DOCUMENTOS DE TRABAJO

Inflation Forecast in Chile with Machine Learning Methods

Felipe Leal
Carlos Molina
Eduardo Zilberman

N° 860 Enero 2020

BANCO CENTRAL DE 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° 860

Working Paper
N° 860

Inflation Forecast in Chile with Machine Learning Methods*

Felipe Leal
Central Bank of Chile

Carlos Molina
University of Chile

Eduardo Zilberman
Central Bank of Chile

Abstract
We apply Machine Learning (ML) methods with Big Data, similar to Medeiros et al. (2019) for the US, to forecast headline and core inflation of the CPI in Chile. We document that the ML methods do not consistently dominates in the inflation forecast for the Chilean case over simple and univariate linear competitors such as AR, the mean and median of past inflation, which have proven to be highly competitive. In fact, these are the best performing methods in many cases. A second contribution of this work is the construction of a large data set with macroeconomic variables related to the Chilean economy, similar to McCracken and Ng (2016), who built (and maintain) a similar data set for the US.

Resumen
En este trabajo, en línea con lo realizado en Medeiros et al. (2019) para EE.UU., aplicamos métodos de Machine Learning (ML) con Big Data para pronosticar la inflación general y subyacente del IPC en Chile. Documentamos que los métodos de ML no ganan en la proyección de inflación para el caso chileno de forma consistente sobre competidores lineales simples y univariados tales como el AR, la media y la mediana de la inflación pasada, que han demostrado ser altamente competitivos. De hecho, estos son los métodos ganadores en muchos casos. Una segunda contribución de este trabajo es la construcción de un gran conjunto de datos con variables macroeconómicas relacionadas con la economía chilena en línea con McCracken y Ng (2016), quienes construyeron (y mantienen) un conjunto similar de datos para Estados Unidos.

* We appreciate the comments of Marcelo Medeiros, Carlos Medel, Pablo Pincheira and Gabriel Vasconelos. The ideas and opinions presented in this work do not necessarily represent the views of the Central Bank of Chile or its authorities. Errors or omissions are the sole responsibility of the authors. Emails: fleal@bcentral.cl, cmolinas@fen.uchile.cl, ezilberman@bcentral.cl.

1 Introducción

Forecasting inflation is an important, yet difficult task. As Stock and Watson (2010) emphasize, "it is extremely difficult to systematically improve inflation forecasting with simple univariate forecasting models". There is ample evidence to support this view[1]. However, this skepticism came before the rise of Big Data and Machine Learning (ML) in economics and finance[2]. In a recent work, Medeiros et al. (2019) show that the combination of Big Data and ML methods can lead to more accurate inflation forecasts in the United States. The gains can be as large as 30 percent in terms of root mean square errors (RMSE). Furthermore, a winning method emerges from its extensive robustness analysis: the Random Forest, which is a highly non-linear method.

Even though it is important to forecast inflation in the United States, it is a particular case study. An open question is whether ML methods can systematically improve forecasts for a variety of macroeconomic series in different countries and contexts. Literature has gradually filled this void. With respect to inflation, Garcia et al. (2017) show that ML methods lead to more accurate forecasts in Brazil, but the winning method is Complete Subset Regressions rather than Random Forest. Similar analyses have been carried out in India (Patrap and Sengupta (2019)), Russia (Baybuza (2018)), and the United Kingdom (Chakraborty and Joseph (2017)), among other countries. The results tend to favor the use of ML methods.

This article considers the Chilean case. Their contributions are twofold. First, we construct a large dataset with macroeconomic variables related to the Chilean economy along the lines of McCracken and Ng (2016), who constructed (and maintain) the dataset for the United States used in Medeiros et al. (2019). Second, we apply a subset of the methods of Medeiros et al. (2019) to forecast headline and core CPI inflation in Chile[3]. The three reference univariate methods are the Random Walk (RW), autoregressive (AR) models and the Unobserved Components of Stochastic Volatility (UCSV) model. We also consider two *naive* methods whose inflation forecast is simply their mean and median. ML methods include the Least Absolute Shrinkage and

1. See Faust and Wright (2013).

2. See Mullainathan and Spiess (2017) and Gu et al. (2018) for recent applications in economics and finance, respectively.

3. In this document, an ML model is any statistical model used for forecasting, which is implemented through an automated computer algorithm and which can handle a large set of predictors.

Selection Operator (LASSO), Adaptive LASSO (adaLASSO), Elastic Net (ElNet), and Adaptive Elastic Net (adaElNet) models. In addition, we consider the following linear methods: Ridge Regression (Ridge), Bayesian Vector Autoregressions (BVAR), Principal Component Factors (Factor), and Complete Subset Regressions (CSR). Finally, we include two non-linear ML methods: Random Forest (RF) and Neural Networks (NN).

We consider direct out-of-sample forecasts for horizons ranging from one month to twelve months ahead, based on a frame of fixed-length moving windows. The set of regressors includes four lags for all variables, as well as four principal components of the data set.

Unlike the aforementioned literature for other countries, we conclude that ML methods cannot consistently beat the benchmark or naive methods. Although some methods work better when variables specific to the Chilean economy are added, they do not represent a significant improvement over the reference or naive methods in terms of RMSE and MAE. It is surprising that, to forecast core CPI inflation (the CPI excluding foods and energy, CPIEFE), simply the mean (mean/median) of past inflation proved to be competitive. In fact, these are the winning methods in many cases. A simple AR (4) process, as well as the complex non-linear RF, are highly competitive when forecasting headline and core inflation from the CPI. Other methods like BVAR, ElNet, and adaElNet are also context-competitive. However, in almost all horizons, for both the CPI and the CPIEFE, there is at least one reference or naive model with similar precision to the best ML methods.

Our main finding that ML does not improve significantly over some *naive* and reference methods leads us to two possible interpretations. One that emphasizes the limitations of the exercise carried out in this document. From database construction to fine tuning of ML model parameters, we tried to stay as close as possible to McCracken and Ng (2016) and Medeiros et al. (2019). This “blind” replication may have favored the reference and/or *naive* models. Perhaps, by excluding some noisy variables from the dataset (or by adjusting some key parameters), ML methods could work better. However, we must bear in mind that these “ex-post” adjustments, even when they aim to capture Chilean specificities, can generate an over-adjustment problem. Furthermore, the data range is very short, which implies a small number of out-of-sample observations compared to the United States. Therefore, the superior predictive tests and the Model Confidence Set (MCS), used to establish the superiority of some ML methods in Medeiros et al. (2019), are not as informative in the

Chilean context. Given the low number of observations, the MCS, for example, is likely to include almost all forecasting methods at the usual significance levels. Therefore, a reasonable concern is whether our main finding that ML methods do not improve inflation forecasts much in Chile will remain valid as the series are lengthened. In this same context of an absence of sufficiently long series, it is that unlike the case studied in Medeiros et al. (2019), the subsample used as out of sample does not present any recession period. These authors showed that much of the RF model's gains in the US come from recessions, where non-linearities and an extended database were important for the RF model to generate better projections than the benchmark models.

Alternatively, the second interpretation emphasizes that Chile is a successful case of inflation targeting, in which both inflation and inflation expectations have been stable and within target during the sample period. Therefore, short-term inflation deviations from the target can be difficult to predict. In this case, sophisticated methods and big data may not add too much (or even generate noise), which could explain why simply using past median or median inflation to forecast inflation is so competitive. As the series are extended, we hope that more research, going beyond our "blind" replication approach, will help to resolve the usefulness of ML techniques for forecasting inflation in Chile.

The document proceeds as follows. Section 2 presents a brief review of the literature for the Chilean case. Section 3 presents the data set. Section 4 describes the empirical strategy, and Section 5 quickly reviews the forecasting methods used in this document. Section 6 presents and discusses the results. Section 7 presents the robustness of the results. Section 8 concludes.

2 Literature review for the Chilean case

The evidence for the Chilean case is extensive and varied at the same time. Pincheira and Medel (2015) analyze different univariate models for inflation-targeting countries, specifically the DESARIMA (Driftless Extended Seasonal ARIMA) class of models. The authors find that this type of model improves the projections in short horizons with respect to the use of benchmark models (autoregressive models of order 6), while the winning models become ambiguous in longer horizons. Along the same lines, but with the objective of incorporating external elements, Pincheira and Gatty (2016) use a SARIMA model of augmented factors, FASARIMA, considering two groups of countries as measures of international inflation (factors): LATAM and OECD. The results show that, by collecting international factors, they contribute to improving inflation projections, that is, they conclude that it is desirable to consider additional elements to those internal to the economy.

On the other hand, the empirical predictive capacity of different specifications for the Phillips curve has been studied. The results are mixed: some are in line with the international evidence of recent years, where the predictive capacity is low and quite unstable (Pincheira and Rubio (2015), Aguirre and Céspedes (2004), Fuentes et al. (2008)). Others do indeed find considerable predictive capacity (Nadal de Simone (2001), Morandé and Tejada (2008)).

However, not only conventional measures such as the CPI have been studied to construct inflation. A recent study explores the use of underlying measures of the price level. Specifically, Nolazco et al. (2019) analyze the use of the CPIEFE, an inflation measure that excludes food and energy prices given their more volatile nature, to predict the overall inflation measure in eight emerging economies. Their results show that there are improvements in short horizons (one to six months) and, especially in the Chilean case, there are also improvements in longer horizons (24 months).

It is also important to mention that it is necessary to consider a wide set of variables when projecting inflation, since there are many mechanisms for transferring different variables that contribute to improving future projections, such as the prices of oil and commodities, among others (Chen et al. (2011), Medel (2015), Medel (2016), Chanut et al. (2018)).

3 The Data

Our database is made up of an extensive number of variables obtained from different sources, including: Central Bank of Chile (BCCh), National Institute of Statistics (INE), FRED, Bloomberg, and Chilean Chamber of Construction (CChC). In particular, we have 64 monthly variables classified into nine groups: Production and income; Working market; Households; Consumption, Orders and inventories; Money and credit; Interest and exchange rate; Prices; Stock market, and External variables[4]. The period covered is from December 2003 to July 2019.

The predictors used in this work are chosen based on their proximity and similarity to those used in McCracken and Ng (2016)[5]. The first four factors obtained from the data from the principal components methodology are included as potential predictors, in addition to the four lags of each independent variable and inflation, accounting for a total of 272 variables[6].

To estimate and evaluate the methods, the sample is divided into two subsamples[7]. The first runs from December 2003 to November 2014 (132 observations) and will be the input to train the models. The second runs from December 2014 to July 2019 (56 observations) and will be used as an independent set of observations against which to contrast and evaluate the predictions of the different methods. Inflation is defined as $\pi_t = \log(P_t) - \log(P_{t-1})$, where π_t is inflation in month t and P_t is a price index. The main price index used is the Consumer Price Index (CPI) calculated by the INE; however, the results for core inflation (CPI excluding food and energy) (CPIEFE) are also reported. The training subsample has a higher and more volatile average inflation than the evaluation subsample, especially marked in the period of the subprime crisis. The average (standard deviation) of monthly inflation for the first subsample is equal to 0.29% (= 0.44%), while for the second period it is 0.23% (= 0.24%). Figure 1 shows both series for the entire period, differentiated by subsamples:

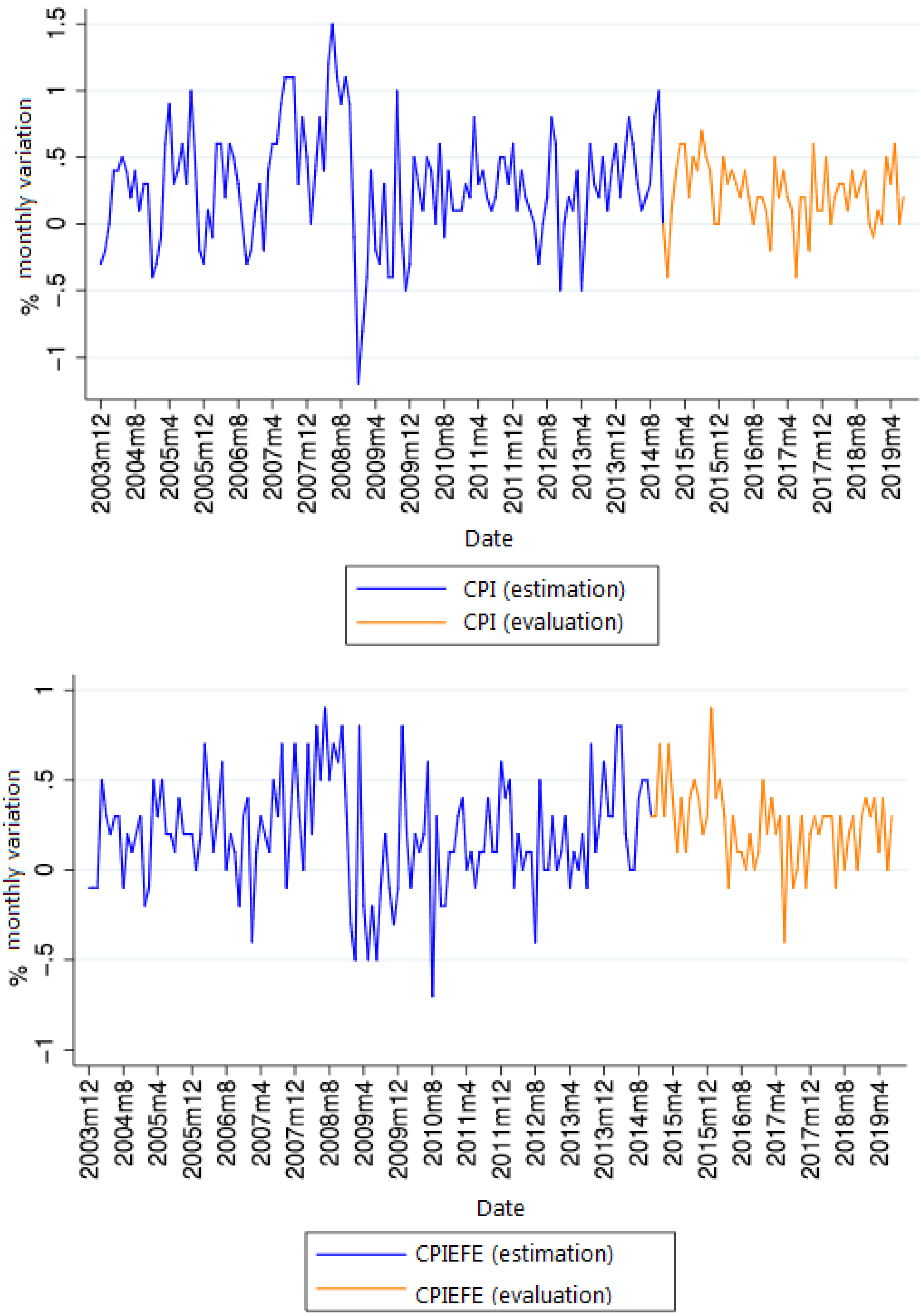
4. Unlike McCracken and Ng (2016), we add a ninth group of external variables taking into account the fact that Chile is a small economy open to the world. We also add variables associated with the mining sector and expectations of economic agents, given the country's great dependence on this sector and the role of expectations in the effectiveness of monetary policy.

5. Out of all the variables available in FRED, Medeiros et al. (2019) take 122 variables that have monthly data from January 1990 to December 2015. For the Chilean case, the data available allow to homologate 64 out of the total 122.

6. See Appendix.

7. The division into training and evaluation subsamples was carried out in a proportion of 70 and 30 percent, respectively. This is a common practice in validation exercises in the absence of a sample large enough to split it in three, following Hastie et al. (2009).

Figure 1: CPI and CPIEFE inflation



Source: Authors' calculations based on Central Bank of Chile's database.

4 Methodology

Monthly inflation is predicted directly under a fixed-size mobile window scheme in order to mitigate the effects of structural breaks and outliers, following Medeiros et al. (2019)[8]. To illustrate the above, the following model is used:

$$\pi_{t+h} = F_h(X_t) + \varepsilon_{t+h} \tag{1}$$

where $t = 1, 2, ..., T$ y $h = 1, 2, ..., H$. Inflation h periods ahead, π_{t+h}, will depend on an n-vector of covariates $X_t = (X_{1t}, X_{2t}, ..., X_{nt})'$ whose effects on inflation will be determined by the function $F_h(\cdot)$; ε_{t+h} is a zero mean random error. In particular, the direct forecast for each model will be given by the following equation:

$$\hat{\pi}_{t+h|t} = \hat{F}_{h,t-R_h+1:t}\left(X_t\right) \tag{2}$$

where $\hat{\pi}_{t+h|t}$ is the inflation forecast h months ahead, considering the information available in t; $\hat{F}_{h,t-R_h+1:t}\left(X_t\right)$ is the objective function estimated from the training sample, which varies according to the method used, the size of the R_h window, the projection horizon h, and the number of lags p. Given that the scheme used is of fixed-size moving windows, $\hat{F}_{h,t-R_h+1:t}$ considers only the information available from $t - R_h + 1$ to t. The size of the window is given by $R_h = 142 - h - p - 1$.

To compare the predictive performance of the different alternatives, the models compete in terms of RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) over the evaluation period. Specifically, the models used correspond to the benchmarks in the literature (RW, AR and UCSV), linear specifications (LASSO, ElNET, RR, adaLASSO, adaElastiNet, Factors, BVAR and CSR). Non-linear (Neural Networks and Random Forest), and the mean and median of past inflation are also added.

8. Unlike the iterative projection, the direct forecast uses the most recent information from the observables. Marcellino et al. (2006) conclude that the direct projection is slightly better than the iterative one; however, the former does not predominate in a uniform way over time and different series. The only exception is the case of BVAR, which considers the joint forecast for covariates and inflation based on the procedure of Banbura et al. (2010).

5 Models

5.1 *Benchmarks*

To quantify the predictive performance of the different methods, the Random Walk (RW)[9] random walk model is taken as a reference point. Simple univariate models such as an autoregressive process (AR) and the UCSV (Unobserved Components Stochastic Volatility) model are also included, as well as the mean and median of past inflation[10].

5.2 Linear Models

5.2.1 Shrinkage Models

These methods estimate the parameters of interest from the minimization of the sum of the squared errors plus a penalty term associated with the number and size of the coefficients. The objective of imposing a cost on the use of additional regressors is to balance out the trade-off between bias and variance[11]. This group includes the LASSO (Least Absolute Shrinkage and Selection Operator), RR (Ridge Regression), ElNet (Elastic Net) models, as well as their adaptive versions adaLASSO (Zou (2006)) and adaElNet (Zou and Hastie (2005)).

In general terms, the punishment term is given by the following equation:

$$\lambda \sum_{p=1}^{p} [(1-\alpha)|b_p| + \alpha |b_p|^2] \tag{3}$$

The sum of the squared errors plus this expression constitute the loss function from which the coefficients in ElNet are estimated; three relevant cases emerge from it. If there is no penalty ($\lambda = 0$), the classic Ordinary Least Squares (OLS) problem will be optimized. If $\alpha = 1$, there are only quadratic

9. For each horizon, the RMSE and MAE values of the Random Walk are normalized to one and the precision measurements of the different methods related to RW are computed.

10. The optimal number of lags in the AR process for each horizon is chosen using the BIC criterion, which determines as the optimal size four lags for all forecasts.

11. The penalty parameter associated with the penalty term determines the complexity of the model by imposing a cost on the use of predictors. In general, a relatively simple model yields stable but inadequate predictions by not fully capturing the underlying structure of the data. On the other hand, the use of a large number of regressors reduces bias by better capturing the behavior of the data at the cost of unstable predictions specific to the training sample, a problem known as overfitting.

restrictions on the parameters, this is Ridge Regression, and if $\alpha = 0$ we have LASSO.

To see the details of how LASSO and their family members work, recommended reads are Hastie et al. (2009), Varian (2014), Mullainathan and Spiess (2017), Driessen et al. (2019), and Coulombe et al. (2019).

5.2.2 Other Linear Models

On the other hand, the factor model (Factors), Bayesian VAR (BVAR), and the CSR (Compete Subset Regressions) are also included. The factor model allows reducing the dimensionality of the problem by estimating a reduced number of components that explain a large part of the variance and covariance between predictors. The first four factors estimated using the principal components methodology are used. For details on the properties of this methodology, see Bai (2003).

The Bayesian VAR (BVAR) is an alternative to factor models and panel VARs for the analysis of large dynamic problems. The regularization parameter is set in relation to the size of the model following Banbura et al. (2010), which makes it possible to mitigate the overfitting problem while retaining the relevant sample information.

CSR is a method that averages the forecasts of various models by reducing the variance and uncertainty associated with each individual model ((Elliot et al. 2015)). The idea behind is to choose a subset of variables and estimate all possible combinations between them to generate a final prediction that will be the average of the individual forecasts. It is important to note that when the set of covariates is too large, the possible combinations grow exponentially, often becoming impractical. To deal with this problem[12], 20 potential variables are chosen through a pre-testing process on which all possible combinations of four regressors are estimated; the final forecast will be the average of all of them. For details on CSR, it is recommended to see Elliot et al. (2013), Elliot et al. (2015).

12. Following Medeiros et al. (2019), inflation h months ahead is regressed against each of the variables with their lags, and the t statistics associated with the predictors are saved. Subsequently, the variables are ordered according to the magnitude of the statistics and the subset of 20 most relevant predictors is selected according to the magnitude of their coefficient.

5.3 No Linear Models

Within the non-linear models we have NN (Neural Networks) and RF (Random Forest). In general terms, NN is a model built from the sum of linear models transformed in a non-linear way ((Hastie et al. 2009)). Its main attraction is related to the flexibility with which it approximates a wide range of functional relationships between variables ((Nakamura 2005)).

RF is a highly nonlinear model constructed from the average of a nonparametric set of regression trees[13]. Each tree is built from a boostrap sample. The aggregation of de-correlated trees with each other through the random selection of predictors makes it possible to reduce the variance of the individual predictions. The logic of the individual trees consists in partitioning the regressor space recursively and generating local predictions from the subspaces; the final prediction is the average of the predictions generated from the multiple trees (hence the name random forest). For details on RF, see Breiman (2001), Hastie et al. (2001), Hastie et al. (2009), and Varian (2014).

6 Results

This section presents the results of the different exercises we carried out[14]. A set of methods (AR, Elnet, adaElnet, BVAR, RF, mean, median) seem to be doing just as well when we review the RMSE and MAE. There is no clear winner, as it varies depending on the horizon, the CPI measure and the evaluation criteria. Also, when one method works particularly well, there is almost always another that does pretty much the same. In general terms, relative to the RW, the models perform better for the CPIEFE projections compared to the CPI.

In each of the tables, the methods with the best results for each horizon are highlighted. RMSEs and MAEs are computed in terms relative to the RW, which is normalized to 1. We highlight in gray the methods that are within a difference of 0.05 points from the winning method that is highlighted in bold. We choose a band of 0.05 because in the case of the US in Medeiros et al. 2019,

13. In the Machine Learning literature, a distinction is made between classification and regression problems depending on the nature of the dependent variable. If the dependent variable is categorical we speak of classification, while if it is continuous we are in the presence of a regression. For details, see Hastie et al. (2009).

14. The codes used to generate the results, which are adaptations of the codes used in Medeiros et al. (2019), as well as the database for Chile, are available upon request.

when the model falls within a 0.05 margin of the winning method, the model is typically included in the 50% Model Confidence Set[15].

Tables 1 and 2 show the results of the different methods for the CPI, considering the RMSE and MAE criteria, respectively.

On average, the winner in RMSE terms is the AR and the mean, followed by the median, adaElNet, BVAR, adaLasso, ElNet, and RF. It is important to note that the difference in the errors of all these models is within an interval of 0.02. BVAR dominates in the first three horizons, adaLASSO dominates the fourth, fifth and sixth is shared by the AR, the media, ElNet, adaElNet and adaLASSO, then the AR alone dominates. On average, the best-performing models show gains close to 23 percent over RW.

Regarding the MAE criterion, the results practically do not vary. On average, AR is the winner followed by RF, BVAR, the median, and the mean. Again, the difference in errors is minimal. The winning models register improvements of the order of 25 percent compared to the RW.

In both cases, LASSO and NN are the worst performing models. Therefore, in the case of inflation measured through the general CPI, there is no one method that remarkably and consistently dominates in all horizons, but rather a broad set that dominates.

---

15. For the Chilean case, due to the short out-of-sample period, the Model Confidence Set could not be informative.

Table 1: RMSE (CPI)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Mean**
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	0.82	0.81	0.80	0.79	**0.77**	**0.77**	0.8	**0.76**	**0.77**	**0.73**	**0.74**	**0.74**	**0.77**
Media	0.78	0.78	0.78	0.78	**0.77**	**0.77**	**0.77**	0.77	**0.77**	0.77	0.77	0.78	**0.77**
Mediana	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.77	0.78	0.78	0.77	0.77	0.78
UCSV	0.85	0.86	0.85	0.85	0.85	0.84	0.86	0.85	0.85	0.86	0.86	0.86	0.85
LASSO	0.76	0.76	0.85	1.37	0.95	1.12	1.51	1.19	0.90	1.09	1.16	1.29	1.08
adaLASSO	0.82	0.78	0.78	**0.75**	0.78	**0.77**	0.78	0.77	0.83	0.86	0.78	0.80	0.79
Ridge	0.76	0.77	0.80	0.85	0.85	0.89	0.91	0.88	0.84	0.86	0.86	0.87	0.85
Elnet	0.76	0.77	0.79	0.78	**0.77**	**0.77**	**0.77**	0.77	0.79	0.78	0.78	0.93	0.79
adaElnet	0.83	**0.75**	0.79	0.78	**0.77**	**0.77**	**0.77**	**0.76**	0.80	0.79	0.78	0.79	0.78
Factor	0.80	0.84	0.86	0.85	0.85	0.87	0.85	0.84	0.84	0.89	0.87	0.90	0.85
BVAR	**0.75**	**0.75**	**0.77**	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78
CSR	0.81	0.8	0.81	0.78	0.78	0.8	0.88	0.85	0.83	0.81	0.83	0.86	0.82
RF	0.76	0.76	0.79	0.81	0.80	0.79	0.79	0.78	**0.77**	0.79	0.81	0.82	0.79
NN	1.06	1.00	0.94	0.93	0.97	1.04	1.19	1.16	1.01	1.01	0.91	1.07	1.02

Source: Authors' calculations.

Table 2: MAE (CPI)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Mean**
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	0.83	0.78	0.77	**0.76**	**0.74**	**0.74**	0.79	**0.75**	0.75	**0.68**	**0.69**	**0.69**	**0.75**
Media	0.77	0.77	0.77	0.77	0.77	0.76	**0.76**	0.76	0.76	0.76	0.76	0.76	0.76
Mediana	0.77	0.77	0.77	**0.76**	0.76	0.76	**0.76**	0.76	0.76	0.76	0.75	0.75	0.76
UCSV	0.83	0.84	0.83	0.84	0.83	0.82	0.85	0.85	0.84	0.84	0.84	0.85	0.84
LASSO	0.75	0.75	0.84	1.10	0.88	0.96	1.35	1.06	0.91	1.02	1.03	1.22	0.99
adaLASSO	0.79	0.76	0.78	0.77	0.77	0.77	0.77	**0.75**	0.82	0.85	0.77	0.79	0.78
Ridge	0.74	0.77	0.78	0.83	0.80	0.85	0.86	0.81	0.77	0.81	0.82	0.81	0.80
Elnet	0.75	0.76	0.77	0.77	0.77	0.76	**0.76**	**0.75**	0.78	0.77	0.77	0.84	0.77
adaElnet	0.80	0.74	0.78	0.77	0.77	0.76	**0.76**	**0.75**	0.79	0.78	0.77	0.77	0.77
Factor	0.79	0.81	0.84	0.82	0.84	0.83	0.84	0.83	0.82	0.87	0.86	0.85	0.83
BVAR	0.74	0.75	**0.76**	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.76
CSR	0.8	0.79	0.8	**0.76**	**0.74**	0.76	0.87	0.82	0.78	0.78	0.77	0.79	0.79
RF	**0.71**	**0.73**	**0.76**	0.8	0.78	0.77	0.77	**0.75**	**0.74**	0.75	0.76	0.78	0.76
NN	1.00	0.97	0.90	0.92	0.94	1.03	1.16	1.07	0.92	0.93	0.87	1.00	0.98

Source: Authors' calculations.

Tables 3 and 4 show the results of the different methods for the CPIEFE (CPI exclusing food and energy), considering the RMSE and MAE criteria, respectively.

In the case of the RMSE, on average, the winner is RF, followed by AR, mean, median, ElNet, adaElNet and BVAR, but the difference between the models is only 0.01 points. In the first horizons the AR is the winner, but from the fifth horizon onwards the RF becomes the most competitive. The best competitors report earnings, on average, close to 25 percent relative to RW.

Regarding the MAE, the results remain quite similar. On average, RF is the winner, followed by AR, Median, and ElNet. Again, the differences are minimal. The winner records average earnings of 34 percent over RW. In general, the RF model has very good results for all horizons, this together with the AR, the mean and the median.

Similar to the case of the headline CPI, for the CPIEFE the results show that models such as the AR, the median or the mean are similar or even more capable in most of the horizons compared to the ML alternatives.

Table 3: RMSE (CPIEFE)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Mean**
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	**0.72**	**0.71**	**0.73**	0.77	**0.74**	0.77	0.74	0.75	0.73	0.74	0.74	0.74	0.74
Mean	0.74	0.74	0.74	**0.74**	**0.74**	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74
Median	0.74	0.74	0.74	**0.74**	**0.74**	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74
UCSV	0.91	0.91	0.91	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.93	0.93	0.92
LASSO	0.76	0.75	0.79	0.76	**0.74**	0.77	1.29	1.48	1.04	1.28	0.90	1.06	0.97
adaLASSO	0.80	0.72	0.94	0.75	**0.74**	0.74	0.75	0.76	0.74	0.77	0.76	0.79	0.77
Ridge	**0.72**	0.72	0.76	0.77	0.78	0.84	0.84	0.85	0.82	0.81	0.82	0.82	0.80
Elnet	0.74	0.74	0.74	**0.74**	**0.74**	0.74	0.74	0.74	0.74	0.75	0.74	**0.73**	0.74
adaElnet	0.74	0.74	0.74	**0.74**	**0.74**	0.74	0.74	0.74	0.74	0.76	0.74	0.74	0.74
Factor	0.78	0.77	0.77	0.78	0.75	0.76	0.80	0.77	0.79	0.8	0.84	0.82	0.79
BVAR	**0.72**	0.73	0.74	**0.74**	**0.74**	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.74
CSR	**0.72**	**0.71**	0.74	0.77	**0.74**	0.82	0.80	0.82	0.78	0.79	0.81	0.80	0.78
RF	0.75	0.73	0.75	0.75	**0.74**	**0.73**	**0.71**	**0.71**	**0.70**	**0.71**	**0.73**	0.74	**0.73**
NN	0.82	0.81	0.98	0.88	0.84	0.92	0.96	0.92	0.87	1.00	0.94	0.94	0.91

Source: Authors' calculations.

Table 4: MAE (CPIEFE)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Mean**
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	0.65	**0.64**	**0.67**	0.71	**0.67**	0.70	0.68	0.69	0.67	0.66	0.67	**0.67**	0.67
Mean	0.68	0.68	0.68	**0.68**	0.68	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69
Median	0.68	0.68	0.68	**0.68**	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68
UCSV	0.88	0.88	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.90	0.90	0.89
LASSO	0.69	0.70	0.74	0.70	0.68	0.71	1.10	1.17	0.87	1.06	0.83	0.98	0.85
adaLASSO	0.74	0.66	0.79	**0.68**	0.68	0.68	0.68	0.70	0.68	0.69	0.68	0.72	0.70
Ridge	0.66	0.66	**0.67**	0.69	0.68	0.77	0.81	0.79	0.74	0.71	0.71	0.72	0.72
Elnet	0.68	0.68	0.68	**0.68**	0.68	0.69	0.69	0.69	0.69	0.69	0.69	0.68	0.68
adaElnet	0.68	0.68	0.68	**0.68**	0.68	0.69	0.69	0.69	0.69	0.70	0.69	0.69	0.69
Factor	0.71	0.68	**0.67**	**0.68**	**0.67**	0.69	0.71	0.70	0.73	0.76	0.77	0.75	0.71
BVAR	0.67	0.67	0.68	0.69	0.68	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69
CSR	**0.64**	0.66	**0.67**	0.71	**0.67**	0.73	0.73	0.73	0.71	0.71	0.70	0.72	0.70
RF	0.67	0.66	0.68	**0.68**	0.68	**0.67**	**0.65**	**0.64**	**0.64**	**0.63**	**0.66**	**0.67**	**0.66**
NN	0.73	0.75	0.95	0.81	0.76	0.85	0.86	0.90	0.80	0.91	0.88	0.84	0.84

Source: Authors' calculations.

7 Robustness

Unlike the United States, Chile is a small economy open to the world that depends largely on the performance of the mining sector and the international context. Therefore, this section includes external variables and the mining sector to capture the specificities of the national economy[16]. Additionally, variables associated with expectations of monetary policy rates, inflation, and rates of sovereign bonds at different terms are added, because a large part of the effectiveness of monetary policy depends on the public's expectations about the future evolution of the economy. In all, 16 variables were added.

Tables 5 and 6 show the relative performance of the different methods for inflation when extending the set of predictors. By construction, the performance of the reference or naive models does not improve with the addition of these new variables. In general terms, the results are maintained, in almost all cases there is at least one reference or naive model with precision similar to ML methods. However, it is worth highlighting the improvement of 8 percentage points in CSR and 5 percentage points for adaLASSO and RF. Since the first two methods are linear, the above suggests that the gains could be partly explained by a more informative data set.

16. See Appendix.

Table 5: RMSE (CPI+Variables)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Mean**
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	0.82	0.81	0.80	0.79	0.77	0.77	0.80	0.76	0.77	**0.73**	**0.74**	**0.74**	**0.77**
Mean	0.78	0.78	0.78	0.78	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.78	**0.77**
Median	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.77	0.78	0.78	0.77	0.77	0.78
UCSV	0.85	0.86	0.85	0.85	0.85	0.84	0.86	0.85	0.85	0.86	0.86	0.86	0.85
LASSO	0.76	0.91	1.36	1.57	1.28	1.50	1.39	1.34	1.37	1.21	1.11	1.31	1.26
adaLASSO	0.76	0.75	**0.76**	0.77	0.81	0.83	**0.76**	**0.71**	0.76	0.74	0.79	0.82	**0.77**
Ridge	**0.74**	0.75	0.78	0.82	0.81	0.84	0.85	0.84	0.79	0.81	0.81	0.82	0.81
Elnet	0.75	0.84	0.78	1.11	0.91	0.77	0.80	0.88	0.77	0.84	0.77	1.04	0.86
adaElnet	0.76	**0.74**	0.77	0.79	0.78	0.77	0.77	0.76	0.77	0.78	0.77	0.80	**0.77**
Factor	0.80	0.80	0.80	0.80	0.81	0.84	0.83	0.83	0.83	0.82	0.83	0.84	0.82
BVAR	**0.74**	0.76	0.77	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	**0.77**
CSR	0.80	0.77	0.78	**0.74**	**0.75**	**0.75**	0.82	0.78	0.77	0.75	0.78	0.82	**0.77**
RF	0.75	0.75	0.79	0.81	0.80	0.78	**0.76**	0.75	**0.74**	0.76	0.77	0.78	**0.77**
NN	0.94	0.92	0.88	0.96	0.88	1.06	1.07	0.91	0.91	1.02	0.94	0.89	0.95

Source: Authors' calculations.

Table 6: MAE (CPI+Variables)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Mean**
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	0.83	0.78	0.77	0.76	0.74	0.74	0.79	0.75	0.75	**0.68**	**0.69**	**0.69**	0.75
Mean	0.77	0.77	0.77	0.77	0.77	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
Median	0.77	0.77	0.77	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.75	0.75	0.76
UCSV	0.83	0.84	0.83	0.84	0.83	0.82	0.85	0.85	0.84	0.84	0.84	0.85	0.84
LASSO	0.73	0.87	1.26	1.51	1.25	1.48	1.38	1.23	1.16	1.23	1.07	1.20	1.20
adaLASSO	0.74	**0.73**	0.74	0.76	0.76	0.80	**0.71**	**0.66**	0.72	0.71	0.78	0.77	**0.74**
Ridge	0.74	0.74	0.76	0.81	0.78	0.82	0.80	0.78	0.74	0.76	0.77	0.77	0.77
Elnet	0.75	0.80	0.76	1.02	0.83	0.76	0.78	0.81	0.76	0.78	0.76	0.91	0.81
adaElnet	0.75	0.74	0.76	0.77	0.76	0.76	0.74	0.73	0.75	0.77	0.76	0.77	0.75
Factor	0.80	0.76	0.77	0.79	0.80	0.81	0.83	0.82	0.82	0.82	0.80	0.79	0.80
BVAR	0.73	0.75	0.76	0.77	0.77	0.77	0.78	0.77	0.77	0.77	0.77	0.77	0.77
CSR	0.78	0.76	**0.73**	**0.71**	**0.71**	**0.71**	0.83	0.75	0.72	0.70	0.71	0.75	**0.74**
RF	**0.72**	**0.73**	0.77	0.79	0.80	0.77	0.73	0.71	**0.70**	0.73	0.74	0.74	**0.74**
NN	0.90	0.87	0.93	0.92	0.92	1.03	1.01	0.91	0.88	1.06	0.92	0.87	0.93

Source: Authors' calculations.

Tables 7 and 8 present the results for the case of core inflation, measured by the CPIEFE. Although the performance of some ML methods improves once we extend the database, these improvements are not enough to position them as the sole winners. Actually, in almost all cases there is at least one naive or reference model with similar precision to ML methods. Therefore, when considering the specificities of the Chilean economy by adding new variables, our conclusions do not change.

Table 7: RMSE (CPIEFE+Variables)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Mean**
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	0.72	0.71	**0.73**	0.77	0.74	0.77	0.74	0.75	0.73	0.74	0.74	0.74	0.74
Mean	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74
Median	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74
UCSV	0.91	0.91	0.91	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.93	0.93	0.92
LASSO	0.84	1.17	1.12	1.26	1.17	1.18	1.45	1.35	1.37	1.53	1.28	1.23	1.25
adaLASSO	0.81	0.79	0.76	0.74	0.76	0.77	0.77	0.78	0.74	0.75	0.75	0.75	0.76
Ridge	**0.71**	0.72	0.74	0.74	0.72	0.78	0.78	0.79	0.77	0.77	0.77	0.76	0.75
Elnet	0.74	0.74	0.74	0.74	0.73	0.74	0.82	0.74	0.80	0.74	0.74	0.76	0.75
adaElnet	0.74	0.74	0.74	0.74	0.75	0.74	0.75	0.74	0.74	0.74	0.74	0.74	0.74
Factor	0.78	0.77	0.76	0.78	0.75	0.76	0.78	0.76	0.77	0.74	0.81	0.75	0.77
BVAR	**0.71**	0.73	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
CSR	0.73	**0.70**	0.74	0.77	0.75	0.82	0.79	0.77	0.72	**0.71**	0.74	0.76	0.75
RF	0.73	0.72	0.74	**0.73**	0.74	**0.72**	**0.71**	**0.71**	**0.70**	**0.71**	**0.72**	**0.71**	**0.72**
NN	0.85	0.79	0.82	0.89	**0.70**	0.91	0.82	0.77	0.85	0.82	0.83	0.94	0.83

Source: Authors' calculations.

Table 8: MAE (CPIEFE+Variables)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Mean**
RW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	0.65	**0.64**	0.67	0.71	0.67	0.70	0.68	0.69	0.67	0.66	0.67	0.67	0.67
Mean	0.68	0.68	0.68	0.68	0.68	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69
Median	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68
UCSV	0.88	0.88	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.90	0.90	0.89
LASSO	0.76	1.07	1.04	1.14	1.12	1.07	1.27	1.18	1.26	1.34	1.17	1.20	1.14
adaLASSO	0.77	0.73	0.70	0.68	0.69	0.71	0.70	0.71	0.68	0.70	0.69	0.68	0.70
Ridge	**0.64**	0.65	**0.65**	**0.66**	0.64	0.72	0.74	0.73	0.70	0.66	0.67	0.65	0.68
Elnet	0.68	0.68	0.68	0.68	0.67	0.69	0.73	0.69	0.72	0.69	0.69	0.71	0.69
adaElnet	0.68	0.68	0.68	0.68	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69
Factor	0.72	0.70	0.69	0.71	0.67	0.68	0.69	0.67	0.69	0.67	0.72	0.67	0.69
BVAR	0.66	0.67	0.68	0.69	0.69	0.70	0.70	0.69	0.69	0.69	0.69	0.69	0.69
CSR	**0.64**	**0.64**	0.66	0.70	0.65	0.70	0.71	0.70	0.65	0.65	**0.63**	0.66	0.67
RF	0.65	0.65	0.67	**0.66**	0.66	**0.65**	**0.66**	**0.65**	**0.64**	**0.64**	0.65	**0.64**	**0.65**
NN	0.73	0.76	0.77	0.87	**0.60**	0.85	0.76	0.68	0.79	0.77	0.79	0.80	0.76

Source: Authors' calculations.

8 Conclusion

We document that ML methods do not consistently outperform simple and univariate linear competitors such as the AR, the mean, and the median of past inflation. Although the RF dominates over inflation and core inflation when the variables are extended, this does not represent a significant improvement in terms of RMSE and MAE. Surprisingly, for forecasting CPI or core CPI inflation, the mean or median of past inflation proved to be competitive. In fact, these are the winning methods in many cases. A simple AR (4) process, as well as the complex non-linear RF, are highly competitive. Other ML methods such as BVAR, ElNet, and adaElNet are also context-competitive. However, in almost all horizons, for both the CPI and the core CPI, there is at least one benchmark model whose performance is very close to (or even better than) the best performing ML method.

We recognize the limitations of the exercise performed in this work. By trying to stay as close to McCracken and Ng (2016) and Medeiros et al. (2019), this "blind" replication may have favored the reference models. Furthermore, the data range is very short, which implies a small number of out-of-sample observations compared to the United States. Therefore, a reasonable concern is whether our main finding that ML methods do not improve inflation forecasts much in Chile would continue to hold as the series lengthens. In this same context of an absence of sufficiently long series, it is that unlike the case studied in Medeiros et al. (2019), the subsample used as an out-of-sample does not present any period of recession, which was when the RF was particularly useful.

Alternatively, Chile is a successful case of inflation targeting, in which both inflation and inflation expectations have been stable and within the target range during the sample period. Therefore, short-term inflation deviations from the target can be difficult to predict. In this case, sophisticated methods and big data may not add too much (or even generate noise), which could explain why simply using past or median inflation to forecast inflation is highly competitive. As the series get longer, we hope that more research, going beyond our "blind" replication approach, will help to resolve the usefulness of ML techniques for forecasting Chilean inflation.

References

Aguirre, A., and L. Céspedes. 2004. "Uso de Análisis Factorial Dinámico para Proyecciones Macroeconómicas." *Banco Central de Chile, Documentos de Trabajo.*

Bai, J. 2003. "Inferential theory for factor models of large dimensions." *Econometrica* 71 (1): 135–171.

Banbura, M., D. Giannone, and L. Reichlin. 2010. "Large Bayesian vector autoregressions." *Journal of Applied Econometrics* 25 (1): 71–92.

Baybuza, I. 2018. "Inflation Forecasting Using Machine Learning Methods." *Russian Journal of Money and Finance* 77 (4): 42–59.

Breiman, L. 2001. "Random forests." *Machine Learning* 45 (1): 5–32.

Chakraborty, C., and A. Joseph. 2017. "Machine learning at central banks." *Bank of England Working Paper Series.*

Chanut, N., M. Marcel, and C. Medel. 2018. "Can Economic Perception Surveys Improve Macroeconomic Forecasting in Chile?" *Banco Central de Chile, Documentos de Trabajo.*

Chen, Y., S. Turnovsky, and E. Zivot. 2011. "Forecasting Inflation using Commodity Price Aggregates." *Journal of Econometrics* 183 (1): 117–134.

Coulombe, P., M. Leroux, D. Stevanovic, and S. Surprenant. 2019. "How is machine learning useful for macroeconomic forecasting?" *CIRANO Working Paper.*

Driessen, A., A. Batch, J. Chen, A. Dunn, and K. Hood. 2019. "Off to the races: A comparison of Machine Learning and alternative data for predicting Economic Indicators." *Big Data for 21st Century Economic Statistics.*

Elliot, G., A. Gargano, and A. Timmermann. 2013. "Complete subset regressions." *Journal of Econometrics* 177 (2): 357–373.

———. 2015. "Complete subset regressions with large-dimensional sets of predictors." *Journal of Economic Dynamics and Control* 54:86–110.

Faust, J., and J. Wright. 2013. "Forecasting Inflation, en G. Elliot y A. Timmermann, eds." *Handbook of Economic Forecasting* 2.

Fuentes, R., F. Gredig, and M. Larraín. 2008. "La brecha de producto en Chile: medición y evaluación." *Economía Chilena* 11 (2).

Garcia, M., M. Medeiros, and G. Vasconcelos. 2017. "Real-time inflation forecasting with high-dimensional models: The case of brazil." *International Journal of Forecasting* 33 (3): 679–693.

Gu, S., B. Kelly, and D. Xiu. 2018. "Empirical Asset Pricing via Machine Learning." *Working paper, University of Chicago.*

Hastie, T., R. Tibshirami, and J. Friedman. 2001. "The Elements of Statistical Learning; Data Minning, Inference and Prediction." *1st ed. New York: Springer.*

———. 2009. "The Elements of Statistical Learning; Data Mining, Inference and Prediction." *2nd ed. New York: Springer.*

Marcellino, M., J.H. Stock, and M.W Watson. 2006. "A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series." *Journal of Econometrics* 135 (1-2): 499–526.

McCracken, M., and S. Ng. 2016. "FRED-MD: A Monthly Database for Macroeconomic Research." *Journal of Business & Economic Statistics* 34 (4): 574–589.

Medeiros, M., G. Vasconcelos, A. Veiga, and E. Zilberman. 2019. "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods." *Journal of Business & Economic Statistics, forthcoming.*

Medel, C. 2015. "Fuelling Future Prices: Oil Price and Global Inflation." *Banco Central de Chile, Documentos de Trabajo.*

———. 2016. "Un Análisis de la Capacidad Predictiva del Precio del Cobre sobre la Inflación Global." *Banco Central de Chile, Documentos de Trabajo.*

Morandé, F., and M. Tejada. 2008. "Fuentes de incertidumbre monetaria en la conducción de la política monetaria en Chile." *Economía Chilena* 11 (3).

Mullainathan, S., and J. Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31 (2): 87–106.

Nadal de Simone, F. 2001. "Proyección de la inflación en Chile." *Economía Chilena* 4 (3).

Nakamura, E. 2005. "Inflation forecasting using a neural network." *Economics Letters* 86 (3): 373–378.

Nolazco, J., P. Pincheira, and J. Selaive. 2019. “Forecasting Inflation in Latin America with Core Measures.” *International Journal of Forecasting, forthcoming.*

Patrap, B., and S. Sengupta. 2019. “Macroeconomic Forecasting in India:Does Machine Learning Hold the Key to Better Forecasts?” *RBI Working Paper Series.*

Pincheira, P., and A. Gatty. 2016. “Forecasting Chilean inflation with international factors.” *Empirical Economics* 51 (3): 981–1010.

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

Pincheira, P., and H. Rubio. 2015. “The low predictive power of simple Phillips curves in Chile.” *Cepal Review* 116.

Stock, J.H., and M.W. Watson. 2010. “Modeling inflation after the crisis.” *NBER.*

Varian, H. 2014. “Big Data: New Tricks for Econometrics.” *Journal of Economic Perspectives* 28 (2): 3–27.

Zou, H. 2006. “The adaptative lasso and its oracle properties.” *Journal of American Statistical Association* 101 (476): 1418–1429.

Zou, H., and T. Hastie. 2005. “Regularization and variable selection via the elastic net.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2): 301–320.

Appendix

Table 9: Transformations of the database variables (1/2)

Transformation	Variable	Source	Group
$\Delta log(x_t)$	Industrial production index	BCCH	Producción e Ingresos
$\Delta log(x_t)$	Production index: Materials	CCHC	Producción e Ingresos
$\Delta log(x_t)$	Production index: manufactures-production	BCCH	Producción e Ingresos
$\Delta log(x_t)$	Production index: Electricity, Water, and Gas	INE	Producción e Ingresos
$\Delta log(x_t)$	Electricity generation index	BCCH	Producción e Ingresos
$\Delta log(x_t)$	Diffusion index: Manufacturing	BCCH	Producción e Ingresos
$\Delta log(x_t)$	Capacity use: Manufacturing	BCCH	Producción e Ingresos
$\Delta log(x_t)$	Diffusion index: Mining	BCCH	Producción e Ingresos
$\Delta log(x_t)$	Vacancy index	BCCH	Mercado Laboral
no transformation	Work force	BCCH	Mercado Laboral
no transformation	Employees	BCCH	Mercado Laboral
Δx_t	Unemployment rate	BCCH	Mercado Laboral
$\Delta log(x_t)$	Employment: Industry	FRED	Mercado Laboral
$\Delta log(x_t)$	Employment: Construction	FRED	Mercado Laboral
$\Delta log(x_t)$	Employment: Services	FRED	Mercado Laboral
no transformation	Wage index	BCCH	Mercado Laboral
no transformation	Labor cost index	BCCH	Mercado Laboral
$\Delta log(x_t)$	Housing permits	FRED	Hogares
$\Delta log(x_t)$	Home / Residential Building Permits	FRED	Hogares
$\Delta log(x_t)$	Sales index: Trade	BCCH	Consumo, ordenes e inventarios
$\Delta log(x_t)$	Sales index: supermarkets	BCCH	Ordenes e inventarios
$\Delta log(x_t)$	Diffusion index: manufactures-general	BCCH	Consumo, ordenes e inventarios
$\Delta log(x_t)$	Diffusion index: manufacturing-inventories	BCCH	Consumo, ordenes e inventarios
$\Delta log(x_t)$	Diffusion index: Construction-demand	BCCH	Consumo, ordenes e inventarios
$\Delta log(x_t)$	Diffusion index: Manufacturing-demand	BCCH	Consumo, ordenes e inventarios
$\Delta log(x_t)$	Diffusion index: retail-inventories	BCCH	Consumo, ordenes e inventarios
$\Delta log(x_t)$	Economy perception index	BCCH	Consumo, ordenes e inventarios
$\Delta^2 log(xt_t)$	M1	BCCH	Dinero y Crédito
$\Delta^2 log(xt_t)$	M2	BCCH	Dinero y Crédito
$\Delta log(x_t)$	Monetary base	BCCH	Dinero y Crédito
$\Delta^2 log(xt_t)$	Reserve reserve banking sector	BCCH	Dinero y Crédito
$\Delta^2 log(xt_t)$	Required reserve banking sector	BCCH	Dinero y Crédito
$\Delta^2 log(xt_t)$	Commercial Loans	BCCH	Dinero y Crédito
$\Delta^2 log(xt_t)$	Consumer Loans	BCCH	Dinero y Crédito
$\Delta^2 log(xt_t)$	Housing Loans	BCCH	Dinero y Crédito
Δx_t	Interbank Rate (IR)	BCCH	Intereses y Tipo de cambio
Δx_t	2-year bonds (pesos)	BCCH	Intereses y Tipo de cambio
Δx_t	5-year bonds (pesos)	BCCH	Intereses y Tipo de cambio
Δx_t	5-year bonds (UF)	BCCH	Intereses y Tipo de cambio
Δx_t	10-year bonds (UF)	BCCH	Intereses y Tipo de cambio
Δx_t	20-year bonds (UF)	BCCH	Intereses y Tipo de cambio
no transformation	5-year bonds (UF) minus IR	BCCH	Intereses y Tipo de cambio
no transformation	10-year bonds (UF) minus IR	BCCH	Intereses y Tipo de cambio

Source: Authors' calculations.

Table 10: Transformations of the database variables (2/2)

Transformation	Variable	Source	Group
no transformation	20-year bonds (UF) minus IR	BCCH	Intereses y Tipo de cambio
$\Delta log(x_t)$	Multilateral exchange rate	BCCH	Intereses y Tipo de cambio
$\Delta log(x_t)$	Exchange rate to dollar	BCCH	Intereses y Tipo de cambio
$\Delta log(x_t)$	Exchange rate to real	BCCH	Intereses y Tipo de cambio
$\Delta log(x_t)$	Exchange rate to euro	BCCH	Intereses y Tipo de cambio
$\Delta log(x_t)$	Exchange rate to yen	BCCH	Intereses y Tipo de cambio
$\Delta log(x_t)$	Producer Price Index: Manufacturing	BCCH	Precios
$\Delta log(x_t)$	Oil Price (WTI)	BCCH	Precios
$\Delta log(x_t)$	Producer Price Index: Mining	BCCH	Precios
$\Delta log(x_t)$	Consumer Price Index: Clothing	BCCH	Precios
$\Delta log(x_t)$	Consumer Price Index: Health	BCCH	Precios
$\Delta log(x_t)$	Consumer Price Index: Transport	BCCH	Precios
$\Delta log(x_t)$	Consumer Price Index: Education	BCCH	Precios
$\Delta log(x_t)$	Consumer Price Index: Services	BCCH	Precios
$\Delta log(x_t)$	Consumer price index: services minus housing	BCCH	Precios
$\Delta log(x_t)$	Consumer Price Index (CPI)	BCCH	Precios
$\Delta log(x_t)$	Consumer Price Index minus food and energy	BCCH	Precios
$\Delta log(x_t)$	IPSA	Bloomberg	Bolsa de valores
$\Delta log(x_t)$	EPU	Cerda, Silva y Valente	Bolsa de valores
$\Delta log(x_t)$	EPUC	Cerda, Silva y Valente	Bolsa de valores
$\Delta log(x_t)$	VXO	Bloomberg	Bolsa de valores
Δx_t	MPR: USA	BCCH	Variables Externas
Δx_t	MPR: Euro	BCCH	Variables Externas
Δx_t	MPR: Japan	BCCH	Variables Externas
$\Delta log(x_t)$	CPI: USA	BCCH	Variables Externas
$\Delta log(x_t)$	CPI (CORE): USA	BCCH	Variables Externas
$\Delta log(x_t)$	Production index Mining	BCCH	Producción e Ingresos
$\Delta log(x_t)$	Mining export index	BCCH	Producción e Ingresos
no transformation	Inflation expectations: current month	BCCH	Precios
no transformation	Inflation expectations: 11 months ahead	BCCH	Precios
no transformation	Inflation expectations: 23 months ahead	BCCH	Precios
no transformation	TPM Expectations: Next Meeting	BCCH	Precios
$\Delta log(x_t)$	Copper price (US / pound)	BCCH	Precios
no transformation	TPM expectations: 11 months ahead	BCCH	Intereses y tipo de cambio
no transformation	Bond expectations (pesos): 2 months ahead	BCCH	Intereses y tipo de cambio
no transformation	Bond expectations (pesos): 11 months ahead	BCCH	Intereses y tipo de cambio
no transformation	Bond expectations (pesos): 23 months ahead	BCCH	Intereses y tipo de cambio

Source: Authors' calculations.

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 Ch$500 si es dentro de Chile y US$12 si es fuera de Chile. Las solicitudes se pueden hacer por fax: +56 2 26702231 o a través del 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 order inside Chile the charge is Ch$500.) Orders can be placed by fax: +56 2 26702231 or by email: bcch@bcentral.cl.

DTBC – 859
A note on currency-hedging
Rodrigo Alfaro, Natan Goldberger

DTBC – 858
Multimarket Contact in Banking Competition in The United States
David Coble

DTBC – 857
Immigration in Emerging Countries: A Macroeconomic Perspective
Agustín Arias, Juan Guerra-Salas

DTBC – 856
Saving Rates in Latin America: A Neoclassical Perspective
Andrés Fernández, Ayşe Imrohoroğlu, Cesar Tamayo

DTBC – 855
Returns to Work Experience in Chile
Rosario Aldunate

DTBC – 854
Hindsight vs. Real time measurement of the output gap: Implications for the Phillips curve in the Chilean Case
Camila Figueroa, Jorge Fornero, Pablo García

DTBC – 853
Prudential Policies and Bailouts - A Delicate Interaction
Ernesto Pasten

DTBC – 852
Capital Controls and Firm Performance
Eugenia Andreasen, Sofia Bauducco, Evangelina Dardati

DTBC – 851
S&P 500 under Dynamic Gordon Model
Rodrigo Alfaro, Andrés Sagner

DTBC – 850
Inflation Globally
Òscar Jordà, Fernanda Nechio

DTBC – 849
Trade Exposure and the Evolution of Inflation Dynamics
Simon Gilchrist, Egon Zakrajsek

DTBC – 848
The link between labor cost and price inflation in the euro area
Elena Bobeica, Matteo Ciccarelli, Isabel Vansteenkiste

DTBC – 847
Trend, Seasonal, and Sectoral Inflation in the Euro Area
James H. Stock, Mark W. Watson

DTBC – 846
Has the U.S. Wage Phillips Curve Flattened? A Semi-Structural Exploration
Jordi Galí, Luca Gambetti

DTBC – 845
The "Supply-Side Origins" of U.S. Inflation
Bart Hobijn

BANCO CENTRAL
DE CHILE