Rethinking the choice between core and headline inflation: Does it matter which one to target?

Chi-Young Choi
University of Texas at Arlington

Jianning Kong*
Shandong University

Donggyu Sul
University of Texas at Dallas

December 3, 2016

Abstract

This study addresses two important questions relating to the appropriate measure of inflation for monetary policy purposes: (1) whether core inflation is statistically different from headline inflation in estimating underlying inflation; and (2) whether core inflation is a better predictor of future headline inflation. Our answer to both questions is 'No'. We find that headline inflation is not statistically different from popular alternative measures of inflation, such as core or trimmed-mean inflation in providing information on underlying inflation. Notwithstanding the statistical equivalence, however, core inflation is still preferred due to its lower volatility than headline inflation, reinforcing the original view. When it comes to forecasting future headline inflation, we find that core inflation is not necessarily superior to headline inflation within the framework of the conventional univariate forecasting model. Moreover, our results suggest that multivariate forecasting models deliver significantly more accurate forecasts of headline inflation than their univariate counterparts.

Keywords: Headline inflation, core inflation, monetary policy, weights, forecasting.

JEL Classification Numbers: E31,E37

*Corresponding author. E-mail: Jianningkong@gmail.com. We benefited from helpful comments from Róisín O'Sullivan, the seminar participants at DePaul University, and the session participants at the Southern Economic Association Meeting in Washington D.C. We are also grateful to Mark Watson for sharing the Matlab codes. Any errors or omissions are the responsibility of the authors.

1 Introduction

What is the more appropriate measure of inflation for monetary policy purposes? Although headline (or overall) inflation has been a typical goal variable of monetary policy with regard to prices[1], it is often claimed in the literature (Bernanke et al. 1999, Bryan and Cecchetti 1994, Mishkin 2007, to cite just a few) that a subset like core inflation that strips of highly volatile food and energy prices is a preferred measure of inflation for monetary policy. Popular rationales for this argument range from the compatibility of core inflation with the basic propositions of monetary theory to its ability to predict future headline inflation. Bryan and Cecchetti (1994), for instance, favor the use of core inflation not just because the highly volatile movements in food and energy prices are typically driven by supply-side shocks that do not constitute underlying monetary inflation, but because the high volatilities of those prices are out of the control of the monetary authorities. Moreover, core inflation is often viewed as a better predictor of future inflation possibly because it measures the more persistent influences of the *underlying* inflation that is of ultimate interest to policymakers (e.g., Clark 2001, Cogley 2002, Smith 2004, Wynne 1999,2008).[2] Given the forward-looking nature of monetary policymaking due to non-negligible time lags between policy action and policy effect, this feature of core inflation has been a particular appeal to policymakers.

Recently, however, this view has been challenged by some researchers (e.g., Bullard 2011, Crone et al. 2013) who maintain that the empirical evidence in support of core inflation is inconclusive, if not misleading. As highlighted by Crone et al. (2013), empirical results vary considerably across studies depending on the forecasting model, the sample period and the measures of inflation used. As such, despite the growing evidence on the basic empirical facts, the debate regarding the appropriate measure of inflation remains largely unsettled. More importantly, previous research has little to say about whether different measures of inflation are directly comparable on an equal footing in a statistical sense. If they are statistically different measures, then the superiority of one measure over another cannot be properly evaluated because they capture different underlying

[1] For example, the Fed has adopted an explicit target of 2% PCE headline inflation since 2012 ($https://www.federalreserve.gov/faqs/economy_{1}4400.htm$).

[2] Although the term 'underlying inflation' is widely used in the literature, there does not seem to be a well-defined or widely agreed upon definition of what exactly underlying inflation is. Instead, there seem to be as many definitions of underlying inflation as there are researchers who have studied the topic. Nevertheless, Sweden's Riksbank defines it as "inflation that measures the more lasting inflation rate, or inflation trend ..." and the Parliament of Australia defines it as "inflation which measures the inflationary pressures in the economy that are predominantly due to market forces, i.e. changes in prices that reflect only the supply and demand conditions in the economy." To our knowledge, neither the Fed nor the ECB has a formal definition of underlying inflation. With this in mind, we focus on the fundamental part of monetary inflation throughout the paper.

inflation rates.

The primary objective of this study is to provide further insight on these unsettled issues from a couple of statistical perspectives. We first explore whether or not core inflation is indeed different from headline inflation in a *statistical* sense. To this end, we develop a tool for testing the statistical equivalence of different measures of inflation by exploiting the information embedded in the correlation between price changes of subaggregate inflation and the weights given to the subcomponents of the overall headline inflation. Since headline inflation is a weighted average of many subcomponents, excluding a subset of it could lead to a mismeasurement of the true underlying inflation if their weights vary with price changes in individual products. For example, in the PCE inflation where the weights of individual components are allowed to change over time by the change in their relative expenditure shares, the weight of a product would rise if the product is purchased relatively more after its price drops. In this case, core inflation will not be statistically equivalent to headline inflation in measuring the true underlying inflation (denoted as π_t throughout the paper) because it excludes some weight-varying subcomponents. In consequence, competing measures of inflation cannot be compared on equal footing. By contrast, if the weights of individual products are uncorrelated with price changes in the lack of sufficient substitution effect, different measures of inflation will not be statistically different as they are consistently estimating π_t. In fact, using monthly PCE data for the period 1977.M1 to 2015.M8, we find little evidence that headline inflation is statistically different either from core inflation or from a simple cross-sectional mean of disaggregate inflation series.[3] The core inflation, however, is still preferred for the monetary policy due to its lower volatility compared to the headline inflation.

Second, in terms of the performance of forecasting the future headline of inflation, we fail to find any compelling evidence on the dominance of core inflation over headline inflation in the conventionally popular univariate forecasting models. Moreover, our analysis reveals that multivariate forecasting models based on rolling regression tend to enhance greatly the precision of forecasting headline inflation, compared to their univariate counterparts. Interestingly, including core inflation in forecasting models along with headline inflation and its subaggregate series helps to deliver more accurate forecasts of future headline inflation, regardless of the subsample periods. This comes as surprise in light of the common view embedded in the literature that there is no reliable multivariate models for forecasting inflation. However, it corroborates the more recent finding by Stock and Watson (forthcoming) that multivariate models are superior to univariate models in the estimation of trend inflation.

[3] To conserve space, we provide here the results for the monthly data only, while leaving the results for the quarterly data to the online Appendix (www.utdallas.edu/~d.sul/papers/online_appendix_July22_2016.xls).

The remainder of the paper proceeds as follows. The next section briefly outlines the statistical concept of headline and core inflation in relation to the underlying inflation. This section also introduces a statistical tool to test whether or not they are statistically different. Section 3 compares the out-of-sample forecasting performance of diverse measures of inflation in both univariate and multivariate frameworks. Section 4 concludes the paper.

2 Testing statistical equivalence of different measures of inflation

Throughout the paper, we focus on the PCE inflation with 178 disaggregate components as the benchmark to investigate whether or not the headline PCE inflation is statistically different from two alternative competing measures of inflation in estimating underlying inflation: (i) PCE inflation excluding food and energy items (henceforth, CORE inflation); and (ii) the Federal Reserve Bank of Dallas's trimmed-mean PCE inflation (hereafter, TRIM inflation). Our scheme here is to compare each of these inflation measures with a simple cross-section mean.

Since headline inflation is a weighted sample average where the weights are based on the real expenditure, it can be written as

$$\pi_t^h = \left(\sum_{i=1}^{n} \omega_{it}\pi_{it}\right)\left(\sum_{i=1}^{n} \omega_{it}\right)^{-1} = \sum_{i=1}^{n} \tilde{\omega}_{it}\pi_{it},$$

where $\pi_{it} = (P_{it} - P_{it-1})/P_{it-1} \simeq \Delta \ln P_{it}$ is the inflation rate for item i at time t and $\tilde{\omega}_{it} = \omega_{it}/\sum_{i=1}^{n}\omega_{it}$ is its weight in the PCE inflation. Notice that the PCE inflation is based on the Fisher-Ideal index, which is a geometric average of the *Laspeyres* and the *Paasche* indexes, and hence higher weights are assigned to more purchased items. The CORE inflation (π_t^{CORE}) is then computed by imposing the restriction of $\omega_{jt} = 0$ for $j \in \{\text{food, energy}\}$ at each t such that the excluded items are largely fixed over time. By contrast, the TRIM inflation (π_t^{TRIM}) removes fat-tailed outliers that can vary in each period.[4]

A critical issue with regard to these alternative measures of inflation is whether or not excluding certain outliers, regardless of the same components over time, leads to a mismeasurement of the true underlying inflation that is unobservable. The answer depends on the correlation between subcomponent inflation rates and their weights. If the weights are uncorrelated with

[4] Although CORE inflation is sometimes broadly defined to include TRIM inflation (e.g., Clark 2001), here we refer CORE inflation to the PCE excluding food and energy. According to the FRB Dallas's website (http://www.dallasfed.org/research/pce/descr.cfm), the current trimmed-mean PCE inflation is calculated by discarding 24 percent of the weight from the lower tail and 31 percent of the weight in the upper tail (see also Dolmas 2009). As such, the measure deals with not just a fat-tail but also an asymmetry in the distribution of price changes.

individual inflation rates, then headline inflation will consistently estimate the underlying inflation like CORE or TRIM inflation (e.g., Bickel and Lehmann 1975). In this vein, testing whether or not headline inflation is statistically different from CORE or TRIM inflation is equivalent to testing whether the weights of subcomponents are correlated with individual inflation rates.

To be specific, if weights ($\tilde{\omega}_{i,t}$) are uncorrelated with subcomponent inflation (π_{it}), then $\mathrm{Cov}(\tilde{\omega}_{i,t}, \pi_{it}) = \mathbb{E}(\tilde{\omega}_{i,t} - \mathbb{E}\tilde{\omega}_{i,t})(\pi_{it} - \mathbb{E}\pi_{it}) = 0$. Given that $\sum_{i=1}^{n} \tilde{\omega}_{i,t} = 1$ and $\mathbb{E}(\sum \tilde{\omega}_{i,t})/n = 1/n$, we can set up the following null hypothesis that weights are uncorrelated with subcomponent inflation,

$$\mathcal{H}_0 : \mathbb{E} \sum_{i=1}^{n} (\tilde{\omega}_{i,t} - 1/n) (\pi_{it} - \frac{1}{n} \sum_{i=1}^{n} \pi_{it}) = 0,$$

which can be rewritten as

$$\mathcal{H}_0 : \mathbb{E} \left(\sum_{i=1}^{n} \tilde{\omega}_{i,t} \pi_{it} \right) = \mathbb{E} \left(\sum_{i=1}^{n} \pi_{it}/n \right). \tag{1}$$

That is, no correlation between weights ($\tilde{\omega}_{it}$) and π_{it} implies that the expected weighted average of subcomponent inflation is equal to the expected cross-sectional mean. To simplify it,

$$\mathcal{H}_0 : \mathbb{E} z_t = 0,$$

where

$$z_t = \sum_{i=1}^{n} \tilde{\omega}_{i,t} \pi_{it} - \sum_{i=1}^{n} \pi_{it}/n = \sum_{i=1}^{n} (\tilde{\omega}_{i,t} - 1/n) \pi_{it}.$$

Under the null hypothesis, the following t-test can be formulated

$$t_z = \frac{z_t}{\sqrt{\hat{\Sigma}_z^2}} \to^d \mathcal{N}(0, 1), \tag{2}$$

where $\hat{\Sigma}_z^2$ is the estimated sample variance of z_t which is defined as

$$\hat{\Sigma}_z^2 = \hat{\Sigma}_{\pi,t}^2 \sum_{i=1}^{n} (\tilde{\omega}_{i,t} - 1/n)^2,$$

where $\hat{\Sigma}_{\pi,t}^2$ is the sample variance of π_{it} at time t. It is straightforward to show that the limiting distribution of t_z is the standard normal if π_{it} is cross-sectionally independent. If either π_{it} or $\tilde{\omega}_{i,t}$ are cross-sectionally dependent, however, the limiting distribution of the t-statistics will not be the standard normal. This leads us to adopt a nonparameteric sieve bootstrap approach which is known to be useful for approximating the sampling distribution (e.g., Chang 2004). This bootstrap method is particularly well suited to our case in which the correlation between weights and individual inflation is unknown on a priori ground.

Before turning to the bootstrap procedure, we assume that individual inflation rates are approximated by the following static factor model

$$\begin{bmatrix} \pi_{it} \\ \Delta \ln W_{it} \end{bmatrix} = \begin{bmatrix} \mu_{\pi,i} \\ \mu_{\omega,i} \end{bmatrix} + \begin{bmatrix} \lambda_i' \theta_t \\ \kappa_i' \eta_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{it} \\ \epsilon_{it} \end{bmatrix}, \tag{3}$$

where W_{it} represents the **real** expenditure for the $i-$th item at time t such that the corresponding weight is calculated as $\omega_{it} = W_{it} / \sum_{i=1}^{n} W_{it}$. $\mu_{\pi,i}$ and $\mu_{\omega,i}$ are the long run growth rates, θ_t and η_t are mean zero vector of the unknown common factors, and λ_i and κ_i are the vector of factor loadings. The remaining approximation errors (ε_{it} and ϵ_{it}), which is dubbed as '*leftover*' term, is assumed to follow an AR(p) process as

$$\begin{bmatrix} \varepsilon_{it} \\ \epsilon_{it} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{p_i} \rho_{ij}^{\varepsilon} \varepsilon_{it-j} \\ \sum_{j=1}^{q_i} \rho_{ij}^{\epsilon} \epsilon_{it-j} \end{bmatrix} + \begin{bmatrix} v_{it} \\ \upsilon_{it} \end{bmatrix}, \tag{4}$$

where v_{it} and υ_{it} may have weak common factors.[5]

Since the common factors may have different dynamic structures, we model it separately as

$$\begin{bmatrix} \theta_{\ell,t} \\ \eta_{\ell,t} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{p_\ell} \rho_{\ell j}^{\theta} \theta_{\ell,t-j} \\ \sum_{j=1}^{q_\ell} \rho_{\ell j}^{\eta} \eta_{\ell,t-j} \end{bmatrix} + \begin{bmatrix} m_{\ell t} \\ \varpi_{\ell t} \end{bmatrix}, \text{ with } \ell = 1, ..., k. \tag{5}$$

The sieve parameters are $\left\{ \mu_{\pi i}, \mu_{\omega i}, \lambda_i, \kappa_i, \rho_{ij}^{\varepsilon}, \rho_{\ell j}^{\epsilon}, \rho_{\ell j}^{\theta}, \rho_{\ell j}^{\eta} \right\}$. The lag lengths are selected by the BIC rule where the maximum lag length is set to be 24. The common factors are estimated by the principal component (PC) method. The resulting residuals are then used in the bootstrap procedure outlined in Appendix A. Note that this sieve bootstrap procedure becomes consistent when both n and $T \to \infty$.

Tables 1 presents the equivalence test results based on the sieve bootstrap for the monthly inflation data that spans from 1977.M1 to 2015.M8, resulting in 463 monthly observations.[6] Recall that the PCE data consist of 178 subcomponents at each time t, or the cross-section dimension of 178. We set the maximum factor number as eight. The number of the common factors to the inflation rates is then selected by using the Bai and Ng (2002)'s IC criteria. IC_1 and IC_2 select one, while IC_3 chooses four. The number of the common factors to the first differenced log

[5] Also note that if the factor loadings are time varying, then the leftover term includes the remaining common factors. To see this, assume that $\pi_{it} = \mu_{\pi,i} + \lambda_{it}' \theta_t + \varepsilon_{it}^* = \mu_{\pi,i} + \lambda_i' \theta_t + (\lambda_{it}' - \lambda_i') \theta_t + \varepsilon_{it}^*$. Hence the leftover term ε_{it} becomes $(\lambda_{it}' - \lambda_i') \theta_t + \varepsilon_{it}^*$.

[6] The disaggregate, subaggregated, and aggregated PCE prices and nominal quantities are collected from the BEA Tables 2.4.4U and 2.4.5U and the real quantity is calculated by the nominal divided by price. The data on TRIM inflation rate come from the FRB Dallas and the CPI data are taken from the BLS's website.

weight is selected as three by IC_1 and IC_2, but five by IC_3. To get more robust results, we assume that θ_t and η_t share the same number of common factors, but we include many common factors as possible. Moon and Weidner (forthcoming) show that including more common factors do not hamper the statistical inference, and more importantly by including more common factors, the estimating results become stablized.

The left panel of Table 1 reports the rejection rates of the equivalence test at the 10% significance level. For each t, we evaluate t_z in eq.(2) by using 10,000 nonparametric sieve bootstrap iterations, and record the number of the rejections over 463 samples. When the number of the common factors (k) is relatively small, say $k \leq 3$, the rejection rates appear to be larger than the nominal size for the all three measures of inflation. When the number of the common factors is more than three, however, the rejection rates for the headline and CORE inflation drop drastically before they stabilized around 0.01 as highlighted by Moon and Weidner (forthcoming).[7] By contrast, the rejection rate for TRIM inflation remains consistently above the nominal significance level for the all ks considered. This is probably because the TRIM inflation is constructed from a unbalanced trimming method (by removing the upper 31 percent and the lower 24 percent of the weights) and hence may mismeasure the underlying inflation if subaggregate inflation rates are distributed symmetrically. Taken together, our results suggest that different measures of inflation are not statistically different from a simple cross-sectional mean of individual inflation once the cross sectional dependence is properly taken into account. The result could also be interpreted as reflecting that the headline inflation is not statistically different from CORE inflation in estimating the underlying true inflation.

Given that the rejection rates reported in Table 1 are below the nominal size of 10% for reasonably large numbers of factors, one may suspect that they could have been driven by low power of the equivalence test. To explore this issue, we investigate the power performance of the test by generating the following pseudo weight functions that is correlated with π_{it},

$$\omega^*_{i,t} = \frac{\pi_{it} - \hat{\pi}_{\min} + \left|\epsilon^j_{it}\right|}{\sum_{i=1}^n \left(\pi_{it} - \hat{\pi}_{\min} + \left|\epsilon^j_{it}\right|\right)},$$

where $\epsilon^j_{it} \sim iid\mathcal{N}(0, \alpha\sigma^2)$, $\hat{\pi}_{\min} = \min\{\hat{\pi}_i\}$ and $\hat{\pi}_i = T^{-1}\sum \pi_{it}$. Note that $\hat{\pi}_{\min}$ is always negative, so that $\omega^*_{i,t} > 0$ for all i and t. We let σ^2 be the estimated variance of π_{it}, which is around 2.5×10^{-3}, and vary α. As α increases, the cross sectional correlation between $\omega^*_{i,t}$ and π_{it} drops and so does the rejection rate. We consider $\alpha \in [2000, 100, 5]$ so that the implied cross sectional correlations

[7] Given that the rejection rates become stabilized after $k = 3$ or 4, it is advisable to add at least three or four common factors in empirical analysis along the similar lines of Moon and Weidner (forthcoming).

between $\omega_{i,t}^{*}$ and π_{it} are approximately $(0.030, 0.130, 0.482)$, respectively. The right panel of Table 1 reports the rejection rates of the pseudo inflation rates at the 10% level for monthly inflation. Inspection of the table suggests that the discriminatory power is overall satisfactory, even when the average cross sectional correlation (ρ) is as small as 0.03. This seems to suggest that the results in Table 1 can hardly be viewed as an artifact of the low-power problem of the test.

In sum, our results in this section convincingly suggest that the PCE headline inflation is not statistically different from CORE or TRIM inflation in consistently estimating the true underlying inflation (π_t). At first, this comes as surprise in view of the fact that the relevant literature on the appropriate measure of inflation is largely split between studies that favor headline inflation and those that advocate the use of core inflation. While the advocates of core inflation claim that the inflation rates of food and energy are not influential to the true underlying inflation but only create noise, an obvious criticism of the focus on core inflation is that it excludes the central items to consumers like food and energy which are more related to their daily purchases (e.g., Bullard 2011). On one hand, our key finding seems at odds with these conflicting views by not taking any side, but on the other hand it can be usefully reconciled with both sides by asserting that the two measures of inflation are basically tracking the same underlying inflation. Even if the prices of food and energy items indeed capture the true underlying inflation, core inflation can still consistently estimate the true underlying inflation. To show this, let π_t^{fe} denote the inflation rate of the food and energy which is assumed to follow the true underlying common factor, and n_1 be the total number of non-food and energy items. Then, the remaining individual inflation rates can be written as

$$\pi_{it}^{c} = a_i + \lambda_i \pi_t^{\mathrm{fe}} + \pi_{it}^{co},$$

so that the weighted mean becomes

$$\begin{aligned} \sum_{i\neq \mathrm{fe}}^{n} \frac{\omega_{it}}{\sum_{i\neq \mathrm{fe}}^{n} \omega_{it}} \pi_{it}^{c} &= \sum_{i\neq \mathrm{fe}}^{n} \frac{\omega_{it}}{\sum_{i\neq \mathrm{fe}}^{n} \omega_{it}} a_i + \sum_{i\neq \mathrm{fe}}^{n} \frac{\omega_{it}}{\sum_{i\neq \mathrm{fe}}^{n} \omega_{it}} \lambda_i \pi_t^{\mathrm{fe}} + \sum_{i\neq \mathrm{fe}}^{n} \frac{\omega_{it}}{\sum_{i\neq \mathrm{fe}}^{n} \omega_{it}} \pi_{it}^{co} \\ &= \hat{\mu} + \hat{\lambda} \pi_t^{\mathrm{fe}} + O_p(n_1^{-1/2}) \\ &\to \mu + \lambda \pi_t^{\mathrm{fe}} \quad \text{as } n_1 \to \infty. \end{aligned}$$

Hence the weighted average of individual inflation consistently captures the key underlying inflation, π_t^{fe}, even when π_t^{fe} is excluded because all the other individual inflation rates are closely linked to the true underlying inflation. That is, to the extent that individual inflation rates follow a factor model and that the true underlying inflation is a function of the linear combination of common factors, the probability limit of any weighted mean of inflation becomes the true underlying inflation irrespective of the non-random weight functions.[8]

[8]In a similar context, modeling the dynamics of inflation within the structure of a common factor model, Stock

That being said, since our interpretation was mainly deduced from the fact that the expenditure weights in PCE data are not much correlated with individual inflation rates, it is illuminating to offer some intuitive explanations. A plausible explanation is that the substitution effect of individual price changes is dominated by the income effect. In that case, a lower price of a product does not necessarily increase the PCE weight for the corresponding product unless the *real* expenditure on the product increases more than the decrease in price. Consequently, the weights of subcomponent inflation will not vary much with individual price changes. Another possibility is that the level of disaggregation of the PCE data is not fine enough to capture the substitution effect satisfactorily. The PCE item of 'New motor vehicles', for instance, comprises cars of different brands that are by nature close substitutes. If the price of a car rises and hence consumers switch from one brand to another brand, the current PCE data cannot capture this substitution effect because they belong to the same item. If the substitution effect mainly takes place within the PCE product level, weights will not be responsive to the price changes of individual products.[9]

3 Out-of-sample forecasting performance

Another important issue is whether or not core inflation is a better predictor of future headline inflation. Advocates of core inflation often argue that core inflation is more useful in forecasting future headline inflation probably because it measures the persistent or durable component of headline inflation. Before proceeding, it is instructive to briefly discuss two popular approaches for forecasting: recursive and rolling methods. While the recursive method utilizes all data observations up to the forecasting point, the rolling regression method exploits a fixed number of observations at each time. In consequence, the recursive method is known to perform better when the underlying parameters are stable over time, whereas the rolling regression approach is preferred when structural change or time-variation is suspected in the key parameters. To explore which approach is better suited to our data, we plot in Figure 1 the AR(1) coefficient estimates of the monthly PCE inflation rate $(p_{t+1} - p_t)$ with a 60-month rolling window. A visual inspection of Figure 1 favors the use of the rolling regression approach because the AR coefficient estimates clearly exhibit time-varying behavior, which cannot be effectively captured by recursive regression.

and Watson (forthcoming) show that the forecasting performance based on the smoothed trend inflation - which is another estimate of the true underlying inflation - is better than a simple random walk model.

[9] Since the substitution effect seems to hinge on the level of disaggregation, one might surmise that use of fine micro-level price data such as the ELI-level data adopted by Nakamura and Steinsson (2008) could change the key results. This is an interesting avenue of research, but we leave it to future research it is beyond the scope of this paper.

To substantiate this point, we plot in Figure 2 the squared forecast errors (SFE) of both rolling and recursive regression methods which are estimated from the forecasting model shown in eq.(6) below. Since the SFE of the rolling method tends to vary with the choice of rolling window with poorer forecasting performance for longer rolling window, we set a rolling window of 60 months at which the variance of SFE appears to stabilize. Not surprisingly, the SFE of the recursive method is highly sensitive to the starting point of sample period. Moreover, the forecasting performance of the recursive method becomes very poor when the sample includes the transition period from the great inflation to the great moderation in the 1970s and 80s, while the performance improves greatly for the subsample periods after 1992 when inflation has been relatively stable. This confirms our prior intuition that the recursive method is not appropriate for data with obvious structural changes. Nevertheless, we find that forecasting based on the rolling regression outperforms the one based on the recursive method in terms of the mean squared forecast error (MSFE) even for the sample period beginning from 1992.M1. This validates our use of the rolling regression method in evaluating the forecasting performance of different measures of inflation.

3.1 Benchmark forecasting model

To compare the out-of-sample forecasting performance of the three measures of inflation under study, we consider the following benchmark forecasting model with 60-month rolling window for the h-horizon.

$$\left(\frac{p_{\mathrm{pce},t+h} - p_{\mathrm{pce},t}}{12}\right) h = \alpha_h + \beta_h \left(p_{jt} - p_{jt-12}\right) + \epsilon_{t+h,t}, \tag{6}$$

where p_{jt} denotes price index of j^{th} inflation measure at t for $j = \{Headline, CORE, TRIM\}$. The left-hand side of eq.(6) therefore represents the annualized PCE headline inflation between t and $t+h$ with the h-month forecasting horizon. Note that inflation rates are converted into percentage changes. The corresponding mean squared forecasting error (MSFE) can then be obtained as

$$MSFE = \frac{1}{P} \sum_{t=T_o+1}^{T_o+P} \left(\hat{p}_{\mathrm{pce},t+h|t} - p_{\mathrm{pce},t+h}\right)^2,$$

where T_o is the end of the sample used in the rolling estimation, and $\hat{p}_{\mathrm{pce},t+h|t}$is the fitted value from eq.(6).

The results from this exercise are summarized in Figure 3 for the three measures of inflation under comparison. The thick solid line represents the average MSFEs for the PCE headline inflation, and the thin solid line and the dotted line respectively depict the average MSFEs for the CORE and TRIM inflation. Each point on the chart represents a 10-year average of the 60-month rolling MSFE for the 24-month forecasting horizon. The numbers on the horizontal axis represent respectively the beginning year (top) and the ending year (bottom) of each 10-year subsample period. For instance, 1985 (top) and 1994 (bottom) on the horizontal axis capture the subsample period of 1985–1994, and so on. As in Crone et al. (2013), we assess the forecasting precision of the three competing measures of underlying inflation by reporting the Giacomini-White (2006) statistic for testing whether there is a statistically significant difference in MSFEs between the models with different measures of inflation at the 5% significance level.[10] Any period in which the MSFEs of headline inflation are statistically different from those of TRIM inflation is noted in symbols (circle).

A few things are of note from Figure 3. The first feature that is apparent in Figure 3 is that the MSFEs of all three measures of inflation takes on a U-shape profile over time, as they decline initially but begins to rise back after the 1995-2004 subsample period. This largely mimics the U-shaped profile seen in the AR coefficient estimates in Figure 1, which is not surprising in light of the close association between the forecasting error and the persistence of inflation series. Forecasting error is likely to be bigger (a larger MSFE) when inflation series is more persistent (a larger value of AR(1) coefficient estimate).

Second, the CORE inflation appears to outperform the PCE headline inflation until around 2005. This confirms the findings by earlier studies (e.g., Bryan et al. 1997, Clark 2001, Cogley 2002) that CORE inflation is a better predictor of future headline inflation. The outperformance of the CORE inflation, however, disappears quickly after 1996-2005 period when it is dominated by the headline inflation. This corroborates the recent finding by Crone et al. (2013) who refute the earlier findings based on a post-1990 dataset. Judging from the Giacomini-White's (2006) statistics, however, there is no statistically significant difference in the MSFE between the CORE inflation and the headline inflation in the entire sample period considered. That is, there is no clear-cut superiority between the two measures of inflation in predicting future headline inflation.

Third, a similar story is told when the headline inflation is compared with the TRIM inflation, except for the subsample period of 1992 and 2003 when the MSFE of the TRIM inflation is significantly smaller than that of the headline inflation. Interestingly, this period coincides with the data span studied by Clark (2001) who documents the dominance of the TRIM inflation over the headline inflation in terms of forecasting precision. Other than this period, however, the headline inflation performs quite comparably, if not marginally better, in forecasting future headline inflation.

[10] See eq.(3) in Crone et al. (2013,p.510) for the GW statistic.

3.2 Out-of-sample performance of competing forecasting models

The overall assessment of our results in the previous section is that headline inflation is not outperformed by CORE or TRIM inflation in terms of forecasting accuracy. The results, however, are not much informative about which specific forecasting model is better in predicting headline inflation. To address this issue, we conduct another set of exercise in a general multivariate forecasting model framework that encompasses conventional univariate models (e.g., Clark 2001) and random-walk models (e.g., Stock and Watson forthcoming) to bivariate forecasting models (e.g., Crone et al. 2013). The novelty of our approach here is to extending the forecasting model to a multivariate model that includes not only conventional regressors such CORE and TRIM, but also subcomponents of PCE inflation that received little attention in the literature. Although it is widely agreed that inflation forecasters have a dearth of reliable multivariate models for forecasting inflation, more recent studies (e.g., Crone et al. 2013) tend to suggest that forecasting performance can be enhanced in a multivariate framework. Crone et al. (2013) also document that CPI inflation is helpful to enhance the precision of forecasting future headline inflation. Since the forecasts from the PCE headline inflation *per se* are at least comparable to those from the alternative measures of inflation considered, we here focus on the headline PCE inflation as a benchmark specification.

To identify the best performing forecasting model, we consider the following general multivariate model specification that nests many popular forecasting models of interest.

$$\left(\frac{p_{pce,t+h} - p_{pce,t}}{12}\right) h = \alpha_h + \beta_h \left(p_{jt} - p_{jt-12}\right) + \sum_{\ell=1}^{k} \gamma_{\ell,h} \left(x_{\ell,t} - x_{\ell,t-12}\right) + \epsilon_{t+h,t}, \tag{7}$$

where p_{jt} denotes price index of j-th inflation measure as before and $x_{\ell t}$ represents other price variables such as CPI inflation and subaggregate inflation series for durable (D), nondurable (ND), and service (S), dubbed as 'Sub' throughout the paper. We also consider the principal component estimated static common factors (PC, hereafter) as well as the smoothed trend with CORE (hereafter, TREND) which bears a close resemblance in spirit to the 'multivariate UCSVO' discussed in Stock and Watson (forthcoming).[11] Note that the multivariate model in eq.(7) embraces (i) random walk models (e.g., Stock and Watson 2007) when $\beta_h = \gamma_{1,h} = \cdots = \gamma_{k,h} = 0$; (ii) single regressor models (e.g., Clark 2001, Smith 2004) when $\gamma_{1,h} = \cdots = \gamma_{k,h} = 0$; and (iii) bivariate regressor models (e.g., Crone et al. 2013) when $\gamma_{2,h} = \cdots = \gamma_{k,h} = 0$.

[11] As an extension of the univariate unobserved components/stochastic volatility outlier-adjustment (UCSVO) model that expresses the rate of inflation as the sum of a permanent component (trend) and a transitory component, the multivariate UCSVO (MUCSVO) model includes a common latent factor in both the trend and idiosyncratic components of inflation by allowing the factor loadings to vary over time. In this regard, the model with smooth trended CORE considered in our analysis takes a similar structure with the MUCSVO.

For the multivariate models, we consider the following seven different model specifications with various combinations of some key variables such as CORE, TRIM, CPI, Sub and TREND, together with the PCE headline inflation.[12] As a result, the total number of models considered here is 20 and the performance of each of these competing models is evaluated relative to the benchmark model (BM) in eq.(6).

Case 1:	PCE + Subs (ND, D and S) + CORE + TRIM + CPI
Case 2:	PCE + PC + CORE + TRIM + CPI
Case 3:	PCE + TREND + CORE + TRIM + CPI
Case 4:	PCE + Sub + CORE + TRIM
Case 5:	PCE + Sub + CORE
Case 6:	PCE + Sub + TRIM
Case 7:	Sub + CORE + TRIM

Table 2 reports the results of such an exercise with the forecasting horizons of h =24 and 36 months for the rolling window of 60 months.[13] The first column of Table 2 presents the results for the 24-month forecasting horizon and those for the 36-month horizon are reported in the third column. The table displays a horse race of the forecasting errors that each of the 20 models make in predicting future headline inflation. Since the entries reported in the table are MSFEs from corresponding forecasting models, forecasting models with lower values in the MSFE can be seen as a better predictor of headline inflation. The MSFEs with the smallest values are denoted in bold-face.

The results in Table 2 illustrate several important points. First, as is obvious from the table, all models make smaller errors in predicting headline inflation when the forecasting horizon is longer. Second, unlike the common perception built in the literature, multivariate models appear to deliver significantly more accurate forecasts of headline inflation than their univariate counterparts, lending credence to the recent finding by Stock and Watson (forthcoming) and Crone et al. (2013). It seems improbable that any one measure of inflation alone has a significant enough forecasting power in predicting future headline inflation. This finding obviously runs counter to one of the most robust findings from the empirical forecasting literature on the superiority of univariate models to their multivariate counterparts. Among the seven multivariate models under study, the

[12] The inclusion of the *three* sub-panels of headline inflation as regressor is mainly guided by the number of common factors chosen by using the Bai and Ng (2002)'s IC criteria. In addition to the three conventional product categories of Durable (D), Non-Durable (ND) and Service (S), we also consider three sub-panels that are randomly selected from 5,000 sub-panel combinations.

[13] While the table only includes the results for the specific forecasting horizons, they are representative of the cases for other horizons which are not reported to conserve the space.

model including both PCE headline inflation and three sub-panels (ND, D and S) together with CORE, TRIM and CPI inflation performs the best. Third, a detailed review of Table 2 suggests the combination of PCE and three sub-panels as superior regressors in predicting future headline inflation. Any forecasting models containing the combination appear to consistently outperform those that do not, implying that those components are essential components in tracking the underlying inflation. However, unlike Stock and Watson (forthcoming) and Crone et al. (2013), we fail to find that including TREND inflation or CPI inflation alone improves the forecasting performance of multivariate models.

Speaking of the role of three sub-panels, it is worth noting that we reach the same conclusion on the best performing forecasting model even when three randomly generated sub-panels are used in place of ND, D and S. As presented in the second column of the table, the average value of MSFE obtained from 5,000 randomly generated combinations of three sub-panels (0.207 for 'PCE + Sub') is smaller than those from many other multivariate models under comparison. However, it is still larger than that from the 'All (sub)' model with the three conventional product categories as sub-panels (0.167). Moreover, although the minimum value of MSFE from the randomly generated sub-panels is slightly smaller than the case with the conventional large product categories, there is no guarantee that the selected subpanels robustly outperform in different subsamples because they are not consistently estimable. Our suggestion to practitioners is therefore to utilize the three conventional sub-components of headline inflation as the proxy for the common factors of underlying inflation. These conclusions carry over to the cases with 36-month forecasting horizon which are presented in the right-hand side panel of Table 2.

The results in Table 2 illustrate already interesting points, but they are based on the entire sample period. To ensure the robustness of our conclusions across subsample periods, we compare the MSFEs of the two leading multivariate models, the one with 'Sub' and the other with 'TREND', for various sub-sample periods based on a rolling regression approach. Figure 4 displays the results with the 36-month horizon, which are very similar to those outlined above in Table 2, clearly demonstrating the dominance of the multivariate models relative to the benchmark model. Notice that the 5% significant cases that the MSFE from multivariate model is different from that of the benckmark model are marked as a circle for the model with 'PCE+Sub+CORE+TRIM+CPI' and as a square for the model with 'PCE+TREND+CORE+TRIM+CPI'. As displayed in Figure 4, the MSFEs of the best performing multivariate model with 'PCE+Sub+CORE+TRIM+CPI' are not only significantly different from those of the benchmark model, but also are consistently below those of the competing multivariate model with 'TREND' in almost all subsample periods considered. That is, our conclusions on the best performing forecasting model are fairly robust to the choice of sample periods.

4 Conclusion

This paper makes two major arguments with regard to the much-debated issue on the measure of inflation for monetary policy purpose. First, the headline inflation is not much statistically different from CORE or TRIM inflation in consistently tracking unobservable underlying inflation. Due to the lower volatility of the latter in practice, however, our results defend the use of CORE inflation as in Bryan and Cecchetti (1994). Second, when it comes to the choice of forecasting model for predicting future headline inflation, our results suggest that a multivariate model with PCE headline inflation and three sub-panels along with CORE, TRIM and CPI inflation has the best forecasting performance among 20 models under study. Its outperformance turns out to be quite robust to the selection of sample periods.

References

[1] Bai, Jushan and Serena Ng. (2002) “Determining the number of factors in approximate factor models.” *Econometrica*, 70(1), 191-221.

[2] Bernanke, Ben S. and Mark Gertler. (1999) “Monetary policy and asset price volatility.” Federal Reserve Bank of Kansas City, *Economic Review*, 84 (Fourth Quarter), 17–51.

[3] Bickel, P.J. and E.L. Lehmann. (1975) “Descriptive statistics for nonparameteric models II. Location.” *The Annals of Statistics*, 3(5), 1045–1069.

[4] Bryan, Michael F., and Stephen G. Cecchetti. (1994) “Measuring Core Inflation.” In Monetary Policy, edited by N. Gregory Mankiw, pp.195–215. Chicago: University of Chicago Press.

[5] Bryan, Michael F., Stephen G. Cecchetti, and Rodney L. Wiggins II. (1997) “Efficient inflation estimation.” NBER Working Papers No. 6183, National Bureau of Economic Research.

[6] Bullard, James. (2011) “Measuring Inflation: The Core is Rotten.” Federal Reserve Bank of St. Louis, *Review*, 93(4), 223–33.

[7] Chang, Yoosoon. (2004) “Bootstrap unit root tests in panels with cross-sectional dependency.” *Journal of Econometrics*, 120(2), 263–293.

[8] Clark, Todd. (2001) “Comparing Measures of Core Inflation.” Federal Reserve Bank of Kansas City, *Economic Review*, 86 (Second Quarter), 5–31.

[9] Cogley, Timothy. (2002) “A Simple Adaptive Measure of Core Inflation.” *Journal of Money, Credit and Banking*, 34(1), 94–113.

[10] Crone, Theodore M., N. Neil K. Khettry, Loretta J. Mester, and Jason A. Novak (2013) “Core Measures of Inflation as Predictors of Total Inflation.” *Journal of Money, Credit and Banking*, 45(2-3), 505–519.

[11] Dolmas, Jim. (2009) “Excluding items from personal consumption expenditures inflation,” Federal Reserve Bank of Dallas, *Staff Papers*, No. 7, June.

[12] Giacomini, Raffaella and Halbert White. (2006) “Tests of Conditional Predictive Ability.” *Econometrica*, 74(6), 1545–1578.

[13] Mishkin, Frederic S. (2007) “Housing and the monetary transmission mechanism.” Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series, 2007-40.

[14] Moon, Roger H. and Martin Weidner. (2015) “Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects.” *Econometrica*, forthcoming.

[15] Nakamura, Emi and Jón Steinsson. (2008) “Five Facts about Prices: A Reevaluation of Menu Cost Models.” *Quarterly Journal of Economics* 123, 1415–1464.

[16] Smith, Julie K. (2004) “Weighted Median Inflation: Is This Core Inflation?” *Journal of Money, Credit and Banking*, 36, 253–63.

[17] Stock, James H., and Mark W. Watson. (2007) “Has Inflation Become Harder to Forecast?” *Journal of Money, Credit and Banking*, 39, 3–33.

[18] Stock, James H., and Mark W. Watson. (2016) “Core Inflation and Trend Inflation.” *Review of Economics and Statistics*, forthcoming.

[19] Wynne, Mark A. (1999) "Core Inflation: A Review of Some Conceptual Issues." Federal Reserve Bank of Dallas, *Review*, research department working paper no. 99-03, June.

[20] Wynne, Mark A. (2008) "Core Inflation: A Review of Some Conceptual Issues." Federal Reserve Bank of St. Louis, *Review*, 90, 205–28.

Appendix

Appendix A: The sieve bootstrap procedure

Denote $\hat{e}_t = [\hat{\varepsilon}_t, \hat{\epsilon}_t]$ where $\hat{\varepsilon}_t = [\hat{\varepsilon}_{1t}, ..., \hat{\varepsilon}_{nt}]$, $\hat{\epsilon}_t = [\hat{\epsilon}_{1t}, ..., \hat{\epsilon}_{nt}]$. Also define $\hat{\xi}_t = [\hat{m}_t, \hat{\varpi}_t]$ where $\hat{m}_t = [\hat{m}_{1t}, ..., \hat{m}_{kt}]$, $\hat{\varpi}_t = [\hat{\varpi}_{1t}, ..., \hat{\varpi}_{kt}]$, and k is the assigned number of the common factors.

1. Estimate the sieve parameters: Constant terms (μ_i^{π} and μ_i^{ω}) and AR(p_i) coefficients. To estimate the factor loading coefficients, first standardize the inflation rates and the first differenced log weights by their standard deviations, and then estimate the k largest eigen-vectors of the $n \times n$ covariance matrixes (Principal Components (PC) estimators). The factor loading coefficients can be estimated by running an individual inflation rate (non-standardized series) on constant and the estimated PC factors.
2. Generate a $(T + M) \times 1$ vector of random integers with the range of $\{1, ..., T - 24\}$. Draw the pseudo errors from the residuals $\hat{e}_t$ and $\hat{\xi}_t$ by selecting the entire n and k-dimensional vectors to preserve the cross-sectional dependence.
3. Generate e_t^* from ξ_t^* using eq. (4) and (5). Discard the first M observations to avoid the initial effects. Here we set $M = 200$.
4. Re-center e_t^* and ξ_t^*. Subtract its time series mean to ensure that the pseudo error e_{it}^* and ξ_t^* have zero means over t.
5. Generate π_{it}^* by combining $\hat{\mu}_i + \hat{\lambda}_i' \hat{\theta}_t^*$ with ε_{it}^* in eq. (3).
6. Construct $\ln \omega_{it}^*$ by taking cumulative sum of $\Delta \ln \omega_{it}^*$. Take exponential to get back ω_{it}^*. Standardize ω_{it}^*.
7. Construct z_t^* by taking the difference in the two pseudo weighted means.
8. Construct t-statistics, $t_z^* = (z_t^* - z_t) / \sqrt{\hat{\Sigma}_z^{*2}}$, where $\hat{\Sigma}_z^{*2}$ is the bootstrapped variance.
9. Repeat the above steps (1 through 7) 10,000 times and construct the critical value at a given significance level.

Appendix B: Common factor estimation

The common factors are often estimated by using two popular approaches: principal component (PC) and cross-sectional averages of the subsamples. Under the assumption that individual inflation follows the following static factor model,

$$\pi_{it} = a_i + \lambda_i' \theta_t + \pi_{it}^o,$$

and the empirical common trend inflation is defined as

$$\hat{\tau}_t = \frac{1}{n} \sum\nolimits_{i=1}^{n} \hat{a}_i + \frac{1}{n} \sum\nolimits_{i=1}^{n} \hat{\lambda}_i' \hat{\theta}_t,$$

where $\hat{\theta}_t$ is an estimate of the common factors θ_t. Rather than including $\hat{\tau}_t$ in the forecasting regression, we include $\hat{\theta}_t$ directly in the forecasting regression because the direct inclusion improves forecasting performance. The PC estimators are required the assumption of the stationarity of

$\pi_{it,t+h}$ meanwhile the cross sectional averages of the subpanels requires the stationarity of π_{it}^{o}. Since we assume that the number of the common factors is three, we split the 178 items into three subpanels. And then we take the cross sectional averages to approximate the common factors. Let $\hat{\theta}_t^{pc}$ and $\hat{\theta}_t^{cs}$ be the PC and the cross sectional averages, respectively. Then it is straightforward to show under the regularity conditions that

$$\begin{aligned} \hat{\theta}_t^{pc} &= H'\theta_t + O_p\left(C_{nT}^{-1/2}\right), \\ \hat{\theta}_t^{cs} &= B'\theta_t + O_p\left(n^{-1/2}\right). \end{aligned}$$

where $C_{nT} = \min\left[n, T\right]$, H is an invertable (3×3) rotating matrix and B is the (3×3) matrix of the sample cross sectional averages λ_i for subpanels.

Table 1: The sieve bootstrap results (10% significance level)

	Rejection rates			Pseudo power		
$\hat{k}$	HDL	CORE	TRIM	$\rho = 0.03$	$\rho = 0.130$	$\rho = 0.482$
1	0.253	0.178	0.379	0.431	0.751	1.000
2	0.171	0.122	0.131	0.424	0.739	1.000
3	0.138	0.106	0.149	0.386	0.711	1.000
4	0.045	0.011	0.159	0.373	0.739	1.000
5	0.010	0.014	0.146	0.333	0.686	1.000
6	0.018	0.013	0.113	0.359	0.673	1.000
7	0.011	0.008	0.144	0.330	0.716	0.999
8	0.014	0.007	0.166	0.328	0.654	1.000

Note: Entries represent the number of times out of 10,000 nonparametric sieve bootstrap iterations that the null hypothesis of the equivalence of inflation measures to a simple cross-section mean is rejected at the 10% significance level. k denotes factor number and 'HDL' represents headline inflation. ρ stands for the average cross sectional correlation between ω^*_{it} and π_{it} over 10,000 bootstrap sample

Table 2: Out-of-sample forecasting performance with 60-month rolling window

Forecasting model	Regressors	24 month-horizon		36 month-horizon	
			random selection mean [min,max]		random selection mean [min,max]
Benchmark	PCE	0.431		0.305	
Random walk model	PCE	1.111		1.066	
	CORE	0.563		0.528	
	TRIM	0.590		0.517	
	TREND	0.468		0.430	
Single regressor model	CORE	0.436		0.286	
	TRIM	0.420		0.285	
	TREND	0.435		0.268	
Bivariate model	PCE+CORE	0.475		0.306	
	PCE+TRIM	0.413		0.274	
	PCE+CPI	0.413		0.264	
	PCE+PC	0.369		0.271	
	PCE+TREND	0.487		0.285	
Multivariate model	All (Sub)	**0.167**†	0.208† [0.134†,0.265†]	**0.115**†	0.132† [0.090†,0.179†]
	All (PC)	0.327†		0.194†	
	All (TREND)	0.344		0.186†	
	PCE+Sub	0.207†		0.157†	
	PCE+Sub+CORE+TRIM	0.174†	0.240† [0.140†,0.304†]	0.123†	0.163† [0.097†,0.216†]
	PCE+Sub+CORE	0.181†	0.277† [0.148†,0.389]	0.139†	0.186† [0.103†,0.266]
	PCE+Sub+TRIM	0.189†	0.279† [0.169†,0.367]	0.134†	0.180† [0.113†,0.244]
	Sub+CORE+TRIM	0.266†	0.258† [0.172†,0.325†]	0.148†	0.174† [0.103†,0.230†]

Note: Entries represent the MSFE for 24-month forecasting horizon with 60-month rolling window. Numbers in curved brackets denote the starting period and the ending period of subsamples. Bold face indicates the minimum MSFE and † denotes the rejection of the null hypothesis that the MSFE from the corresponding forecasting model is equal to that from the benchmark forecasting model (BM) based on the Giacomini-White's (2006) statistic at the 5% level. 'random selection' represents the multivariate model with three subpanels that are randomly generated from 5,000 subgroups with the sizes of group1 = 59, group2 = 59, and group3 = 60.

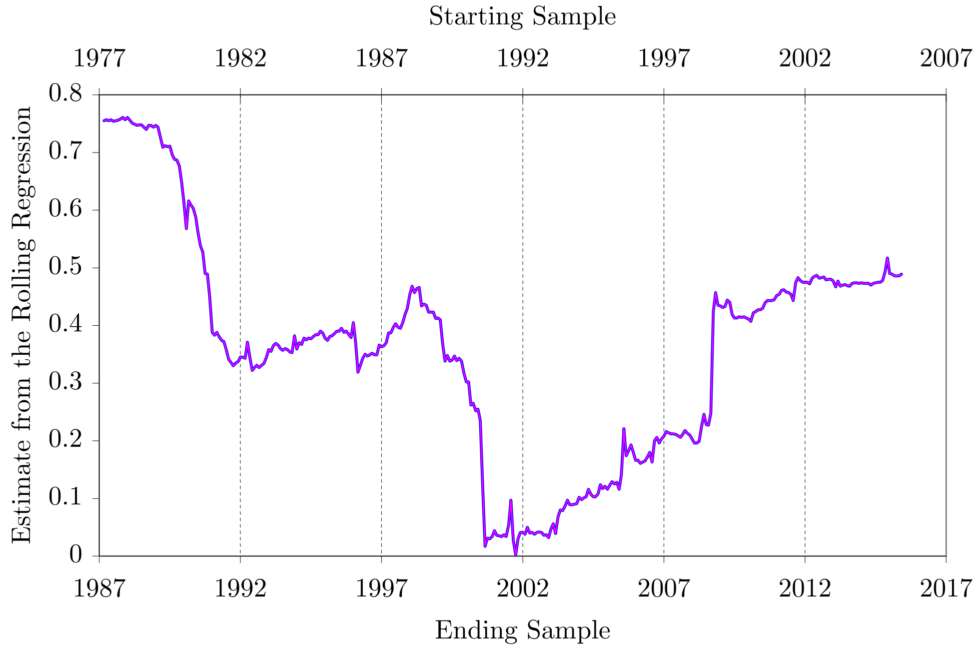

Figure 1: Rolling Estimated AR(1) Parameters with Monthly PCE Inflation

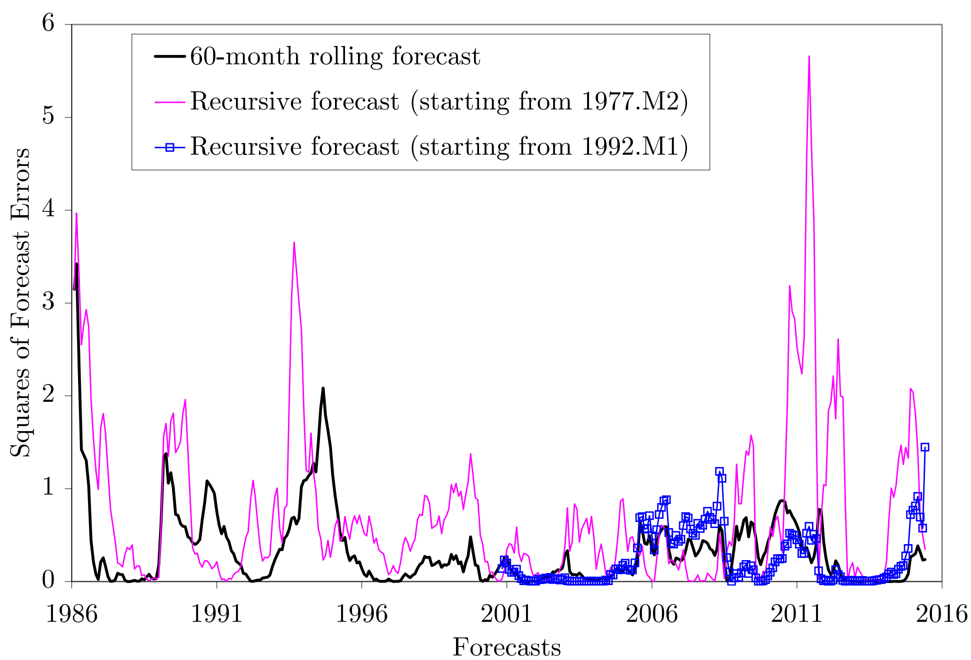

Figure 2: The Rolling and Recurisve MSFEs with 36-month Horizon

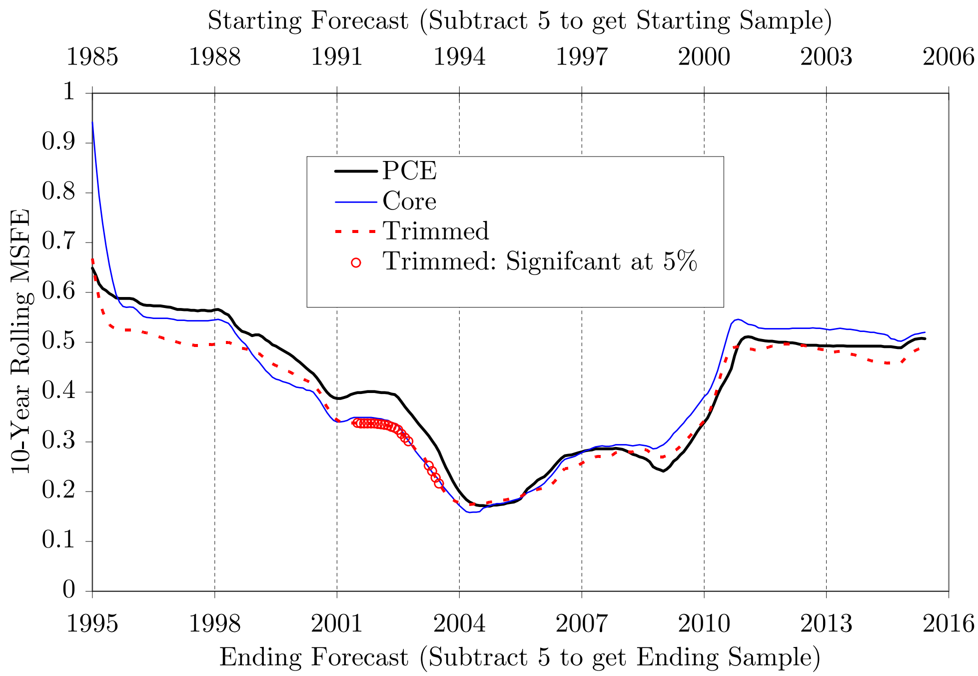

Figure 3: 10-Year Rolling MSFEs with 24-month horizon

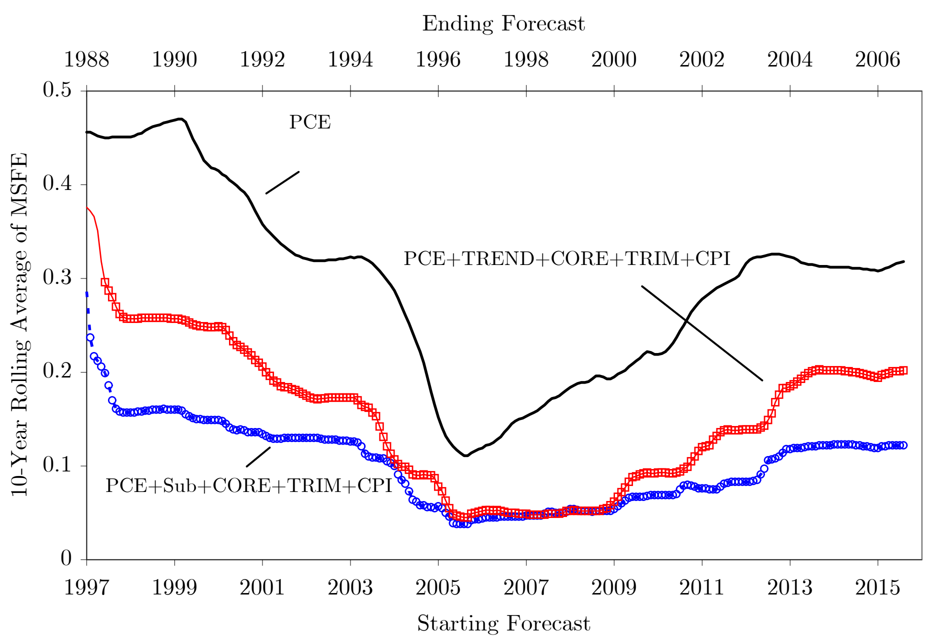

Figure 4: Performance of some selected forecasting models
for rolling subsamples (36-month horizon)