Measuring Core Inflation in Australia with Disaggregate Ensembles

Francesco Ravazzolo*
(Norges Bank)

Shaun P. Vahey†
(Melbourne Business School)

August 9, 2009

Abstract

We construct ensemble predictives for inflation in Australia based on the out of sample forecast performance of many component models, where each component model uses a particular disaggregate inflation series. Following Ravazzolo and Vahey (2009), the disaggregate ensemble can be interpreted as a measure of core inflation. We demonstrate that the ensemble forecast densities for measured inflation using disaggregate information by city and by sector are well calibrated. The resulting density forecasts outperform considerably those from a benchmark autoregressive model. And the point forecasts are competitive. From a structural perspective, the disaggregate ensemble core inflation measure suggests that the more traditional weighted median and trimmed mean measures periodically understate and overstate inflationary pressures in Australia.

Keywords: core inflation; underlying inflation; density combination; ensemble forecasting

JEL codes: C11; C32; C53; E37; E52

*Norges Bank, Research Department. francesco.ravazzolo@norges-bank.no
†University of Melbourne. spvahey@gmail.com

1 Introduction

Since the introduction of Inflation Targeting, many central banks have focused greater attention on the behaviour of measured inflation. Unfortunately, the theoretical concept of inflation is conceptually mismatched with the headline CPI measure; see, for example, the arguments in Quah and Vahey (2005). In particular, relative price movements are confounded with general price movements. For example, should we think of recent increases in commodity prices as part of inflation? Or as a relative price movement?

A number of central banks examine regularly disaggregate inflation series for less volatile and leading evidence of the inflationary process. The aim in using a 'core' or 'underlying' measure to communicate inflationary pressures is that the influence of relative prices can be removed, or at least moderated. (Hereafter, we use the terms core and underlying interchangeably.) One popular approach truncates (and averages) the disaggregate inflation (or price) cross-sectional distribution to provide a 'core' measure. A second approach excludes or zero-weights particular disaggregates; the resulting measure is commonly referred to as an 'Ex' core measure. In practice, the identity of the discarded series varies across central banks and through time. Although theoretical considerations are often advanced as a justification for both of these approaches, there is considerable uncertainty over which disaggregates or what proportion of the cross-section should be discarded. Faced with this uncertainty, practitioners often propose that the acid test of candidate core inflation measures should be the ability to predict measured inflation at a given horizon; see, for example, Roger (1998), Wynne (1999) and Smith (2004).

In this paper, we reformulate the measuring underlying inflation problem. We start by focusing directly on the forecasting problem, limiting our attention to candidate disaggregate series as forecasting variables. In contrast to the earlier literature on core inflation, we assess forecasting performance based on the complete density for inflation. Tests of point forecast accuracy provide no guidance on the usefulness of core measures for general (but unknown) loss functions.

Our ensemble methodology follows the analysis of US core inflation by Ravazzolo and Vahey (2009). We construct ensemble predictives based on the out of sample forecast performance of many component models, where each component model uses a particular disaggregate series. We demonstrate that the ensemble predictives provide well-calibrated forecast densities for measured inflation in Australia. Combining the evidence from two sources of disaggregation, by city and by sector, yields considerable improvements in density performance. The resulting density forecasts are preferable to those from a benchmark autoregressive model, with competitive point forecast performance. The core inflation measure defined as the h-step ahead disaggregate ensemble forecast uses time varying weights on the components, but typically does not discard individual disaggregates.

In our application, we focus entirely on one quarter ahead forecasts. Within the underlying inflation literature, the horizon of interest varies, typically between one and 8 quarters ahead. Although longer horizon ensemble forecasts are possible with our methodology, we prefer to focus on horizons much shorter than the focal range of many Inflation Targeting central banks. Our hope is that the disaggregate ensemble core measure picks up the inflation already in the pricing pipeline, and does not respond to future changes in policy stance. For further discussion of the choice of forecasting horizon in the core inflation literature, see Brischetto and Richards (2006).

The remainder of this paper is as follow. We provide a brief review of the core inflation literature in section 2. We discuss our ensemble modeling strategy in section 3. We describe our component models and our ensemble predictives in section 4. We summarize out Australian data set in section 5, and present our results in section 6. We conclude and make some suggestions for subsequent research in the final section.

2 A brief review of the core inflation literature

It is widely recognized by the public and central bankers that movements in the CPI do not always capture "inflation". Although measured (CPI) inflation assesses the cost of acquiring a representative basket of particular goods and services, inflation is usually defined as a "sustained increase in the general price level". The weights on the disaggregates in the cost of living index reflect the preferences and budget constraint of the representative consumer. But those weights can lead to a misleading assessment of inflation pressures because relative price changes are confounded with sustained general price movements.

In the core inflation literature, the aim is to measure the part of general price increase that is sustained over time. Most central banks consider a variety of measures of underlying inflation. Many of these are derived by removing the unwanted component, which is often treated as 'noise'; see, for further discussion, Brischetto and Richards (2006).

Traditional methods for measuring core inflation include smoothing and structural time series modeling. The first of these takes a moving average of measured inflation and labels this the core. The second makes specific assumptions about the functional form for underlying inflation (such as taking it to be a Gaussian random walk) and produces an estimate with the Kalman filter. Neither approach has any particular theoretical appeal. There is no economic justification for assuming that either, say, a 4-quarter moving average or an ARIMA(1,1,3) process removes the influence of relative prices.

Partly as a result of dissatisfaction with applications of these statistical techniques, many central banks consider measures of core inflation obtained by zero weighting particular components. Bryan and Cecchetti (1994) took this approach one step further by zero-weighting the disaggregates in the tails of the cross section. Although Bryan and Cecchetti (1994) offer a menu cost model as a rationale for truncating the distribution, that theory does not imply any particular truncation factor for the disaggregate distribution.

A related problem blights 'Ex' core measures that always exclude particular components. The argument for using 'Ex' measures is that if you know that one or two (or more) disaggregate series contain a great deal of 'noise', then they should be dropped from measured inflation to form the core measure. For example, Personal Consumption Expenditures chained price index excluding food and energy in the US; see Ravazzolo and Vahey (2009). But there is no theory that says food and energy do not matter for inflation. Casual observation may support the hypothesis that these are volatile components of the measured inflation. But so are many other disaggregate series. Should they be excluded too? And how many disaggregate series should we exclude?

The uncertainties involved in the selection of truncation factors, or the identification of series to exclude, affect the usefulness of the candidate underlying measures as communication tools. The public often suspect that the central bank exploiting these communication devices prefers to ignore inconvenient data. For example, the December 1997 Reserve Bank of New Zealand background briefing for the PTA mentions this difficulty as a motivation for discontinuing its use.[1]

In our ensemble approach described below, we avoid using strong off-model or prior information about which disaggregates signal future values of measured inflation. Instead, we formulate the measuring core inflation problem as one of combining component forecast densities for measured inflation, where each component is based on a particular disaggregate series. In this sense, we let the data speak clearly about which disaggregates are important. If particular disaggregates do not matter for inflation in the next quarter, they receive a small weight (bounded at zero). In so doing, we formally account for the uncertainty over which disaggregate should be included, and also over the type of disaggregation.

[1] See http://www.rbnz.govt.nz/monpol/pta/0055243.html.

3 Modeling strategy

Garrat, Mitchell and Vahey (2009) drew attention to the antecedents of density combination in macroeconomic forecast applications.[2] Outside of the econometrics literature, the benefits of the ensemble approach to forecasting have been recognized for around 15 years. Meteorologists and statisticians have focused a great deal of attention on analyzing statistical ensembles. The idea behind the ensemble approach is to consider a large number of models, each of which is a variant or component of the 'preferred' specification. Each component could be viewed as an approximation of the current state of the 'true' but unknown specification, an considered together, the ensemble approximates the truth.

In the meteorological forecasting literature, the ensemble methodology is a response to what macroeconometricians sometimes call 'uncertain instabilities'; see, for example, Clark and McCracken (2009). Namely, that individual empirical specifications tend to exhibit instabilities and which can be difficult to isolate with short runs of real-time macroeconomic data.

In the ensemble approach, density forecasts are generated from a common theoretical framework with slightly different initial conditions (measurements, auxiliary assumptions). The framework from which the component specifications are derived might allow for data, parameter, and/or model uncertainty. Ensemble predictive methods are commonly used by the majority of weather prediction institutions worldwide.[3]

Bache, Mitchell, Ravazzolo and Vahey (2009) list four common characteristics of an ensemble modeling strategy for macro modeling.

1. Generation of forecasting densities, rather than point forecasts
2. Predictive density construction from a large number of component macro-econometric models
3. Forecast density evaluation and combination based on out of sample performance, rather than in-sample analysis
4. Component model weights vary through evaluation—ensemble densities have time varying weights

[2] Bache, Mitchell, Ravazzolo and Vahey (2009) summarize the recent literature in economics using the ensemble methodology.

[3] For example, the "Ensemble Prediction System" developed by the European Centre for Medium-Range Weather Forecasts.

Papers in the economics literature that satisfy these criteria include (among others): Jore, Mitchell and Vahey (2009), Kasha and Ravazzolo (2009), Gerard and Nimark (2008) and Garratt, Mitchell and Vahey (2009).[4] In these cases, the out of sample densities from many macro-econometric component models are directly combined into the ensemble using an 'opinion pool'.[5] These papers differ in the design of the model space and the number of components considered, as well as the applied problem of interest.

Another strand of the ensemble economics literature uses informative priors and Markov chain Monte Carlo methods to produce ensembles. Maheu and Gordon (2008) and Geweke (2009) use mixture models to give non-Gaussian predictives; Andersson and Karlsson (2007) produce symmetric Gaussian predictive densities from many vector autoregressions.[6] Geweke (2009) discusses the relationships between density pooling and mixture modeling, and argues that the former presents a more coherent approach for incomplete model spaces. Clearly, both variants can be effective methods for combining densities in forecasting applications. (In a related literature, Patton (2004), Maheu and McCurdy (2009) and Amisano and Geweke (2009) consider ensembles in various financial applications.)

Before we move onto discuss the model space and ensembles for our core inflation application, it is worth considering whether we want a core measure to forecast the entire density of measured inflation. In our view, restricting attention to point forecast accuracy makes no sense. There is no reason to believe that the inflation process is Gaussian; and there is nothing particulary compelling about the quadratic loss function. In the absence of either assumption, RMSFE forecast accuracy has no justification. And with an unknown loss function, the forecast densities can only be evaluated by considering their calibration properties; see the discussion in Mitchell and Wallis (2009). Put simply, if we want to evaluate a core measure on forecast performance, the whole forecast density seems natural.

A second issue worthy of reflection is whether we want a core measure to forecast at all. For example, Quah and Vahey (1995) argue for a structural interpretation of underlying measures. Brischetto and Richards (2006) also take a structural view. While we think

[4]Smith *et al.* (2009) consider the performance of the Norges Bank nowcasting system which also adopts the ensemble methodology.

[5]Wallis (2005) uses opinion pools to average (model free) survey forecasts, rather than those from macro-econometric models. Mitchell and Hall (2005) use opinion pools to combine forecasts from two institutions.

[6]Frequentist approaches to mixture model estimation are also feasible but practitioners have tended to prefer Bayesian simulation methods with scope for informative priors.

structural issues are important, in this paper we start with the perspective that the core should be informative about inflation probabilities. We shall return to the structural interpretation of our preferred core measure at the end of our application.

4 Component model space and ensembles

For each observation in the policymaker's out of sample 'evaluation period', we use forecast density combination to compute the weight on each component model. The component models use a common time series structure, namely an autoregressive specification with four lags, AR(4).[7] Each component model uses a particular disaggregate inflation measure. The weights on the individual components are based on the 'fit' of the component predictive densities for measured inflation. Given these weights, we construct ensemble forecast densities for measured inflation.

More formally, consider a policymaker aggregating forecasts supplied by 'experts', each using a unique component forecasting model. Given $i = 1, \ldots, N$ components (where N could be a large number), we define the ensemble core by the convex combination also known as a linear opinion pool:

$$DE_\tau = p(\pi_{\tau,h}) = \sum_{i=1}^{N} w_{i,\tau,h}\ g(\pi_{\tau,h} \mid I_{i,\tau}), \qquad \tau = \underline{\tau}, \ldots, \overline{\tau}, \tag{1}$$

where $g(\pi_{\tau,h} \mid I_{i,\tau})$ are the h-step ahead forecast densities from component model i, $i = 1, \ldots, N$, conditional on the information set I_τ.

Each component model forecasts disaggregate inflation. Then in each recursion, we center the component forecasts on measured inflation. In effect, this step restricts the ensemble to be uni-modal but not symmetric.[8].

After this centering procedure, each component model produces h-step ahead forecasts for measured inflation. Each component model uses data, dated $\tau - h$ or earlier, to produce an h-step ahead forecast density for τ. The non-negative weights, $w_{i,\tau,h}$, in this finite mixture sum to unity, are positive, and vary by recursion in the evaluation period $\tau = \underline{\tau}, \ldots, \overline{\tau}$.

We emphasize that the ensemble forecast density could be non-Gaussian even if the component models produce Gaussian predictives. The linear opinion pool ensemble (1)

[7] Ravazzolo and Vahey (2009) consider time varying parameter components.

[8] Bao *et al* (2009) discuss the common practice of centering ensemble forecast densities prior to combination.

accommodates skewness and kurtosis. The flexible structure resulting from linear pooling allows the data to reveal whether, for example, the ensemble should have fat tails, or asymmetries.[9]

We construct the ensemble forecast density for measured inflation using equation (1), which we label core inflation. Implementation of the density combination requires a measure of component density fit to provide the weights. A number of recent applications in the economics literature have used density scoring rules. In this application, we utilize the Continuous Ranked Probability Score (CRPS), which as (among others) Gneiting and Raftery (2007), Panagiotelis and Smith (2008) and Ravazzolo and Vahey (2009) note, rewards predictive densities from component models with high probabilities near (and at) the outturn.[10]

The weights for the h-step ahead DE CPI densities are:

$$w_{i,\tau,h} = \frac{\left[\sum_{\underline{\tau}}^{\tau-1-h} X(g(\pi_{\tau,h} \mid I_{i,\tau}))\right]}{\sum_{i=1}^{N}\left[\sum_{\underline{\tau}}^{\tau-1-h} X(g(\pi_{\tau,h} \mid I_{i,\tau}))\right]}, \qquad \tau = \underline{\tau}, \ldots, \overline{\tau}. \tag{2}$$

where $g(\pi_{\tau,h} \mid I_{i,\tau})$ is the centered predictive density for measured aggregate inflation $\pi_{\tau,h}$ given by model i; and, where X is the CRPS-based measure of density performance as in Ravazzolo and Vahey (2009).

Using (1) and (2), we construct two disaggregate ensembles (DE) that combine predictive densities from sectors, and cities, respectively. The sector DE, denoted "DE_s", contains 10 components (sector disaggregates); the city DE, "DE_c", contains 8 components (city disaggregates). We will also use the "grand ensemble" technique proposed by Garratt, Mitchell and Vahey (2009) to compare and combine the two ensembles based on different types of disaggregation. We denote this "DE_cs" and we give equal weights (e.g. 0.5) to the ensembles, DE_c and DE_s.[11]

As a benchmark for our forecast evaluations, we use an AR(4) model for measured inflation. We use a noninformative priors for the AR(4) parameters, with an expanding window for estimation—so that forecasts are recursive. The predictive densities follow the

[9] Kascha and Ravazzolo (2009) compare and contrast logarithmic and linear pooling. Logarithmic opinion pools force the ensemble predictives to be symmetric, but accommodate fat tails; see also, Smith *et al.* (2009).

[10] See Panagiotelis and Smith (2008) for an explanation of how CRPS is calculated for each component density.

[11] Garratt, Mitchell and Vahey (2009) explore the use of recursively estimated weights to construct their grand ensembles.

t-distribution, with mean and variance equal to OLS estimates; see, for example, Koop (2003) for details.

To assess the calibration properties of the ensemble densities we follow Diebold *et al.* (1998) and compute *PITS*, probability integral transforms, and apply the Berkowitz (2001) likelihood ratio test for independence, zero mean and unit variance of the *PITS*. The test statistic is distributed $\chi^2(3)$ under the null hypothesis of no calibration failure, with a maintained hypothesis of normality. We also report the average (over the evaluation period $T = \overline{\tau} - \underline{\tau}$) logarithmic score. The logarithmic score of the i-th density forecast, $\ln g(\pi_{\tau,h} \mid I_{i,\tau})$, is the logarithm of the probability density function $g(. \mid I_{i,\tau})$, evaluated at the outturn $\pi_{\tau,h}$. Hence, the log score evaluates the predictives at the outturn only. We investigate relative predictive accuracy by considering a Kullback-Leibler information criterion (KLIC)-based test, based on the expected difference in two models' log scores; see Bao *et al.* (2007), Mitchell and Hall (2005) and Amisano and Giacomoni (2007). Suppose there are two density forecasts, $g(\pi_{\tau,h} \mid I_{1,\tau})$ and $g(\pi_{\tau,h} \mid I_{2,\tau})$, so that the KLIC differential between them is the expected difference in their log scores: $d_{\tau,h} = \ln g(\pi_{\tau,h} \mid I_{1,\tau}) - \ln g(\pi_{\tau,h} \mid I_{2,\tau})$. The null hypothesis of equal density forecast accuracy is $\mathcal{H}_0 : E(d_{\tau,h}) = 0$. A test can then be constructed since the mean of $d_{\tau,h}$ over the evaluation period, $\overline{d}_{\tau,h}$, under appropriate assumptions, has the limiting distribution: $\sqrt{T}\overline{d}_{\tau,h} \to N(0, \Omega)$, where Ω is a consistent estimator of the asymptotic variance of $d_{\tau,h}$.[12] Mitchell and Wallis (2009) explain the importance and practical difficulties of using information-based methods to discriminate between competing density forecasts.

5 Data

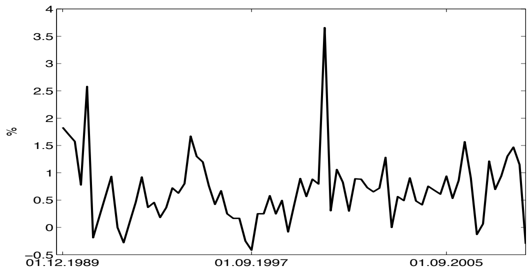
We apply our ensemble methodology to combine Australian disaggregate inflation forecasts for quarter on quarter growth of the Consumer Price Index (CPI). We assess the accuracy of the disaggregate ensembles, and other core measures, using an evaluation period from 1997Q1 to 2008Q4 (48 observations). The period 1994Q4 to 1996Q4, we use as a 'training period' to initialize the ensemble weights.[13]

As mentioned above, the Australian CPI gives a breakdown by sectors and cities. The first decomposes CPI into 10 disaggregates representing sectors. In our empirical

[12] When evaluating the density forecasts we treat them as primitives, and abstract from the method used to produce them. Amisano and Giacomoni (2007) and Giacomini and White (2006) discuss more generally the limiting distribution of related test statistics.

[13] Data are available from Australian Bureau of Statistics http://www.abs.gov.au.

Figure 1: Data



(a) CPI

Note: The graph show Australian quarterly growth CPI inflation over the sample period 1989:Q4-2008Q4.

analysis, we exclude the sector 'financial and insurance services' for which there are data from 2005Q3 only. The second form of disaggregation decomposes CPI into 8 cities.

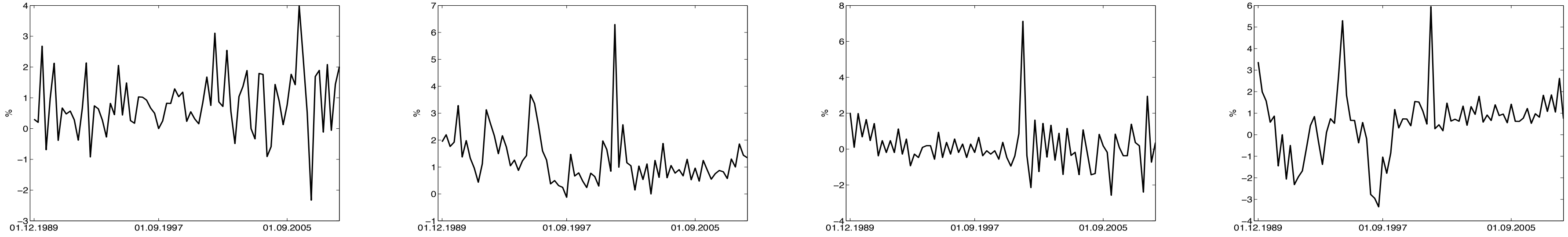
Figures 1, 2 and 3 plot respectively the CPI, its sector disaggregates and its city disaggregates over the sample 1989Q4 - 2008Q4. One striking feature is the high degree of contemporaneous dependence across cities. In contrast, the sectors display more heterogeneity, with differences in means and volatility.

6 Results

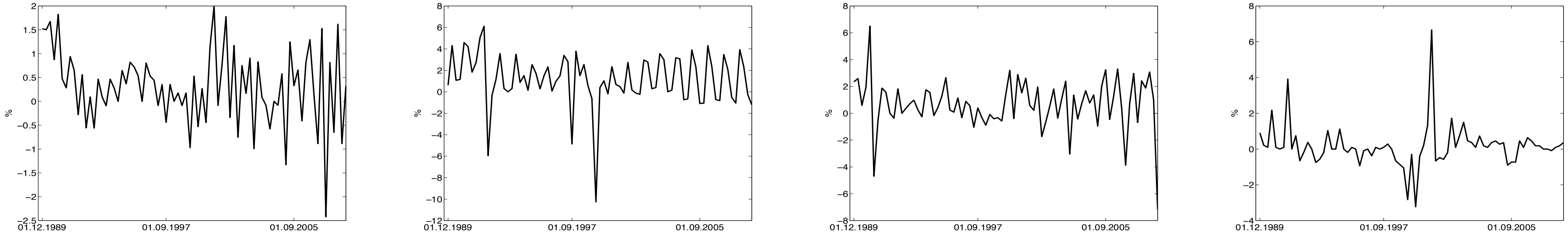
Recall that we construct the disaggregate core inflation measure, DE, by combining the predictive densities from the disaggregate component models. We compare and contrast the ensembles using disaggregation by sector DE_s, by city DE_c, and the grand ensemble of the two, DE_cs. Below we report evaluations for the one step (one quarter) ahead horizon.[14]

[14]We also computed but do not report forecasts for two, three and four step ahead horizons. Results are qualitatively similar and available upon request.

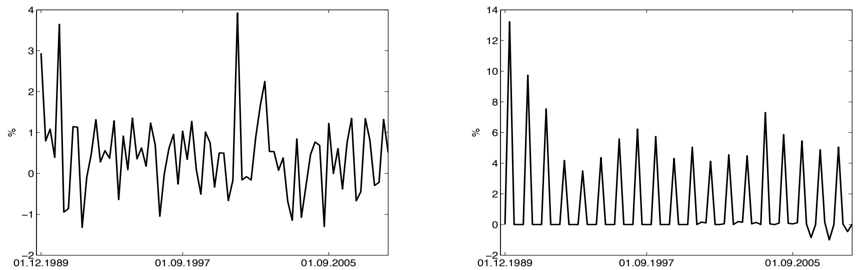
Figure 2: Disaggregate Inflation: Sectors



(a) Food - Alcoholic bev. & tobacco - Clothing & footwear - Housing, water, el. & fuels



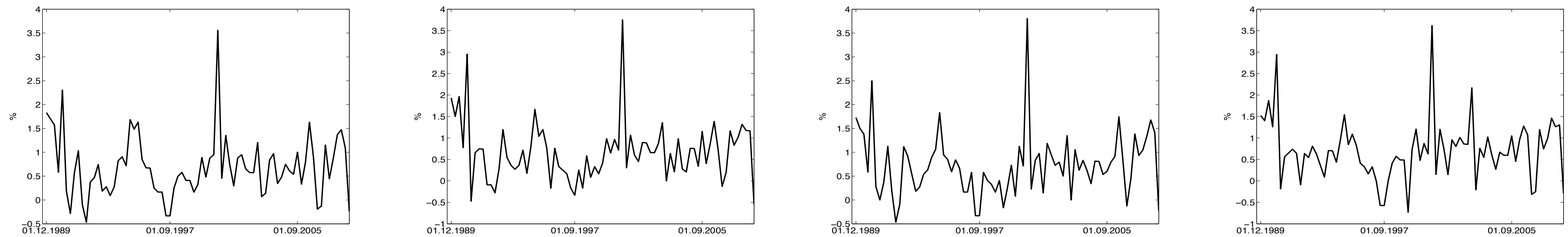
(b) Furnishings & house equipment - Health - Transport - Communications



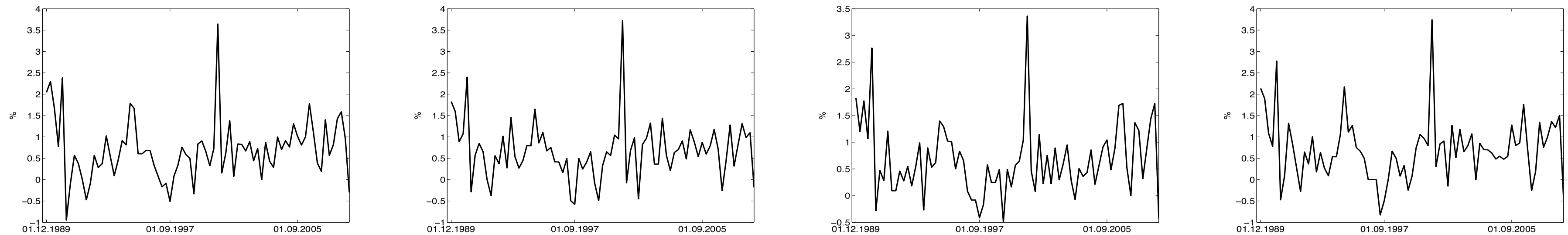
(c) Recreation - Education

Note: The graphs in these figures show Australian quarterly growth CPI sector disaggregates over the sample period 1989:Q4-2008Q4.

Figure 3: Disaggregate Inflation: Cities



(a) Sydney - Melbourne - Brisbane - Adelaide



(b) Perth - Hobart - Darwin - Canberra

Note: The graphs in these figures show Australian quarterly growth CPI city disaggregates over the sample period 1989:Q4-2008Q4.

Table 1: Forecast performance

	LR	LS	LS-test
AR	0.185	-1.078	
DE_s	0.222	-0.940	0.148
DE_c	0.170	-0.128	0.065
DE_cs	0.215	-0.857	0.037

Note: The column LR is the Likelihood Ratio p-value of the test of zero mean, unit variance and independence of the inverse normal cumulative distribution function transformed *PITS*, with a maintained assumption of normality for transformed *PITS*. LS is the average logarithmic score, averaged over the evaluation period. LS-test is the p-value of the KLIC-based test for equal density forecasting performance of AR and DE12 over the sample 1997Q1 to 2008Q4.

Before turning to the density evaluations for our various ensembles, we summarize point forecast performance. The root mean squared prediction error (RMSPE) of DE_s, DE_c, DE_cs, and the benchmark AR(4) are 0.520, 0.548, 0.471 and 0.378, respectively. The Clark and West (2006) test for superior predictive accuracy (against the null of equal accuracy) indicates that the ensembles are competitive with the AR benchmark with test statistics of 1.562, 1.619, 1.596 for the DE_s, DE_c, and DE_cs, respectively. The critical value for rejection of the null for a 95% interval is 1.645.[15]

We turn now to the ex post (end of period) evaluation of the forecast densities from the ensemble core measures and the benchmark. Table 1 has four rows; one for each ensemble and the benchmark. The columns report (reading from left to right) the Berkowitz likelihood ratio test (based on the $PITS$), the log scores (averaged over the evaluation period), and the p-values for the equal predictive density accuracy test (based on the log scores), respectively. Whereas DE_s, DE_c, and DE_cs appear to be well calibrated on the basis of the Berkowitz likelihood ratio, the final column shows that the AR is rejected in favor of DE_cs only using the KLIC-based test. The core measure DE_cs delivers a statistically significant improvement in the log score (reported in the second column) based on a 95 percent confidence interval.

The weights in DE_s and DE_c display some variation through time. Tables 2 and

[15] Smith (2004) and Kiley (2008) discuss the point forecasting properties of various core inflation measures. Most fail to outperform simple AR benchmarks.

Table 2: Disaggregate weights, DE_s

	1997Q1	2002Q4	2008Q4
Food	0.166	0.178	0.156
Alc. bev. and tobacco	0.100	0.098	0.110
Cloth. and footwear	0.095	0.068	0.080
Housing, water, el. and fuel	0.077	0.086	0.100
Furnishings and house equip.	0.150	0.159	0.128
Health care	0.066	0.068	0.070
Transport	0.094	0.112	0.104
Communications	0.122	0.095	0.111
Recreation	0.085	0.093	0.101
Education	0.046	0.044	0.041

Note: The columns reports disaggregate weights for the ensemble DE_s in three observations, 1997Q1, 2002Q4 and 2008Q4.

Table 3: Disaggregate weights, DE_c

	1997Q1	2002Q4	2008Q4
Sydney	0.182	0.175	0.182
Melbourne	0.085	0.070	0.079
Brisbane	0.120	0.143	0.138
Adelaide	0.135	0.120	0.110
Perth	0.096	0.116	0.110
Hobart	0.074	0.111	0.104
Darwin	0.108	0.135	0.148
Canberra	0.200	0.131	0.130

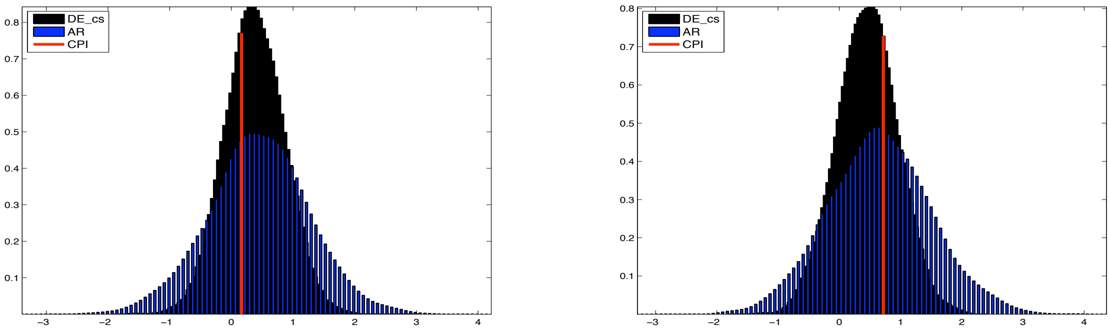
Note: The columns reports disaggregate weights for the ensemble DE_c in three observations, 1997Q1, 2002Q4 and 2008Q4.

3 report the weights on the sector and city disagregates, respectively, for three specific observations. It can be seen from Table 2 that generally all disaggregate components have a non-zero weight.[16] There does not seem to be a case for excluding the information on individual disaggregates, or groups of particular disaggregates, on the basis of these weights.

To provide insight into the probability of tail events for inflation, figure 4 provides the ensemble predictive densities from DE_sc at particular observations, namely 1997Q1 and 2002Q4, the first and the middle sample values in our evaluation period. We see that the AR(4) benchmark produces density forecasts that are too wide, with a high probability

[16]Geweke (2009) argues that even a zero weight is not sufficient to conclude that a component model has zero value for the linear opinion pool.

Figure 4: AR and DE_cs density forecasts



(a) 1997Q1-2002Q4

Note: The figures plot the histogram of the density forecasts given by AR benchmark and by the disaggregate ensemble DE_cs for two different periods, the first and middle sample forecasts. The realized value for CPI is also provided.

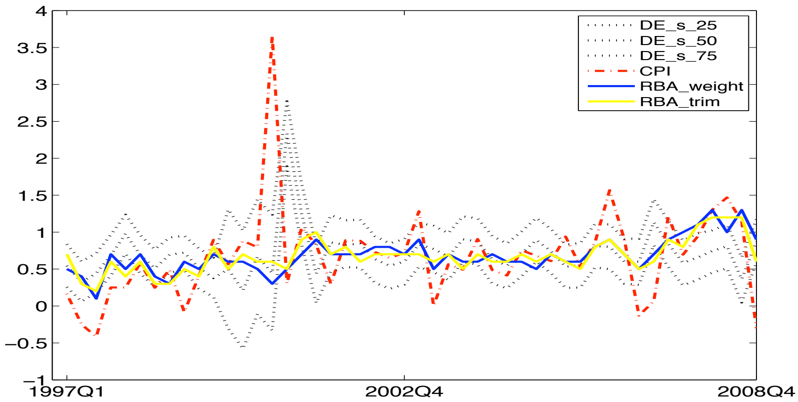
mass attributed to (quarterly) inflation of greater than two percent in absolute value for both observations. The core predictives contain more mass in the regions around the outturn than the AR(4) benchmark, with relatively minor departures from symmetry.

Returning to the issue of structural interpretation of our preferred core measure, in figure 5, we plot the median from our grand ensemble core, DE_sc, together with the 25 and 75 percentiles. The plot shows that the median of the DE_sc core ignores several extreme values in the actual measured inflation series. Typically, the probability of inflation being less than zero is well below 25 percent.

This figure also plots lines for the trimmed mean core and weighted median used by the Reserve Bank of Australia; see the appendix for details. The disaggregate ensemble core inflation measure suggests that both more traditional underlying measures periodically understate and overstate inflationary pressures in Australia. The year 2008 saw several outturns above the 75th percentile for both of these underlying measures. The DE core implies that inflationary pressures were more moderate. We should note also that the traditional measures of underlying inflation plotted here are less timely than the DE core—there is a lag in the data release.

One advantage of our probabilistic approach to measuring core inflation is that we can calculate the probability of specific events for measured inflation of interest to policymakers. As an example, we calculate the (one step ahead) probability that measured inflation exceeds the upper bound and midpoint of the inflation target. Strictly speaking, the target for monetary policy in Australia is to achieve a four-quarter inflation rate of

Figure 5: Inflation interval forecasts



(a) DE_s

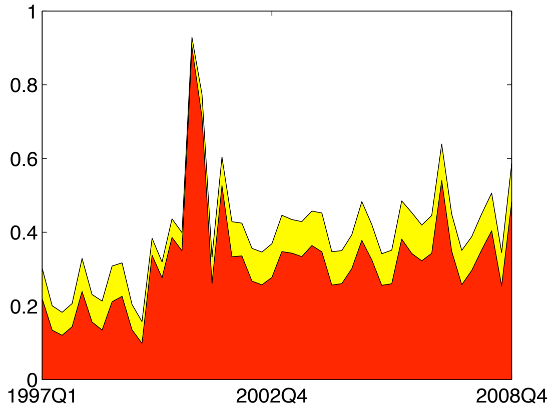
Note: The figure shows the posterior median, the 25th and 75th percentiles of the predictive density given by disaggregate ensemble DE_cs in panel (c) (black dotted lines), actual inflation (red dashed line), RBA "weighted median" and "trimmed mean" inflation (blue and yellow solid lines respectively).

2-3 per cent on average over the cycle. However, we work with analogous thresholds for the one step ahead horizon, interpreted at a quarterly frequency. That is the events of interest are: (1) measured inflation greater than 0.74 percent (upper bound), and; (2) measured inflation great than 0.62 percent (midpoint). The time series for the two event probabilities are plotted below in figure 6. As a visual aid, we label the first event "red", and the second event "yellow" and shade the plot appropriately. The figure suggests that the probability of exceeding the upper threshold generally remains less than 50 percent.

7 Conclusions

We conclude from our analysis that the ensemble approach provides a means of generating well-calibrated forecast densities for Australian measured inflation from disaggregate information. Our ensemble core uses information from disaggregation both by city and by sector. In future work, we hope to explore ensemble measures based on a larger number of disaggregates.

Figure 6: Measured inflation probabilities



Note: The area plots the probability that measured inflation is greater than 0.74 % (yellow color), and measured inflation is greater than 0.62 % (red color).

Appendix: Alternative Core Measures

Quote from the RBA "notes to tables".

The 'Weighted median' and 'Trimmed mean' are calculated using the component level data of the consumer price index. Both measures exclude interest charges prior to the September quarter 1998 and are adjusted for the tax changes of 1999-2000. The 'Trimmed mean' is calculated by ordering all the CPI components by their price change in the quarter and taking the expenditure-weighted average of the middle 70 per cent of these price changes. The 'Weighted median' is the price change in the middle of this ordered distribution, taking also expenditure weights into account. Annual rates of 'Weighted median' and 'Trimmed mean' inflation are calculated based on compounded quarterly rates. For calculating the 'Weighted median' and 'Trimmed mean', where CPI components are identified as having a seasonal pattern, quarterly price changes are estimated on a seasonally adjusted basis. Seasonal adjustment factors are calculated as concurrent factors, that is using the history of price changes up to and including the current CPI release. There is a series break at September 2002 due to the ABS publishing the 'Weighted median' and 'Trimmed mean' on behalf of the RBA from that point forward, using data to a higher level of precision than is publicly available. For further information on the various measures of underlying consumer price inflation, refer to 'Box D: Underlying Inflation', Statement on Monetary Policy, May 2002; 'Box D: Measures of Underlying Inflation', Statement on Monetary Policy, August 2005; and Roberts (2005), 'Underlying Inflation: Concepts,

Measurement and Performance', Reserve Bank of Australia Research Discussion Paper No 2005-05.

References

[1] Andersson, M. and Karlsson, S. (2007), "Bayesian Forecast Combination for VAR Models", Sveriges Riksbank Working Paper Series No. 216.

[2] Amisano, G. and R. Giacomini (2007), "Comparing Density Forecasts via Likelihood Ratio Tests", *Journal of Business and Economic Statistics*, 25, 2, 177-190.

[3] Amisano, G. and J. Geweke (2009), "Evaluating the Predictive Distributions of Bayesian Models of Asset Returns", *International Journal of Forecasting* (forthcoming). Available as Working Paper 969, European Central Bank.

[4] Bache, I.W., J. Mitchell, F. Ravazzolo and S.P. Vahey (2009), "Macro Modeling with Many Models", NSIER working paper.

[5] Bao, Y., T-H. Lee and B. Saltoglu (2007), "Comparing Density Forecast Models", *Journal of Forecasting*, 26, 203-225. First circulated as 'A test for density forecast comparison with applications to risk management', University of California, Riverside, 2004.

[6] Bao, L., T. Gneiting, E.P. Grimit, P. Guttop, and A.E. Raftery (2007), "Bias Correction and Bayesian Model Averaging for Ensemble Forecasts of Surface Wind Direction", University of Washington, Department of Statistics, Technical Report, No 557.

[7] Berkowitz, J. (2001), "Testing Density Forecasts, with Applications to Risk Management", *Journal of Business and Economic Statistics*, 19, 465-474.

[8] Brischetto, A. and A. Richards (2006), "The Performance of Trimmed Mean Measures of Underlying Inflation", RBA Research Discussion Papers 2006-10, Reserve Bank of Australia.

[9] Bryan, M.F. and S.G. Cecchetti (1994), "Measuring core inflation", in Mankiw (ed.) *Monetary Policy*, University of Chicago Press.

[10] Clark, T.E. and M. W. McCracken (2009), "Averaging Forecasts from VARs with Uncertain Instabilities", *Journal of Applied Econometrics* (forthcoming). Available as Federal Reserve Bank of Kansas City Working Paper 06-12.

[11] Clark, T.E. and K.D. West (2006), "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models", *Journal of Econometrics*, 138, 291-311.

[12] Diebold, F.X., Gunther, T.A. and A.S. Tay (1998), "Evaluating Density Forecasts; with applications to financial risk management", *International Economic Review,* 39, 863-83.

[13] Garratt, A., J. Mitchell and S.P. Vahey (2009), "Measuring Output Gap Uncertainty" unpublished manuscript, Birkbeck College, University of London.

[14] Gerard, H. and K. Nimark (2008), "Combining Multivariate Density Forecasts Using Predictive Criteria", RBA Research Discussion Papers 2008-02, Reserve Bank of Australia.

[15] Geweke, J. (2009), *Complete and Incomplete Econometric Models.* Princeton University Press (forthcoming).

[16] Giacomini, R. and H. White (2006), "Tests of Conditional Predictive Ability", *Econometrica*, 74, 1545-1578.

[17] Gneiting, T. and A. Raftery (2007), "Strictly Proper Scoring Rules, Prediction, and Estimation", *Journal of the American Statistical Association*, 102, 359-378.

[18] Jore, A. S., J. Mitchell and S.P. Vahey (2008), "Combining Forecast Densities from VARs with Uncertain Instabilities", *Journal of Applied Econometrics* (forthcoming). Available as NIESR Discussion Paper No. 303.

[19] Kascha, C. and F. Ravazzolo (2009), "Combining Inflation Density Forecasts", *Journal of Forecasting* (forthcoming). Available as Norges Bank Working Paper 2008/22.

[20] Kiley, M. (2008), "Estimating the Common Trend Rate of Inflation for Consumer Prices and Consumer Prices Excluding Food and Energy Prices", Federal Reserve Board, FEDS 2008-38, July.

[21] Koop, G. (2003), *Bayesian Econometrics.* John Wiley and Sons.

[22] MacKenzie, D. (2003), "Ensemble Kalman Filters Bring Weather Models Up to Date", *SIAM News*, 36, 8, October.

[23] Maheu, J.M. and S. Gordon (2008), "Learning, Forecasting and Structural Breaks", *Journal of Applied Econometrics*, 23, 553-583.

[24] Maheu, J.M. and T.H. McCurdy (2009), "How Useful are Historical Data for Forecasting the Long-Run Equity Return Distribution?", *Journal of Business and Economic Statistics*, 27, 95-112.

[25] Mitchell, J. and S.G. Hall (2005), "Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR "Fan" Charts of Inflation", *Oxford Bulletin of Economics and Statistics*, 67, 995-1033.

[26] Mitchell, J. and K. F. Wallis (2009), "Evaluating Density Forecasts: Forecast Combinations, Model Mixtures, Calibration and Sharpness", NIESR Discussion Paper No. 320 and Warwick University Discussion Paper.

[27] Patton, A. (2004), "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation", *Journal of Financial Econometrics*, 2, 130-168.

[28] Panagiotelis, A. and M. Smith (2008), "Bayesian Density Forecasting of Intraday Electricity Prices using Multivariate Skew t Distributions", *International Journal of Forecasting*, 24, 710-727.

[29] Quah, D.T. and S.P. Vahey (2005), "Measuring core inflation", *Economic Journal*, 105, 1130-44.

[30] Ravazzolo, F. and S.P. Vahey (2009), "Ditch the Ex! Measure Core Inflation with a Disaggregate Ensemble", Norges Bank unpublished manuscript.

[31] Roger, S. (1998), "Core Inflation: Concepts, Uses and Measurement", Reserve Bank of New Zealand Discussion Paper G98/9.

[32] Smith, J.K. (2004), "Weighted Median Inflation: is this core inflation?", *Journal of Money, Credit and Banking*, 36, 253-263.

[33] Smith, C., K. Gerdrup, A.S. Jore, and L.A: Thorsrud (2009), "More than One Weight to Skin a Cat: combining densities at Norges Bank", Norges Bank unpublished manuscript.

[34] Svensson, L. and N. Williams (2007), "Monetary Policy with Model Uncertainty: Distribution Forecast Targeting ", unpublished manuscript, Princeton University.

[35] Wallis, K.F (2005), "Combining Density and Interval Forecasts: a Modest Proposal", *Oxford Bulletin of Economics and Statistics*, 67, 983-994.

[36] Wynne, M.A. (1999), "Core Inflation: A Review of Some Conceptual Issues", ECB Working Paper Series, no 5.