized linear regression, and edge regression in particular, can outperform the best standard SARMA models calculated through a grid search when forecasting inflation. Thus, a new and effective technique for forecasting inflation based on past values is provided for use by financial analysts and investors. The results indicate that more attention should be paid to automatic learning techniques for forecasting inflation time series, even as basic as penalized linear regressions, because of their superior empirical performance.

Keywords: Ridge regression; penalised linear model; ARMA, SARMA; inflation forecasting

Resumen

Este artículo relaciona los Modelos Autorregresivos Estacionales de Media Móvil (SARMA) con la regresión lineal. Sobre la base de esta relación, el documento muestra que los modelos lineales penalizados pueden superar la precisión del pronóstico fuera de la muestra de los mejores modelos SARMA al pronosticar la inflación en función de valores pasados, debido a la penalización y a la validación cruzada. El artículo construye un ejemplo funcional mínimo utilizando la regresión de arista para comparar ambos enfoques que compiten al pronosticar la inflación mensual en 35 países seleccionados de la Organización para la Cooperación y el Desarrollo Económico y en tres grupos de países. Los resultados verifican empíricamente la hipótesis de que la regresión lineal penalizada, y la regresión de arista en particular, puede superar a los mejores modelos estándar SARMA calculados a través de una búsqueda de cuadrícula cuando se pronostica la inflación. Así, se proporciona una técnica nueva y efectiva para pronosticar la inflación basada en valores pasados para el uso de analistas financieros e inversores. Los resultados indican que se debe prestar más atención a las técnicas de aprendizaje automático para el pronóstico de series de tiempo de la inflación, incluso tan básicas como las regresiones lineales penalizadas, debido a su rendimiento empírico superior.

Palabras clave: Regresión de arista; modelo lineal penalizado; ARMA, SARMA; pronóstico de la inflación

JEL: C53, C52

INTRODUCTION

When forecasting inflation, central banks tend not to rely exclusively on time series methods because their primary goal is not only short-term forecasting but also long-term inflation control by means of monetary policy. Consequently, they frequently complement (or even avoid) time series methods, favouring instead structural models that are based on macroeconomic variables or alternatives such as indicator analysis (Quinn, Kenny, & Meyler, 1999). Meanwhile, financial market agents often require updated and frequent short-term forecasts without the need for insights into the underlying causes of inflation. For these agents, time series methods are a convenient and expedite option, given their relative simplicity. There are more than a dozen inflation forecasting methods (Faust & Wright, 2013; Gómez, Sánchez & Millán, 2019; Gil, Castellanos & González, 2019), and here provide one more, related to the recursive autoregression method and the direct method. An method based on the forecasting improvements that are achieved by standard machine learning techniques applied to penalised linear regressions, which relate to the more standard econometric technique of (seasonal) autoregressive moving average models.

This paper argues that a simple penalised linear model can surpass the forecasting performance of standard econometric models for time series, such as ARMA or SARMA models, when forecasting inflation based on past values. To illustrate this, compare a minimal example of a penalised linear regression with standard (seasonal) autorregresive moving average models (SARMA) for forecasting monthly inflation in 35 selected countries of the Organisation for Economic Co-operation and Development (OECD) and 3 groups of countries. In the minimal example, the out-of-sample predictive performance of a ridge regression with one-fold cross-validation is compared with that of the best SARMA model obtained through a grid search. The results support our argument.

Penalised linear regression is shown to be mathematically related to (S)ARMA modeling, yet empirically more effective for forecasting inflation. The increase in effectiveness seems to come from the penalisation procedure and the empirical regularisation hyperparameter tuning through cross-validation, something which (S)ARMA modeling lacks. In empirical tuning, an out-of-sample experiment inside the original sample is created, where one part of the data is fitted and the level of regularisation that leads to the best performance on the other part of the data is chosen (Mullainathan & Spiess, 2017).

The recent popularisation of machine learning has brought with it the introduction of several new algorithms for predicting time series, from general models such as supervised learning models applied to sliding windows of the data to more specific models such as long short-term memory (LSTM) neural networks. The basic idea behind supervised machine learning in the context of prediction lies in improving its performance by using training data during its construction. As expressed by Mullainathan and Spiess (2017) supervised machine learning seeks to predict well out of sample by taking a loss function $L(\hat{y}, y)$ as an input, where y is an independent variable and \hat{y} its prediction, to look for a function \hat{f} that has a low expected loss

function $E_{(y, x)} [L(\hat{f}(x), y)]$ on new data points of the same distribution, where x is the vector of predictors.

One of the supervised machine learning techniques that has not received so much attention for the prediction of time series are the penalised linear models. In these models, a penalty is added to the loss function in order to decrease the complexity of the model and achieve greater parsimony. The purpose is to attain greater generality in the model in a such way that it will better capture the signal or the implicit pattern in the data above the noise that the data itself contains. The “regularisation” that is performed by adding the penalty should achieve a better fit out of sample even while the fit within the sample deteriorates due to the very same penalty.

DEVELOPMENT AND ANALYSIS:

RELATION BETWEEN THE SARMA MODEL AND PENALISED LINEAR REGRESSION

SARMA models are among those preferred for modelling and forecasting stationary processes that are also seasonal. SARMA models can be expressed as higher order ARMA models (in turn, an ARMA model can be approximated by a higher order AR model). The AR part attempts to model the variable in question as the result of a linear combination of its own past values. The MA part models the error term as a linear combination of contemporaneous and lagging white noise terms. The ARMA model attempts to express the variable in question as a parsimonious combination of an AR part and an MA part.

Penalised regression models, on the other hand, attempt to model and predict any dependent variable as a linear combination of independent variables. But unlike using the least squares estimation of the standard linear regression, the estimation in penalised regressions adds a term in the minimisation that has the general effect of decreasing the values of the coefficients through a method called shrinkage or regularisation. The amount of ideal shrinkage is controlled in practice by cross-validation, that is, by maintaining part of the sample as a validation sample in which the generality of prediction achieved by the method in the remaining (training) sample is examined and corroborated. Penalised regression models are now part of the basic arsenal of machine learning.

The observation that a SARMA model can be approximated by a higher order AR model, and therefore by a linear regression of lagged terms, raises the question of whether the use of penalised regression can improve the predictive capacity of an original SARMA model. This is question address empirically in inflation data from several countries and groups of countries and leaves open the debate of which method is superior in other samples.

SARMA models

SARMA models, introduced by **Box and Jenkin (1976)**, are a generalisation of ARMA models that consider seasonality in the data generating process. Seasonality is defined as a periodic pattern in the time series. Used here the multiplicative

SARMA $SARMA(p, q) \times (P, Q)_s$, defined as a model for a stationary time series y_t which obeys (1):

$$\Phi(L^s)\phi(L)(y_t - \mu) = \Theta(L^s)\theta(L)\varepsilon_t \quad (1)$$

Where (2),

$$Ly_t = y_{t-1} \quad (2)$$

Is the lag operator and the AR, MA, seasonal AR and seasonal MA polynomials are defined as (3):

$$\begin{aligned} \text{AR} : \phi(L) &= 1 - \phi_1 L - \dots - \phi_p L^p, \\ \text{MA} : \theta(L) &= 1 - \theta_1 L + \dots + \theta_q L^q, \\ \text{Seasonal AR} : \Phi(L^s) &= 1 - \Phi_1 L^s - \dots - \Phi_p L^{sP}, \\ \text{Seasonal MA} : \Theta(L^s) &= 1 - \Theta_1 L^s + \dots + \Theta_q L^{sQ}. \end{aligned} \quad (3)$$

• *A SARMA model approximated as an AR model*

It should be noted that the $SARMA(p, q) \times (P, Q)_s$ can always be expressed as an $ARMA(p, q)$ model of the form (4):

$$\tilde{\phi}(L)(y_t - \mu) = \tilde{\theta}(L)\varepsilon_t \quad (4)$$

Where (5),

$$\tilde{\phi}(L) = \Phi(L^s)\phi(L) \text{ and } \tilde{\theta} = \Theta(L^s)\theta(L). \quad (5)$$

By the Wold representation theorem, if $y_t - \mu$ is covariance-stationary, $y_t - \mu$ can be expressed as (6):

$$y_t - \mu = \sum_{k=0}^{\infty} \pi_k \varepsilon_{t-k} + \eta_t, \quad (6)$$

Where η_t is a deterministic time series. In the context of the $ARMA(p, q)$ model of the form (7):

$$\tilde{\phi}(L)(y_t - \mu) = \tilde{\theta}(L)\varepsilon_t \quad (7)$$

Indeed (8),

$$(y_t - \mu) = \frac{\tilde{\theta}(L)}{\tilde{\phi}(L)} \varepsilon_t = \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}. \quad (8)$$

If $y_t - \mu$ is invertible, it can also introduced (9):

$$\pi(L) = 1 - \sum_{k=1}^{\infty} \pi_k L^k \quad (9)$$

In order to write (10):

$$\tilde{\phi}(L)(y_t - \mu) = \tilde{\theta}(L)\varepsilon_t = \tilde{\theta}(L)\pi(L)(y_t - \mu) \quad (10)$$

Or (11),

$$\pi(L)(y_t - \mu) = \frac{\tilde{\theta}(L)}{\tilde{\theta}(L)}\varepsilon_t = \varepsilon_t \quad (11)$$

In this way, $(y_t - \mu)$ can be expressed as an $AR(\infty)$ (12):

$$(y_t - \mu) = \sum_{k=1}^{\infty} \pi_k L^k (y_t - \mu) + \varepsilon_t. \quad (12)$$

This expression can in turn be truncated to yield an approximation of $(y_t - \mu)$ as an $AR(m)$ process (13):

$$(y_t - \mu) = \sum_{k=1}^m \pi_k L^k (y_t - \mu) + \varepsilon_t. \quad (13)$$

This proves that the $SARMA(p, q) \times (P, Q)_s$ can be approximated as an $AR(m)$ process which is no more than a linear regression of m lagged terms. In this sort of approximation, it would be convenient to have m at least as big as s to approximate at least a $SARMA(p, q) \times (1, 1)_s$ model.

Penalised linear regression

It has just been shown that $(y_t - \mu)$ can be expressed as a linear regression of m lagged values in approximation (14):

$$\begin{aligned} (y_t - \mu) &= \sum_{k=1}^m \pi_k (y_{t-k} - \mu) + \varepsilon_t, \\ y_t &= (1 - \sum_{k=1}^m \pi_k)\mu + \sum_{k=1}^m \pi_k y_{t-k} + \varepsilon_t, \\ y_t &= a + \sum_{k=1}^m \pi_k y_{t-k} + \varepsilon_t, \end{aligned} \quad (14)$$

Where $a = (1 - \sum_{k=1}^m \pi_k)\mu$. The representation of y_t as a linear regression motivates the use of machine learning methods to predict y_t , in particular the use of a penalised linear regression using regularisation and cross-validation. In this paper, use Tikhonov regularisation based on the L_2 -norm.

- *Tikhonov regularisation*

Tikhonov regularisation (Hoerl & Kennard, 1970; Tikhonov & Arsenin, 1977) consists in adding a new penalty term to the standard quadratic loss function in the least squares minimisation problem. So, instead of minimising just the sum of squares of the residuals (15):

$$\min_{\tilde{\pi}} \mathbf{e}'\mathbf{e} = \min_{\tilde{\pi}} (\mathbf{y} - \mathbf{X}\tilde{\pi})' (\mathbf{y} - \mathbf{X}\tilde{\pi}) \quad (15)$$

In order to find the vector $\tilde{\pi}$ of estimated coefficients of $\{\pi_k\}_{k=1}^m$, the following minimisation is performed (16):

$$\begin{aligned} & \min_{\tilde{\pi}} \mathbf{e}'\mathbf{e} + \lambda \|\tilde{\pi}\| \\ & = \min_{\tilde{\pi}} (\mathbf{y} - \mathbf{X}\tilde{\pi})' (\mathbf{y} - \mathbf{X}\tilde{\pi}) + \lambda \|\tilde{\pi}\|, \end{aligned} \quad (16)$$

Where λ is a non-negative hyperparameter, $\|\cdot\|$ is the Euclidean or L_2 -norm defined as $\|\tilde{\pi}\| = \sqrt{\pi_1^2 + \dots + \pi_m^2}$, \mathbf{e} is a vector of estimated residuals of the linear regression, \mathbf{y} is a vector of y_t values in time, and \mathbf{x} is the matrix of predictors made of the column vectors $(\mathbf{1} \ y_{-1} \ \dots \ y_{-m})$, where $\mathbf{1}$ is a vector of ones, and y_{-k} is a vector of lagged values y_{t-k} .

This kind of regression is also referred to as ridge regression (Hoerl & Kennard, 1970; Anzola, Vargas & Morales, 2019). Its purpose is to shrink the size of the coefficients to avoid their being excessively large. This tends to deteriorate the in-sample performance for the sake of a better out-of-sample performance, improving the fit to the signal in the data, instead of its noise.

Other more general regularisation schemes exist, such as elastic net (Zou & Hastie, 2005), but focus on a minimal working example. In the elastic net, the minimisation involves two simultaneous penalty terms (17):

$$\min_{\tilde{\pi}} \mathbf{e}'\mathbf{e} + \lambda(1-\rho)\|\tilde{\pi}\| + \frac{1}{2}\lambda\rho\|\tilde{\pi}\|, \quad (17)$$

Where $|\cdot|$ is the L_1 -norm defined as $|\tilde{\pi}| = |\pi_1| + \dots + |\pi_m|$. At the same time, the elastic net regularisation contains as a special case the L_1 -regularised linear regression or LASSO (Santosa & Symes, 1986; Tibshirani, 1996) when $\rho = 0$ which entails a variable selection method, since it attempts to make one or more of the π_k equal to zero.

- *Cross-validation in the ridge regression*

This present paper restricts ourselves to one-fold cross-validation for the choice of the right λ parameter in the ridge regression. The data to estimate the linear model is divided into two subsets: a training subset of data with time indices

$\tau_1 = \{1, \dots, t_1\}$ and a validation subset of data with time indices $\tau_2 = \{t_2 = t_1 + 1, \dots, t_3 - 1\}$. The least squares minimisation problem is progressively done in the training subset with time indices τ_1 with a sequence of values of λ . The performance in terms of each value in the minimisation problem is evaluated both in τ_1 and in τ_2 . The potential values of λ to be explored are computed using Bayesian optimisation (Mockus, 1989), a kind of optimisation that attempts to minimise the number of evaluations while preserving the breadth of the exploration of the range of possible parameters. It is expected that at first, the minimum found decreases both in τ_1 and in τ_2 . But soon enough, the minimisation progresses only in τ_1 , but not out of sample (in τ_2), where the performance deteriorates. At that step, the search for a better λ value is stopped. In effect, the cross-validation mechanism validates that the minimisation problem in τ_1 is generalisable (to the validation sample τ_2). This avoids overfitting the parameters and selection bias, since the performance of the vector $\tilde{\pi}$ for predicting the estimated \hat{y}_t is evaluated in data that was not used to construct the model, just as in real life.

METHODOLOGY

The inflation rates of 35 OECD countries and three groups of countries (OECD, 2019) were selected for the out-of-sample performance comparison of the SARMA and ridge regression predictions. These are all the inflation rates of the OECD sample which exhibited no seasonal integration, according to the OCSB test (Osborn, Chui, Smith, & Birchenhall, 2009), allowing us to apply a SARMA model more appropriately since a SARIMA model would be more suitable in the case of seasonal integration. The inflation rates were all stationary according to standard tests such as the ADF test (Dickey & Fuller, 1979; MacKinnon, 1996). Since want to include an annual seasonality, at least $s = 12$ and include at least 12 lagged predictors in the ridge regression.

In order to evaluate completely out-of-sample the performance of the ridge regression, the model is used to generate one step ahead forecasts. A realistic forecaster does not use the forecasting model once for the sample $\tau_1 \cup \tau_2$ as a whole in order to predict a new value for the next period, but every new period he expands τ_1 with the next value in the time sequence and τ_2 is moved one new value to the right in time. This method of evaluating forecasts was reproduced with an expanding window of one-step forecasts with re-estimation. That is, every new period, the whole model is re-estimated in order to produce the one-step ahead forecast. This sequence of one-step ahead forecasts \hat{y}_t can be indexed with a new set $\tau_3 = \{t_3, \dots, T\}$, so that the true out-of-sample evaluation uses this set of indices τ_3 . The performance of the SARMA model is evaluated similarly, except that the whole sample $\tau_1 \cup \tau_2$ is used to generate the model expanding it one period to the right in time when forming every forecast, since no cross-validation is performed. The initial τ_1 sample starts in February of 2010 and ends in January 2015, while the final τ_1 sample starts in February of 2010 and ends in October 2016. The τ_2 always has a fixed length of 25 months so the initial τ_2 sample starts in February 2015 and ends in February 2016, and the final τ_2

sample starts in October 2016 and ends on November 2017. Finally, the τ_3 sample starts on March 2016 and ends on December 2017. One has to bear in mind that for the ridge regression the twelve past months are used for every one-step ahead forecast, so, for example, for every cross-validation performed in the τ_2 sample 13 ridge regression forecasts are computed.

The predictive performance of both models in all countries and group of countries is assessed via the overall out-of-sample (18) (Gu, Kelly, & Xiu, 2018):

$$R_{\text{oos}}^2 = 1 - \frac{\sum_{(i,t \in \tau_3)} (y_{i,t+1} - \hat{y}_{i,t+1})^2}{\sum_{(i,t \in \tau_3)} (y_{i,t+1} - \bar{y}_{i,t+1})^2}, \quad (18)$$

Where i denotes the country/group of countries and $\bar{y}_{i,t+1} = \frac{1}{2} \sum y_{i,t}$ is the naive forecast made of the average inflation so far for each period since the beginning of τ_1 .

Also calculated the out-of-sample R^2 for each country or group of countries (19) (Kvalseth, 1985):

$$R_{\text{oos},i}^2 = 1 - \frac{\sum_{t \in \tau_3} (y_{i,t+1} - \hat{y}_{i,t+1})^2}{\sum_{t \in \tau_3} (y_{i,t+1} - \bar{y}_{i,t+1})^2}. \quad (19)$$

In order to compare each predictive model, used the Diebold and Mariano (2002) test for differences in out-of-sample predictive accuracy between two models. The sample of inflation rates should have strong cross-sectional correlations due to the global macroeconomic factors influencing the OECD countries, so one of the assumptions of weak dependence of the test is violated. Thus, used the version of the test of Gu et al. (2018) to compare two methods of forecasting where the test statistic is $DM_{12} = \bar{d}_{12} / \hat{\sigma}_{12}$, where (20):

$$d_{12,t+1} = \frac{1}{N} \sum_{i=1}^N \left((\hat{e}_{i,t+1}^{(1)})^2 - (\hat{e}_{i,t+1}^{(2)})^2 \right) \quad (20)$$

$\hat{e}_{i,t+1}^{(1)}$ and $\hat{e}_{i,t+1}^{(2)}$ are the prediction error for the inflation of country/group of countries i at time t using each method, the first (1) or the second (2), N is the number of countries/groups of countries, and \bar{d}_{12} and $\hat{\sigma}_{12}$ are the mean and Newey-West standard errors of $d_{12,t}$ over the τ_3 sample. According to (Gu et al., 2018), due to the low potential autocorrelation in the $d_{12,t+1}$ time series, the asymptotic normality of the statistic is more likely to be guaranteed, which in turn allows for appropriate p-values in the comparisons.

In order to select the SARMA model, a grid search was conducted to select the best model. The grid search looked for the SARMA model with minimal Akaike

information criterion with a correction for small samples (AICc) (Burnham & Anderson, 2004) among all SARMA models with $p \in \{0, 1, 2, 3, 4, 5\}$, $q \in \{0, 1, 2, 3, 4, 5\}$, $P \in \{0, 1, 2\}$ and $Q \in \{0, 1, 2\}$ and $p + q + P + Q \leq 5$. It is worth noting that, asymptotically, optimising the AIC criterion is equivalent to minimising the out-of-sample one-step forecast mean square error MSE (Hyndman, 2013). The grid search is supposed to be more comprehensive than popular stepwise procedures such the Hyndman & Khandakar (2008) selection procedure of the *R auto.arima* function of the *forecast* package, for example.

Two common error measures in the sample countries were also computed. These measures were the Mean Absolute Error (MAE) (21) and Root Mean Square Error (RMSE) (22), defined as:

$$\text{MAE}_i = \frac{1}{n(\tau_3)} \sum_{t \in \tau_3} |y_{i,t+1} - \hat{y}_{i,t+1}| \quad (21)$$

$$\text{RMSE}_i = \sqrt{\frac{1}{n(\tau_3)} \sum_{t \in \tau_3} (y_{i,t+1} - \hat{y}_{i,t+1})^2} \quad (22)$$

Where $n(\tau_3)$ is the cardinality of τ_3 . The mean absolute percentage error (MAPE) was not computable because the monthly inflation rate is often zero in different months for several countries.

RESULTS

The overall out-of-sample R^2 was greater for the forecasts based on the ridge regression (19.9%) than for the forecasts based on the best SARMA models (10.6%). Furthermore, in 75.8% of the cases that the out-of-sample R^2_{OOS} for each country or group of countries was positive for the ridge regression forecasts, the R^2_{OOS} was better than in the forecasts by the best SARMA models (Table 1). The value of DM12 was 2.73. Method 1 was the best SARMA model and method 2 was the ridge regression. This rejects the one-sided null hypothesis that the SARMA model has better out-of-sample predictive accuracy than the ridge regression, with $p = 0.00311$. Although the behaviour of the Diebold-Mariano test is oversized in small samples, the Appendix performs a Monte Carlo analysis which shows that $p < 0.02$ can even be guaranteed in an oversized scenario. Table 2 and Table 3 show the mean absolute error and the root mean squared error per country or group of countries for both forecasting methods. The mean absolute error of the ridge regression forecasts was better (lower) in 57.9% of the countries/group of countries examined, while the root mean squared error was better in 63.2% of the countries/group of countries. The results clearly show the superiority of the ridge regression in out-of-sample forecasting terms in the illustrative case chosen.

TABLE 1.
 $R^2_{00s, i}$ by country/group of countries i for the best SARMA model and ridge regression

Country	SARMA	Ridge
Belgium		•
Canada	11%	13%•
Czech Republic	•	
Denmark	39%•	
Finland	8%•	
Hungary	•	
Iceland	27%•	
Ireland	40%•	
Italy	•	
Japan	22%•	
South Korea	24%	24%•
Mexico	1%	20%•
Norway	5%•	
Poland	12%•	
Slovakia	•	
Sweden	27%•	9%
Switzerland	31%•	29%
Turkey	27%•	14%
United Kingdom	0%	23%•
United States	18%•	
Brazil	21%•	13%
Chile	•	
China	0%	31%•
Colombia	61%•	42%
Costa Rica	•	
Estonia	11%•	
India	3%	14%•
Indonesia	8%	20%•
Israel	5%•	
Latvia	12%	34%•
Lithuania	13%	45%•
Russia	57%•	39%
Saudi Arabia	43%•	20%
Slovenia	26%	49%•
South Africa	7%	26%•
OECD	3%•	
G-7	7%•	
G-20	0%•	

Note: Missing numbers denote negative out-of-sample coefficients of determination. The symbol (•) denotes the better method (the one with higher coefficient of determination) for that particular country or group of countries.

TABLE 2.

Mean absolute error of the out-of-sample forecasts in τ_3 by country/group of countries i for the best SARMA model and ridge regression.

Country	SARMA	Ridge
Belgium	0.0020	0.0018•
Canada	0.0018•	0.0020
Czech Republic	0.0024•	0.0024
Denmark	0.0025	0.0017•
Finland	0.0019	0.0019•
Hungary	0.0023	0.0022•
Iceland	0.0025	0.0023•
Ireland	0.0034	0.0025•
Italy	0.0021•	0.0023
Japan	0.0019	0.0014•
South Korea	0.0023	0.0021•
Mexico	0.0032	0.0026•
Norway	0.0032	0.0030•
Poland	0.0021•	0.0023
Slovakia	0.0017•	0.0017
Sweden	0.0022•	0.0023
Switzerland	0.0015•	0.0016
Turkey	0.0050•	0.0061
United Kingdom	0.0017	0.0016•
United States	0.0020	0.0016•
Brazil	0.0022•	0.0024
Chile	0.0019•	0.0019
China	0.0038	0.0030•
Colombia	0.0019•	0.0025
Costa Rica	0.0032•	0.0034
Estonia	0.0032	0.0031•
India	0.0058	0.0057•
Indonesia	0.0027	0.0026•
Israel	0.0025•	0.0025
Latvia	0.0033	0.0030•
Lithuania	0.0038	0.0033•
Russia	0.0020•	0.0026
Saudi Arabia	0.0024•	0.0029
Slovenia	0.0035	0.0025•
South Africa	0.0028	0.0025•
OECD	0.0011•	0.0012
G-7	0.0013	0.0012•
G-20	0.0010	0.0009•

Note: The symbol (•) denotes the better method (the one with lower mean absolute error) for that particular country or group of countries.

TABLE 3.

Root mean squared error of the out-of-sample forecasts in τ_3 by country/group of countries i for the best SARMA model and ridge regression

Country	SARMA	Ridge
Belgium	0.0027	0.0026•
Canada	0.0026	0.0025•
Czech Republic	0.0029•	0.0029
Denmark	0.0034	0.0024•
Finland	0.0025	0.0023•
Hungary	0.0029	0.0028•
Iceland	0.0031	0.0027•
Ireland	0.0040	0.0030•
Italy	0.0024•	0.0026
Japan	0.0023	0.0019•
South Korea	0.0029	0.0029•
Mexico	0.0045	0.0040•
Norway	0.0041	0.0036•
Poland	0.0026•	0.0029
Slovakia	0.0020•	0.0021
Sweden	0.0027•	0.0030
Switzerland	0.0019•	0.0019
Turkey	0.0063•	0.0068
United Kingdom	0.0021	0.0019•
United States	0.0023	0.0020•
Brazil	0.0027•	0.0029
Chile	0.0025•	0.0026
China	0.0049	0.0041•
Colombia	0.0025•	0.0031
Costa Rica	0.0043•	0.0044
Estonia	0.0041	0.0036•
India	0.0073	0.0069•
Indonesia	0.0035	0.0033•
Israel	0.0032	0.0030•
Latvia	0.0041	0.0035•
Lithuania	0.0047	0.0038•
Russia	0.0029•	0.0034
Saudi Arabia	0.0029•	0.0034
Slovenia	0.0042	0.0035•
South Africa	0.0035	0.0031•
OECD	0.0014•	0.0014
G-7	0.0016	0.0015•
G-20	0.0012	0.0011•

Note: The symbol (•) denotes the better method (the one with lower root mean squared error) for that particular country or group of countries.

CONCLUSIONS

The hypothesis of this paper was that a SARMA model is mathematically related to a linear regression, so that using penalised linear regression for forecasting out-of-sample should surpass the forecast performance of the best SARMA models when forecasting inflation. The mathematical relation between both kinds of models was shown and a minimal working example based on ridge regression was built and applied to selected OECD countries and groups of countries as an empirical illustration of our hypothesis. The illustrative case decisively showed the better forecasting performance of the ridge regression for forecasting inflation, introducing a new forecasting method. Our work can also be seen as a new way of estimating SARMA models with machine learning methods, by first expressing the SARMA model as an AR model, and then thinking of it as a penalised regression which uses an optimization penalty in the least squares minimisation and cross-validation to fine-tune the penalty. Our results indicate that more attention should be given to machine learning techniques for time series forecasting of inflation, even as basic as penalised linear regressions, due to their superior empirical performance.

APPENDIX

Here replicate the Monte Carlo analysis that was developed by **Diebold and Mariano (2002)** for Gaussian forecast errors for the sample size of the forecast series $T = 22$. The original Diebold-Mariano DM test was found to be oversized for small sample sizes. Nevertheless, this simulation shows that in our results the Diebold-Mariano test between the best SARMA model forecast and the ridge regression is significant with $p < 0.002$, even in a bad oversized scenario, if Gaussian errors are assumed.

After simplification, the original Diebold-Mariano test statistics DM for one-step ahead forecasts is equal to (23):

$$DM = \frac{\overline{d_t}}{\sqrt{\frac{\hat{\gamma}_d(0)}{T}}}, \quad (23)$$

Where $d_t = (e_t^1)^2 - (e_t^2)^2$ is the difference between the squares of the residuals of the two kind of forecasts, 1 and 2, the bar represents the sample mean of the loss differential, and $\hat{\gamma}_d(0)$ is the sample autocovariance of the loss differential at displacement 0 (i.e. its sample variance).

For the Monte Carlo analysis, drawing realisations of $\{e_t^1, e_t^2\}_{t=1}^T$ following a bivariate forecast-error process with different degrees of contemporaneous and serial correlation in the forecast errors. First, generate $\{v_t^1, v_t^2\}_{t=1}^T$, with $v_t \sim N(\mathbf{0}, \mathbf{R})$, where (24):

$$\mathbf{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \quad (24)$$

Is the desired contemporaneous correlation matrix. Second, introduce MA(1) serial correlation with parameter θ as (25):

$$\begin{bmatrix} e_t^1 \\ e_t^2 \end{bmatrix} = \frac{1}{\sqrt{1+\theta^2}} \begin{bmatrix} 1+\theta L & 0 \\ 0 & 1+\theta L \end{bmatrix} \begin{bmatrix} v_t^1 \\ v_t^2 \end{bmatrix}, t=1, \dots, T, \quad (25)$$

Where the scalar normalises the unconditional variance to 1. Using $v_0^1 = v_0^2 = 0$, it can be shown that the correlation between v^1 and v^2 is the same as the correlation between e^1 and e^2 . Given this procedure, calculates the empirical size of the *DM* test for all combinations of $\rho = 0, 0.5$ and 0.9 , and also $\theta = 0, 0.5$ and 0.9 at the same level of the p value obtained in the Diebold-Mariano test between the best SARMA model forecast and the ridge regression (i.e. at the 0.311% level). This shows us the expected p values for the chosen sample size of $T = 22$ and the chosen level under different degrees of contemporaneous and serial correlation. Table 4 illustrates that a p value of at least under 0.02 is guaranteed in a bad oversized scenario for the the Diebold-Mariano test between the best SARMA model forecast and the ridge regression.

TABLE 4.

Empirical size of the DM statistics for a sample size of $T = 22$

	$\theta = 0$	$\theta = 0.5$	$\theta = 0.9$
$p = 0$	0.0046	0.01008	0.01388
$p = 0.5$	0.00454	0.00942	0.014
$p = 0.9$	0.0042	0.00914	0.0148

Note: ρ is the contemporaneous correlation between the innovations underlying the forecast errors, and θ is the parameter of the MA(1) forecast error. All of the tests are at the same level as the original Diebold-Mariano test between the best SARMA model forecast and the ridge regression (i.e. at the 0.311% level). A total of 50,000 Monte Carlo replications are performed.

REFERENCES

- Anzola, C., Vargas, P. & Morales, A. (2019). Transición entre sistemas financieros bancarios y bursátiles. Una aproximación mediante modelo de Swithing Markov. *Económicas CUC*, 40(1), 123–144. <https://doi.org/10.17981/econcuc.40.1.2019.08>
- Box, G. E. P. & Jenkin, G. M. (1976). *Time series analysis, forecasting and control*. San Francisco: Holden-Day.
- Burnham, K. P. & Anderson, D. R. (2004). Multimodel inference. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), 427–431. <https://doi.org/10.2307/2286348>

- Diebold, F. X. & Mariano, R. S. (2002). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 20(1), 134–144. <https://doi.org/10.1198/073500102753410444>
- Faust, J. & Wright, J. H. (2013). Forecasting Inflation. In, G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Vol. 2. Part. A (pp. 2–56). Amsterdam: Elsevier. <https://doi.org/10.1016/B978-0-444-53683-9.00001-3>
- Gil, J., Castellanos, D. & Gonzalez, D. (2019). Margen de intermediación y concentración bancaria en Colombia: un análisis para el periodo 2000-2017. *Económicas CUC*, 40(2), 9–30. <https://doi.org/10.17981/econcuc.40.2.2019.01>
- Gómez, C., Sánchez, V. & Millán, E. (2019). Capitalismo y ética: una relación de tensiones. *Económicas CUC*, 40(2), 31–42. <https://doi.org/10.17981/econcuc.40.2.2019.02>
- Gu, S., Kelly, B. T. & Xiu, D. (december, 2018). Empirical Asset Pricing Via Machine Learning [Paper 18-04]. *31st Australasian Finance and Banking Conference*, AFBC, Sydney, Australia, 1–79. <https://doi.org/10.2139/ssrn.3159577>
- Hoerl, A. E. & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.2307/1267351>
- Hyndman, R. J. (july, 2013). Facts and fallacies of the AIC. [Online]. Available from <https://robjhyndman.com/hyndsight/aic/>
- Hyndman, R. J. & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(1), 1–22. <https://doi.org/10.18637/jss.v027.i03>
- Kvalseth, T. O. (1985). Cautionary Note about R 2. *The American Statistician*, 39(4), 279–285. <https://doi.org/10.2307/2683704>
- MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, 11(6), 601–618. [https://doi.org/10.1002/\(SICI\)1099-1255\(199611\)11:6<601::AID-JAE417>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<601::AID-JAE417>3.0.CO;2-T)
- Mockus, J. (1989). *Bayesian approach to global optimization*. Dordrecht: Kluwer Academic Publishers.
- Mullainathan, S. & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 31(2), 87–106. <https://doi.org/10.1257/jep.31.2.87>
- OECD. (2019). Inflation (CPI). [indicator]. <https://doi.org/10.1787/eee82e6e-en>
- Osborn, D. R., Chui, A. P. L., Smith, J. P. & Birchenhall, C. R. (2009). Seasonality and the order of integration for consumption. *Oxford Bulletin of Economics and Statistics*, 50(4), 361–377. <https://doi.org/10.1111/j.1468-0084.1988.mp50004002.x>
- Quinn, T., Kenny, G. & Meyler, A. (1999). Inflation analysis: An overview. [MPRA Paper No. 11361]. Munich: UTC. Retrieved from https://mpra.ub.uni-muenchen.de/11361/1/MPRA_paper_11361.pdf
- Santosa, F. & Symes, W. W. (1986). Linear Inversion of Band-Limited Reflection Seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4), 1307–1330. <https://doi.org/10.1137/0907087>

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>

Tikhonov, A. N. & Arsenin, V. Y. (1977). *Solution of ill-posed problems*. Washington: Winston & Sons.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>

Javier Humberto Ospina-Holguín holds a PhD in Administration, a Master of Science in Organizations and a Bachelor of Science in Physics from Universidad del Valle, as well as a Master of Science in Economics from the University of Amsterdam. Javier is currently Associate Professor in the Department of Accounting and Finance at Universidad del Valle (Cali, Colombia). <https://orcid.org/0000-0002-0103-3280>

Ana Milena Padilla-Ospina holds a PhD in Administration, a Master of Science in Organizations and a Bachelor of Business Administration from Universidad del Valle. She is currently Assistant Professor in the Department of Administration and Organizations at Universidad del Valle (Cali, Colombia). <https://orcid.org/0000-0003-3859-8741>