Trimmed-Mean Inflation Statistics: Just Hit the One in the Middle

Brent Meyer and Guhan Venkatu

Trimmed-Mean Inflation Statistics: Just Hit the One in the Middle
Brent Meyer and Guhan Venkatu

This paper reinvestigates the performance of trimmed-mean inflation measures some 20 years since their inception, asking whether there is a particular trimmed-mean measure that dominates the median CPI. Unlike previous research, we evaluate the performance of symmetric and asymmetric trimmed-means using a well-known equality of prediction test. We find that there is a large swath of trimmed-means that have statistically indistinguishable performance. Also, while the swath of statistically similar trims changes slightly over different sample periods, it always includes the median CPI—an extreme trim that holds conceptual and computational advantages. We conclude with a simple forecasting exercise that highlights the advantage of the median CPI relative to other standard inflation measures.

Brent Meyer is at the Federal Reserve Bank of Cleveland and can be reached at Brent.H.Meyer@clev.frb.org. Guhan Venkatu is at the Federal Reserve Bank of Cleveland and can be reached at Guhan.Venkatu@clev.frb.org. The authors thank Todd Clark and Tim Dunne for their helpful criticisms and guidance.

Introduction

Trimmed-mean inflation statistics diagnose the most volatile monthly price changes as noise and "trim" them from the price-change distribution, leaving a clearer inflation signal behind. These measures systematically remove sources of noise on a monthly basis, rather than ad hoc exclusionary measures such as the ex food and energy ("core") CPI— which implicitly suggests that relative price changes in all other retail price components are inflation signal, even though food prices are no more volatile than the core CPI itself (Clark 2001).

It's been roughly 20 years or so since trimmed-mean inflation statistics were first examined by Bryan and Pike (1991), and then more rigorously by Bryan and Cecchetti et al (1994, 1997). Bryan, Cecchetti, and Wiggins (1997) documents that the retail price-change distribution is leptokurtic (fat-tailed), arguing that in the presence of excess kurtosis the mean is an inefficient measure of location compared to a trimmed-mean approach. They find that the "optimal" symmetric trimmed-mean CPI measure is an 18% percent trim using the benchmark of the 36-month centered moving average in the headline CPI. In similar work, Bryan and Cecchetti (1994) also focus on the more extreme trim, the median CPI, and tout its use as a "core" inflation measure because it is more strongly correlated with changes in the money supply and a better forecaster than the ex food and energy CPI.[1]

In this paper we investigate whether or not the median CPI is still the appropriate trimmed-mean inflation statistic to use as a measure of underlying inflation. Rather than just focusing on symmetric trims like Bryan and Cecchetti et al, we open our investigation up to the full set of symmetric and asymmetric trims. In somewhat of a departure from other research on underlying inflation, we use gauge the usefulness of a particular trimmed-mean inflation measure by its ability to forecast future inflation. Importantly, we test whether modest differences in forecasting ability are statistically different using the Diebold-Mariano (DM) equality of prediction test. Our sample includes data from 1967 through 2011, and we test for changes in the "optimal" trim by splitting the sample and allowing for rolling-windows. We close with a simple forecasting exercise that highlights the advantage of the median CPI relative to other standard inflation measures.

Others have investigated trimmed-mean inflation statistics since the work of Bryan and Cecchetti et al in the mid-1990s. Smith (2004), using both conditional and unconditional forecasting models, finds that the weighted median CPI outperforms the core CPI. Clark (2001) evaluates a handful of core inflation measures' ability to track the current inflation trend and forecast future inflation, and finds the16% trimmed-mean CPI and CPI ex energy to be superior "core" measures. Meyer and Pasaogullari (2010) find the median and the 16% trimmed-mean CPI forecast year-ahead headline inflation about as

---

[1] In that paper, the optimal trim is the 15 percent trimmed-mean, which is chosen because it was the minimum variance estimator during this sample period.

well as inflation expectations do, and outperform simple forecasting models. Crone, Khettry, Mester, and Novak (2011) found that over longer-horizons (i.e. 24-months), the median CPI yields a forecast significantly superior to that of the headline or ex food and energy CPI index.

Dolmas (2005) applies the trimmed-mean procedure to Personal Consumption Expenditures Price Index (PCE) data. He allows for asymmetric trims and ties their use to the shape of the PCE price change distribution over his sample period. Interestingly, through a series of forecasting exercises, Dolmas points to a "cost" of imposing symmetry in terms of a higher root-mean-squared error (RMSE). However, he does not test to see if these losses in forecast accuracy are significant.

Many studies in this area of research—such as Clark (2001) or Cogley (2002)—start with a candidate's ability to track an in-sample trend, and then evaluate its ability to forecast future inflation in an out-of-sample setting. This paper proceeds in a similar fashion, except that instead of a centered or backward looking trend, our benchmark is annualized inflation over the next 3 years.

As a preview of our results, we find that since 1982, the trimmed-mean CPI measure with the lowest RMSE is the 30-36 percent trimmed-mean—which trims 30 percent off the lower tail and 36 percent off the upper tail. However, the forecast stemming from this trim is not statistically different from that of a wide selection of trimmed-means over a variety of time-periods. Interestingly, this wide-swath of trims with statistically equal forecasting power tends to include aggressive and roughly symmetric trimming points, such as the median CPI.

In general, we find aggressive trimming (close to the median) that is not too asymmetric appears to deliver the best forecasts over the time periods we examine. However, these "optimal" trims vary slightly across periods and are never statistically superior to the median CPI. Given that the median CPI is conceptually easy for the public to understand and is easier to reproduce, we conclude that it is arguably a more useful measure of underlying inflation for forecasters and policymakers alike. We close with evidence that the median CPI generally outperforms the headline CPI and core CPI in an out-of-sample forecasting test.

Data

The Consumer Price Index (CPI) is one of the two major retail prices indexes constructed for the United States. Its data are collected and assembled by the Bureau of Labor Statistics (BLS). In its broadest form, the CPI tracks the inflation experienced by urban consumers, or roughly 87 percent of the US population. [2, 3] The index is currently

---

[2] There are 38 separate areas that form the full urban sample.

divided into 211 categories called *item strata*, which are associated with 8 major groups: Food and Beverages, Housing, Apparel, Transportation, Medical Care, Recreation, Education and Communication, and Other Goods and Services. Also included in the index are taxes and government-charged user fees, such as auto-registration fees.

To collect price information on item strata, BLS employees call or visit thousands of retailers, rental units, and doctors' offices in 38 urban areas across the US every month. In particular, these employees collect data for 305 items, called *entry-level items*, associated with the item strata. If an item in this set is no longer available, or has changed in quantity or quality, the field worker selects a substitute, and notes the change. Analysts at the BLS's national offices then make any adjustments necessary to preserve consistency and comparability in the price data on an item across time.

The weights for the item strata in the CPI are based on information from the Consumer Expenditure Survey (CEX). In this survey, some households provide information on their quarterly purchases, while others provide more detailed bi-weekly diaries, to track more frequently purchased items, such as food and personal care products. Over a 2 year period, the BLS estimates that the CEX provides it with approximately 28,000 weekly diaries and 60,000 quarterly interviews. Since 2002, the BLS has updated CPI expenditure weights every two years. For example, as of January 2012, expenditure weights in the CPI are based on data from the 2008-2010 CEX. In the two years prior to that, beginning in January 2010, expenditure weights were based on data from the 2006-2008 CEX. Before 2002, weights were updated approximately every 10 years.

In between biennial updates of expenditure weights, the BLS assumes that the quantity of items in the CPI *market basket* does not vary. Expenditure shares, however, will change because of changes in item prices. As such, the biennial weights originally drawn from the CE are updated monthly as follows:

$$ w_{i,t+1} = \frac{w_{i,t} * \left(\frac{p_{i,t+1}}{p_{i,t}} \right)}{\left(\frac{P_{t+1}}{P_t} \right)} \, , $$

where $w_{i,t}$ is the weight of item i at time t, and $\{w_{i,t} \mid 0 \leq w_{i,t} < 1\}$, $p_{i,t}$ is the index value of item i at time t, and P_t is the index value of the CPI at time t, where $P_t = \sum w_{i,t} p_{i,t}$.

In this analysis, we use data from January 1967 to September 2011, which are divided into 3 periods: January 1967 to December 1982, January 1983 to December 1997, and January 1998 to September 2011. Though the entire CPI is used in each of these three periods, each employs different degrees of aggregation of the item strata. The first period, 1/1967-12/1982, uses 36 components. The second period, 1/1983-12/1997, uses 42

---

[3] Strictly speaking, this measure is called the Consumer Price Index for All Urban Consumers, or the CPI-U. A related measure, called the Consumer Price Index for Urban Wage Earners and Clerical Workers, or the CPI-W, tracks the inflation experienced by a subset of all urban consumers, who represent about 32 percent of the US population.

components, and incorporates a change in the way that owner-occupied housing is treated in the CPI. The last period 1/1998-9/2011, uses 45 components, and incorporates a major revision to the CPI's item structure, as well as regional indexes for owner-occupied housing.[4,5] In all cases, we use seasonally adjusted indexes for components where they exist.

The largest change between the 1967-1982, and 1983-forward time periods is a change in the BLS' methodology for measuring owner-occupied housing costs. In December 1982, the BLS switched to a rental equivalency approach. This change in methodology created a new component—Owners' Equivalent Rent (OER).[6] OER is the largest single component in the CPI, accounting for roughly 25 percent of the overall index by expenditure weight.

Having one oversized component relative to the rest of the market basket creates a potential issue for trimmed-mean price statistics. As the aggressiveness of the trim increases (toward the median) the influence of this component could grow disproportionately. For example, from January 1998 to July 2007 OER was the median component roughly two-thirds of the time. Brischetto and Richards (2007) recognized this and proposed a solution to reduce the influence of OER on the median CPI, by breaking it up into smaller regional components. The Federal Reserve Bank of Cleveland has adopted this procedure and it has lessened the frequency that an OER component is the median component.[7] [8]

Choice of Benchmark

The choice of inflation benchmark is a key aspect of research on underlying inflation measures. Some attempts focus on a candidate measure's ability to track or forecast inflation horizons of a year or less. In our view, this isn't a long enough time period to allow relative price shocks to unwind, and much of what is expressed in the near-term growth rate of headline inflation is noise. We prefer to measure inflation over longer, monetary

---

[4] There are actually 44 indexes from January 1998 to December 2001. The 45th index, leased cars and trucks, begins in January 2002.

[5] For information on the construction of regional indexes for owner-occupied housing, see [http://www.clevelandfed.org/research/data/us-inflation/revmcpi.cfm].

[6] In 2010, the BLS shifted the relative importance value (weight) of an unpriced component— 'secondary residences'—from 'lodging away from home' to OER, and changed the new component name to 'owners' equivalent rent of residences.' The BLS stated the reason for the shift was to lessen the volatility of the shelter. This move did not change the price index for OER, it just added a little less than 1.0 percentage point to its weight (potentially contributing to an entirely different issue for extreme trimmed-mean measures like the median CPI).

[7] See http://www.clevelandfed.org/research/data/us-inflation/revmcpi.cfm for more details.

[8] In some sense, the results we present later in the paper may ease some discomfort with this potential issue, as the median CPI is in the set of statistically similar forecasts in every time horizon we investigate.

policy relevant, time horizons that have a greater propensity to exude true inflation—the effect on prices in general that is due to directly to the impact of monetary policy. In this sense, a more appropriate benchmark would be to measure inflation over a 2 or 3 year time horizon.

Many studies of core inflation treat a longer-run centered moving average as their trend proxy. In fact, much of Bryan and Cecchetti's work uses the 36-month centered moving average in the headline CPI as its benchmark. However, using a centered moving average would leave this paper open to Blinder's critique of Cecchetti (1997):

> *Regarding the criterion, notice that a 36-month centered moving average treats the past and the future symmetrically. That is natural from the viewpoint of a scholar analyzing historical data. But it is very unnatural from the viewpoint of a central banker living in real time. Historically, t – 1 and t + 1 look more or less the same. But if you must make a decision at time t, there is a world of difference between t – 1 and t + 1. The past is known; the future is not. More important, the past is the dead hand of history, but the central bank must worry about the future.*

We take Blinder's advice and focus on trimmed-means' ability to track the trend in future inflation, specifically the annualized percent change in the headline CPI over the next 36 months. This choice of benchmark allows enough time for most relative price changes to work themselves out, leaving just the monetary impulse that is inflation behind.

We could have also chosen the 24-month ahead annualized percent change in the CPI as our inflation benchmark. The results of our empirical tests would largely be unaffected by switching to this benchmark. However, as the gasoline price shock of 2008 unwound, the 24-month annualized percent change in the CPI actually turned negative in early 2010. Given that this "deflation" was largely driven by falling energy prices, it suggests that a 24-month growth rate isn't quite long enough to dampen the effect of relative price shocks.

Evaluating the full set of trimmed-means

As in Bryan, Cecchetti, and Wiggins (1997), we calculate weighted $\alpha, \beta\%$-trimmed means in the following fashion. We first begin by sorting the monthly price-change data, $\{x_1, x_2, \dots x_i\}$, and the associated weights, $\{w_1, w_2, \dots w_i\}$. The percent trimmed from the lower tail is $\alpha = jn$, while the percent trimmed from the upper tail is $\beta = kn$; α and $\beta = \{0, 1, 2, \dots 50\}$. The $\alpha, \beta\%$-trimmed mean is given by:

$$\bar{x}_{\alpha,\beta} = \frac{1}{n-j-k} \sum_{i=j+1}^{n-k} w_i x_i$$

We trim the weighted distribution in 1-percent increments, to create 2601 unique trimmed-mean measures. To make comparisons to the forecast target, the subsequent 36-month percent change in the CPI, we annualize all figures. For example,

$$\bar{x}_{\alpha,\beta}^{ann} = \left(\left(\frac{\bar{x}_{\alpha,\beta}+100}{100}\right)^{12} - 1\right) * 100$$

We then compare each trimmed mean, $\bar{x}_{\alpha,\beta}^{ann}$, at time t to the forecast target at time $t+36$, and generate a summary measure for forecast accuracy, root mean square error ($RMSE$), as follows:

$$RMSE_{\alpha,\beta} = \sqrt{\frac{1}{t}\sum_t\left(\pi_{(t+36)}^{ann} - \bar{x}_{\alpha,\beta\,(t)}^{ann}\right)} \ ,$$

where $\pi_{(t+36)}^{ann}$ is the annualized, 36-month percent change in the CPI from time t to $t+36$.

In attempting to determine which trimmed-mean measure will be the best forecaster of future inflation, we take two approaches:

1) Non-rolling sample: Uses the entire time series, or a specific subsample, to find the trimmed-mean, $\bar{x}_{\alpha,\beta}$, with the lowest $RMSE$.
2) Rolling samples: Uses overlapping, 5-year windows. We calculate each $RMSE_{\alpha,\beta}$ for data from the initial 15-year window, t to $t+60$. The trimmed-mean, $\bar{x}_{\alpha,\beta}$, with the lowest $RMSE$ in this initial period is saved as the best trimmed-mean forecaster for time t, along with its trimming percentages α and β. This process is then repeated for the next 5-year window, $t+1$ to $t+61$, with the relevant information assigned to time $t+1$, and so on. In this way, we construct a time-varying, best trimmed-mean forecaster.

We then take these lowest-RMSE trimmed-mean measures, and compare their squared forecast errors to those from trimmed-means with higher RMSEs, to determine whether the differences are statistically significant. To do this, we use a Diebold-Mariano (DM) test with the pre-whitened quadratic spectral kernel introduced by Andrews and Monahan (1992) to correct the standard errors for heteroskedasticity and autocorrelation.[9] A common approach is to use a DM test with Newey-West standard errors that employs the Bartlett kernel. However, Clark and McCracken (2011) show that as the prediction horizon increases, this variance estimator becomes biased, overstating the rejection region.

Using the CPI component data described in the above section, we now compute the entire set of symmetric and asymmetric trimmed-mean price statistics. The least aggressive trimmed-mean is mean or, in this case, the headline CPI; and the most aggressive trimmed-mean measure is the median (which trims 49.99 percent from each tail). Unlike Bryan and

---

[9] We also tried the Harvey-Leybourne-Newbold (HLN) approach and found qualitatively similar results. However, as noted in Clark and McCraken (2011), the HLN test isn't guaranteed to yield a positive variance, which leads to incomplete results over a few time horizons.

Cecchetti, we evaluate our candidates for "optimal" trim on the basis of their ability to forecast our benchmark for inflation—that is the annualized percent change in the headline CPI over the next 3 years.

Figure 1 plots the RMSEs for all the various trimmed-mean measures over our full sample—from January 1967 to September 2011. The lowest RMSE belongs to $\bar{x}_{43,51}$, which is the trimmed-mean CPI that excludes 43 percent of the lower tail and nearly all (49 percent) of the upper tail. The contour plot reveals a small turquoise area, just shy of the symmetric trim (the dotted black line) where all of these trimmed-mean measures have a RMSE between 2.0 and 2.5.[10] That next swath (burnt orange) ranges from a RMSE of 2.5 and 3.0. This grouping is roughly symmetric and is fairly large—encapsulating the median CPI all the way to about a 6 percent symmetric trim. Going further out on the contour plot, the deterioration in forecast accuracy tends to follow the same shape; venturing too far away from a symmetric trim leads to a poorer RMSE. In fact, the worst performing trim, the $\bar{x}_{49,100}$ excludes the lower-half of the distribution leaving the just the upper-half remaining, and carrying a RMSE of 7.45 (roughly three times worse than the best performing trim).

Next, we'd like to pay particular attention to the period following what's commonly referred to as the "Volker Disinflation.[11]" Figure 2 shows the post-1983 contour plot. Post-1983, the $\bar{x}_{30,64}$ owns the best forecast accuracy, with a RMSE of 1.02. That said, there are nearly 80 or so trims (of the 2500 total) that have a RMSE within 0.1 percentage point of the $\bar{x}_{30,64}$. This contour plot differs somewhat from figure 1. While it is still the case that, in general terms, asymmetric trims are associated with worse forecasting performance; the lowest RMSE swath is much larger that in figure 1.[12]

To determine the statistical significance of these modest differences in forecasting performance, we employ the DM equality of prediction test that utilizes the pre-whitened quadratic-spectral kernel to ensure HAC standard errors. Figure 4 shows the results from our equality of prediction tests for the full sample period (1967-2011). This procedure tests for the mean difference in squared forecast errors between the trim with the lowest RMSE during that time period and all others. The orange areas show (with 90% and 95% confidence) where we fail to reject the null hypothesis that the mean difference forecast

---

[10] The lowest RMSE over the whole time period, belonging to $\bar{x}_{43,51}$ is 2.45, so the scale may be overstating the relative forecast accuracy a little.

[11] If the monetary regime, inflation process, or some other structural factor has changed since then, the results in Figure 1 may not hold. Also, the item strata changed in level of aggregation and component definition starting in January 1983. Bryan, Cecchetti, Wiggins (1997) note that the level of kurtosis positively correlated with the number underlying components. The addition of 6 additional components may have increased the kurtosis in the underlying price change distribution, changing the "optimal" trim.

[12] Figures 3a and 3b further illustrate the difference between the two time periods. The blue dots signify the minimum RMSE over these sample periods. Over both periods, the RMSE falls as the trimming point increases, say, to about a 10 percent symmetric trim. After that, the gains are much smaller.

errors between the "optimal" trim and the candidate trim are equal to zero. The boxes outlined in black represent the symmetric trims.

Figure 4 reveals a large swath of statically indistinguishable forecasts that include a majority of the symmetric trims. Interestingly, over this time period—one that includes the high inflation episode of the late 1970s—some extremely asymmetric cases could not be rejected. In fact, one could exclude the entire lower tail of the price-change distribution, as long as they trimmed at least 14 percent from the upper tail, and remain in the non-rejection area.

However, once we split the sample and focus on the post-1983 period, that extreme asymmetry disappears and the area of statistically indistinguishable predictions shrinks precipitously (illustrated in Figure 5). If anything, there appears to be a penalty in terms of forecasting accuracy as the trimming points become too asymmetric. More importantly, more aggressive trimming is needed to reach the edge of the non-rejection area. Focusing on just the symmetric cases for example, a 14% symmetric (7% from each tail) trim is statistically indistinguishable from the lowest RMSE trim in the full sample. Yet, once the 1967-1982 period is excluded from the sample, the amount of trimming necessary to reach the minimum threshold doubles to a 28% symmetric trim (or 14% from each tail). In either sample period, the median CPI appears in the non-rejection region.[13]

In general our results suggest:

1) There is a wide range of trimmed-mean measures that deliver roughly "equal" forecasting accuracy.
2) That range of trims almost always includes the symmetric trims (after trimming roughly 14 percent or so of the most dramatic price changes from the monthly distribution).
3) No single trimmed-mean measure can be declared unequivocally "the best." However, over nearly every time period we examined, the median CPI is included in the set of trims that has statistically equal forecasting accuracy as the lowest RMSE trim in that time period. Given that the median CPI is conceptually and computationally simple, we'd advocate for its continued use as an underlying inflation indicator.

---

[13] Given the modest differences between figures 4 and 5, we were curious about how much the "optimal" trim changed across sample periods. Figure 6 (in the appendix) illustrates the changing nature of "optimal" trim in a 5-year rolling-windows framework. Interestingly, it appears that the upper and lower trimming points move in concert, which hints at the symmetric nature of the process. This modest instability taken together with the wide area of non-rejection in figures 3 and 4, suggests that more aggressive (and roughly symmetric) trimming reduces the likelihood of producing poor forecasts.

An Illustrative Forecasting Test

Given the results above—that there is no single trimmed-mean measure that strictly dominates the median CPI—we still feel it necessary to illustrate its usefulness of the median CPI as an inflation forecaster relative to the headline and "core" CPI. To perform this illustrative forecasting test, we follow the procedure set forth by Crone, Khettry, Mester, and Novak (2011).

In arguing against the use of "core" inflation measures as useful predictors of future inflation (and therefore appropriate guideposts for monetary policymakers to follow), they use a fixed-window rolling regression technique to test whether the year-over-year growth rate in the CPI ex food and energy, CPI ex energy, or the median CPI (produced by FRBC) outperforms forecasts based on the trend in headline inflation. We proceed using the same framework.

The forecasting equation is:

$$\pi_{t,t+h} = \alpha + \beta x_{t-l,t} + \varepsilon_t \, ,$$

where $\pi_{t,t+h}$ is the annualized growth rate in headline inflation over the horizon 't' to 't+h.' We allow 't+h' = 6-, 12-, 24-, and 36-months ahead. Unlike Crone, Khettry, Mester, and Novak (2011) that restrict their independent "core" measure to its respective 12-month growth rate, in $x_{t-l,t}$, we allow l = 1,3,6,9,or 12 months.

Allowing for shorter growth rates in the various core measures is motivated largely by the work of Bryan and Cecchetti and is also suggested by Bryan and Meyer (2011). This research shows that the trimmed-means, by efficiently excluding noisy relative prices changes can more accurately forecast future inflation over shorter time horizons (i.e. 3-months). Bryan and Meyer even point out that the 3-month annualized growth rate in the median CPI can more accurately forecast 3-year ahead inflation than the 12-month growth rate in the headline CPI. They also note that as the length of the trend increases—say, looking at the 24-month annualized growth rate in the candidate forecasters—gains from using core inflation dissipate. To us, this actually makes intuitive sense. If, in fact, trimmed-means (and other core measures) are eliminating relative price noise appropriately, then as these noisy price movements unwind over longer horizons, the trend in core measures and headline inflation should converge.

We start by estimating over a 120-month fixed window from January 1984 to December 1993, and then use those coefficients to forecast inflation over the next 6-, 12-, 24-, and 36-months ahead.[14] The first forecast error for each forecasting horizon and

---

[14] Given our view that the growth rate in headline CPI over longer time horizons is more apt to reflect actual inflation, we are primarily concerned with the 24- and 36-month horizons.

specification is computed. We then roll the sample forward 1 month (to February 1984-January 1994) and repeat the forecast until our last period of 120 months that ends in May 2009, gathering up forecast errors at each step.[15]

As in Crone, Khettry, Mester, and Novak (2011), we use the forecast errors we collected from the rolling-windows procedure to calculate out-of-sample RMSEs for each specification and forecasting horizon. Again, to be comparable and consistent with Crone, Khettry, Mester, and Novak (2011), we also evaluate the statistical significance of the core measures' forecast accuracy relative to the headline CPI using the Giacomini-White (2003) equality of prediction test. This is the appropriate test because we are evaluating forecasts from two competing models with a fixed window length in a rolling-window scheme.

We calculate the Giacomini-White (GW) statistic as follows:

Let $d_t = e_{1t}^2 - e_{2t}^2$, where e_{1t}^2 is the squared prediction error from the baseline model at time t. In our case, it is the model that uses lagged growth rates of headline inflation has regressors. e_{2t}^2 is the squared prediction error for the competing models that use underlying inflation measures as regressors. The associated test statistic is $S_{GW} = \frac{\bar{d}}{\sqrt{\sigma^2 / n}}$, where $\bar{d}$ is the mean of d and σ^2 is the variance of the loss differential. The GW statistic is a two-sided t statistic estimated using heteroskedasticy and autocorrelation consistent (HAC) standard errors with $h\text{-}1$ lags to correct for autocorrelation. A useful property of this statistic is that negative values that the forecast from the baseline model is superior to the alternative model (and vice versa).

Tables 1 through 4 report the results of the out-of-sample rolling-window forecasting exercise that compares how well lags of headline CPI forecast future inflation relative to the core CPI, 16% trimmed-mean CPI, and median CPI.

Given our prior that the appropriate CPI inflation benchmark is 36-months ahead, we report those results first, in Table 1. We find that the models with the near-term trends in the core measures as independent variables all generally outperform the model with headline inflation as the independent variable in terms of RMSE (except for the model with the 1-month annualized growth rate in the core CPI). The GW statistics confirm that those differences are statistically significant, especially for the 12-month growth rate in median CPI (which carries a p-value of less than 0.003).

These findings are somewhat contrary to the results set forth by Crone, Khettry, Mester, and Novak (2011). We attribute much of this to a difference in sample periods. Our study included the forecasts stemming from the 2008 energy price shock, during which the 12-month growth rate in the headline CPI touched 5.0 percent. To us, this is exactly the

---

[15] We stop in May 2009 to maintain a consistent sample period across all time horizons, given that the last 36-month ahead forecast is evaluated on data through May 2012.

type of episode that illustrates the usefulness of core inflation measures, particularly the median CPI.

Tables 2, 3, and 4 repeat this forecasting exercise in forecasting headline inflation over the next 24-, 12-, and 6-month horizons. Consistent with Crone, Khettry, Mester, and Novak (2011), gains in forecasting accuracy of the different core measures start to disappear over shorter, not-policy-relevant, forecasting horizons and using longer-term (12-month) trends in the candidate measures. Though, in our tests, it is never the case that models based on the headline CPI significantly outperform any of the core measures.

Conclusion

While we originally set out to find a single superior trimmed-mean measure, we could not conclude as such. In fact, it appears that a large swath of candidate trims hold statistically indistinguishable forecasting ability. That said, in general, the best performing trims over a variety of time periods appear to be somewhat aggressive and almost always include symmetric trims. Of this set, the median CPI stands out, not for any superior forecasting performance, but because of its conceptual and computational simplicity—when in doubt, hit the one in the middle.

Interestingly, and contrary to Dolmas (2005) we were unable to find any convincing evidence that would lead us to choose an asymmetric trim. While his results are based on components of the PCE chain-price index, a large part (roughly 75% of the initial release) of the components comprising the PCE price index are directly imported from the CPI. It could be the case that the imputed PCE components are creating the discrepancy. The trimmed-mean PCE series currently produced by the Federal Reserve Bank of Dallas trims 24 percent from the lower tail and 31 percent from the upper tail of the PCE price-change distribution. This particular trim is relatively aggressive and is not overly asymmetric—two features consistent with the best performing trims in our tests.

Finally, even though we failed to best the median CPI in our first set of tests, it remains the case that the median CPI is generally a better forecaster of future inflation over policy-relevant time horizons (i.e. inflation over the next 2-3 years) than the headline and core CPI.
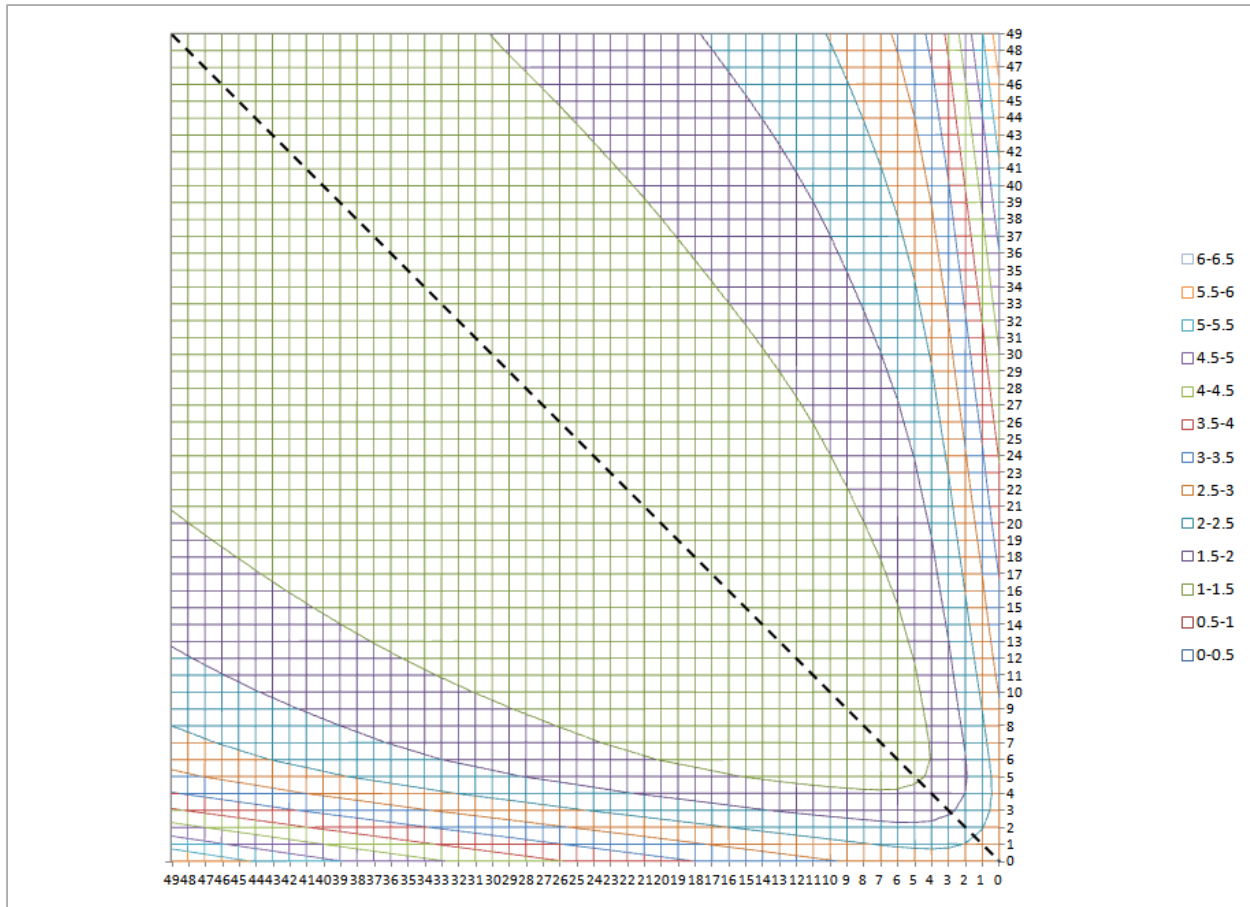
Figure 1: Contour plot of RMSEs for Trimmed Means, Full Sample

The following plots RMSEs for monthly CPI trimmed-means from forecasts of the 36-month ahead, 36-month percent change in the CPI, $RMSE_{\alpha,\beta} = \sqrt{\frac{1}{t}\Sigma_t(\pi_{(t+36)}^{ann} - \bar{x}_{\alpha,\beta\,(t)}^{ann})}$, for the period from January 1967 to September 2011. The lower-tail trim, α, is shown on the vertical axis, while the upper-tail trim, β, is shown on the horizontal axis. The symmetric trimmed-means, where $\alpha = \beta$, are shown by the dashed black line. The lowest RMSE area of the plot is in the upper-left corner (shown in turquoise), with an RMSE range of 2.0-2.5. Each successive contiguous area increases the RMSE range by 0.5.
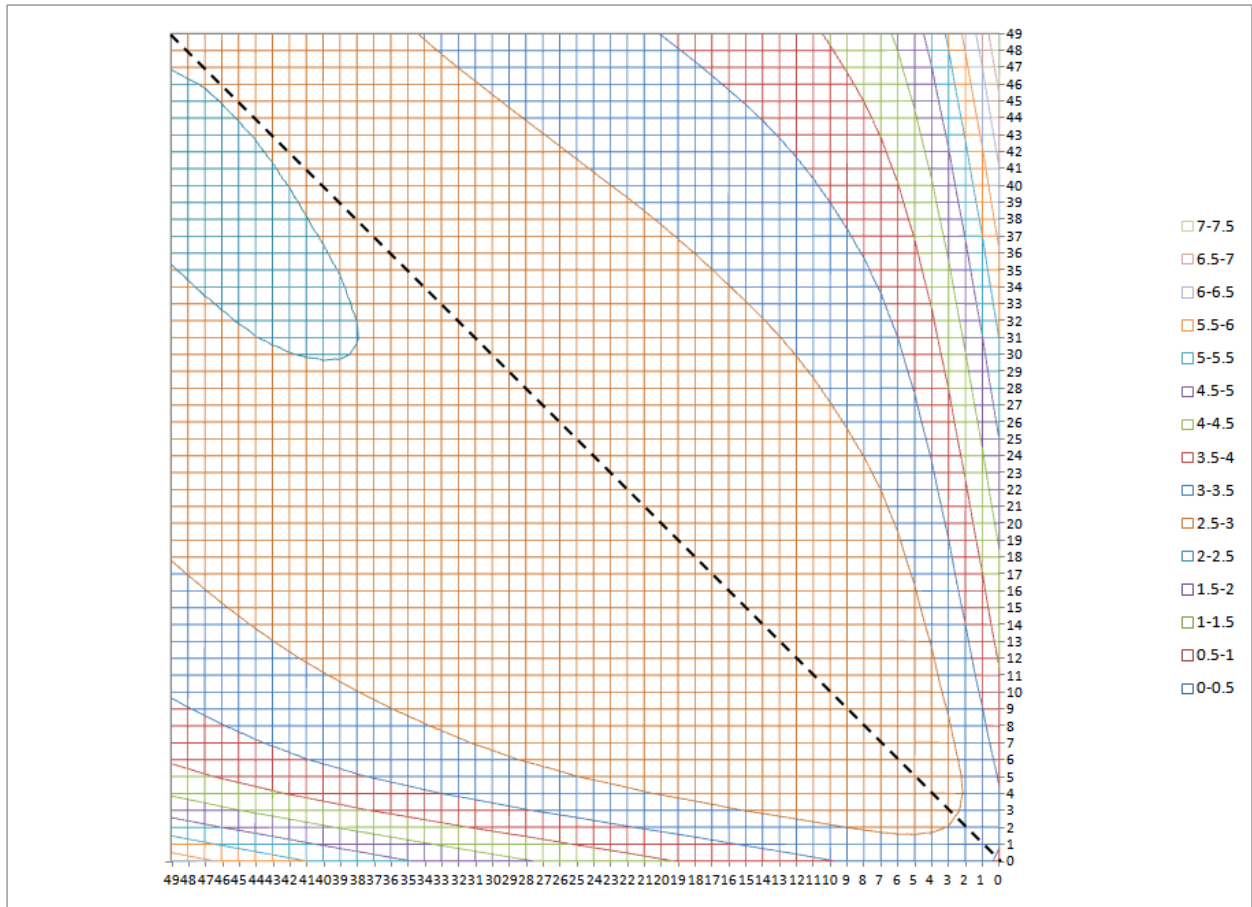
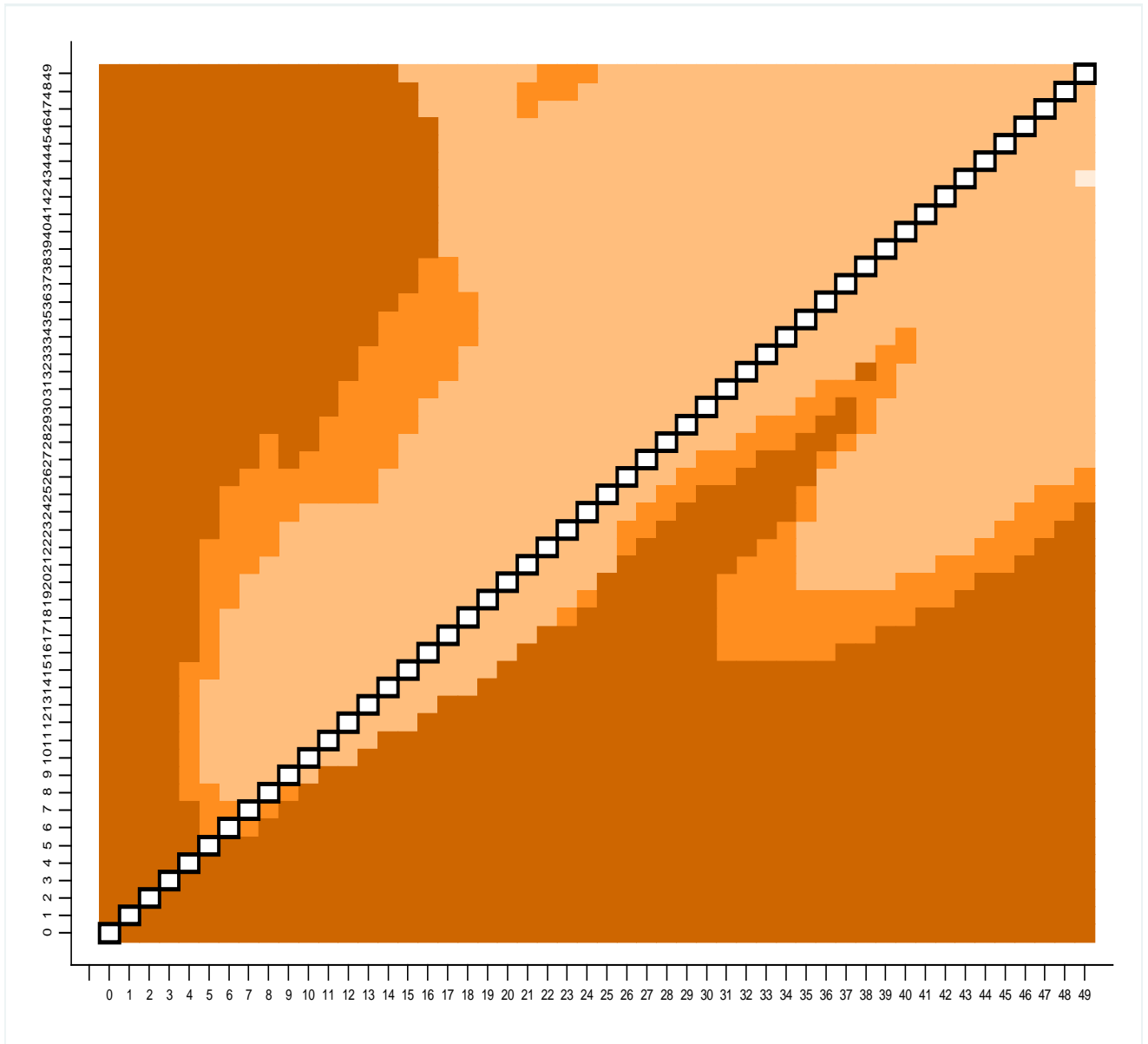**Figure 2:** Contour plot of RMSEs for Trimmed Means, Post-1983

The following plots RMSEs for monthly CPI trimmed-means from forecasts of the 36-month ahead, 36-month percent change in the CPI, $RMSE_{\alpha,\beta} = \sqrt{\frac{1}{t}\Sigma_t(\pi^{ann}_{(t+36)} - \bar{x}^{ann}_{\alpha,\beta\,(t)})}$, for the period from January 1982 to September 2011. The lower-tail trim, , is shown on the vertical axis, while the upper-tail trim, β , is shown on the horizontal axis. The symmetric trimmed-means, where $\alpha = \beta$, are shown by the dashed black line. The lowest RMSE area of the plot is roughly in the center (shown in light green), with an RMSE range of 1.0-1.5. Each successive contiguous area increases the RMSE range by 0.5.

Figure 3: Contour plots of Statistically Indistinguishable RMSEs for Trimmed Means, Full Sample

The following shows the trimmed-means with RMSEs that are statistically indistinguishable from the lowest-RMSE trimmed-mean of $\bar{x}_{43,49}$ (shown in white) for the period from January 1967 to December 2008, using the Diebold-Mariano (DM) test with the quadratic spectral kernel and Andrews (1991) optimal bandwidth . The 90% confidence interval is shown in light orange, while the 95% confidence interval includes both the light and dark orange areas. Brown areas show trimmed-means with RMSEs that are statistically different from that for $\bar{x}_{43,49}$. The lower-tail trim, α, is shown on the vertical axis, while the upper-tail trim, β, is shown on the horizontal axis. The symmetric trimmed-means, where $\alpha = \beta$, are highlighted with black boxes.

Figure 4: Contour plots of Statistically Indistinguishable RMSEs for Trimmed Means, Post-1983 Sample

The following shows the trimmed-means with RMSEs that are statistically indistinguishable from the lowest-RMSE trimmed-mean of $\bar{x}_{30,64}$ (shown in white) for the period from January 1983 to December 2008, using the Diebold-Mariano (DM) test with the quadratic spectral kernel and Andrews (1991) optimal bandwidth. The 90% confidence interval is shown in light orange, while the 95% confidence interval includes both the light and dark orange areas. Brown areas show trimmed-means with RMSEs that are statistically different from that for $\bar{x}_{30,64}$. The lower-tail trim, α, is shown on the vertical axis, while the upper-tail trim, β, is shown on the horizontal axis. The symmetric trimmed-means, where $\alpha = \beta$, are highlighted with black boxes.
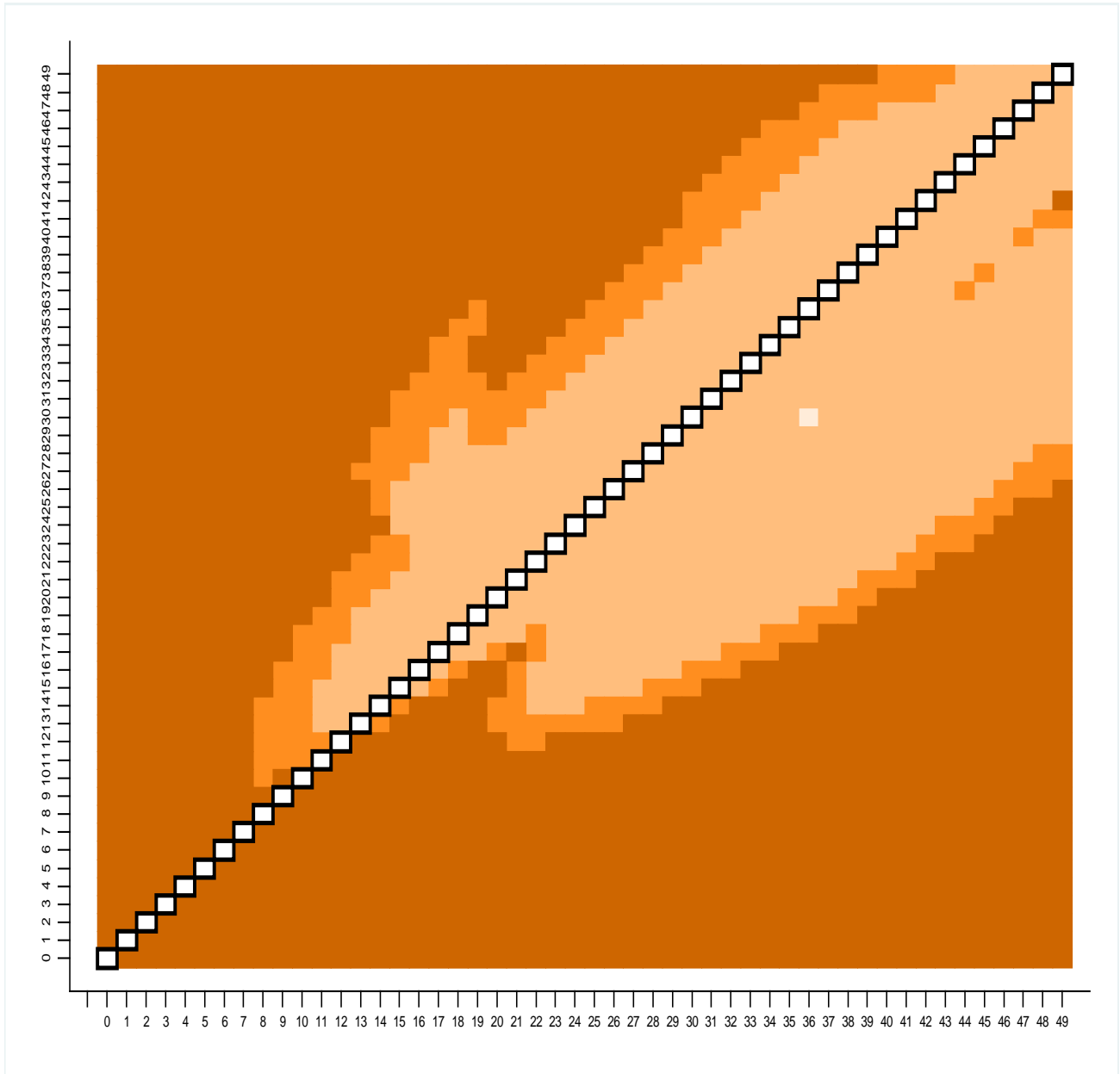
Table 1: Out-of-Sample Test of Forecast Accuracy: 36 months-ahead

Forecast Accuracy: 36-months ahead RMSES						
Independent variable						
Annualized percent change over the last;	**1 month**	**3 months**	**6 months**	**9 months**	**12 months**	
CPI	0.768	0.758	0.750	0.747	0.745	RMSE
Core CPI	0.712	0.636	0.602	0.600	0.602	RMSE
	1.930	*2.03**	*2.073**	*2.104**	*1.982**	GW statistic
16% trimmed-mean	0.692	0.632	0.597	0.589	0.558	RMSE
	2.085*	*2.156**	*2.427**	*2.566**	*2.8**	GW statistic
Median CPI	0.689	0.615	0.581	0.574	0.543	RMSE
	2.714*	*2.728**	*3.052**	*3.123**	*3.174**	GW statistic

* denotes significance at the 5% level

Note: The estimated equation is of the form : $\pi_{t,t+h} = \alpha + \beta x_{t-l,t} + \varepsilon_t$; where $\pi_{t,t+h}$ is the annualized growth rate in headline inflation over the horizon 't' to 't+h.' $x_{t-l,t}$ is the annualized growth rate in the independent, where l = 1,3,6,9,or 12 months. The out-of-sample forecast horizon is January 1994 through May 2009. The forecast errors are generated from 120 month fixed-window rolling regressions starting in January 1984 to December 1993. Giacomini-White (GW) equality of prediction test is estimated with HAC standard errors using *h-1* lags to control for autocorrelation. A negative value for the GW statistic indicates that the model with headline CPI as the regressor is more accurate than those using underlying inflation measures as regressors.

Table 2: Out-of-Sample Test of Forecast Accuracy: 24 months-ahead

Forecast Accuracy: 24-months ahead RMSES						
Independent variable						
Annualized percent change over the last;	**1 month**	**3 months**	**6 months**	**9 months**	**12 months**	
CPI	0.945	0.924	0.914	0.905	0.905	RMSE
Core CPI	0.904	0.853	0.837	0.824	0.805	RMSE
	1.218	*1.087*	*1.034*	*1.139*	*1.433*	GW statistic
16% trimmed-mean	0.873	0.826	0.795	0.779	0.716	RMSE
	1.5625	*1.297*	*1.383*	*1.608*	*2.003**	GW statistic
Median CPI	0.893	0.842	0.818	0.805	0.748	RMSE
	1.383	*1.312*	*1.646*	*2.001**	*3.629**	GW statistic

* denotes significance at the 5% level

Note: The estimated equation is of the form : $\pi_{t,t+h} = \alpha + \beta x_{t-l,t} + \varepsilon_t$; where $\pi_{t,t+h}$ is the annualized growth rate in headline inflation over the horizon 't' to 't+h.' $x_{t-l,t}$ is the annualized growth rate in the independent, where l = 1,3,6,9,or 12 months. The out-of-sample forecast horizon is January 1994 through May 2009. The forecast errors are generated from 120 month fixed-window rolling regressions starting in January 1984 to December 1993. Giacomini-White (GW) equality of prediction test is estimated with HAC standard errors using *h-1* lags to control for autocorrelation. A negative value for the GW statistic indicates that the model with headline CPI as the regressor is more accurate than those using underlying inflation measures as regressors.

Table 3: Out-of-Sample Test of Forecast Accuracy: 12 months-ahead

Forecast Accuracy: 12-months ahead RMSES						
Independent variable						
Annualized percent change over the last;	**1 month**	**3 months**	**6 months**	**9 months**	**12 months**	
CPI	1.313	1.285	1.271	1.254	1.215	RMSE
Core CPI	1.287	1.232	1.220	1.220	1.216	RMSE
	1.314	*0.961*	*0.696*	*0.404*	*-0.007*	GW statistic
16% trimmed-mean	1.250	1.174	1.155	1.162	1.123	RMSE
	1.5714	*1.307*	*1.250*	*1.195*	*1.029*	GW statistic
Median CPI	1.284	1.242	1.237	1.239	1.224	RMSE
	1.105	*0.699*	*0.471*	*0.187*	*-0.100*	GW statistic

* denotes significance at the 5% level

Note: The estimated equation is of the form : $\pi_{t,t+h} = \alpha + \beta x_{t-l,t} + \varepsilon_t$; where $\pi_{t,t+h}$ is the annualized growth rate in headline inflation over the horizon 't' to 't+h.' $x_{t-l,t}$ is the annualized growth rate in the independent, where l = 1,3,6,9,or 12 months. The out-of-sample forecast horizon is January 1994 through May 2009. The forecast errors are generated from 120 month fixed-window rolling regressions starting in January 1984 to December 1993. Giacomini-White (GW) equality of prediction test is estimated with HAC standard errors using *h-1* lags to control for autocorrelation. A negative value for the GW statistic indicates that the model with headline CPI as the regressor is more accurate than those using underlying inflation measures as regressors.

Table 4: Out-of-Sample Test of Forecast Accuracy: 6 months-ahead

Forecast Accuracy: 6-months ahead RMSES						
Independent variable						
Annualized percent change over the last;	**1 month**	**3 months**	**6 months**	**9 months**	**12 months**	
CPI	1.910	1.889	1.846	1.820	1.791	RMSE
Core CPI	1.865	1.831	1.823	1.813	1.804	RMSE
	1.304	*0.784*	*0.542*	*0.138*	*-0.237*	GW statistic
16% trimmed-mean	1.861	1.805	1.775	1.773	1.720	RMSE
	2.383*	*1.334*	*1.221*	*0.955*	*1.115*	GW statistic
Median CPI	1.850	1.835	1.837	1.837	1.817	RMSE
	1.401	*0.679*	*0.188*	*-0.358*	*-0.522*	GW statistic

* denotes significance at the 5% level

Note: The estimated equation is of the form : $\pi_{t,t+h} = \alpha + \beta x_{t-l,t} + \varepsilon_t$; where $\pi_{t,t+h}$ is the annualized growth rate in headline inflation over the horizon 't' to 't+h.' $x_{t-l,t}$ is the annualized growth rate in the independent, where l = 1,3,6,9,or 12 months. The out-of-sample forecast horizon is January 1994 through May 2009. The forecast errors are generated from 120 month fixed-window rolling regressions starting in January 1984 to December 1993. Giacomini-White (GW) equality of prediction test is estimated with HAC standard errors using *h-1* lags to control for autocorrelation. A negative value for the GW statistic indicates that the model with headline CPI as the regressor is more accurate than those using underlying inflation measures as regressors.

References

Andrews, D. W., & Monahan, J. (1992). An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica, Vol. 60, No. 4*, 953-966.

Blinder, A. S. (1997, May/June). Commentary. *Federal Reverse Bank of St. Louis, Review*.

Brischetto, A., & Richards, A. (2006). The Performance of Trimmed Mean Measures of Underlying Inflation. *Reserve Bank of Australia, Research Discussion Paper*.

Bryan, M. F., & Cecchetti, S. G. (1994). Measuring Core Inflation. In e. N. Gregory Mankiw, *Monetary Policy* (pp. 195-215). Chicago: University of Chicago Press.

Bryan, M. F., & Meyer, B. H. (2011, June 1). *Should we even read the monthly inflation report? Maybe not. Then again...* Retrieved from Macroblog: http://macroblog.typepad.com/macroblog/2011/06/should-we-even-read-the-monthly-inflation-report-maybe-not-then-again.html

Bryan, M. F., & Pike, C. J. (December 1, 1991). Median Price Changes: An Alternative Approach to Measuring Current Monetary Inflation. *Federal Reserve Bank of Cleveland, Economic Commentary*.

Bryan, M. F., Cecchetti, S. G., & Wiggins, R. L. (1997). Efficient Inflation Estimation. *National Bureau of Economic Research Working Paper no. 6183*.

Clark, T. E. (Second Quarter 2001). Comparing Measures of Core Inflation. *Federal Reserve Bank of Kansas City Economic Review*.

Clark, T. E., & McCracken, M. W. (2011). Advances in Forecast Evaluation. *Federal Reserve Bank of Cleveland, working paper no. 11-20*.

Cogley, T. (2002). A Simple Adaptive Measure of Core Inflation. *Journal of Money, Credit and Banking. Vol 34, No. 1*, 94-113.

Crone, T. M., Khettry, N. K., Mester, L. J., & Novak, J. A. (2011). Core Measures of Inflation as Predictors of Total Inflation. *Federal Reserve Bank of Philadelphia Woking Papers no. 11-24*.

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics, Vol. 13, No. 3*, 253-263.

Dolmas, J. (2005). Trimmed Mean PCE Inflation. *Federal Reserve Bank of Dallas, Research Department Working Paper 0506*.

Giacomini, R., & White, H. (2006). Tests of Conditional Predictive Ability. *Econometrica, Vol. 74*, 1545-1578.

Harvey, D. I., Leybourne, S. J., & Newbold, P. (1998). Tests for Forecast Encompassing. *Journal of Business and Economic Statistics, Vol. 16, No. 2*, 254-259.
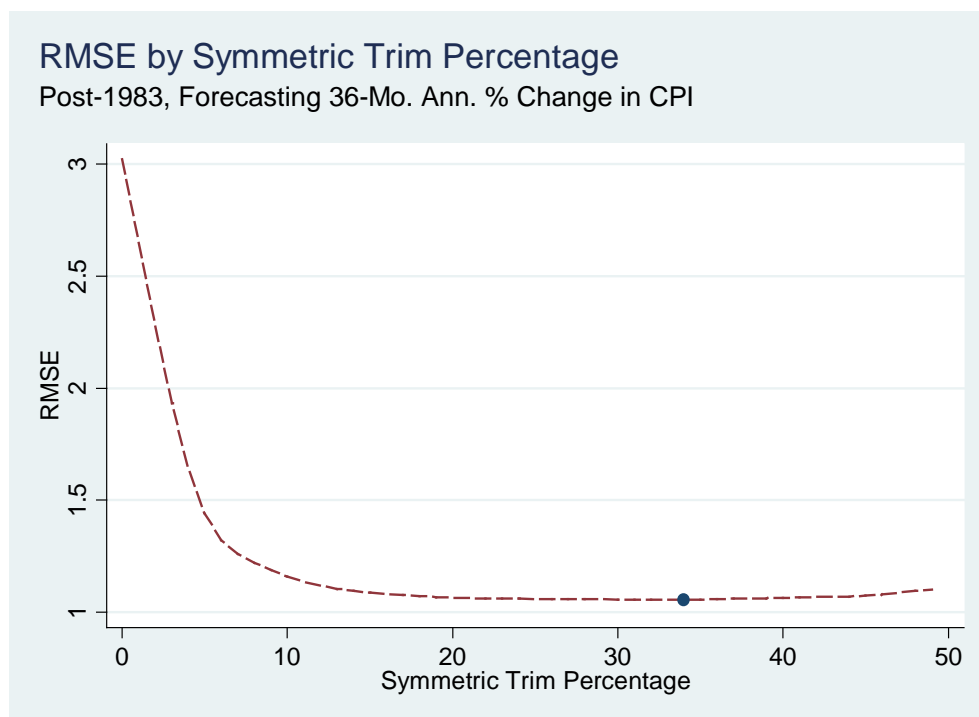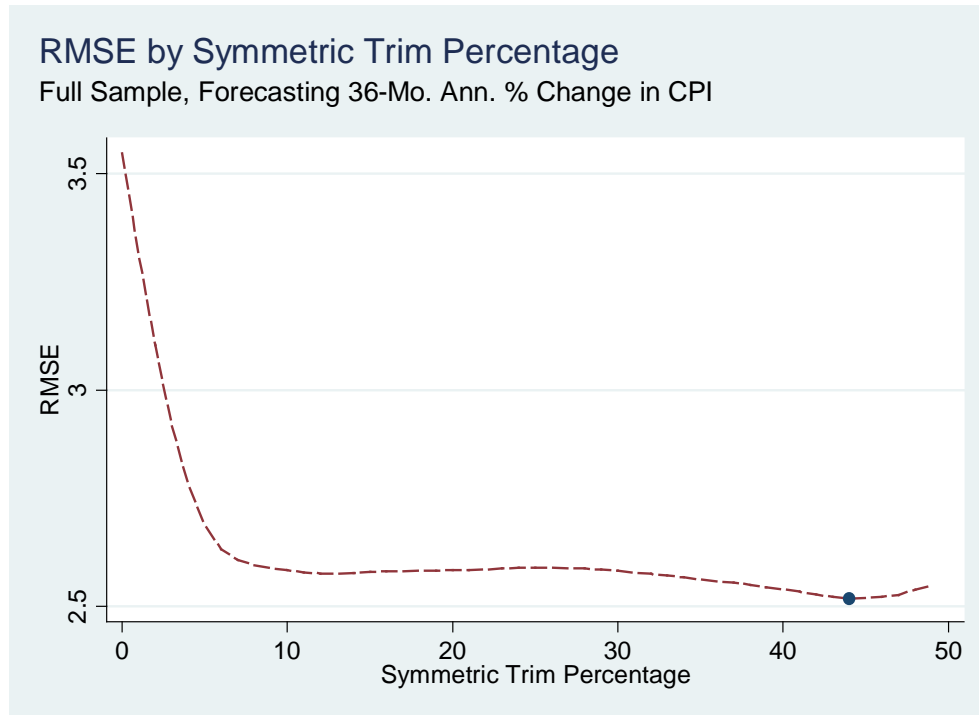
Meyer, B. H., & Pasaogullari, M. (2010). Simple Ways to Forecast Inflation: What Works Best? *Federal Reserve Bank of Cleveland, Economic Commentary, No. 2010-17*.

Smith, J. K. (2004). Weighted Median Inflation: Is This Core Inflation. *Journal of Money, Credit and Banking, Vol 36.*, 253-63.

APPENDIX

Figures 3a and 3b: RMSEs for Symmetric Trimmed Means, Full Sample and Post-1983

The following show the information presented in Figures 1 and 2, but only for symmetric trimmed-means (corresponding to the dashed black lines above), i.e., where $\alpha = \beta$. The blue dots identify the lowest-RMSE symmetric trimmed-means.

RMSE by Symmetric Trim Percentage
Full Sample, Forecasting 36-Mo. Ann. % Change in CPI

RMSE by Symmetric Trim Percentage
Post-1983, Forecasting 36-Mo. Ann. % Change in CPI

Figure 6: Optimal Trimming Points, Using 5-Year Rolling Windows

The following shows the trimming points, α and β, of the trimmed-mean, $\bar{x}_{\alpha,\beta}$, that minimize the RMSE for a forecast of the 36-month percent change in the CPI over overlapping 5-year intervals. For example, the first observations, for January 1967, show the trimming points of the RMSE-minimizing trimmed-mean for the 5-year period from January 1967 to December 1976. The second observations, for February 1967, show the trimming points of the RMSE-minimizing trimmed-mean for the 5-year period from February 1967 to January 1977, and so on. The blue line shows the lower-tail trim, α, while the red line shows 100 minus the upper-tail trim, β.

Optimal Trimming Points

Forecast of 36-Mo. % change in CPI using 5-year rolling windows