The Evasive Predictive Ability of Core Inflation

Pincheira, Pablo and Selaive, Jorge and Nolazco, Jose Luis

School of Business, Adolfo Ibáñez University, BBVA Research
School of Business, University of Chile, University of Chile

6 January 2016

The Evasive Predictive Ability of Core Inflation

Pablo Pincheira[1]

School of Business, Adolfo Ibáñez University

Jorge Selaive

BBVA Research & School of Business, University of Chile

&

Jose Luis Nolazco

Universidad de Chile

January 2016

Abstract

We explore the ability of traditional core inflation −consumer prices excluding food and energy− to predict headline CPI annual inflation. We analyze a sample of OECD and non-OECD economies using monthly data from January 1994 to March 2015. Our results indicate that sizable predictability emerges for a small subset of countries. For the rest of our economies predictability is either subtle or undetectable. These results hold true even when implementing an out-of-sample test of Granger causality especially designed to compare forecasts from nested models. Our findings partially challenge the common wisdom about the ability of core inflation to forecast headline inflation, and suggest a careful weighting of the traditional exclusion of food and energy prices when assessing the size of the monetary stimulus.

JEL Codes: E31, E17, E37, E52, E58
Key Words: Inflation, Forecasting, Time Series, Monetary Policy, Core Inflation.

---

[1]The views expressed in this paper do not necessarily represent those of the BBVA. All remaining errors are ours. Address: Diagonal Las Torres 2640, office 521, building C, Peñalolén, Santiago, Chile. Phone: (562) 23311489. E-mail: pablo.pincheira@uai.cl (corresponding author).

I. Introduction

The objective of this paper is to evaluate the ability of core inflation to forecast headline inflation. Differing from the existing literature, we take a global perspective, analyzing this predictive ability using a common methodology and a common sample period for 33 different countries. To our knowledge, a thorough study aimed at quantifying this predictability for a number of countries has not been written yet.

The point that we address in this paper is important because, in words of Bullard (2011a), the "core predicts headline" argument is fairly popular. In a context in which inflation is not easy to forecast (Stock and Watson, 2008) the idea that core inflation may be a useful predictor in principle is very appealing, especially for central banks that are responsible for maintaining overall price stability and need to know where inflation is heading.

There is no unique way to define a core inflation measure. In fact there are several articles comparing and analyzing the behavior of different core measures. See for instance Robalo, Duarte and Morais (2003), Clark (2001) and Rich and Steindel (2007) just to mention a few. Despite the number of different core inflation definitions, one of the most widely used is based on the CPI excluding "food" and "energy" components (Robalo, Duarte and Morais, 2003). In the particular case of the US, where emphasis is given to the Personal Consumer Expenditure (PCE) price index, the usual core measure is PCE excluding the same components.

The emphasis on core measures of inflation relies on the hope that by removing volatile components, we may end up with a clearer indicator about future developments of headline inflation. According to Crone, Khettry, Mester and Novak (2013) this is the prevailing view. In fact, food and energy components have been historically highly volatile (for example, due to temporary supply disruptions), and their large price fluctuations are usually expected to correct themselves within a relatively short period of time. As Freeman (1998) explains, since inflation may be either too sensitive to exogenous variables or vulnerable to a few particular volatile components, it is common to use ''core'' or "underlying" inflation measures to capture trends in total inflation. Nevertheless, hopes are not facts, and an empirical evaluation about the information that core inflation may have to predict headline inflation is required. In fact, challenging the prevailing view, there are some interesting arguments suggesting that emphasis on core inflation might not be a good idea. First, one might think that core measures have lower predictive ability than inflation itself because the exclusion of items on which people spend a nontrivial portion of their income and which prices might propagate to others in the economy might weaken the ability of core to predict total inflation. This might be particularly relevant if the persistence of energy and food prices is high. Second, another economic reason for paying attention to core is the belief that is more demand than supply driven and, consequently, more affected by monetary policy actions. Nevertheless, the

crystal clear distinction between demand and supply shocks is at least thin. The incorporation of further processed food in the Consumer Price Index (CPI) baskets with more labor and non-tradable components may have ruined that distinction. Third, and following Bullard (2011b), the logic of relative prices also suggests that changes in energy consumption triggered by price changes could put pressure on all the other prices in the economy. Accordingly, if energy prices go up for a long time, it is plausible to expect that the other prices will go down for a while, which means that core will underestimate total inflation during that period. This implies that core inflation may not be a good predictor of future headline inflation after all. Under these arguments, headline inflation should probably have more weight on policymaking decisions than core inflation.

Central bankers around the world have taken both sides of the debate. The European Central Bank and the Bank of England have an explicit focus on headline measures, and their policymakers pay less attention to core inflation. In contrast, the Federal Open Market Committee (FOMC) focuses on core measures (see Smith, 2012), particularly on core PCE. Following the common practice of the FOMC, many other central banks have taken for granted the importance of core inflation, paying less attention to the truly forecasting ability of those measures on headline inflation.

Differing from our paper, articles exploring the predictive relationship between core and headline inflation focus on one particular country. Besides, different papers use different methodologies and sample periods, which makes it hard to draw a conclusion at an international level. For instance, Le Bihan and Sédillot (2000) analyze the ability of four indicators of underlying inflation to forecast inflation in France. The authors conclude that their out-of-sample results are not very compelling. A fairly similar conclusion is achieved by Freeman (1998) for the US, mentioning that measures of underlying inflation are not very useful for forecasting headline inflation. Bermingham (2007) also addresses the same topic but for the case of Ireland. Differing from the previous two papers, Bermingham does find evidence supporting the usefulness of a core measure when forecasting headline inflation. A similar result is shown by Song (2005) using Australian data. Crone et al. (2013) also analyze this topic for the US finding evidence of long term predictability from core to headline CPI, but not in the case of PCE. More efforts in these directions can be found in Cogley (2002), Khettry and Mester (2006), Kiley (2008), Meyer and Pasaogullari (2010), Smith (2012), Stock and Watson (2015) and Faust and Wright (2013) but also with a focus solely on the US[2].

Our main results from the analysis of 33 different economies indicate that core inflation does have the ability to predict headline inflation in about two thirds of our countries. This share of countries reduces to 40% when

---

[2] Crone et al present a section in which they analyze the different predictive results between core and headline inflation for the US. They mention that the different papers analyzing the topic are fairly heterogeneous but, in general terms, models based on some alternative measure of CPI seems to be a better predictor of future total CPI than models based on total CPI.

predictability is analyzed at policy relevant forecasting horizons. Furthermore, this predictive ability is sizable only for about one quarter of the countries in our sample. For many countries instead, reductions in out-of-sample Root Mean Squared Prediction Errors (RMSPE) are only marginal. These results hold true even when implementing an out-of-sample test of Granger causality especially designed to compare forecasts from nested models. Our results indicate that for some countries we should widen our view to look for other variables that may help to predict inflation, beyond the traditional core measure based on the exclusion of food and energy prices. Multivariate trend inflation measures allowing the inclusion of some food and energy items as in Stock and Watson (2015) might be worth pursuing.

The rest of the paper is organized as follows. In section 2 we describe our data and introduce the econometric setup. In section 3 we present our main empirical results and section 4 concludes.

II.   Data and Econometric Setup

1. Our Data

For our main analysis we consider the Consumer Price Index (CPI) of a total of 33 countries at a monthly frequency. As a measure of core inflation we consider the CPI excluding food and energy. We do not show results for PCE in the US, simply because we want to preserve a uniform analysis to make international comparisons. The data cover the sample period from January 1994 to March 2015. Our set of countries includes all OECD economies with monthly CPI information, plus Peru and Colombia. The set of countries is displayed in Table 1. We obtain the CPI for Chile directly from the National Statistics Institute, which is the government agency in charge of the construction of the CPI. For the rest of the 30 OECD countries we obtain CPI series from the Main Economic Indicators section of the OECD web page. For Peru and Colombia we use their respective Central Banks as source for the data. Our series are not seasonally adjusted.

Table 1
Sample of Countries

Austria	Hungary	Peru
Belgium	Iceland	Poland
Canada	Ireland	Portugal
Colombia	Israel	Slovak Republic
Czech Republic	Italy	Slovenia
Denmark	Japan	Spain
Estonia	Korea, Rep.	Sweden
Finland	Luxembourg	Switzerland
France	Mexico	Turkey
Germany	Netherlands	United Kingdom
Greece	Norway	United States

Our basic unit of analysis corresponds to year-on–year (y-o-y) inflation rate computed according to the following simple expression:

$$\pi_t = 100[Ln(CPI_t) - Ln(CPI_{t-12})]$$

We depart from Stock and Watson (2002), Ciccarelli and Mojon (2010) and others, in that we focus only on forecasting year-on-year inflation rate at different horizons. We also depart from those articles in the construction of multistep ahead forecasts because we use a dynamic or iterated forecasting set up instead of a direct approach. Our reading of Marcellino, Stock and Watson (2005) is that there is no reason to expect superiority of either approach: direct or iterative. The authors point out that under correct specification of the exogenous variables, the iterated approach should be more efficient. Nevertheless, the direct approach should be more robust to model misspecification[3].

With the year-on-year transformation we end up with a total of 243 observations for most of our countries spanning the period January 1995 to March 2015. The only two exceptions are Colombia and Estonia. The sample for Colombia starts in December 1999 and for Estonia starts in January 1999. Consequently, for Colombia we have a total of 184 observations, whereas for Estonia we have 195 observations.

2. Forecast Evaluation Framework

Our basic evaluation strategy considers a couple of univariate specifications for headline inflation that we call benchmark models, and that are described in detail in subsection 3. We evaluate the predictive ability of our benchmark models against their augmented versions with core inflation both in-sample and out-of-sample. To describe the out-of-sample exercise, let us assume that we have a total of T+1 observations of headline inflation (π_t) for a given country. We generate a sequence of P(h) h-step-ahead forecasts estimating the models in either rolling windows of fixed size R or expanding windows of size equal or greater than R. For instance, to generate the first h-step-ahead forecasts using rolling windows, we estimate our models with the first R observations of our sample. Then, these forecasts are built with information available only at time R and are compared to observation π_{R+h}. Next, we estimate our models with the second rolling window of size R that includes observations through R+1. These h-step-ahead forecasts are compared to observation π_{R+h+1}. We iterate until the last forecasts are built using the last R available observations for estimation. These forecasts are compared to observation π_{T+1}. When recursive or expanding windows are used instead, the only difference with the procedure described in previous lines relies on the size of the estimation windows. In the recursive scheme,

---

[3] Pincheira and Gatty (2016) show that the iterative approach generates much more accurate forecasts than the direct approach when forecasting Chilean inflation.

estimation windows size grows with the number of available observations for estimation. For instance, the first h-step ahead forecast is constructed estimating the models in a window of size R, whereas the last h-step-ahead forecasts is constructed based on models estimated in a window of size T+1-h.

We generate a total of P(h) forecasts, with P(h) satisfying R+(P(h)-1)+h=T+1. So

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

Being more specific, we have a total of 243 observations for most of our countries, so T+1= 243. We also set R to either 80 or 100, which means that the total number of forecasts is either P(h)=164-h or P(h)=144-h. In particular we construct a total of 143 one-step-ahead forecasts when R=100 and a total of 163 one-step-ahead forecasts when R=80[4].

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

$$\text{SRMSPE} = \sqrt{\frac{1}{P(h)} \sum_{t=R}^{T+1-h} \left(\pi_{t+h} - \widehat{\pi}_{t+h|t}\right)^2}$$

where SRMSPE stands for "Sample Root Mean Squared Prediction Error" and $\widehat{\pi}_{t+h|t}$ represents the forecast of π_{t+h} made with information known up until time t.

We carry out inference about predictive ability by considering pairwise comparisons between each model and its augmented version. Inference is carried out within the framework developed by Clark and West (2007) (henceforth CW). Their test statistic is mainly aimed at evaluating models in an out-of-sample fashion. With the CW test we evaluate whether the core measures of inflation provide additional information to that already contained in our benchmarks.

The CW test can be considered either as an encompassing test or as an adjusted comparison of Mean Squared Prediction Errors (MSPE). The adjustment is made in order to make a fair comparison between nested models. Intuitively, this test removes a term that introduces noise when a parameter, that should be zero under the null hypothesis of equal MSPE, is estimated.

The core statistic of the Clark and West (2007) test is constructed as follows

---

[4]Let us recall that we have fewer observations for Colombia and Estonia. Notice that we construct the same number of forecasts for these countries: either 164-h or 144-h. The adjustment comes from the number of observations used in the first estimation windows.

$$\hat{z}_{t+h} = \left(\hat{e}_{1,t+h}\right)^2 - \left[\left(\hat{e}_{2,t+h}\right)^2 - \left(\hat{\pi}_{1,t+h|t} - \hat{\pi}_{2,t+h|t}\right)^2\right]$$

Where $\hat{\pi}_{1,t+h|t}$ and $\hat{\pi}_{2,t+h|t}$ denote the h-step ahead forecasts generated from the two models under consideration. Model 1 is the parsimonious or "small" model that is nested in the larger model 2. In other words, model 2 would become model 1 if some of its parameters would be set to zero.

Similarly, $\hat{e}_{1,t+h} = \pi_{t+h} - \hat{\pi}_{1,t+h|t}$ and $\hat{e}_{2,t+h} = \pi_{t+h} - \hat{\pi}_{2,t+h|t}$ represent the corresponding forecast errors.

With some little algebra it is straightforward to show that $\hat{z}_{t+h}$ could also be expressed as follows

$$\text{SMSPE} - \text{Adjusted} = \frac{2}{P(h)} \sum_{t=R}^{T+1-h} \hat{e}_{1,t+h}\left(\hat{e}_{1,t+h} - \hat{e}_{2,t+h}\right) \text{ (1)}$$

This statistic is used to test the following null hypothesis

$$H_0: E(\text{SMSPE} - \text{Adjusted}) = 0$$

against the alternative

$$H_A: E(\text{SMSPE} - \text{Adjusted}) > 0$$

Clark and West (2007) suggest a one sided test for a t-type statistic based upon the core statistic in (1). They recommend asymptotically normal critical values for their test.

It is important to emphasize here that the Clark and West (2007) test is fairly different from the traditional Diebold and Mariano (1995) and West (1996) test (henceforth DMW). One of the most important differences is that they are designed for different purposes. While the DMW test is comparing the accuracy of two different forecasting methods, the CW test evaluates model adequacy. In other words, it is testing whether the larger model is more appropriate than the smaller model. Put differently, the most important difference between the DMW and CW tests relies on the fact that the DMW test is a standard normal test for the differences in MSPE between two models, whereas the CW test is a standard normal test comparing the same MSPE differences but after a very specific adjustment is made. Clark and West (2007) show via simulations that their adjustment generates a test with adequate size and much more power than normal tests comparing unadjusted differences in MSPE, like the DMW test does. They also show that unadjusted tests are severely undersized when comparing nested models.

As we already mentioned, the CW test can be also considered as an encompassing test. This means that it is evaluating whether a particular combination between the model with and without core generates forecasts with the lowest RMSPE between the following strategies: A) forecasting with a univariate benchmark, B) forecasting

with the univariate benchmark augmented with core inflation, or C) forecasting with an average between the strategies in A) and B). Let us elaborate. For a given scalar λ we could build the following convex forecast combination

$$\pi_{t+h|t}^{C} = \lambda \pi_{2,t+h|h} + (1-\lambda)\pi_{1,t+h|h}$$

with forecast error given by

$$e_{t+h|t}^{C} = \lambda e_{2,t+h|h} + (1-\lambda)e_{1,t+h|h} = \lambda\left(e_{2,t+h|h} - e_{1,t+h|h}\right) + e_{1,t+h|h}$$

The corresponding MSPE is given by

$$E(e_{t+h|t}^{C})^2 = \lambda^2 E\left(e_{2,t+h|h} - e_{1,t+h|h}\right)^2 + E(e_{1,t+h|h})^2 + 2\lambda E\left(e_{2,t+h|h} - e_{1,t+h|h}\right)e_{1,t+h|h} \quad (2)$$

As long as

$$E\left(e_{2,t+h|h} - e_{1,t+h|h}\right)^2 > 0$$

Expression 2) is a strictly convex quadratic function with a unique global minimum given by:

$$\lambda^* = \frac{E\left(e_{1,t+h|h} - e_{2,t+h|h}\right)e_{1,t+h|h}}{E\left(e_{2,t+h|h} - e_{1,t+h|h}\right)^2}$$

We notice that under mild conditions the numerator of the CW statistic converges in probability to twice the numerator of λ^*. As long as λ^* is different from either one or zero, the MSPE of the optimal combination should be lower than the MSPE of the two individual forecast in the combination. Rejection of the null hypothesis of the CW statistic indicates that a combination with a positive weight on the forecast incorporating a core measure should be preferable to either individual forecast[5]. In section III we also report empirical results involving the following optimal forecast combination:

$$\pi_{t+h|t}^{C} = \lambda^* \pi_{2,t+h|h} + (1-\lambda^*)\pi_{1,t+h|h}$$

3. Forecasting Approach

Our basic approach considers the comparison of forecasts coming from a benchmark model with forecasts coming from the same benchmark model but augmented with a core inflation measure. We consider the following two main specifications:

---

[5] If the CW statistic cannot reject the null then we have three possibilities: combination gains are negligible, small or they might be obtained with a negative weight on the forecasts with core, which has no simple interpretation.

$$\pi_{t+1} = \alpha + \beta(L)\pi_t^{core} + \varepsilon_t \quad (3.a)$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t - \theta u_{t-1} - \tau u_{t-12} + \tau\theta u_{t-13} \quad (3.b)$$

$$\Delta\pi_{t+1} = \alpha + \vartheta(L)\Delta\pi_t^{core} + \omega_t \quad (4.a)$$

$$\omega_t = \gamma\omega_{t-1} + v_t - \varphi v_{t-1} - \delta v_{t-12} + \delta\varphi v_{t-13} \quad (4.b)$$

Where $\beta(L) = \sum_{j=0}^{q}\beta_j L^j$ and $\vartheta(L) = \sum_{j=0}^{p}\vartheta_j L^j$ represent lag polynomials, L represents the lag operator such that

$$L^j X_t = X_{t-j}$$

and Δ represents the "difference operator" such that

$$\Delta X_t = X_t - X_{t-1}$$

The lag order *q and p* are sometimes set at 12, or in some other exercises are estimated by Bayesian Information Criterion (BIC) with $1 \leq q, p \leq 12$.

For our in-sample analyses we consider both specifications (3a, 3b and 4a, 4b) setting both lag length parameters *p* and *q* to 12 (*p=q=12*). In these in-sample analyses we are only evaluating the relationship between headline inflation at time t+1 and headline and core inflation at time t, so we are only explicitly evaluating one step-ahead-forecasts.

In our out-of-sample analyses we consider more variations. In particular we consider both rolling and expanding windows, two initial estimation windows sizes R=80 and R=100 and also two options for the choice of our lag length parameters *p* and *q*. First we set both lag length parameters to 12, as in our in-sample analysis, but also we allow for an automatic lag selection strategy in each of the estimation windows according to the Bayesian Information Criterion (BIC). Adding up all our different variations we end up with a total of 16 different out-of-sample exercises: 8 with specifications in levels and 8 with specifications in first differences. Finally, notice that in our out-of-sample exercises we explore predictability at several horizons: 1,3,6,9,12,18 and 24 months ahead. To create multi-step ahead forecasts we use the iterated method relying on the following ARIMA specification for core inflation:

$$\pi_t^{core} - \pi_{t-1}^{core} = \alpha(\pi_{t-1}^{core} - \pi_{t-2}^{core}) + \mu_t - a\mu_{t-1} - b\mu_{t-12} + ab\mu_{t-13} \quad (5)$$

Where μ_t is a white noise process.

We notice that univariate versions of the models 3-5 were evaluated in their ability to predict headline inflation for a number of countries in Pincheira and Medel (2015). These models were extremely competitive at short horizons yet similar to the usual benchmarks available in the literature at longer horizons. We also notice that in (4a) and (5) we have imposed a unit root in the models used to generate forecasts for headline and core inflation. This is also in line with important papers in the forecasting literature, see Stock and Watson (2002) and Atkeson and Ohanian (2001) for instance. Besides, Clements and Hendry (2001) and Pincheira and Medel (2012a) provide interesting insights regarding the use of models with unit roots to generate forecasts for stationary variables. We notice also that our specifications in first differences are driftless expressions. That is done on purpose to avoid the presence of deterministic trends in long-run forecasts. To give a simple example, let us consider the case in which we add a drift "c" to expression (5) so we obtain the following new expression (5'):

$$\pi_t^{core} - \pi_{t-1}^{core} = c + \alpha(\pi_{t-1}^{core} - \pi_{t-2}^{core}) + \mu_t - a\mu_{t-1} - b\mu_{t-12} + ab\mu_{t-13} \quad (5')$$

Following Box, Jenkins and Reinsel (2008) the eventual or explicit form of the forecast function for (5') is given by

$$\hat{f}_t(h) = \left[\frac{c}{1-\alpha}\right]h + b_t + a_t(\alpha^h), \text{for } h > 11$$

Here $\hat{f}_t(h)$ is the best linear forecast of f_{t+h} based on information available at time t. Furthermore a_t and b_t represents adaptive coefficients, that is, coefficients that are stochastic and functions of the process at time t. See Box, Jenkins and Reinsel (2008) for details. Expression (5') shows that whenever the drift "c" is different from zero, optimal linear forecasts have a linear deterministic trend with slope c/(1-α). This means that long term forecasts will be divergent, which is not desirable in the case of a process like inflation, which is relatively stable in most of the countries in our sample. This is the reason why we are considering driftless expressions, or expressions with the drift set to zero.

III. Empirical Results

1. In Sample Analysis

Tables 2 and A1 in appendix A show statistics coming from our expressions 3a, 3b and 4a, 4b when estimated in-sample using the whole sample period. In these exercises we set the lag length parameters *p* and *q* to 12. Table A1 shows the F statistic for the null hypothesis that all the coefficients associated to core inflation are zero. It also indicates the P-value associated to these tests. We notice that F-statistics are calculated using the Newey-West (1987) HAC estimator. Table 2 shows aggregated results for these in-sample exercises.

Table 2
Share of countries for which the marginal predictability from core to headline inflation is statistically significant at the 5% level
In Sample-Analysis

Specification in Differences	Specification in Levels
79%	76%

Source: Authors' elaboration

Results in Table 2 strongly support the common wisdom that core is a useful predictor for headline inflation in most of our countries. In fact, in only three countries the null hypothesis of no predictability cannot be rejected at usual significance levels for both specifications (differences and levels): Colombia, Korea and the US. For the rest of the countries, the null of no predictability is either rejected in one or both specifications, typically at extremely high confidence levels. See Table A1 in appendix A for details.

Our in-sample analysis clearly indicates that core inflation does help to predict headline inflation in most of our countries. Nevertheless, results in Table 2 and Table A1 are only analyzing a predictive relationship between time t and time t+1. In other words they analyze one-step-ahead forecasts, which might not be the most important horizon from the monetary policy point of view. Furthermore, in-sample analyses are usually criticized because they are relatively different from a real time forecasting exercise and also because they have shown a tendency to overfit the data. To mitigate these shortcomings, we move next to a multistep ahead out-of-sample analysis.

2. Out-of-Sample Analysis

As a first step, in this section we show results of one out-of-sample test of Granger causality that emerged in recent years. This test is due to Clark and West (2007). Then, as a second step, we show results about forecast accuracy.

a. Granger causality

The fundamental question we are trying to answer in this paper is whether core inflation has the ability to predict headline inflation. Let us be more precise about the way in which we answer this question. First, in principle we are studying predictability for a very wide set of forecasting horizons. We are considering h=1, 3, 6, 9, 12, 18 and 24 months ahead. This means that if we could find evidence of predictability for any of these forecasting horizons, the answer to our fundamental question would be yes: there is predictability. Therefore, we are being

fairly tolerant with respect to the forecasting horizon because evidence of predictability for h=1 would be treated equivalently to evidence of predictability when h=12, unless stated otherwise. Second, our question about the predictive content that core inflation may have to forecast headline inflation is a question about predictive ability beyond that contained in traditional, not naive, univariate benchmarks. Third, the Clark and West (2007) test assumes a null hypothesis of no predictability for core inflation. This means that if we cannot reject the null hypothesis it is either because core inflation has no predictive ability, or because our test has no power to detect it. If this is the case, a possible scenario is one in which predictability of core inflation is subtle, small or unstable, so that our test is not able to adequately process the evidence against the null hypothesis. Fourth, predictability is an exercise-dependent statement. By this we mean that our findings may depend on several aspects of our out-of-sample exercise like: the econometric specification, our dependent variable (expressed in levels or first differences), the number of observations used in our estimation windows, the method we use to update our parameter estimates (rolling or expanding) and the way we are selecting the number of lags in our specifications. Given that in principle our results may depend on all these variations, we consider a total of 16 out-of-sample exercises for each country in our sample. These 16 exercises are the result of considering 8 exercises for inflation in levels, and the same exercises for inflation expressed in first differences. These 8 exercises are divided in two groups of 4. In one group the selection of the number of lags is set fixed to 12 whereas in the other group is selected automatically with BIC. Finally this smaller group of 4 exercises is the result of the combinations of two different estimation window sample sizes (80 and 100) and two different updating schemes: rolling and recursive or expanding.

Table 3 shows two columns with figures. In the first column we see the percentage of countries for which core inflation is statistically significant at the 5% level according to the Clark and West (2007) t-statistic. In this first column we consider all forecasting horizons. In the second column we consider only predictability for horizons between 9 and 18 months, which we consider more relevant in terms of monetary policy decisions. In different rows we have the results for the sixteen out-of samples exercises. So, for instance let us analyze the figure in the upper left corner which is 62.50%. This means that in 62.50% of the countries in our sample the Clark and West (2007) test rejects the null hypothesis of no predictability when we consider a model with specification in levels, with automatic lag selection according to BIC and when an initial estimation window of 80 observations is used jointly with a recursive or expanding window strategy to update the parameter estimates. Next to the 62.50% we see a 46.88% which represents the percentage of countries in which the Clark and West (2007) test rejects the null hypothesis of no predictability when focusing only in our policy relevant forecasting horizons.

The first column in Table 3 averages 63.87%. The second column averages 39.60%. These figures are interesting. They say that in about two thirds of our countries we do find evidence to reject the null of no predictability from core to headline. This figure reduces to about 40% when we focus only on medium term

forecasts ($9 \leq h \leq 18$). While these figures are relatively high, they are lower than those obtained in our in-sample analysis.

To take a wider view of our results we present Table 4. This table shows, for each forecasting horizon and each out-of-sample exercise, the percentage of countries for which core inflation is statistically significant according to the Clark and West (2007) test. Results in Table 4 indicate that when focusing on particular forecasting horizons, core inflation has the ability to predict headline inflation at most in two thirds of our countries. This happens when forecasting one month ahead. If we place our attention on more important forecasting horizons from the policy point of view (like 9, 12 or 18 months ahead) it is in less than a half of our countries that core inflation seems to have predictive ability for headline inflation.

Table 3
Share of countries for which core inflation is statistically significant
at any forecasting horizon

Specification in level	Any horizon	$9 \leq h \leq 18$
Automatic lag selection of core inflation according to BIC		
Recursive Window= 80	62.50%	46.88%
Recursive Window=100	63.64%	42.42%
Rolling Window=80	46.88%	25.00%
Rolling Window=100	63.64%	39.39%
12 lags of core inflation		
Recursive Window= 80	78.13%	46.88%
Recursive Window=100	75.76%	51.52%
Rolling Window=80	59.38%	37.50%
Rolling Window=100	57.58%	30.30%

Specification in first differences	Any horizon	$9 \leq h \leq 18$
Automatic lag selection of core inflation according to BIC		
Recursive Window= 80	69.70%	50.00%
Recursive Window=100	72.73%	48.48%
Rolling Window=80	56.25%	28.13%
Rolling Window=100	66.67%	33.33%
12 lags of core inflation		
Recursive Window= 80	71.88%	46.88%
Recursive Window=100	66.67%	42.42%
Rolling Window=80	50.00%	25.00%
Rolling Window=100	60.61%	39.39%

Column Average					63.87%	39.60%	

Table 4: Share of countries for which core inflation is statistically significant when predicting headline inflation in different out-of-sample exercises

Specification in level	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Automatic lag selection of core according to BIC							
Recursive Window= 80	46.88%	31.25%	31.25%	34.38%	34.38%	40.63%	31.25%
Recursive Window=100	48.48%	24.24%	36.36%	30.30%	27.27%	30.30%	18.18%
Rolling Window=80	31.25%	25.00%	25.00%	25.00%	21.88%	18.75%	18.75%
Rolling Window=100	45.45%	21.21%	33.33%	27.27%	27.27%	30.30%	21.21%
12 lags of core inflation							
Recursive Window= 80	65.63%	43.75%	40.63%	40.63%	40.63%	34.38%	28.13%
Recursive Window=100	60.61%	39.39%	39.39%	45.45%	48.48%	39.39%	27.27%
Rolling Window=80	31.25%	31.25%	28.13%	28.13%	31.25%	25.00%	28.13%
Rolling Window=100	39.39%	21.21%	21.21%	24.24%	30.30%	15.15%	15.15%

Specification in first differences	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Automatic lag selection of core according to BIC							
Recursive Window= 80	50.00%	37.50%	31.25%	40.63%	43.75%	37.50%	28.13%
Recursive Window=100	48.48%	36.36%	30.30%	36.36%	36.36%	42.42%	27.27%
Rolling Window=80	40.63%	31.25%	15.63%	18.75%	18.75%	15.63%	18.75%
Rolling Window=100	48.48%	30.30%	21.21%	24.24%	30.30%	18.18%	21.21%
12 lags of core inflation							
Recursive Window= 80	62.50%	43.75%	28.13%	34.38%	34.38%	40.63%	37.50%
Recursive Window=100	57.58%	45.45%	36.36%	36.36%	36.36%	39.39%	30.30%
Rolling Window=80	34.38%	31.25%	28.13%	25.00%	18.75%	12.50%	18.75%
Rolling Window=100	45.45%	27.27%	24.24%	30.30%	33.33%	30.30%	21.21%
Column Average	**47.28%**	**32.53%**	**29.41%**	**31.34%**	**32.09%**	**29.40%**	**24.45%**

Results in tables 3 and 4 are important. In sharp contrast with our in-sample results, we now see that at policy relevant forecasting horizons, it is in about 40% of the countries that the predictive contribution of core inflation seems to be significant. More interestingly is to notice that results in these tables may be considered benevolent to the potential predictive ability of core inflation. This is so for two reasons. First, Monte Carlo simulations that are detailed in appendix C indicate that simulated critical values at the 10% level are in general higher than our asymptotically normal critical value 1.645. This means that using these simulated critical values we would find even less evidence in favor of the predictive contribution from core to headline. Second, and following Bullard (2011a), as we are only considering univariate benchmarks, it is fairly possible that in some countries core inflation may be contributing significantly to these simple benchmarks for the only reason that is a proxy of some potentially omitted relevant variables.

b. Forecast Accuracy

As we mentioned in section 2, the Clark and West (2007) test can be considered either as an encompassing test or as an adjusted comparison of MSPE. In other words, the Clark and West (2007) test evaluates potential but not raw gains in forecast accuracy. To have a notion of precision, Table 5 shows the percentage of countries in our sample for which the RMSPE of the models with core inflation is lower than the RMSPE of the same model but just excluding core inflation. Figures in Table 5 are similar to those in Table 4 and indicate that at most in two thirds of our countries the RMSPE ratio between models with and without core inflation favors the inclusion of this measure.

Table 5
Share of countries for which models with core inflation display lower RMSPE than the same model but without core inflation

Specification in level	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Automatic lag selection of core according to BIC							
Recursive Window= 80	51.52%	45.45%	48.48%	51.52%	42.42%	45.45%	42.42%
Recursive Window=100	45.45%	39.39%	54.55%	54.55%	54.55%	45.45%	42.42%
Rolling Window=80	51.52%	36.36%	42.42%	45.45%	48.48%	39.39%	39.39%
Rolling Window=100	51.52%	48.48%	57.58%	57.58%	57.58%	48.48%	45.45%

12 lags of core inflation

	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Recursive Window= 80	36.36%	33.33%	42.42%	39.39%	33.33%	30.30%	27.27%
Recursive Window=100	39.39%	33.33%	45.45%	48.48%	45.45%	36.36%	39.39%
Rolling Window=80	12.12%	21.21%	30.30%	33.33%	42.42%	30.30%	48.48%
Rolling Window=100	24.24%	21.21%	27.27%	27.27%	48.48%	33.33%	42.42%

Specification in first differences	**h=1**	**h=3**	**h=6**	**h=9**	**h=12**	**h=18**	**h=24**

Automatic lag selection of core according to BIC

	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Recursive Window= 80	54.55%	57.58%	57.58%	57.58%	54.55%	51.52%	51.52%
Recursive Window=100	54.55%	57.58%	45.45%	60.61%	57.58%	48.48%	45.45%
Rolling Window=80	45.45%	39.39%	30.30%	33.33%	39.39%	33.33%	39.39%
Rolling Window=100	54.55%	51.52%	36.36%	54.55%	63.64%	51.52%	45.45%

12 lags of core inflation

	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Recursive Window= 80	48.48%	39.39%	33.33%	36.36%	42.42%	42.42%	48.48%
Recursive Window=100	39.39%	45.45%	42.42%	51.52%	48.48%	45.45%	45.45%
Rolling Window=80	15.15%	12.12%	18.18%	18.18%	18.18%	12.12%	24.24%
Rolling Window=100	24.24%	18.18%	15.15%	30.30%	36.36%	30.30%	33.33%
Column Average	**40.53%**	**37.50%**	**39.20%**	**43.75%**	**45.83%**	**39.02%**	**41.29%**

Source: Authors' computations

Table 5 shows the percentage of countries for which the inclusion of core inflation renders a lower RMSPE. Nevertheless, a tiny reduction of 0.01% may not be relevant from an economic point of view. We fully take into consideration this observation in Table 6, where we show the share of countries for which the inclusion of core inflation yields a reduction in RMSPE of at least 5%. We acknowledge that the choice of a 5% is totally arbitrary, but the choice of a significance level of 5% or 1% is also entirely subjective. Having said that, let us go back to Table 6. Figures in this table are overwhelmingly low. The highest number is lower than 37%. The average across the table is 19.43% indicating that, on average, one country out of five gets reductions in RMSPE greater than 5% when including core inflation.

Table 6

Share of countries for which models with core inflation display reductions of at least 5% in RMSPE with respect to the same model but without core inflation

Specification in level	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Automatic lag selection of core according to BIC							
Recursive Window= 80	15.15%	18.18%	21.21%	15.15%	18.18%	15.15%	15.15%
Recursive Window=100	12.12%	18.18%	18.18%	21.21%	21.21%	27.27%	24.24%
Rolling Window=80	12.12%	18.18%	24.24%	27.27%	27.27%	27.27%	21.21%
Rolling Window=100	12.12%	15.15%	15.15%	24.24%	24.24%	30.30%	27.27%
12 lags of core inflation							
Recursive Window= 80	21.21%	21.21%	21.21%	21.21%	24.24%	24.24%	21.21%
Recursive Window=100	21.21%	21.21%	15.15%	27.27%	36.36%	27.27%	24.24%
Rolling Window=80	9.09%	15.15%	21.21%	21.21%	24.24%	24.24%	24.24%
Rolling Window=100	12.12%	12.12%	18.18%	21.21%	24.24%	24.24%	24.24%

Specification in first differences	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Automatic lag selection of core according to BIC							
Recursive Window= 80	15.15%	18.18%	27.27%	27.27%	30.30%	21.21%	21.21%
Recursive Window=100	18.18%	15.15%	24.24%	24.24%	24.24%	15.15%	15.15%
Rolling Window=80	15.15%	15.15%	12.12%	12.12%	12.12%	12.12%	12.12%
Rolling Window=100	15.15%	9.09%	9.09%	12.12%	15.15%	12.12%	12.12%
12 lags of core inflation							
Recursive Window= 80	18.18%	24.24%	24.24%	30.30%	27.27%	27.27%	21.21%
Recursive Window=100	18.18%	24.24%	27.27%	30.30%	30.30%	21.21%	18.18%
Rolling Window=80	9.09%	12.12%	15.15%	15.15%	18.18%	12.12%	15.15%
Rolling Window=100	12.12%	9.09%	12.12%	15.15%	18.18%	15.15%	12.12%
Column Average	**14.77%**	**16.67%**	**19.13%**	**21.59%**	**23.48%**	**21.02%**	**19.32%**

Source: Authors' computations

It is also interesting to compare the "best specifications" with and without core inflation. By "best specification" we mean the out-of-sample exercise providing the lowest RMSPE. While in appendix B we show full tables of

the lowest RMSPE across our sixteen out-of-sample exercises with and without core, Table 7 next shows the ratio of RMSPE from the "best specifications" with and without core. Figures below 1 favor specifications including core inflation. Shaded cells highlight situations in which reductions in RMSPE by including core inflation are equal or greater than 5%.

Table 7
RMSPE ratio between the best specification with and without core
Figures below 1 favor specifications with core

Countries	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Austria	0.91	0.90	0.96	0.95	0.98	1.01	1.00
Belgium	0.99	1.00	1.00	1.01	1.01	0.99	1.00
Canada	0.99	1.00	0.98	0.97	0.99	0.96	1.00
Chile	1.01	0.99	0.99	0.99	0.99	1.01	1.00
Colombia	1.02	1.03	1.02	1.03	1.02	1.02	1.03
Czech Republic	1.06	1.05	1.01	1.08	1.14	0.98	0.88
Denmark	1.00	1.00	1.01	1.01	1.00	1.00	1.00
Estonia	1.04	1.04	1.04	1.05	1.06	1.03	1.04
Finland	1.02	1.03	1.03	1.04	1.03	1.00	0.99
France	0.96	1.00	1.00	1.04	1.06	1.03	1.01
Germany	0.99	0.99	0.98	0.99	0.97	0.98	0.99
Greece	0.96	0.97	1.01	0.96	0.95	0.98	0.93
Hungary	1.01	1.01	1.00	1.00	0.98	1.02	0.98
Iceland	1.01	1.00	1.00	0.98	0.98	0.98	0.96
Ireland	1.02	1.01	0.97	0.97	0.95	0.93	0.93
Israel	1.02	1.03	1.01	1.00	0.96	0.94	0.96
Italy	0.98	0.99	0.98	1.00	1.00	1.03	1.04
Japan	1.01	1.03	1.01	1.00	1.00	1.01	1.01
Korea	1.01	1.00	1.01	1.01	1.01	0.99	0.98
Luxembourg	1.01	1.03	1.07	1.04	0.99	1.00	1.03
Mexico	0.90	0.82	0.65	0.61	0.60	0.56	0.53
Netherlands	1.00	1.01	1.01	0.98	0.98	0.98	0.97
Norway	0.97	1.00	1.01	1.01	1.00	1.00	0.97
Peru	0.27	0.31	0.42	0.58	0.73	0.92	1.21
Poland	1.02	0.99	0.97	0.97	1.00	0.98	0.96
Portugal	0.96	0.92	0.97	0.99	1.01	1.01	1.00
Slovak Republic	1.07	1.04	0.99	0.93	0.98	0.88	0.87
Spain	0.99	0.99	1.02	1.01	1.02	0.96	0.96
Sweden	1.01	1.00	1.00	0.95	0.95	1.00	1.00
Switzerland	0.90	0.96	1.06	1.04	1.09	1.07	1.07
Turkey	0.95	0.89	0.99	1.20	1.36	1.53	1.37
United Kingdom	1.01	0.97	0.94	0.93	0.93	0.98	0.98
United States	1.01	1.00	1.00	1.00	1.00	1.00	0.99

Source: Authors' elaboration

A nice summary of the results in Table 7 is given in Table 8 below, which indicates the percentage of countries for which the "best specification" with core yields reductions in RMSPE beyond a given threshold. For clarity of

exposition, let us analyze the column in Table 8 in which we show results when forecasting one year ahead (h=12). The first figure in that column is 55% indicating that in 55% of our countries the RMSPE of the "best specification" with core is lower than the RMSPE of the "best specification" without core. The second figure in that column is 42%, indicating that in 42% of the countries the RMSPE of the "best specification" with core is as least 1% lower than the RMSPE of the "best specification" without core. Finally the last number in that column is 15%, indicating that in 15% of the countries the RMSPE of the "best specification" with core is as least 5% lower than the RMSPE of the "best specification" without core.

From Table 8 we see that in the road from plain reductions in RMSPE to "sizable" reductions in RMSPE we lose most of our countries. In fact, reductions of 5% or more are only achieved by a handful of countries.

Table 8
Share of countries for which the RMSPE ratio between the best specifications with and without core is below a given threshold

Threshold	h=1	h=3	h=6	h=9	h=12	h=18	h=24
%<1	48%	52%	42%	52%	55%	58%	67%
%<0.99	36%	33%	33%	45%	42%	45%	42%
%<0.98	27%	24%	21%	36%	27%	30%	33%
%<0.97	25%	25%	16%	28%	22%	22%	31%
%<0.96	21%	15%	9%	18%	21%	18%	24%
%<0.95	15%	15%	9%	15%	15%	15%	15%

Notes: Table 8 indicates the percentage of countries for which the best specification including core outperforms the best specification without core by a given margin. Different margins are implicitly provided in the first column under the title: Threshold. We consider margins of at least 0%, 1%, 2%, 3%, 4% and 5%.
Source: Authors' elaboration

The analysis in Tables 5-8 is important but somewhat unfair with specifications including core inflation. It is well known that out-of-sample comparisons in RMSPE between nested models are biased towards the model with fewer parameters. This bias is at the center of the discussion in Clark and West (2006, 2007). To overcome this shortcoming we present Table 9. In this table we show the percentage of countries in which the optimal combination between the models with and without core inflation generates reductions in RMSPE of at least 5%. Notice that the optimal combination will always have lower or equal RMSPE than both of the individual models in the combination, which means that reductions in RMSPE are guaranteed (unless the optimal combination is achieved with a factor λ^* either zero or one). Nevertheless, optimal reductions could be either high or fairly low, that is why we explore whether an important number of countries show reductions of at least 5%. As expected, figures in Table 9 are higher than in Table 6. In fact the total average in Table 9 is 26.8% indicating that one out of four countries, on average, could potentially benefit "importantly" in terms of forecast accuracy (reductions in RMSPE of 5% or more) by using the model with core inflation in an optimal combination with the model without core inflation. Despite this improvement, again these results are in sharp contrast with those obtained in

our in-sample analysis and suggest caution at monetary policy makers when using traditional core inflation to anticipate future developments in headline inflation.

Table 10 shows the countries displaying reductions of 5% of more in RMSPE when comparing the lowest RMSPE across our 16 out-of-sample exercises with and without core. In principle the group of countries is fairly heterogeneous. Its members are: Austria, Mexico, Peru, Switzerland, Turkey, Portugal, UK, Slovak Republic, Ireland, Sweden, Israel, Czech Republic and Greece. Only thirteen countries spread in different forecast horizons. If we restrict the analysis to forecasting horizons between six and eighteen months we end up with only eight countries. Basically Switzerland, Turkey, Ireland, Greece and Czech Republic do not make the cut.

Table 9
Share of countries for which the optimal combination between models with and without core inflation display reductions of at least 5% in RMSPE with respect to the model but without core inflation

Specification in level	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Automatic lag selection of core according to BIC							
Recursive Window= 80	21.2%	24.2%	27.3%	33.3%	30.3%	39.4%	0.242
Recursive Window=100	12.1%	21.2%	24.2%	30.3%	45.5%	51.5%	0.364
Rolling Window=80	18.2%	21.2%	24.2%	33.3%	39.4%	42.4%	0.394
Rolling Window=100	12.1%	18.2%	21.2%	27.3%	45.5%	51.5%	0.394
12 lags of core							
Recursive Window= 80	33.3%	27.3%	27.3%	33.3%	39.4%	39.4%	0.333
Recursive Window=100	30.3%	21.2%	24.2%	33.3%	48.5%	39.4%	0.303
Rolling Window=80	15.2%	18.2%	30.3%	33.3%	36.4%	33.3%	0.364
Rolling Window=100	18.2%	15.2%	21.2%	24.2%	36.4%	33.3%	0.364

Specification in first differences	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Automatic lag selection of core according to BIC							
Recursive Window= 80	21.2%	21.2%	27.3%	27.3%	33.3%	33.3%	0.364
Recursive Window=100	18.2%	18.2%	24.2%	24.2%	30.3%	33.3%	0.273
Rolling Window=80	15.2%	21.2%	12.1%	12.1%	15.2%	15.2%	0.152
Rolling Window=100	15.2%	15.2%	12.1%	12.1%	15.2%	15.2%	0.152

12 lags of core

Recursive Window= 80	24.2%	30.3%	24.2%	33.3%	30.3%	39.4%	0.303
Recursive Window=100	27.3%	27.3%	27.3%	36.4%	33.3%	39.4%	0.273
Rolling Window=80	18.2%	15.2%	21.2%	24.2%	24.2%	21.2%	0.212
Rolling Window=100	21.2%	18.2%	15.2%	18.2%	27.3%	21.2%	0.212
Column Average	**20.1%**	**20.8%**	**22.7%**	**27.3%**	**33.2%**	**34.3%**	**29.4%**

Source: Authors' elaboration

Table 10
Countries for which the best specification with core displays reductions in RMSPE of at least 5% with respect to the best specification without core

h=1	h=3	h=6	h=9	h=12	h=18	h=24
Austria	Austria	Mexico	Austria	Mexico	Mexico	Mexico
Mexico	Mexico	Peru	Mexico	Peru	Peru	Slovak Republic
Peru	Peru	UK	Peru	UK	Slovak Republic	Ireland
Switzerland	Portugal		UK	Ireland	Ireland	Czech Republic
Turkey	Turkey		Slovak Republic	Sweden	Israel	Greece

Notes: The best specification is the specification providing the lowest RMSPE across our 16 exercises.
Source: Authors' elaboration

IV. Summary and Concluding Remarks

In this paper we use monthly CPI data for 33 countries during the sample period January 1995-March 2015 to explore whether core inflation has some predictive power for year-on-year headline inflation. Our findings are fairly interesting, considering the fact that the common wisdom posits that core inflation is a natural predictor for headline inflation.

Our out-of-sample results indicate that core inflation does have the ability to predict headline inflation in about two thirds of our countries. This share of countries reduces to 40% when predictability is analyzed at policy relevant forecasting horizons. Furthermore, this predictive ability is sizable only for about one quarter of the countries in our sample. For many countries instead, reductions in out-of-sample RMSPE are only marginal. These results hold true even when implementing an out-of-sample test of Granger causality especially designed to compare forecasts from nested models. Our results indicate that for some countries we should widen our view to look for other variables that may help to predict inflation, beyond the traditional core measure based on the exclusion of food and energy prices.

We think that the set of findings reported in this paper are both interesting and useful for monetary policymakers. Our results confirm that core inflation is an important predictor of headline inflation for a subset of countries in our sample, but also indicate that core inflation does not add much information for prediction in many other countries. This is very important, especially in the context of recent years, in which variables of economic activity, traditionally used to predict inflation in Phillips curve type of models, have lost their predictive power. For countries in which core is not an important predictor for headline inflation, the search for accurate predictors must continue.

Overall, our findings challenge the common wisdom about the ability of core inflation to forecast headline inflation, and suggest, for several countries, a careful weighting of the traditional exclusion of food and energy prices when assessing the size of the monetary stimulus.

V.   References

1.   ATKESON A. and L. OHANIAN (2001). "Are Phillips Curves Useful for Forecasting Inflation" Federal Reserve Bank of Minneapolis, *Quarterly Review* 25 (1), pp. 2-11.

2.   BLANCHARD O., G. DELL'ARICCIA and P. MAURO (2010). "Rethinking Macroeconomic Policy" *Journal of Money, Credit and Banking,* Supplement to Vol. 42, N° 6 pp. 199-215.

3.   BERMINGHAM C. (2007). "How Useful is Core Inflation for Forecasting Headline Inflation?" *The Economic and Social Review*, Vol. 38, N° 3, Winter, pp. 355-377.

4.   BLINDER, A S., and R. REIS. (2005). "Understanding the Greenspan Standard," Prepared for the Federal Reserve Bank of Kansas City symposium, The Greenspan Era: Lessons for the Future, Jackson Hole, Wyoming, August 25-27, 2005.

5.   BOX, G., G. JENKINS and G. REINSEL (2008). "Time Series Analysis: Forecasting and Control". 4th edition, Wiley.

6.   BULLARD, J. (2011a). "Measuring Inflation: The Core is Rotten". Federal Reserve Bank of St. Louis Review, July/August 2011, 93(4), pp. 223-33.

7.   BULLARD, J. (2011b). "President's Message: Headline vs. Core Inflation: A Look at Some Issues". Federal Reserve Bank of St. Louis.

8.   CICCARELLI, M. and B. MOJON, (2010). "Global Inflation." *TheReview of Economicsand Statistics* 92(3), pp. 524-535.

9.   CLARK, T. and K. WEST (2006). "Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis." *Journal of Econometrics* 135(1-2), pp. 155-186

10.   CLARK, T. and K. WEST (2007). "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models." *Journal of Econometrics* 138, pp. 291-311.

11.   CLARK, T. and M. McCRACKEN (2001). "Tests of equal forecast accuracy and encompassing for nested models". *Journal of Econometrics* 105, 85–110.

12. CLARK, T. and M. McCRACKEN (2005). "Evaluating direct multistep forecasts". *Econometric Reviews* 24, 369–404.

13. CLARK, T. (2001). "Comparing Measures of Core Inflation". Federal Reserve Bank of Kansas City Economic Review. Second Quarter 2001. pp. 5-31.

14. CLEMENTS, M. and D. HENDRY (2001). "Forecasting with difference-stationary and trend-stationary models". *Econometrics Journal* (4), pp. S1-S19.

15. COGLEY, T. (2002) "A Simple Adaptive Measure of Core Inflation." *Journal of Money, Credit, and Banking*, 34, 94–113.

16. CRONE T., N. KHETTRY, L. MESTER and J. NOVAK (2013). "Core Measures of Inflation as Predictors of Total Inflation". *Journal of Money, Credit and Banking* Vol. 45, N° 2-3, pp. 505-519.

17. DIEBOLD, F. and R. MARIANO (1995). "Comparing Predictive Accuracy". *Journal of Business and Economic Statistics* 13(3). pp. 253-263.

18. FAUST, J. and J. WRIGHT (2013). "Forecasting Inflation". Chapter 1 in Handbook of Economic Forecasting Vol. 2. pp. 2-56.

19. FREEMAN, D. (1998). "Do core inflation measures help forecast inflation?" *Economics Letters* 58, 143–147.

20. KHETTRY N. and L. MESTER (2006). "Core Inflation as a Predictor of Total Inflation". Research Rap-Special Report, Research Department, Federal Reserve Bank of Philadelphia. April, 2006.

21. KILEY, M. (2008) "Estimating the Common Trend Rate of Inflation for Consumer Prices and Consumer Prices Excluding Food and Energy Prices." Finance and Economics Discussion Series Paper 2008-38, Board of Governors of the Federal Reserve System.

22. LE BIHAN H. and F. SÉDILLOT (2000). "Do core inflation measures help forecast inflation? Out-of-sample evidence from French data" *Economics Letters* 69 (2000) 261–266.

23. LEI LEI SONG (2005), "Do underlying measures of inflation outperform headline rates? Evidence from Australian Data. *Applied Economics, Taylor & Francis Journals*, vol. 37(3), pp. 339-345.

24. MARCELLINO, M., J. STOCK and M. WATSON (2006). "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series," *Journal of Econometrics* 127 (1-2): 499-526.

25. MEYER, B. and M. PASAOGULLARI. (2010) "Simple Ways to Forecast Inflation: What Works Best?" Economic Commentary No. 2010–17, Federal Reserve Bank of Cleveland.

26. NEWEY, W.K. and K. WEST (1987). "A Simple, Positive, Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703-8

27. PINCHEIRA, P. and C.A. MEDEL (2012a). "Forecasting Inflation with a Random Walk," Working Paper N°669 Central Bank of Chile.

28. PINCHEIRA, P. and C.A. MEDEL (2015). "Forecasting Inflation with a Simple and Accurate Benchmark: The Case of the US and a Set of Inflation Targeting Countries". *Finance a uvěr-Czech Journal of Economics and Finance*, 65, 2015, no. 1.

29. PINCHEIRA, P. and. A. GATTY (2015). "Forecasting Chilean Inflation with International Factors" *Empirical Economics*, forthcoming.

|23

30. RICH, R. and C. STEINDEL (2007). "A Comparison of Measures of Core Inflation". *FRBNY Economic Policy Review*. December.

31. ROBALO C., P. DUARTE and L. MORAIS (2003). "Evaluating core inflation indicators". *Economic Modelling* 20 (2003) pp. 765-775.

32. SMITH, J. K. (2012) "PCE Inflation and Core Inflation." Working Paper 1203, Federal Reserve Bank of Dallas, January 2010.

33. STOCK J. and M. WATSON (1999). "Forecasting Inflation". *Journal of Monetary Economics* 44, pp. 293-335.

34. STOCK J. and M. WATSON (2002). "Macroeconomic forecasting using diffusion indexes". *Journal of Business and Economic Statistics* 20, pp. 147-162.

35. STOCK J. and M. WATSON (2008). "Phillips Curve Inflation Forecasts". NBER Working Papers 14322. National Bureau of Economic Research, Inc.

36. STOCK J. and M. WATSON (2015). "Core Inflation and Trend Inflation." NBER Working Paper N°21282

37. WEST, K. (1996), "Asymptotic Inference about Predictive Ability" *Econometrica* 64(5) pp.1067-84.

Appendix A: In Sample Analysis

Table A1 below shows the F statistics and their P-values corresponding to the null hypothesis of no marginal predictability from core to headline. The first two columns correspond to expressions 3a and 3b in the text, whereas the last two columns correspond to expressions 4a and 4b. Differing from Table 2 in the text, which presents aggregated results, we present in Table A1 detailed results for each country.

Table A1
The Ability of Core Inflation to Predict Headline Inflation (1995:01 – 2015:03)
In-Sample Analysis

Country	Specification in Differences		Specification in Levels	
	F-Stat	P-Value	F-Stat	P-Value
Austria	8.640	0.000	6.674	0.000
Belgium	2.995	0.001	1.991	0.026
Canada	0.712	0.739	2.367	0.007
Chile	1.486	0.131	1.967	0.028
Colombia	1.118	0.349	0.965	0.484
Czech Republic	4.573	0.000	4.270	0.000
Denmark	6.789	0.000	5.430	0.000
Estonia	1.583	0.101	1.698	0.071
Finland	2.200	0.013	2.504	0.004
France	13.181	0.000	13.982	0.000
Germany	3.139	0.000	2.498	0.004

Greece	**16.103**	**0.000**	**15.441**	**0.000**
Hungary	**3.367**	**0.000**	**1.589**	**0.096**
Iceland	**4.437**	**0.000**	**4.517**	**0.000**
Ireland	**3.198**	**0.000**	**2.793**	**0.002**
Israel	**2.636**	**0.003**	**1.363**	**0.186**
Italy	**4.984**	**0.000**	**5.272**	**0.000**
Japan	**2.276**	**0.010**	**2.738**	**0.002**
Korea	0.859	0.589	0.901	0.547
Luxembourg	**11.083**	**0.000**	**11.503**	**0.000**
Mexico	**36.280**	**0.000**	**7.424**	**0.000**
Netherlands	**4.567**	**0.000**	**4.370**	**0.000**
Norway	**5.700**	**0.000**	**5.488**	**0.000**
Peru	**249.258**	**0.000**	**216.229**	**0.000**
Poland	**3.998**	**0.000**	**3.694**	**0.000**
Portugal	**14.753**	**0.000**	**12.868**	**0.000**
Slovak Republic	**3.235**	**0.000**	1.421	0.158
Spain	**15.376**	**0.000**	**17.595**	**0.000**
Sweden	**1.849**	**0.042**	**2.254**	**0.011**
Switzerland	**14.315**	**0.000**	**16.062**	**0.000**
Turkey	**129.347**	**0.000**	**1137.095**	**0.000**
United Kingdom	**1.582**	**0.098**	1.080	0.378
United States	0.607	0.835	0.688	0.763

Notes: Table A1 presents the F-statistics for the null hypothesis of no predictability from core to headline inflation. The sample period is (1995:01 – 2015:03). The dependent variable is: either the level of inflation or its first difference: $\pi_t - \pi_{t-1}$. The F-statistics are constructed using HAC standard errors according to Newey and West (1987).

Source: Authors' elaboration

Appendix B: Forecast Accuracy

Tables B1, B2 display the lowest RMSPE across our sixteen out-of-sample exercises for specifications with core (Table B1) and without core (Table B2).

Table B1
Lowest RMSPE across sixteen out-of-sample exercises with core inflation

Countries	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Austria	0.23	0.40	0.60	0.78	0.90	0.92	0.90
Belgium	0.27	0.56	0.91	1.22	1.45	1.42	1.44
Canada	0.31	0.56	0.72	0.82	0.89	0.85	0.86
Chile	0.39	0.94	1.62	2.12	2.51	2.62	2.62
Colombia	0.22	0.58	0.94	1.23	1.49	1.68	1.85
Czech Republic	0.37	0.73	1.10	1.42	1.71	1.67	1.65

	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Denmark	0.21	0.41	0.62	0.79	0.94	0.95	0.94
Estonia	0.48	1.06	1.82	2.63	3.40	3.67	3.54
Finland	0.27	0.52	0.83	1.09	1.34	1.43	1.44
France	0.20	0.38	0.58	0.76	0.89	0.91	0.89
Germany	0.24	0.38	0.51	0.65	0.76	0.78	0.80
Greece	0.42	0.72	1.08	1.37	1.68	1.90	2.02
Hungary	0.43	0.91	1.52	2.02	2.50	2.72	2.78
Iceland	0.53	1.22	2.07	2.84	3.71	4.07	4.20
Ireland	0.31	0.73	1.36	2.01	2.57	3.01	3.22
Israel	0.37	0.84	1.30	1.65	1.86	1.84	1.82
Italy	0.16	0.32	0.56	0.80	0.99	1.02	1.05
Japan	0.27	0.54	0.79	0.99	1.17	1.22	1.24
Korea	0.27	0.53	0.76	0.93	1.12	1.16	1.17
Luxembourg	0.28	0.53	0.81	0.98	1.07	1.07	1.12
Mexico	0.22	0.54	0.72	0.89	1.15	1.31	1.43
Netherlands	0.26	0.45	0.62	0.73	0.85	0.87	0.85
Norway	0.38	0.71	0.90	0.99	1.09	1.07	1.00
Peru	0.19	0.43	0.84	1.39	1.97	2.57	3.46
Poland	0.31	0.72	1.18	1.52	1.89	1.99	2.04
Portugal	0.32	0.58	0.90	1.22	1.52	1.70	1.81
Slovak Republic	0.45	0.93	1.45	1.85	2.26	2.03	2.19
Spain	0.30	0.67	1.02	1.30	1.54	1.55	1.57
Sweden	0.30	0.55	0.88	1.10	1.32	1.45	1.44
Switzerland	0.26	0.50	0.73	0.89	0.98	0.95	0.96
Turkey	0.72	1.57	2.97	4.82	6.56	8.68	10.29
United Kingdom	0.24	0.48	0.74	0.93	1.09	1.15	1.22
United States	0.33	0.78	1.12	1.30	1.43	1.41	1.41

Notes: Table B1 presents the lowest RMSPE coming from sixteen different out-of-sample strategies. All these strategies are based on univariate models augmented with core inflation. These exercises are described in detail in section 3.

Source: Authors' elaboration

Table B2
Lowest RMSPE across sixteen out-of-sample exercises without core inflation

Countries	h=1	h=3	h=6	h=9	h=12	h=18	h=24
Austria	0.25	0.44	0.62	0.82	0.92	0.91	0.89
Belgium	0.27	0.56	0.91	1.21	1.44	1.43	1.44
Canada	0.32	0.57	0.73	0.84	0.90	0.88	0.86
Chile	0.39	0.95	1.63	2.15	2.52	2.59	2.64
Colombia	0.22	0.57	0.93	1.20	1.46	1.64	1.79
Czech Republic	0.35	0.70	1.09	1.31	1.51	1.71	1.87
Denmark	0.21	0.41	0.61	0.79	0.94	0.95	0.94
Estonia	0.46	1.02	1.76	2.50	3.20	3.56	3.39
Finland	0.26	0.50	0.81	1.05	1.31	1.42	1.44
France	0.21	0.38	0.58	0.73	0.84	0.89	0.88
Germany	0.24	0.38	0.52	0.66	0.78	0.80	0.80
Greece	0.43	0.74	1.08	1.43	1.76	1.93	2.18
Hungary	0.43	0.91	1.51	2.02	2.54	2.66	2.84
Iceland	0.53	1.22	2.07	2.90	3.78	4.14	4.37
Ireland	0.30	0.72	1.40	2.09	2.71	3.23	3.47

Israel	0.37	0.81	1.28	1.65	1.94	1.95	1.90
Italy	0.16	0.32	0.57	0.80	0.99	0.99	1.00
Japan	0.27	0.52	0.78	0.98	1.17	1.20	1.22
Korea	0.27	0.53	0.75	0.91	1.10	1.18	1.20
Luxembourg	0.28	0.52	0.76	0.94	1.08	1.07	1.09
Mexico	0.25	0.66	1.11	1.47	1.92	2.36	2.69
Netherlands	0.26	0.44	0.61	0.74	0.86	0.89	0.87
Norway	0.40	0.71	0.88	0.99	1.09	1.06	1.03
Peru	0.70	1.40	2.00	2.41	2.69	2.78	2.85
Poland	0.30	0.73	1.22	1.57	1.88	2.02	2.13
Portugal	0.33	0.63	0.92	1.24	1.51	1.69	1.81
Slovak Republic	0.42	0.89	1.46	2.00	2.30	2.31	2.51
Spain	0.30	0.68	1.00	1.29	1.52	1.62	1.64
Sweden	0.30	0.55	0.88	1.15	1.39	1.45	1.44
Switzerland	0.29	0.52	0.69	0.85	0.90	0.89	0.90
Turkey	0.76	1.77	2.99	4.01	4.80	5.67	7.49
United Kingdom	0.24	0.50	0.79	1.00	1.18	1.17	1.24
United States	0.33	0.77	1.12	1.30	1.42	1.41	1.42

Notes: Table B2 presents the lowest RMSPE coming from sixteen different out-of-sample strategies. All these strategies are based on univariate models. These exercises are described in detail in section 3.
Source: Authors' elaboration

Appendix C: Simulation of Critical Values

In this paper we carry out inference about predictive ability according to the test developed by Clark and West (2007). While the asymptotic distribution of the test is not normal, Clark and West (2007) advocate the work of Clark and McCracken (2001, 2005) to argue that critical values coming from a standard normal distribution work well in sufficiently large samples. Furthermore, they provide a number of simulations which confirm the original claim that standard normal critical values are indeed adequate. Both the simulation evidence and the work by Clark and McCracken (2001, 2005) are based on the construction of direct multistep forecasts so, in principle, the extension of these results to multistep ahead forecasts coming from the iterative method is not guaranteed.

With this in mind we carry out our own simulations to check the behavior of normal critical values in this environment (iterated multistep forecasts). We generate pseudo inflation observations under the null hypothesis that core inflation is not part of the data generating process for headline inflation. Accordingly, we use the univariate expressions (C1, C2, C3, C4) to generate both inflation processes (headline and core) as follows:

$$\pi_{t+1} = \alpha + \varepsilon_t \quad (C1)$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t - \theta u_{t-1} - \tau u_{t-12} + \tau\theta u_{t-13} \quad (C2)$$

$$\pi_t^{core} = \delta + \omega_t \quad (C3)$$

$$\omega_t = \varphi\omega_{t-1} + \mu_t - a\mu_{t-1} - b\mu_{t-12} + ab\mu_{t-13} \quad \text{(C4)}$$

We calibrate these two processes to match in-sample estimates for one of the countries in our sample: UK. Table C1 shows estimates of our parameters.

Table C1
Parameters Used in our Simulations

Headline Parameters	Core Parameters
$\alpha=2.5804$	$\delta=1.4738$
$\rho=0.993$	$\varphi=0.987$
$\theta=-0.122$	$a=0.067$
$\tau=0.920$	$b=0.667$

Finally, innovations u_t and μ_t are assumed to be normally distributed with the following variance-covariance matrix

$$\text{V=}\begin{pmatrix} 0.046027 & 0.027833 \\ 0.027833 & 0.044598 \end{pmatrix}$$

which is consistent with a correlation coefficient of 0.61 between headline and core innovations.

In each of our 2500 independent replications, we generate pseudo headline and core inflation series of 743 observations each. We remove the first 500 values to finally work with the last 243 observations, which is the actual sample size in most of our empirical exercises. With this pseudo sample we mimic the exercise described in section 2 and compute out-of-sample forecasts generated with and without core inflation according to expressions (3a, 3b, 4a, 4b) and (5). We use recursive windows of initial sample size equal to 100 observations. Next, we compute the Clark and West (2007) t-statistic at each forecasting horizon h=1, 2, 3, 4, 5, 6, 9, 12 and 24. We do this for 2500 replications. We sort the corresponding t-statistics and define the 10% critical value as the 90th percentile of their empirical distribution, the 5% critical value as the 95th percentile of their empirical distribution and the 1% critical value as the 99th percentile of their empirical distribution. We construct these critical values when forecasting using expressions (3a, 3b) and (5) (specification in levels for headline inflation) and also when using expressions (4a, 4b) and (5) (specification in differences for headline inflation). Tables C2, C3 show the critical values for both specifications: levels and first differences.

Table C2
Simulated Critical Values for the Clark and West (2007) test
Specification in Levels

Forecasting horizon	10% Significance Level	5% Significance Level	1% Significance Level
h=1	1.677799	2.020880	2.621016
h=2	1.701031	2.004402	2.660423
h=3	1.724257	2.094449	2.841970
h=4	1.895335	2.311611	3.032057
h=5	2.077425	2.521734	3.298822
h=6	2.241740	2.744843	3.451766
h=9	2.291249	2.792350	3.629083
h=12	2.277567	2.730035	3.613442
h=24	2.100612	2.537843	3.296637

Table C3
Simulated Critical Values for the Clark and West (2007) test
Specification in Differences

Forecasting horizon	10% Significance Level	5% Significance Level	1% Significance Level
h=1	1.755990	2.129223	2.747964
h=2	1.657149	2.026913	2.707084
h=3	1.647164	2.023477	2.773199
h=4	1.692398	2.090552	3.023605
h=5	1.790443	2.260685	3.201712
h=6	1.953402	2.474727	3.342456
h=9	1.823973	2.283940	3.055948
h=12	1.818356	2.214600	3.077975
h=24	1.848447	2.263118	3.132431

Figures in tables C2 and C3 are greater than the corresponding asymptotically normal critical value at the 10%, 5% and 1% significance levels. Let us recall that these asymptotical critical values are 1.282, 1.645 and 2.32 respectively. Consequently, our simulated critical values suggest that usage of standard normal critical values is not entirely adequate because they might generate an oversized test. Furthermore, simulated critical values seem to be a function of the forecasting horizon, which suggest that an adequate critical value for h=1, for instance, might not be adequate for h=24.

Interestingly, the lowest simulated critical value in tables C2 and C3 is close to 1.645. Consequently, we have decided to use that critical value in our empirical exercises. There are two ways in which we could justify this choice. First, under the assumption of asymptotic normality for the Clark and West (2007) test, 1.645 corresponds to a 5% critical value in one-sided tests. Second, simulations in Tables C2 and C3 indicate that the type I error is greater or equal than 10% when 1.645 is used, so we probably have more rejections of the null of no predictability than in the case of a normally distributed test. Let us recall that the common wisdom posits that core measures do have the ability to predict headline inflation. Given our simulations, it seems to us that the choice of 1.645 as a critical value places a lot of weight in favor of this common belief and from that point of view it provides insurance against easy calls against the common wisdom.

Simulations in tables B2 and B3 use data generating processes calibrated to UK data. Nevertheless, similar results are found when the same DGPs are calibrated to US data. (results are available upon request). We acknowledge, however, that much more research needs to be done to fully understand the behavior of the Clark and West (2007) test in small samples when multistep ahead forecasts are computed using the iterated strategy.