# Probability Forecasting for Inflation Warnings from the Federal Reserve*

Anthony Garratt
(University of Warwick)

James Mitchell
(University of Warwick)

Shaun P. Vahey
(Australian National University)

June 27, 2013

**Abstract**

Forecasting inflation constitutes a primary task of monetary policymakers. The US Federal Reserve and other central banks occasionally publish verbal warnings about extreme inflation events. In this paper, we set out a formal framework in which monetary policymakers communicate inflation event warnings to the public. We show that these warnings require probabilistic forecasts in the presence of an asymmetry between the costs of anticipated and unanticipated inflation. In our application, we combine the predictive densities produced from Vector Autoregressive (VAR) models utilizing prediction pools and evaluate our ensemble probabilistic forecasts for quarterly US data from 1990:1 to 2012:1. We adopt a cost-loss ratio for forecast evaluation consistent with our policymaking communication framework. The incidence of low inflation events has increased during the Great Recession and, for these events, economic loss could be reduced by up to 30% relative to a benchmark model in which inflation follows an integrated moving average process. Calibrating our framework to the actual verbal statements from the Federal Reserve during the Great Recession implies a performance advantage of around 5% to 10%. Our findings indicate considerable scope for using formal forecasting methods to forewarn the public of extreme inflation events.

**Keywords**: Vector autoregression; Forecasting inflation; Probability forecasts; Cost-Loss ratio.

*Financial support from the ESRC (grant No. RES-062-23-1753), the ARC (grant No. LP0991098), Norges Bank, the Reserve Bank of New Zealand, and the Reserve Bank of Australia is gratefully acknowledged. First presented at the Computational and Financial Econometrics Conference in London, December 2011. We are grateful to seminar participants from the University of New South Wales and the University of Sydney for comments. Contact: Anthony Garratt, Warwick Business School, University of Warwick, Anthony.Garratt@wbs.ac.uk.

# 1 Introduction

Bernanke (2003), Greenspan (2004) and Kilian and Manganelli (2007), among others, highlight the importance of probabilistic assessments as the basis for the communication of monetary policy to the public. Central banks, including the US Federal Reserve, publish verbal warnings on occasions when extreme inflation outcomes are likely. Despite this, the inflation forecasting systems currently utilized by (almost all) central banks do not aim to produce ex ante probabilities for these events of interest.

Our analysis differs from preceding forecasting inflation studies in two respects. First, we adopt a formal framework in which monetary policymakers communicate inflation warnings to the public. We demonstrate that inflation warnings from central banks require probabilistic forecasts where there is an asymmetry in the economic costs associated with anticipated and unanticipated inflation events. Second, in our application using econometric models for forecasting, we utilize a cost-loss ratio based approach for forecast evaluation that is consistent with our policymaking communication framework. In applying the cost-loss methodology adopted in (for example) Granger and Pesaran (2000a,b) and Berrocal, Raftery, Gneiting, and Steel (2010) our study demonstrates the considerable scope for using formal forecasting methods to forewarn the public of inflation events.

In our application, we deploy a standard macroeconometric model space, based on a large number of bivariate Vector Autoregressive (VAR) models for inflation and the output gap, differentiated by the lag order of the VAR and the measure of the output gap. Orphanides and van Norden (2005) utilize a similar model space for forecasting inflation (evaluating models individually), as do Garratt, Mitchell, Vahey and Wakerly (2011) (examining forecast combinations). We combine the VAR forecast densities using two types of prediction pool: the Linear Opinion Pool (LOP) and the Logarithmic Opinion Pool (LogOP). With symmetric forecast densities from each VAR, the former allows scope for asymmetric forecast densities, whereas the latter does not. Using US inflation data, over the out of sample period from 1990:1 to 2012:1, we compare the forecasting performance with a benchmark model in which inflation follows an integrated moving average (IMA) process. Stock and Watson (2007) and Clark and Ravazzolo (2012) have deployed this specification as a benchmark in earlier studies forecasting US inflation. The IMA benchmark produces poorly-calibrated forecast densities. In contrast, both variants of the VAR ensemble reveal good calibration, as well as superior performance in terms of the logarithmic score. There is little to chose between the two ensembles in terms of forecast performance; the LOP forecast densities are close to symmetric.

Turning to the economic significance of the forecast performance, utilizing a cost-loss ratio, we gauge the gain from the adoption of our framework during the low inflation events associated with the recent Great Recession. We find that economic loss from inflation event warnings by the Federal Reserve would be reduced by up to 30%, relative to the IMA benchmark model. (Our approach also performs better than a 'climatological' forecast, based on the unconditional probability of inflation events.) Calibrating our framework to the actual verbal statements from the Federal Reserve during the Great Recession implies a performance advantage of around 5% to 10% over the IMA.

The remainder of this paper is structured as follows. In Section 2, we set out background information about both monetary policy forecasting and communication in practice and our probabilistic framework for the communication of inflation event warnings.

In Section 3, we describe the empirical methods, model space and data used in our application. In Section 4, we report our results for probability forecasting US inflation. In Section 5, we conduct a cost-loss ratio based analysis of forecast accuracy, given the asymmetric costs of anticipated and unanticipated inflation events. In Section 6, we re-examine the recent Great Recession from the perspective of our monetary policy communication framework. We draw some conclusions and ideas for extending the analysis in the final section.

# 2 Communicating inflation event warnings

## 2.1 Background

Central banks devote considerable resources to the assessment of inflationary pressures. Most routinely produce reports on the state of the macroeconomy which contain explicit forecasts for inflation (along with other key macroeconomic variables). However, there are differences in the way the forecast information is presented to the public. For example, the Federal Reserve publishes projections four times a year, in conjunction with the minutes of the Federal Open Market Committee (FOMC). Their published tables detail summary statistics for the 'central tendency' and the 'range' of projections.[1] In contrast, some central banks, including the Bank of England, Sveriges Riksbank and Norges Bank utilize probabilistic forecasts to communicate.

Regardless of the way forecasts are presented to the public, monetary policymakers typically issue explicit verbal warnings about forthcoming extreme inflation events. For example, the FOMC publishes (at least) eight statements per year about monetary conditions. During the Great Recession, which resulted in weak inflationary pressures, the FOMC signalled the danger of low inflation. Specifically, on March 18, 2009, the FOMC press release warned that inflation could drop 'below rates that best foster economic growth and price stability . . .'.[2] The financial press focus primarily on these statements—including the occasional verbal inflation event warnings—when appraising monetary policy strategy.

In contrast to the approach adopted in this paper, the Federal Reserve does not currently link the assessment of inflation event risks formally to a single model (or combination of models) for probability forecasting; see the discussions by Bernanke (2003) and Kilian and Manganelli (2007). Bernanke (2007) describes the forecasting process in the following terms.

> [T]he forecasts of inflation . . . provided to the Federal Open Market Committee are developed through an eclectic process that combines model-based projections, anecdotal and other 'extra-model' information, and professional judgment. In short . . . practical forecasting continues to involve art as well as science.
>
> Bernanke (2007)[3]

[1] Download from http://www.federalreserve.gov/monetarypolicy/fomccalendars.htm.
[2] Download from http://www.federalreserve.gov/newsevents/press/monetary/20090318a.htm.
[3] Download from http://www.federalreserve.gov/newsevents/speech/bernanke20070710a.htm

## 2.2 A probabilistic framework

In this paper, we link formally the decision to issue an inflation warning by the central bank to probabilistic forecasts for inflation. The decision to communicate the inflation risk to the public is contingent on the probability that inflation is less than (or greater than) a given threshold, $Z$.

Our approach is based upon the asymmetry between the costs of anticipated and unanticipated inflation. The notion that the economic cost to society of unanticipated inflation exceeds the cost of anticipated inflation has a long tradition in macroeconomics. Since relative prices determine the allocation of resources in the economy, if inflation is unexpectedly high or low, the price signals to private agents become distorted, leading to a misallocation of resources. For this reason, central banks traditionally devote considerable efforts to informing the private sector about the outlook for inflation, and also to justifying stable and predictable inflation as the basis for monetary policy. For example, Bank of England Governor, Eddie George, summarizes the costs of anticipated and unanticipated inflation, noting that the latter exceed considerably the former; see George (1992). The Governor buttresses the case for stable inflation with quotations from influential American economist and statistician Irving Fisher, delivered during a lecture at the London School of Economics in 1921. Fisher argued that unanticipated inflation 'makes a gamble of every contract', and that the impacts of unanticipated inflation are so widespread that there is 'scarcely a nook or cranny in human life which has not been touched'. Even though anticipated inflation may be preferable to unanticipated inflation, anticipated fluctuations in inflation also impose costs on society. Some contracts specify prices in nominal terms, and cannot be easily renegotiated intra-contract if anticipated inflation fluctuates (e.g. most labor contracts).

With the typical central bankers view of the relative costs of anticipated and unanticipated inflation in mind, we assume that if the central bank issues a preemptive inflation event warning, then the private sector adjusts contracts relatively easily to the anticipated inflation. In our framework, this results in a one-period economic cost to society of $C$, regardless of whether the inflation event occurs or not. On the other hand, if the central bank does not issue an inflation event warning, and the event occurs, the unanticipated inflation results in a one-period loss, $L$, where $0 < C < L$.[4]

Although as Driffill, Mizon and Ulph (1990) note, there is little direct empirical evidence on the relative cost of anticipated inflation, a high incidence of inflation event warnings implies that the cost-loss ratio, $R$, defined as $C/L$, lies towards the lower end of the feasible range for $R$, namely (0, 1), for a given inflation event. To see this, consider how the optimal rule for publishing the warning can be derived for a given inflation threshold, $Z$. Society incurs the per period cost $C$ regardless of whether the event occurs, so the expected cost of the warning is $C$. The expected cost of the alternative communication strategy, 'no warning', is $p_Z L$, where $p_Z$ is the probability that the inflation event occurs. If the policymaker minimizes the expected cost, an event warning is issued provided that $R = C/L < p_Z$. On the other hand, if $R = C/L \geq p_Z$, then the central bank issues no warning. For low values of $R$, where the costs of anticipated inflation are small relative to those of unanticipated inflation, smaller probabilities will trigger an inflation event warning.[5]

[4]The framework could be extended to make the costs to society contingent on realized inflation.

[5]Although the relative cost $R$ affects the decision to issue the inflation event warning, in principle

It should be emphasized that, in this framework, a rational policymaker requires the (ex ante) probability in order to decide whether to issue a warning about an inflation event. Preemptive inflation event warnings from a central bank require probabilistic forecasts because of the asymmetry in the economic costs associated with anticipated and unanticipated inflation events.

Similar frameworks have been deployed in previous research to study applied forecasting problems in meteorology, finance and economics. See, for examples, Murphy (1977), Katz and Murphy (1997), Granger and Pesaran (2000a,b), Clements (2004) and Berrocal, Raftery, Gneiting and Steel (2010).[6]

### 2.3 Discussion

The decision to communicate with the public is contingent on the probability that an inflation event occurs. We define an inflation event as when inflation is less than (or greater than) a given threshold. Unfortunately, economic theory is silent over which thresholds matter for inflation warnings in practice. Many central banks operate with a range of desired inflation outcomes specified in an explicit inflation interval target. The upper and lower bounds constitute natural candidates for inflation event warnings for explicit inflation targeting central banks. For others, such as the Federal Reserve during our sample period, the thresholds of interest are less clear. In the application that follows we take a pragmatic approach to the unknown threshold, $Z$, presenting results for a variety of feasible values. To simplify the analysis, we set the cost of inflation to zero for 'normal' inflationary conditions—when the central bank correctly believes extreme inflation events will not happen outside the target range.

A central bank also has discretion over the horizon of interest for inflation event warnings. As Bernanke (2007) discusses, short-term projections by central banks tend to be more accurate than longer-term forecasts. In part, this reflects the feedback between monetary policy and the forecasts in the longer term. If the Federal Reserve believes that there is a significant risk of inflation in the medium to long run, then monetary policy actions (over time) should aim to ensure that the path of inflation is consistent with its mandate. In the short-term, however, since policy changes affect the economy with a considerable time lag—something in the region of four (or more) quarters constitutes a conventional assumption by policymakers—the path of the economy is effectively pre-determined. In our subsequent application, we focus our analysis on short-term forecasting, considering a range of forecast horizons less than one year. In the text, we emphasize the forecast for the current period, sometimes referred to as the 'nowcast'

## 3 Application: model space, data and methods

### 3.1 Statistical model space

We begin with a bivariate VAR model space for inflation, $\pi_t$, and the output gap (the deviation of real output from potential), $y_t$, of the type explored by Orphanides and van

both $C$ and $L$ could be time varying; hence there could be variations in $R$. In the application that follows we calibrate the model to a low and stable inflation environment.

[6]Clements (2004) utilizes a cost-loss ratio to evaluate inflation forecasts. But, unlike this paper, Clements considers neither event warnings nor econometric models for probability forecasting.

Norden (2002, 2005). The theory that aggregate demand, captured by the output gap, predominately determines the movements in inflation (with unknown time lags), provides foundation for the empirical specification, which allows for simultaneity. The $j$-th VAR model takes the form:

$$\begin{bmatrix} \pi_t \\ y_t^j \end{bmatrix} = \begin{bmatrix} a_{\pi\pi}^j & a_{\pi y}^j \\ a_{y\pi}^j & a_{yy}^j \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ y_{t-1}^j \end{bmatrix} + \begin{bmatrix} \epsilon_{\pi t}^j \\ \epsilon_{yt}^j \end{bmatrix}, \qquad t = 1, \ldots, T, \tag{1}$$

where $[\epsilon_{\pi t}^j, \epsilon_{yt}^j]' \sim i.i.d.\ N(\mathbf{0}, \mathbf{\Sigma}^j)$. That is, we consider a baseline VAR specification in which the measure of interest for the output gap has been varied to give $J$ linear and Gaussian VAR models, indexed $j = 1, \ldots, J$. For expositional ease, we ignore the intercept and restrict the lag order of the $J$ VARs to one.

Following Garratt, Mitchell and Vahey (2011), our baseline VAR setup uses seven output gap measures, $J = 7$, derived from the set of univariate 'off-model' filters deployed by Orphanides and van Norden (2002, 2005). Federal Reserve researchers Edge and Rudd (2012) also deploy similar univariate detrending methods.

We define the output gap as the difference between observed output and unobserved potential (or trend) output. Let $q_t$ denote the (logarithm of) actual output in period $t$, and $\mu_t^j$ be its trend using definition $j$, where $j = 1, \ldots, J$. Then the output gap, $y_t^j$, is defined as the difference between actual output and its $j^{th}$ trend measure. That is, we assume the following trend-cycle decomposition:

$$q_t \ = \ \mu_t^j + y_t^j. \tag{2}$$

The seven methods of univariate trend extraction in our baseline VAR are: quadratic, Hodrick-Prescott (HP), forecast-augmented HP, Christiano and Fitzgerald, Baxter-King, Beveridge-Nelson and Unobserved Components. We summarize these seven well-known univariate filters in Appendix 1.

In our application, we vary a single auxiliary assumption to generate the model space. Namely, we vary the lag length in the VAR (which for ease of exposition we fixed at one in equation (1)). If we have $J$ output gap measures, and for any given $y_t^j$ we have $K$ different variants defined over different values of the maximum lag length, then in total we have $N = J \times K$ models, and $N$ associated forecasts of inflation (and the output gap) from the ensemble system. In our application, we restrict $K$ to a maximum of four and therefore we consider $N = 28$ models.

In practice, of course, the Federal Reserve does not utilize exactly this model space. The motivation for deploying these models stems from their common usage by central banks around the world. Nevertheless, as Orphanides and van Norden (2005) emphasize, individually VAR models of this type often perform little better than simple univariate autoregressive benchmarks in (pseudo) real-time out of sample evaluations of point forecasting accuracy. In contrast, Garratt, Mitchell, Vahey and Wakerly (2011) note that using an ensemble of VAR models delivers well-calibrated forecast densities for inflation in the US and Australia. Nevertheless, with empirical analysis limited to just two variables, the scope for predicting the inflationary impacts of macroeconomic developments unrelated to aggregate demand is limited. We shall return to this point subsequently.[7]

[7]We experimented by augmenting the baseline VAR to include additional exogenous variables to reflect energy costs. The forecasting performance was very similar.

## 3.2 The data

Orphanides and van Norden (2002, 2005) stress that our output gap measures are subject to considerable data revisions. Failing to account for this by using heavily-revised data can mask real-time predictive content. Since we are interested in real-time prediction, parameter estimation uses a rolling window of data, with each window using a different vintage of data. A vintage of data is the vector of time series observations available from a data agency at the most recent observation in the information set of forecasters.

The quarterly real-time real Gross Domestic Product (GDP) dataset has 188 vintages, with the first vintage (dated 1965:4) containing time series observations from 1954:3 to 1965:3. The last vintage used for parameter estimation (dated 2012:3) starts in 1954:3 and ends in 2012:2. US GDP deflator data are released with a one quarter lag. The raw data for GDP (in practice, Gross National Product, GNP, for some vintages) are taken from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists. The data comprise successive vintages from the National Income and Product Accounts, with each vintage reflecting the information available around the middle of the respective quarter. Croushore and Stark (2001) provide a description of the real-time GDP database. The GDP deflator price series used to construct inflation is constructed analogously.

Figures 1 displays the inflation data in our evaluation period, from 1990:1 to 2012:1.[8] Figure 2 plots the first-difference of real GDP for comparison. We define inflation (output growth) as the first difference in the logarithm of the GDP deflator (GDP), multiplied by 400.[9]

The period of particular interest in this study, the Great Recession, comprises the period from 2008:1 through to the end of sample. Arguably, the Great Recession commenced with the fall of Lehman Brothers in 2008:3, or perhaps even earlier, when liquidity dropped in financial markets. However, 2009:1 marks the point where the Federal Reserve took the extraordinary step of cutting nominal interest rates to (effectively) zero. The era immediately prior to the Great Recession, from 1984 to 2007, is often referred to by economists as the 'Great Moderation'. During this period, inflation and output growth data exhibit lower volatility than the pre-1984 data, and the unconditional mean inflation rate is notably lower. The degree to which this period of relative stability can be attributed to monetary policy has been debated within the macroeconomic literature. Since 2008, through the Great Recession, as Figures 1 and 2 show, the data have exhibited greater volatility, with sharp drops in inflation and, at times, negative output growth. In terms of inflation, the striking feature of the Great Recession is the increased threat of low inflation from 2009. For example, using an inflation event threshold of 1%, there are four observations in which inflation falls below this threshold between 2008:1 and 2012:1. In the period between 1990:1 and 2007:4, the number of incidents is just two.[10]

In is important to note that in the run up to the Great Recession, in the period between 2003 and 2006, there are several realizations of high inflation. The upward spikes apparent in various inflationary measures for the US during this period are often regarded as (the response to) relative price movements, and, in particular, energy and food costs.

[8] Unlike the empirical analysis which follows, the chart plots a single vintage of data dated 2012:3.

[9] We also experimented with a Personal Consumption Expenditure based measure of inflation which gave very similar results to those reported below.

[10] The second measurements of inflation, which we utilize for forecast evaluation, indicate seven observations below this threshold since 2008:1.

See, for example, the discussion in Yellen (2006) and Ravazzolo and Vahey (2013).[11]

## 3.3 Generating forecasts

Armed with our many VAR specifications, it is straightforward to produce forecast densities for both inflation and the ($j$-th) output gap through our evaluation period: $\tau = \underline{\tau}, \ldots, \overline{\tau}$ where $\underline{\tau}$ =1990:1 and $\overline{\tau}$ =2012:1 (89 quarterly observations). Our focus is forecasting inflation, of direct relevance to monetary policymakers, given the output gap data as well as lagged inflation data. In addition the output gap is latent, with $J$ proxies, and published 'outturn' data do not exist.

To implement our prediction pool approach to forecast combination through the evaluation period requires an additional assumption about which measurement of inflation is to be forecast. Following Clark (2011) and others, we use the second estimate as the 'final' data to be forecast. For consistency, we report results for the same definition of final data for all forecast evaluations; see the discussion in Corradi, Fernandez and Swanson (2009).

For each VAR in our model space, we estimate the parameters using Ordinary Least Squares over a rolling sample window. A number of researchers have argued that a rolling window for model fitting provides a pragmatic approach to handling structural change; see, for example, Clark (2011). However, the forecasting performance of macroeconometric models can be sensitive to the choice of the length of the rolling window. Following Clark (2011), we present results below for a window length of 40 for parameter estimation for all VARs (and our benchmark specification).[12]

Given each VAR specification is estimated by Ordinary Least Squares, the predictive densities for inflation (and the output gap), for a given VAR specification, (1), are (multivariate) Student-$t$. Appendix 2 provides details of how the predictive densities are constructed from each VAR.

We combine the out of sample forecasts from the $N$ VAR models using two types of prediction pool: the Linear Opinion Pool (LOP) and the Logarithmic Opinion Pool (LogOP). Given $i = 1, \ldots, N$ VAR specifications, the ensemble density for inflation defined by the LOP is:

$$p^{LOP}(\pi_\tau) = \sum_{i=1}^{N} w_{i,\tau}\ g(\pi_\tau \mid I_{i,\tau}), \qquad \tau = \underline{\tau}, \ldots, \overline{\tau}, \tag{3}$$

where $g(\pi_\tau \mid I_{i,\tau})$ are the one step ahead forecast densities from model $i$, $i = 1, \ldots, N$ for inflation $\pi_\tau$, conditional on the information set $I_{i,\tau}$. The publication delay in the production of real-time macroeconomic data ensures that this information set contains lagged variables, here assumed to be dated $\tau - 1$ and earlier. The non-negative weights, $w_{i,\tau}$, in this finite mixture sum to unity.[13] Furthermore, the weights may change with each recursion in the evaluation period $\tau = \underline{\tau}, \ldots, \overline{\tau}$.

[11] Yellen (2006) is downloadable from http://www.frbsf.org/news/speeches/2006/0315.html. Clark and Terry (2010) discuss the absence of pass through from energy prices to other prices during the period.

[12] We investigated alternative benchmark window lengths of 60 and 80 periods but found the qualitative nature of our findings to be robust. The magnitude of the performance differential over the benchmark varied somewhat. Shorter (longer) rolling windows for in-sample parameter estimation delivered a larger (smaller) forecast performance differential.

[13] The restriction that each weight is positive could be relaxed; for discussion see Genest and Zidek (1986).

In contrast, the ensemble density for inflation defined by the LogOP, can be expressed in its geometric form as:

$$p^{LogOP}(\pi_\tau) = \frac{\prod_{i=1}^{N} g(\pi_\tau \mid I_{i,\tau})^{w_{i,\tau}}}{\int \prod_{i=1}^{N} g(\pi_\tau \mid I_{i,\tau})^{w_{i,\tau}} d\pi_\tau}, \quad \tau = \underline{\tau}, \ldots, \overline{\tau}, \tag{4}$$

where the denominator is a constant that ensures that the ensemble density is a proper density. The LogOP is linear in its logarithmic form, where a feature of the logarithmic combination is that it preserves the $t$-distribution form of the combination given $t$ component densities. Regardless of the type of prediction pool, we assume that the various output gap measures differ from the 'true' output gap by more than conventional (Gaussian) white noise measurement error. That is, the true output gap is never observed, even ex post, although, in principle, a combined density for the unobserved output gap might be constructed following Garratt, Mitchell and Vahey (2011).

We utilize weights based on the fit of the individual VAR forecast densities for inflation. Following Amisano and Giacomini (2007), Hall and Mitchell (2007), Jore, Mitchell and Vahey (2010) and Garratt, Mitchell, Vahey and Wakerly (2011) we use the logarithmic score to measure density fit for each model through the evaluation period. The intuitive appeal of the logarithmic scoring rule stems from the high score assigned to a density forecast with high probability at the realized value. The logarithmic score of the $i^{\text{th}}$ density forecast, $\ln g(\pi'_\tau \mid I_{i,\tau})$, is the logarithm of the probability density function $g(. \mid I_{i,\tau})$, evaluated at the realized inflation outturn, $\pi'_\tau$. Specifically, the recursive weights for the one step ahead densities from both type of prediction pool take the form:

$$w_{i,\tau} = \frac{\exp\left[\sum_{\tau-\kappa}^{\tau-1} \ln g(\pi'_\tau \mid I_{i,\tau})\right]}{\sum_{i=1}^{N} \exp\left[\sum_{\tau-\kappa}^{\tau-1} \ln g(\pi'_\tau \mid I_{i,\tau})\right]}, \qquad \tau = \underline{\tau}, \ldots, \overline{\tau} \tag{5}$$

where $\tau - \kappa$ to $\underline{\tau} - 1$ comprises the rolling window used to construct the weights. It is important to note that the weight on the various specifications varies through time. Hence, the ensemble exhibits greater flexibility than any single linear VAR specification (in which the individual model parameters are recursively updated).

# 4 Assessing out of sample forecast performance

Although our primary interest is the Great Recession, from 2008:1, there are very few observations for forecast evaluation. Hence, we begin with a statistical assessment of inflation forecasting performance over the evaluation period 1990:1 to 2012:1 (89 observations).[14] In assessing forecast performance, using both a LOP and a LogOP of VAR forecasts, we present and discuss results for the one step ahead forecasts only. Results for longer horizons, up to one year ahead, presented in Appendix 3, show that (as expected) performance drops with the forecast horizon.

Since forecast performance can be sensitive to the length of the rolling window, $\kappa$, used to construct the weights, we present statistical assessments of forecast performance for a variety of values of $\kappa = 5, 10, 15, 20, 25, 30, 35$.

[14]Recall that we treat the output gap as a latent variable for which there are no observations for forecast evaluation.

We present out of sample forecast metrics in the four columns of Table 1a for the LOP, and Table 1b for LogOP. The first column in each table reports how the root mean square forecast errors (RMSFE) vary with $\kappa$. The second column in each displays the Brier Score (Brier, 1950) for inflation below its unconditional mean:

$$BS \;=\; (1/M)\sum_{\tau=\underline{\tau}}^{\overline{\tau}}(o_\tau - f_\tau)^2, \tag{6}$$

where $M$ is the number of periods in the out of sample evaluation, $\overline{\tau}$ minus $\underline{\tau}$, $o_\tau$ is the outcome (1 if inflation below mean, 0 otherwise), and $f_\tau$ is the real-time probability, generated by each prediction pool of VAR forecasts, that inflation is below its mean. The unconditional mean inflation rate over our evaluation period 1990:1-2012:1 is 2.2%. Low Brier scores are preferred.

The third columns of Table 1a and 1b report the respective calibration statistics based on the Likelihood Ratio (LR) test proposed by Berkowitz (2001). Following Rosenblatt (1952), Dawid (1984) and Diebold, Gunther and Tay (1998), we use the probability integral transforms (*pits*) of the realization, $\pi'_\tau$ , of the variable with respect to the forecast densities. We gauge calibration by examining whether the *pits* $z_\tau$, where:

$$z_\tau = \int_{-\infty}^{\pi'_\tau} p(u)du, \tag{7}$$

are uniform and, for our one step ahead forecasts, independently and identically distributed; see Diebold et al. (1998). We use a three degrees of freedom variant with a test for independence, where under the alternative $z_\tau$ follows an autoregressive first-order process. (Appendix 3 contains results for a battery of additional one-shot tests of calibration for the VAR systems, and the benchmark model used in our subsequent analysis.) The probability values of the LR test reported in the table do not allow for parameter estimation uncertainty, given the large number of models under consideration. The final columns of Table 1a and 1b display the average logarithmic score over the evaluation period.

The first two columns of Table 1a suggest a relatively flat response over $\kappa$ for both RMSFE and the Briers score for the VAR LOP. For example, the RMSFE varies by less than 3% over the range considered for the rolling window length. A rolling window length of 5 gives the 'best' Brier Score, 0.232, and the score deteriorates only slightly as $\kappa$ increases to 40. The third column indicates that the forecast densities are well calibrated, regardless of the $\kappa$ value. That is, the null hypothesis of no calibration failure can be rejected only at the 59% level of significance or greater. The fourth column shows that the average logarithmic score varies by no more than 3% of the 'best', with the optimized value attained at $\kappa = 10$. Given the fairly robust performance across rolling window lengths, we present results below for a representative case, where $\kappa = 10$; that is, a rolling window for combination of two and a half years.

Turning to Table 1b, the VAR LogOP shares many of the characteristics exhibited by the VAR LOP (described in Table 1a). There is no evidence of calibration failure by the LR test, regardless of $\kappa$. The RMSFE, the Brier score and the average logarithmic score vary little with $\kappa$. For simplicity, we match the LogOP rolling window length to that of the VAR LOP—we again set $\kappa = 10$.

**Table 1a: VAR LOP forecast performance, by window $\kappa$**
**1990:1-2012:1**

| $\kappa$ | RMSFE | Brier Score | LR | Log Score |
|---|---|---|---|---|
| 5 | 0.897 | 0.232 | 0.69 | -1.363 |
| 10 | 0.897 | 0.234 | 0.72 | -1.350 |
| 15 | 0.910 | 0.239 | 0.64 | -1.369 |
| 20 | 0.914 | 0.242 | 0.59 | -1.386 |
| 25 | 0.914 | 0.244 | 0.67 | -1.391 |
| 30 | 0.921 | 0.246 | 0.59 | -1.387 |
| 35 | 0.910 | 0.244 | 0.69 | -1.375 |
| 40 | 0.929 | 0.246 | 0.63 | -1.378 |

**Table 1b: VAR LogOP forecast performance, by window $\kappa$**
**1990:1-2012:1**

| $\kappa$ | RMSFE | Brier Score | LR | Log Score (LS) |
|---|---|---|---|---|
| 5 | 0.905 | 0.233 | 0.83 | -1.350 |
| 10 | 0.901 | 0.235 | 0.74 | -1.350 |
| 15 | 0.908 | 0.241 | 0.56 | -1.358 |
| 20 | 0.912 | 0.243 | 0.45 | -1.373 |
| 25 | 0.915 | 0.244 | 0.48 | -1.377 |
| 30 | 0.917 | 0.246 | 0.40 | -1.374 |
| 35 | 0.911 | 0.244 | 0.47 | -1.366 |
| 40 | 0.921 | 0.246 | 0.43 | -1.368 |

Given that the choice of window for combination, $\kappa$, has little impact on forecast performance for the VAR LOP and VAR LogOP, we turn to the evaluation of the ensemble inflation nowcasts. We adopt a commonly deployed benchmark in the inflation forecasting literature, namely an integrated time series model for inflation, with a first order moving average error process. The IMA model for inflation takes the form: $\Delta\pi_t = \alpha + \varepsilon_t + \theta\varepsilon_{t-1}$. An IMA model has been found by Stock and Watson (2007) to provide competitive point forecasts for US inflation.[15]

Table 2 reports out of sample forecast performance statistics, based on sequences of one step ahead inflation forecasts, evaluated at the end of period, for both VAR ensembles and for the benchmark. The first column of Table 2 reports RMSFE statistics. The second column provides the LR test for forecast calibration. The RMSFE reported in column 1 of Table 2 indicate a gain for both the VAR LOP and the VAR LogOP (first two rows) of around 10% over the IMA benchmark (third row), with the VAR LOP the most preferred. The forecast densities for both ensembles are well calibrated, with large $p$-values for the LR calibration test. But the IMA benchmark is poorly calibrated by the same LR test, at the 1% significance level.

[15]We also explored a variant of IMA with stochastic volatility, following Clark and Ravazzolo (2012). Despite reasonable performance in terms of point forecasting, the densities displayed poor calibration, with an inferior average logarithmic score, and hence are not reported.

The upper panel of Table 3 reports the average logarithmic scores of the inflation forecast densities over the evaluation period in column 1, together with the equivalent statistics for a selection of 'low' inflation events in the next three columns. The events are outcomes less than 1%, less than 1.25% and less than 1.5%. Diks, Panchenko and van Dijk (2011) discuss the importance of gauging forecast performance for events of economic importance. The average inflation rate for the evaluation period is 2.2%; Figure 1 shows that inflation falls the below 1% threshold 6 times out of 89 observations; and, for the 1.25% and 1.5% thresholds, 11 and 18 times out of 89 observations, respectively. The lower panel of Table 3 reports the average logarithmic scores for a selection of 'high' inflation events. Namely, inflation above 3%, 3.2% and 3.5%. In Figure 1, these events occur 16, 12 and 8 times out of 89 observations, respectively.[16]

The average logarithmic scores for inflation reported in column 1 of the upper panel of Table 3 suggest that both types of prediction pool (for the VAR models) are preferred to (that is, give higher values than) the IMA benchmark. To gauge the statistical importance of this ordering, we consider a Kullback-Leibler information criterion (KLIC)-based test, utilizing the expected difference in the logarithmic scores of the candidate densities; see Bao, Lee, and Saltoglu (2007), Mitchell and Hall (2005) and Amisano and Giacomini (2007). Suppose there are two forecast densities, $p(\pi_\tau \mid I_{1,\tau})$ and $p(\pi_\tau \mid I_{2,\tau})$, so that the KLIC differential between them is the expected difference in their log scores: $d_\tau = \ln p(\pi_\tau \mid I_{1,\tau}) - \ln p(\pi_\tau \mid I_{2,\tau})$. The null hypothesis of equal forecast performance is $\mathcal{H}_0 : E(d_\tau) = 0$. A test can then be constructed since the mean of $d_\tau$ over the evaluation period, $\overline{d}_\tau$, under appropriate assumptions, has the limiting distribution: $\sqrt{T}\overline{d}_\tau \rightarrow N(0, \Omega)$, where $\Omega$ is a consistent estimator of the asymptotic variance of $d_\tau$. Amisano and Giacomini (2007) and Giacomini and White (2006) discuss the limiting distribution of related test statistics. Mitchell and Wallis (2011) discuss the value of information-based methods for evaluating well-calibrated forecast densities. The KLIC-based test for the VAR LOP (VAR LogOP) ensemble relative to the IMA gives a $p$-value of 0.13 (0.10). Regardless of the method for taking forecast combinations, the resulting VAR ensemble outperforms the benchmark at the 13% significance level.

Turning to the remaining columns of the upper panel of Table 3, we see a consistent pattern in the orderings, with the VAR LogOP performing marginally better than the VAR LOP, and both approaches preferred to the IMA benchmark. That is, the ordering of the logarithmic scores, starting with the highest, is VAR LogOP, VAR LOP, IMA. The lower panel of Table 3, which relates to 'high' inflation events, gives a similar pattern in two cases. However, for the highest inflation threshold, 3.5%, last column, the IMA performs best. Hence, the evidence indicates that the forecast performance of the VAR models is weaker for the highest inflation threshold considered.

To summarize the results so far, the VAR ensembles typically outperform the IMA benchmark both in terms of RMSFE and logarithmic scores, except for extremely high inflation outcomes. There is little to choose between the two methods for taking forecast density combinations. And, the benchmark IMA specification exhibits poor calibration.

[16] In the (revised) second measurements used for forecast evaluation, however, there is greater incidence of low inflation events. For example, there are 13 inflation events below the 1% threshold.

**Table 2: RMSFE and *pits*-based forecast evaluation, 1990:1-2012:1**

| | RMSFE | LR |
|---|---|---|
| LOP | 0.897 | 0.72 |
| LogOP | 0.901 | 0.74 |
| IMA | 1.014 | 0.01 |

Notes: RMSFE is Root Mean Squared Forecast Error; LR is the $p$-value for the Likelihood Ratio test of zero mean, unit variance and zero first order autocorrelation of the inverse normal cumulative distribution function transformed *pits*, with a maintained assumption of normality for the transformed *pits*.

**Table 3: Average logarithmic scores, various inflation events, 1990:1-2012:1**

| | $\pi$ | $\pi < 1.0\%$ | $\pi < 1.25\%$ | $\pi < 1.5\%$ |
|---|---|---|---|---|
| LOP | -1.360 | -0.384 | -0.506 | -0.667 |
| LogOP | -1.350 | -0.378 | -0.500 | -0.665 |
| IMA | -1.495 | -0.454 | -0.637 | -0.787 |

| | $\pi > 3.0\%$ | $\pi > 3.2\%$ | $\pi > 3.5\%$ |
|---|---|---|---|
| LOP | -0.558 | -0.457 | -0.273 |
| LogOP | -0.554 | -0.452 | -0.262 |
| IMA | -0.569 | -0.469 | -0.246 |

To gain further insight into the forecast performance across inflation intervals, we compute the Briers score for each VAR ensemble and the IMA benchmark. We also utilize the well-known decomposition of the Brier score into uncertainty, reliability and resolution; see Murphy (1973). The Brier score can be written as:

$$BS \;=\; \overline{o}(1-\overline{o}) + \frac{1}{M}\sum_{k=1}^{K} n_k(p_k - \overline{o}_k)^2 - \frac{1}{M}\sum_{k=1}^{K} n_k(\overline{o}_k - \overline{o})^2, \tag{8}$$

where the first term measures uncertainty, and $\overline{o}$ is the observed frequency of the event in question over $M$ evaluation periods. This component is the same for all specifications, being a function only of the observations and not the forecasts. The second and third terms measure reliability and resolution, respectively, where $K$ is the number of bins over which we assign the forecast probabilities $p_k$, $n_k$ is the number of times $p_k$ is forecast and $\overline{o}_k$ is the observed frequency of the occurrence of the event in bin $k$. As described in Galbraith and van Norden (2012), reliability measures the difference between the actual forecast probabilities of an event and the observed frequency of the event (the observed frequency in each of the $K$ bins). For example, suppose 20% is the probability of a specific inflation event (say inflation less say 1%), and the model (system) predicts that inflation falls below this same threshold for 20% of the observations, then we have a perfect match with the data, and reliability (calibration) is perfect, with this term being zero. In contrast, resolution measures the ability to distinguish between relatively high-probability and low-probability events. If resolution is high, then the conditional expectation of the outcomes

differs substantially from its unconditional mean. The resolution enters negatively into the decomposition as high resolution lowers the Brier score. Low Brier scores are preferred.[17]

Table 4 reports the Brier score and the decomposition for the two ensembles and the benchmark, utilizing the same six inflation event thresholds as in Table 3. The VAR LOP and VAR LogOP ensembles have approximately the same Briers score for all events. Moreover, both outperform the benchmark IMA for low inflation events. That is, the Brier score is highest for the IMA in the upper panel of Table 4. The major contribution to the superior performance comes from the third term, measuring resolution. For high inflation events, the three events in the lower panel of Table 4, the two prediction pools match the IMA benchmark. And the resolution is similar across all three specifications.

The results so far indicate that both the VAR LOP and VAR LogOP ensembles are well calibrated, whereas the IMA benchmark is not. Furthermore, both ensembles typically outperform the IMA in terms of RMSFE, the logarithmic score and the Brier score. For low inflation events—which predominate in the Great Recession—the prediction pools perform well. However, for high inflation events, the prediction pools give lower resolution, and the benchmark is tougher to beat in these circumstances.

**Table 4: Briers score and decomposition, 1990:1-2012:1**

| | $\pi < 1.0\%$ | $\pi < 1.25\%$ | $\pi < 1.5\%$ |
|---|---|---|---|
| LOP | 0.104 [*0.125; 0.020; 0.041*] | 0.140 [*0.168; 0.003; 0.031*] | 0.186 [*0.207; 0.015; 0.036*] |
| LogOP | 0.104 [*0.125; 0.011; 0.032*] | 0.141 [*0.168; 0.005; 0.032*] | 0.189 [*0.207; 0.018; 0.036*] |
| IMA | 0.124 [*0.125; 0.016; 0.017*] | 0.169 [*0.168; 0.029; 0.028*] | 0.215 [*0.207; 0.029; 0.021*] |
| | $\pi > 3.0\%$ | $\pi > 3.2\%$ | $\pi > 3.5\%$ |
| LOP | 0.153 [*0.155; 0.024; 0.026*] | 0.122 [*0.133; 0.022; 0.033*] | 0.055 [*0.072; 0.021; 0.038*] |
| LogOP | 0.152 [*0.155; 0.022; 0.025*] | 0.120 [*0.133; 0.017; 0.032*] | 0.052 [*0.072; 0.018; 0.038*] |
| IMA | 0.153 [*0.155; 0.027; 0.029*] | 0.121 [*0.133; 0.027; 0.039*] | 0.055 [*0.072; 0.019; 0.036*] |

Notes: Briers score, followed in parenthesis by Uncertainty, Reliability and Resolution, respectively

# 5 Cost-Loss ratio based evaluation

Recall that we aim to use our forecasting framework for the communication of inflation warnings. As discussed earlier, in order to decide whether to issue an inflation event warning, the policymaker requires the probability of inflation falling below (above) the appropriate threshold, in the presence of asymmetric economic costs. In general, the cost to society of anticipated inflation is taken to be much less than that from unanticipated inflation.

[17]The verification sample is partitioned into deciles of forecast probability—all 1 to 10% forecasts are put together with their corresponding observations, 11 to 20% and so on. We take the average of the probabilities in each bin to get a single number for the term $p_k$, $\overline{o}_k$ is the observed frequency of the occurrence of the event in bin $k$, where $n_k$ is the number of forecasts which fall into that bin. Reliability is the consistency between the a priori predicted probabilities (the observed frequency in that bin) and a posterior observed frequency. Therefore, lower scores indicate high consistency between these two terms. Resolution is the difference between the observed frequency in bin $k$ and the overall sample average frequency, $\overline{o}$.

We assume that there are four possible combinations of action and occurrence, where Total Economic Loss, $TEL$, depends on both. If action is taken, the Federal Reserve issues a preemptive event warning. A one-period (time-invariant) cost of $C$ is then incurred by society, irrespective of whether the event occurs. On the other hand, in the absence of a warning, an economic loss, $L$, is incurred by society only if the inflation event occurs. The relative cost ratio, $R = C/L$, will lie in the region $0 < R < 1$, since the costs are asymmetric. In evaluating the forecasts, we work with a number of thresholds defining inflation events, and consider a range of relative cost ratios.

We summarise the contingencies in Table 5. We denote the number of observations in which action was taken (the warning issued), but the inflation event did not occur, as $n_{01}$. We denote the number of observations in which action was taken and the event did occur $n_{00}$. In both of these cases, society incurs the cost $C > 0$. We denote the number of times that no action was taken and the event did not happen $n_{11}$. In this case, society incurs no cost. We denote the number of observations in which no action was taken but the event occurs $n_{10}$, giving a loss to society of $L$.

**Table 5: Contingencies for an inflation event**

| | **Event Occurs** | |
|---|---|---|
| **Action Taken** | Yes | No |
| Yes | $n_{00}$ $[C]$ | $n_{01}$ $[C]$ |
| No | $n_{10}$ $[L]$ | $n_{11}$ |

The Total Economic Loss, $TEL$, can be expressed as:

$$TEL = Ln_{10} + C(n_{01} + n_{00}). \tag{9}$$

Since we found little difference overall between the performance resulting from the two types of prediction pools (taking into account RMSFE, calibration, logarithmic score, and the Brier score), we report results only for the VAR LOP case.

Figures 3, 4 and 5 plot $TEL$ against $R$ for the VAR LOP and our IMA benchmark and, in addition, using the unconditional probability of the event (calculated as the number of times the event occurred prior to our evaluation period, divided by the number of observations). This is sometimes referred to as a 'climatological' forecast in the meteorological literature. Figures 3, 4 and 5 refer to the same low inflation events as in our tables; that is, inflation less than 1.0%, less than 1.25% and less than 1.5%, respectively. (The unconditional mean of the inflation rate outcomes over our evaluation period is 2.2%, with a standard deviation of 0.962.) Each plot shows a range of $R$ values, less than 0.4. For higher values of $R$, with small asymmetries in costs, the various specifications, and the climatological benchmark perform similarly.

With inflation events less than 1.0%, Figure 3, there is a considerable performance advantage to the VAR LOP over the IMA benchmark (normalized to 100 in the Figure) across a range of values of $R$, from nearly zero to 0.4. The size of the performance differential varies with $R$, peaking at approximately 25 percentage points with $R$ at 0.1, and falling non-monotonically as $R$ increases. For $R$ around 0.3, the IMA just beats the VAR LOP. For $R$ values in excess of that level, the VAR LOP again dominates the IMA, with the performance differential steadily increasing with $R$ to around 20 percentage points as $R$ reaches 0.4. For the lower values of $R$ shown, the climatological forecast

dominates the IMA, but the IMA is preferred for $R$ values between 0.05 and 0.3. The improvement in forecast performance of the VAR LOP over the climatological forecast peaks at around 50 percentage points, for $R$ equals 0.15, but falls to nearly zero as $R$ increases to 0.3. For values of $R$ greater than 0.3, the performance advantage in favor of the VAR LOP widens to around 15 percentage points.

Turning to Figure 4, for inflation events below 1.25%, we see the same general pattern as in Figure 3. The performance advantage of the VAR LOP over the IMA benchmark peaks at around 30 percentage points for $R = 0.05$, and declines (with reversals) as $R$ increases. However, unlike for the lower threshold event, the VAR LOP outperforms the benchmark over the entire range of $R$ displayed. As in Figure 3, the IMA is inferior to the climatological forecast for low values of $R$, in this case below 0.15. The performance advantage of the VAR LOP over the climatological forecast peaks at around 45 percentage points with $R$ at 0.2. The performance differential in favour of VAR LOP declines as $R$ rises beyond that level but never falls below 10 percentage points for the higher values of $R$ displayed.

Looking at the inflation events below 1.5%, Figure 5, the performance differential favors the VAR LOP over the IMA for almost the entire range of $R$ displayed. For $R$ greater than 0.2, the VAR LOP is marginally preferred, but never with more than a 10 percentage point differential. The best relative performance of the VAR LOP is achieved at $R = 0.1$, with a saving of about 20 percentage points. For this 1.5% threshold, the VAR LOP struggles to beat the climatological forecast for $R$ values below 0.2. Elsewhere the VAR LOP delivers a performance advantage over the climatological specification, which peaks at around 30 percentage points when $R$ equals 0.25. In general, for this inflation event, the climatological forecast dominates the IMA for $R$ below 0.2.

Although low inflation events are more common for the Great Recession, for completeness, we repeated our analysis for the following high inflation events: inflation above 3.0%, inflation above 3.2%, and inflation above 3.5%. For 3.0% and 3.2% thresholds, the VAR LOP typically gave similar performance to the climatological forecast, for low values of $R$. This result is consistent with the finding that the VAR LOP forecasts lack resolution for these inflation events. For higher values of $R$, with these two events, and for all values of $R$ (below 0.4), the VAR LOP ensemble outperformed the climatological forecast. But even in these cases, the IMA benchmark dominated both.

Overall, we draw the following conclusions from our cost-loss based analysis for the inflation events considered. For low inflation warnings, the VAR LOP ensemble generally gives a superior forecast performance to both the IMA benchmark and the climatological forecast. The size of the improvement over the IMA varies with $R$, but is greatest for values of $R$ less than 0.10—where unanticipated inflation is relatively costly to society. For high inflation events, the climatological forecast is harder to beat and the IMA model performs relatively better than for low inflation events.[18]

To check the robustness of the forecast performance of the VAR LOP ensemble, we repeated our analysis with the 'core PCE' inflation measure, which excludes food and energy prices. The results confirmed the superiority of the VAR LOP over the IMA for low inflation events. For both low and high inflation events, the VAR LOP gave a similar performance to the climatological benchmark. We emphasize, however, that the number

[18]The bivariate VAR models have difficulty predicting high inflation events in the 2000s, thought to stem from fuel, energy and food price shocks. Ravazzolo and Vahey (2013) discuss the sources of inflationary shocks in recent US data.

of extreme inflation events is much smaller if food and energy costs are stripped out. The lack of resolution of the VAR LOP ensemble with the core PCE inflation measure owes something to the low incidence of extreme inflation events in that series.

# 6 Recent events in monetary policy

In this section, we re-examine the path of the macroeconomy through the recent Great Recession from the perspective of our monetary policy communication framework.

Beginning with the FOMC Press Release of January 28, 2009, the Committee repeatedly warned the public about the risk of low inflation events, by predicting that inflation would be 'below rates that best foster economic growth and price stability'. Subsequent warnings, using identical wording, were made in April and May of that year. Figures 6, 7 and 8 plot the probability of inflation events between 2008:1 and 2012:1, for the three events inflation less than 1%, 1.25% and 1.5%, respectively. Recall that these are one step ahead probabilities based on quarterly data. For example, the 2009:1 probability shown refers to the forecast for the first quarter of 2009, based on the data vintage 2009:1, in which the most recent observation for GDP is for the fourth quarter of 2008. So the forecast for 2009:1 could have been made mid-quarter and an inflation warning issued during 2009:1. In effect, this is a nowcast.

Figure 6 shows the probability of inflation less than 1% for 2009:1 to be considerably higher than the corresponding probability for 2008:4 according to the VAR LOP. Specifically, the probability increased to around 0.31 from about 0.10. In 2009:2, the probability rises to 0.44. The probability of this low inflation event peaked for 2009:3 at approximately 0.7, before declining to around 0.42 for the final quarter of 2009. Thereafter, the probability of this event dropped back below 0.31, with oscillations, for the remainder of the evaluation period. The inflation observations used for evaluation for these four quarters of 2009 where 1.85, -0.02, 0.39 and 0.51, respectively. Hence, three of these realized values actually fell below the threshold.[19]

What range of values for $R$ in our theoretical framework would be consistent with the actual FOMC communication strategy? Recall that the FOMC forewarned of low inflation from the beginning of 2009. In our theoretical framework, the warning will be issued if the (one step ahead) probability of the event exceeds $R$. Hence, from Figure 6, a value of $R$ around of 0.3 would result in a low inflation warning from early 2009, given the forecasts from the VAR LOP.[20]

Figure 6 also plots the probabilities of the same inflation event for the IMA benchmark and the climatological forecast. The IMA probabilities typically react less quickly to falls in inflationary pressures than the VAR LOP ensemble during 2008 and 2009. The IMA also indicates a considerably higher risk of low inflation events than the VAR LOP for 2010 and the first half of 2011, with probabilities fluctuating around 0.5. The climatological forecast is flat at just over 0.1 (indicating the percentage of observations in the pre-evaluation sample with realizations below the 1% threshold). Notice that for a critical

[19]Four other low inflation events by this threshold during the Great Recession were 2008:4, 2010:1 2010:4 and 2011:4. The respective realised values for inflation were 0.55, 0.97, 0.38 and 0.88.

[20]The actual FOMC language did shift slightly during 2009. An FOMC Press Release warned that inflation would be 'below rates that best foster economic growth and price stability' from late January: the language softened from June 24, 2009 to warn that inflation would 'remain subdued'—perhaps reflecting slightly weaker risk later in 2009.

value of $R$ of around 0.3, the IMA would imply a longer sequence of warnings from the FOMC (stretching into 2011); and the climatological forecast would imply none. Neither approach seems easy to reconcile with the FOMC's actual communication strategy.

Turning to Figures 7 and 8, displaying probabilities for the less extreme low inflation events, using thresholds of 1.25% and 1.5%, the same general patterns are apparent as in Figure 6. The risk of low inflation increases sharply at the start of 2009 based on the VAR LOP. And the risk remains elevated through 2009. The IMA gives a less timely indicator of low inflation risk with probabilities evaluated in late 2009 through into 2011. Since the climatological forecasts make no use of macroeconomic variables during the evaluation, the probabilities are constant throughout the evaluation.[21]

Looking across all three inflation events, Figures 6 through 8, the VAR LOP implies a communication strategy broadly consistent with the behavior of the Federal Reserve during 2009, for values of $R$ in the range 0.3 to 0.4. With the relative cost of anticipated inflation in this region, the VAR LOP consistently outperforms the IMA benchmark, with an economic cost saving averaging around 9%, but typically in the range 5% to 10%; see Figures 3 through 5. This seems like a reasonable ballpark estimate of the scope for using formal probability forecasting methods to forewarn the public of inflation events.

# 7 Conclusions

The US Federal Reserve and other central banks around the world published warnings about low inflation events during the Great Recession. In this paper, we have proposed a probability forecasting framework specifically for the purpose of issuing warnings to the public. In our application, we have combined the predictive densities produced from VAR models, utilizing prediction pools. The resulting probabilistic forecasts for US inflation were well calibrated. Furthermore, by deploying a cost-loss ratio for forecast evaluation, we have shown that the approach has a considerable performance advantage over the benchmark IMA model for the low inflation events associated with the Great Recession. Furthermore, we have demonstrated that the IMA fails to outperform a climatological forecast for some feasible parameter values.

Our finding indicates considerable scope for using formal forecasting methods to forewarn the public of extreme inflation events. Further research in this area should encompass macroeconometric forecasting models with explicit consideration of the contribution of food and energy price shocks. A common finding is that more comprehensive macroeconometric models struggle to beat simple autoregressive benchmarks in terms of point forecasting. But, as this paper has demonstrated, preemptive warnings of extreme inflation events by monetary policymakers require probability forecasts.

[21] All three approaches indicate higher probabilities than for the 1% threshold because the 1.25% and 1.5% thresholds are closer to the unconditional mean.

# References


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

[2] Bao, Y., T-H. Lee and B. Saltoglu (2007), "Comparing Density Forecast Models", *Journal of Forecasting*, 26, 203-225.

[3] Baxter, M, and R.G. King (1999), "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series", *Review of Economics and Statistics*, 81, 594-607.

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

[5] Bernanke, B.S. (2003), "An Unwelcome Fall in Inflation?", Remarks Before the Economics Roundtable, University of California, San Diego, La Jolla, California, July 23, 2003.

[6] Bernanke, B.S. (2007), "Inflation Expectations and Inflation Forecasting", Speech delivered at the Monetary Economics Workshop of the National Bureau of Economic Research Summer Institute, Cambridge, Massachusetts, July 10, 2007.

[7] Berrocal, V.J., A.E. Raftery, T. Gneiting and R.C. Steel (2010), "Probabilistic Weather Forecasting for Winter Road Maintenance", *Journal of American Statistical Association*, 105, 490, 522-537.

[8] Beveridge, S., and C.R. Nelson (1981), "A New Approach to Decomposition of Time Series into Permanent and Transitory Components with Particular Attention to Measurements of the Business Cycle", *Journal of Monetary Economics*, 7, 151-174.

[9] Brier, G.W. (1950), "Verification of Forecasts Expressed in Terms of Probability", *Monthly Weather Review*, 78, 1-3.

[10] Christiano, L. and T.J. Fitzgerald (2003), "The Band Pass Filter", *International Economic Review*, 44, 2, 435-465.

[11] Clark, T.E. and S.J. Terry (2010), "Time Variation in the Inflation Passthrough of Energy Prices", *Journal of Money Credit and Banking*, 42, 7, 1419-1433.

[12] Clark, T.E. (2011), "Real-Time Density Forecasts From Bayesian Vector Autoregressions with Stochastic Volatility" *Journal of Business and Economic Statistics*, 29, 327-341.

[13] Clark, T.E. and F. Ravazzolo (2012), "Comparing Models of Time-Varying Macroeconomic Volatility", Mimeo.

[14] Clements, M.P. (2004), "Evaluating the Bank of England Density Forecasts of Inflation", *Economic Journal*, 114, 855-877.

[15] Corradi, V., A. Fernandez and N.R. Swanson (2009), "Information in the Revision Process of Real-Time Data Sets", *Journal of Business and Economic Statistics*, 27, 455-467.

[16] Corradi, V. and N.R. Swanson (2006), "Predictive Density Evaluation", G. Elliot, C.W.J. Granger and A. Timmermann, eds, *Handbook of Economic Forecasting,* North-Holland, 197-284.

[17] Croushore, D. and T. Stark (2001), "A Real-time Data Set for Macroeconomists", *Journal of Econometrics*, 105, 111-130.

[18] Dawid, A.P. (1984), "Statistical Theory: the Prequential Approach", *Journal of the Royal Statistical Society, Series A*, 147, 278-290.

[19] Diebold, F.X., T.A. Gunther and A.S. Tay (1998), "Evaluating Density Forecasts; with Applications to Financial Risk Management", *International Economic Review,* 39, 863-83.

[20] Diks, C., V. Panchenko, V. and D. van Dijk (2011), "Likelihood-Based Scoring Rules for Comparing Density Forecast in Tails", *Journal of Econometrics*, 163(2), 215-230.

[21] Driffill, E.J., G. Mizon and A. Ulph (1990), "Costs of Inflation". In B.M. Friedman and F.H. Hahn (eds.), *Handbook of Monetary Economics*, Elsevier, Edition 1, Vol. 2, Number 2.

[22] Edge, R.M and J.B. Rudd (2012), "Real-Time Properties of the Federal Reserve's Output Gap", Finance and Economics Discussion Series 2012-86. Board of Governors of the Federal Reserve System (U.S.).

[23] Galbraith, J.W., and S. van Norden (2012), "Assessing GDP and Inflation Probability Forecasts Derived from the Bank of England Fan Charts", *Journal of the Royal Statistical Society, Series A*, 175(3), 713-727.

[24] Garratt, A., Lee, K., Mise, E. and K. Shields (2008), "Real-Time Representations of the Output Gap", *Review of Economics and Statistics*, 90(4), 792-804.

[25] Garratt, A., J. Mitchell, S.P. Vahey (2011), "Measuring Output Gap Nowcast Uncertainty", *CAMA Working Paper 2011-16.*

[26] Garratt, A., J. Mitchell, S.P. Vahey and E. Wakerly (2011), "Real-time Inflation Forecast Densities from Ensemble Phillips Curves", *North American Journal of Economics and Finance*, 22, 77-87.

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

[28] George, E. (1992), "The Case for Price Stability", *Bank of England Quarterly Bulletin*, November, 441-448.

[29] Genest, C. and J. Zidek (1986), "Combining Probability Distributions: a Critique and an Annotated Bibliography", *Statistical Science*, 1, 114-135.

[30] Granger, C. W. J. and M.H. Pesaran (2000a), "A Decision-Based Approach to Forecast Evaluation." In W.S. Chan, W.K. Li and H. Tong. (eds.), *Statisitcs and Finance: An Interface*, London: Imperial College Press.

[31] Granger, C. W. J. and M.H. Pesaran (2000b), "Economic and Statisitcal Measures of Forecast Accuracy", *Journal of Forecasting*, 19, 537-560.

[32] Greenspan, A. (2004) "Risk and Uncertainty in Monetary Policy", *American Economic Review*, Papers and Proceedings, May, 33-40.

[33] Hall, S.G. and J. Mitchell (2007), "Density Forecast Combination", *International Journal of Forecasting*, 23, 1-13.

[34] Harvey, A.C. (2006), "Forecasting with Unobserved Components Time Series Models". In G. Elliot, C.W.J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting,* North-Holland, 327-412.

[35] Hodrick, R. and E. Prescott (1997), "Post-War U.S. Business Cycles: An Empirical Investigation", *Journal of Money, Banking and Credit*, 29, 1-16.

[36] Jore, A.S., J. Mitchell and S.P. Vahey (2010), "Combining Forecast Densities from VARs with Uncertain Instabilities", *Journal of Applied Econometrics*, 25, 621-634.

[37] Katz, R.W. and A.H. Murphy (eds.) (1997), *Economic Value of Weather and Climate Forecasts*, Cambridge: Cambridge University Press.

[38] Kilian, L. and S. Manganelli (2007), "Quantifying The Risk of Deflation", *Journal of Money Credit and Banking*, 39, No. 2-3, 561-590.

[39] Marcellino, M., J. Stock and M.W. Watson (2006), "A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Series h-steps Ahead", *Journal of Econometrics*, 135, 499-526.

[40] Mise, E., T-H. Kim and P. Newbold (2005), "On the Sub-Optimality of the Hodrick-Prescott Filter", *Journal of Macroeconomics*, 27, 1, 53-67.

[41] 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.

[42] Mitchell, J. and K.F. Wallis (2011), "Evaluating Density Forecasts: Forecast Combinations, Model Mixtures, Calibration and Sharpness", *Journal of Applied Econometrics*, 26, 1023-1040.

[43] Murphy, A. H. (1973), "A New Vector Partition of the Probability Score", *Journal of Applied Meteorology* , 12, 595-600.

[44] Murphy, A. H. (1977), "The Value of Climatological, Categorial and Probabilistic Forecasts in the Cost-Loss Situation", *Monthly Weather Review*, 105, 803-816.

[45] Noceti, P., J. Smith and S. Hodges (2003), "An Evaluation of Tests of Distributional Forecasts", *Journal of Forecasting*, 22, 447-455.

[46] Orphanides, A. and S. van Norden (2002), "The Unreliability of Output-Gap Estimates in Real Time", *Review of Economics and Statistics*, 84, 4, 569-583.

[47] Orphanides, A. and S. van Norden (2005), "The Reliability of Inflation Forecasts Based on Output-Gap Estimates in Real Time", *Journal of Money Credit and Banking*, 37, 3, 583-601.

[48] Ravazzolo, F. and S.P. Vahey (2013), "Forecast Densities for Economic Aggregates From Disaggregates Ensembles", *Studies in Nonlinear Dynamics and Econometrics.* Forthcoming.

[49] Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation", *The Annals of Mathematical Statistics*, 23, 470-472.

[50] Stock, J.H. and M.W. Watson (2007), "Why Has Inflation Become Harder to Forecast?", *Journal of Money, Credit, and Banking*, 39, 3-34.

[51] Wallis, K.F. (2003), "Chi-squared Tests of Interval and Density Forecasts, and the Bank of England's fan charts", *International Journal of Forecasting*, 19, 165-175.

[52] Watson, M.W. (2007), "How Accurate are Real-Time Estimates of Output Trends and Gaps?", *Federal Reserve Bank of Richmond Economic Quarterly*, Spring 2007.

[53] Yellen, J.L. (2006), "The US Economy in 2006", Speech before an Australian Business Economists luncheon Sydney, Australia, March 15, 2006.

[54] Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York: John Wiley and Sons.

## Appendix 1: Output trend definitions

We summarize the seven univariate detrending specifications below.

1. For the quadratic trend based measure of the output gap we use the residuals from a regression (estimated recursively) of output on a constant and a squared time trend.

2. Following Hodrick and Prescott (1997, HP), we set the smoothing parameter to 1600 for our quarterly US data.[22]

3. Since the HP filter is two-sided filter it relates the time-$t$ value of the trend to future and past observations. Moving towards the end of a finite sample of data, the HP filter becomes progressively one-sided and its properties deteriorate with the Mean Squared Error (MSE) of the unobserved components increasing and the estimates ceasing to be optimal in a MSE sense. We therefore follow Mise et al. (2005) and seek to mitigate this loss near and at the end of the sample by extending the series with forecasts. At each recursion the HP filter is applied to a forecast-augmented output series (again with smoothing parameter 1600), with forecasts generated from an univariate AR(8) model in output growth (again estimated recursively using the appropriate vintage of data). The implementation of forecast augmentation when constructing real-time output gap measures for the US is discussed at length in Garratt, Lee, Mise and Shields (2008).[23]

4. Christiano and Fitzgerald (2003) propose an optimal finite-sample approximation to the band-pass filter, without explicit modeling of the data. Their approach implicitly assumes that the series is captured reasonably well by a random walk model and that, if there is drift present, this can be proxied by the average growth rate over the sample.

5. We also consider the band-pass filter suggested by Baxter and King (1999). We define the cyclical component to be fluctuations lasting no fewer than six, and no more than thirty two quarters—the business cycle frequencies indicated by Baxter and King (1999) - and set the truncation parameter (the maximum lag length) at three years. As with the HP filter we augment our sample with AR(8) forecasts to fill in the 'lost' output gap observations at the end of the sample due to truncation. Watson (2007) reviews band-pass filtering methods.

[22] We could, of course, allow for uncertainty in the smoothing parameter. We reduce the computational burden in this application by fixing this parameter at 1600.

[23] There is always a question empirically about what model should be used to produce the forecasts used to augment the sample. Here, in following Garratt Lee, Mise and Shields (2008), we deliberately select a high lag order to ensure no important lags are omitted—we favor unbiasedness over efficiency when estimating the AR. But, of course, this is not the only modeling option. We did experiment with use of a model selection criterion, specifically the BIC, which is known to favor more parsimonious models. In fact, inference in our application is robust to either choice with near identical output gap estimates. In principle, our methodology can avoid any selection by simply treating each variant of the HP filter, or indeed any of the other filters, as a new output gap estimator; i.e., we can increase $J$. But in order both to keep this application manageable and to stick to a well-known representative model space we focus on these $J = 7$ filters.

6. The Beveridge and Nelson (1981) decomposition relies on a priori assumptions about the correlation between permanent and transitory innovations. The approach imposes the restriction that shocks to the transitory component and shocks to the stochastic permanent component have a unit correlation. We assume the ARIMA process for output growth is an AR(8), the same as that used in our forecast augmentation.

7. Finally, our Unobserved Components model assumes $q_t$ is decomposed into trend, cyclical and irregular components

$$q_t = \mu_t^7 + y_t^7 + \xi_t, \quad \xi_t \sim i.i.d.\ N(0, \sigma_\xi^2), \quad t = 1, \ldots, T \tag{A1.1}$$

where the stochastic trend is specified as

$$\begin{aligned} \mu_t^7 &= \mu_{t-1}^7 + \beta_{t-1} + \eta_t, \quad \eta_t \sim i.i.d.\ N(0, \sigma_\eta^2) \\ \beta_t &= \beta_{t-1} + \zeta_t, \quad \zeta_t \sim i.i.d.\ N(0, \sigma_\zeta^2). \end{aligned}$$

The cyclical component is assumed to follow a stochastic trigonometric process:

$$\begin{bmatrix} y_t^7 \\ y_t^{7*} \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \begin{bmatrix} y_{t-1}^7 \\ y_{t-1}^{7*} \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} \tag{A1.2}$$

where $\lambda$ is the frequency in radians, $\rho$ is a damping factor such that $0 \leq \rho \leq 1$ and $\kappa_t$ and $\kappa_t^*$ are two independent white noise Gaussian disturbances with common variance $\sigma_\kappa^2$. We estimate this model by maximum likelihood, exploiting the Kalman filter, and estimates of the trend and cyclical components are obtained using the Kalman smoother. For a detailed description of the unobserved components approach see Harvey's (2006) literature review and the references therein.

## Appendix 2: Predictive densities from the VAR model

This appendix provides the formula for the predictive density from the VAR models. In what follows, we use $\mathbf{Z}'$ to denote the transpose of $\mathbf{Z}$.

Stack the VAR model, (1), as

$$\mathbf{Y}^j = \mathbf{Z}^j \mathbf{A}^j + \mathbf{E}^j, \tag{A2.2}$$

where $\mathbf{Y}^j = \begin{bmatrix} \mathbf{y}_1^j \\ \mathbf{y}_2^j \\ \\ \mathbf{y}_T^j \end{bmatrix}$, $\mathbf{y}_t^j = \left[\pi_t \; y_t^j\right]$, $\mathbf{z}_t^j = \mathbf{y}_{t-1}^j$, $\mathbf{Z}^j = \begin{bmatrix} \mathbf{z}_1^j \\ \mathbf{z}_2^j \\ \\ \mathbf{z}_T^j \end{bmatrix}$ with $\mathbf{E}^j$ defined from $\epsilon_t^j$ conformably.

Denote the usual Ordinary Least Squares coefficient estimators as $\widehat{\mathbf{A}}^j$ and $\widehat{\mathbf{\Sigma}}^j$. Integrating out the coefficients, the one step ahead posterior predictive density for $\mathbf{y}_{T+1}^j$ follows a multivariate $t$ with mean $\mathbf{z}_{T+1}^j \widehat{\mathbf{A}}^j$, covariance $\left[1 + \mathbf{z}_{T+1}^j (\mathbf{Z}^{j\prime}\mathbf{Z}^j)^{-1} \mathbf{z}_{T+1}^{j\prime}\right] \widehat{\mathbf{\Sigma}}^j$, with $T$ degrees of freedom; see Zellner (1971), pp. 233-236.

We note that $h$-step ahead densities ($h > 1$) could also be produced in a similar manner utilizing the direct forecast methodology; see Marcellino, Stock and Watson (2006).

## Appendix 3: Calibration Tests

A common approach to forecast density evaluation provides statistics suitable for one-shot tests of (absolute) forecast accuracy, relative to the "true" but unobserved density. Following Rosenblatt (1952), Dawid (1984) and Diebold, Gunther and Tay (1998), evaluation can use the probability integral transforms (*pits*) of the realization of the variable with respect to the forecast densities. We gauge calibration by examining whether the *pits* $z_\tau$, where:

$$z_\tau = \int_{-\infty}^{\pi'_\tau} p(u)du, \tag{10}$$

are uniform and, for our one step ahead forecasts, independently and identically distributed; see Diebold et al (1998). In practice, therefore, density evaluation with the *pits* requires application of tests for goodness of fit and independence at the end of the evaluation period. Given the large number of component densities under consideration in the ensemble, we do not allow for parameter estimation uncertainty in the components when evaluating the *pits*. Corradi and Swanson (2006) review *pits* tests computationally feasible for small $N$.

The goodness of fit tests employed include the Likelihood Ratio (LR) test proposed by Berkowitz (2001); we use a three degrees of freedom variant with a test for independence, where under the alternative $z_\tau$ follows an AR(1) process. In addition, we consider the Anderson-Darling (AD) test for uniformity, a modification of the Kolmogorov-Smirnov test, intended to give more weight to the tails, and advocated by Noceti, Smith and Hodges (2003). Finally, following Wallis (2003), we employ a Pearson chi-squared test ($\chi^2$) which divides the range of the $z_\tau$ into eight equiprobable classes and tests for uniformity in the histogram. Turning to the test for independence of the *pits*, we use a Ljung-Box (LB) test, based on (up to) fourth-order autocorrelation.

Examining the goodness of fit and independence *pits* tests presented in the first two rows of the top panel of Table A1, we see that the one step ahead ensemble forecast densities are well calibrated for a 10 percent $p$-value. (Instances of appropriate calibration are marked in boldface). To control the joint size of the four evaluation tests applied would require the use of a stricter $p$-value. For example, the Bonferroni correction indicates a $p$-value threshold of $(100 - 90)/4 = 2.5$ percent, rather than 10 percent. The bottom row of Table A1 displays the corresponding results from the battery of *pits*-based tests for the benchmark univariate IMA model. The IMA fails the LR test for a 5 percent $p$-value and has lower logarithmic scores than either ensemble. The IMA model for inflation takes the form: $\Delta\pi_t = \alpha + \varepsilon_t + \theta\varepsilon_{t-1}$.

Tables A2 and A3 display the calibration test results for longer horizons, $h = 2$ and $h = 4$. In these cases, at least one tests reveals evidence of poor calibration for all specifications.

**Table A1: Forecast Density Evaluation for Inflation, $h = 1$**

| | LR | AD | $\chi^2$ | LB | Log score |
|---|---|---|---|---|---|
| LOP | **0.72** | **0.59** | **0.85** | **0.16** | -1.360 |
| LogOP | **0.74** | **0.55** | **0.82** | **0.16** | -1.397 |
| IMA | 0.01 | **0.21** | **0.82** | **0.12** | -1.495 |

**Table A2: Forecast Density Evaluation for Inflation, $h = 2$**

| | LR | AD | $\chi^2$ | LB | Log score |
|---|---|---|---|---|---|
| LOP | 0.01 | **0.25** | **0.52** | 0.01 | -1.466 |
| LogOP | 0.00 | 0.09 | **0.44** | 0.00 | -1.474 |
| IMA | 0.00 | 0.09 | 0.04 | 0.01 | -1.544 |

**Table A3: Forecast Density Evaluation for Inflation, $h = 4$**

| | LR | AD | $\chi^2$ | LB | Log score |
|---|---|---|---|---|---|
| LOP | 0.00 | 0.04 | **0.35** | 0.00 | -1.616 |
| LogOP | 0.00 | 0.01 | **0.25** | 0.00 | -1.573 |
| IMA | 0.00 | 0.03 | 0.02 | 0.00 | -1.665 |

Notes: LR is the $p$-value for the Likelihood Ratio test of zero mean, unit variance and zero first order autocorrelation of the inverse normal cumulative distribution function transformed *pits*, with a maintained assumption of normality for the transformed *pits*; AD is the small-sample (simulated) $p$-value from the Anderson-Darling test for uniformity of the *pits* assuming independence of the *pits*. $\chi^2$ is the $p$-value for the Pearson chi-squared test of uniformity of the *pits* histogram in eight equiprobable classes. LB is the $p$-value from a Ljung-Box test for independence of the *pits* based on autocorrelation coefficients up to four. Log Score is the average logarithmic score over the evaluation period.

**Figure 1: Inflation (annualized, %)**

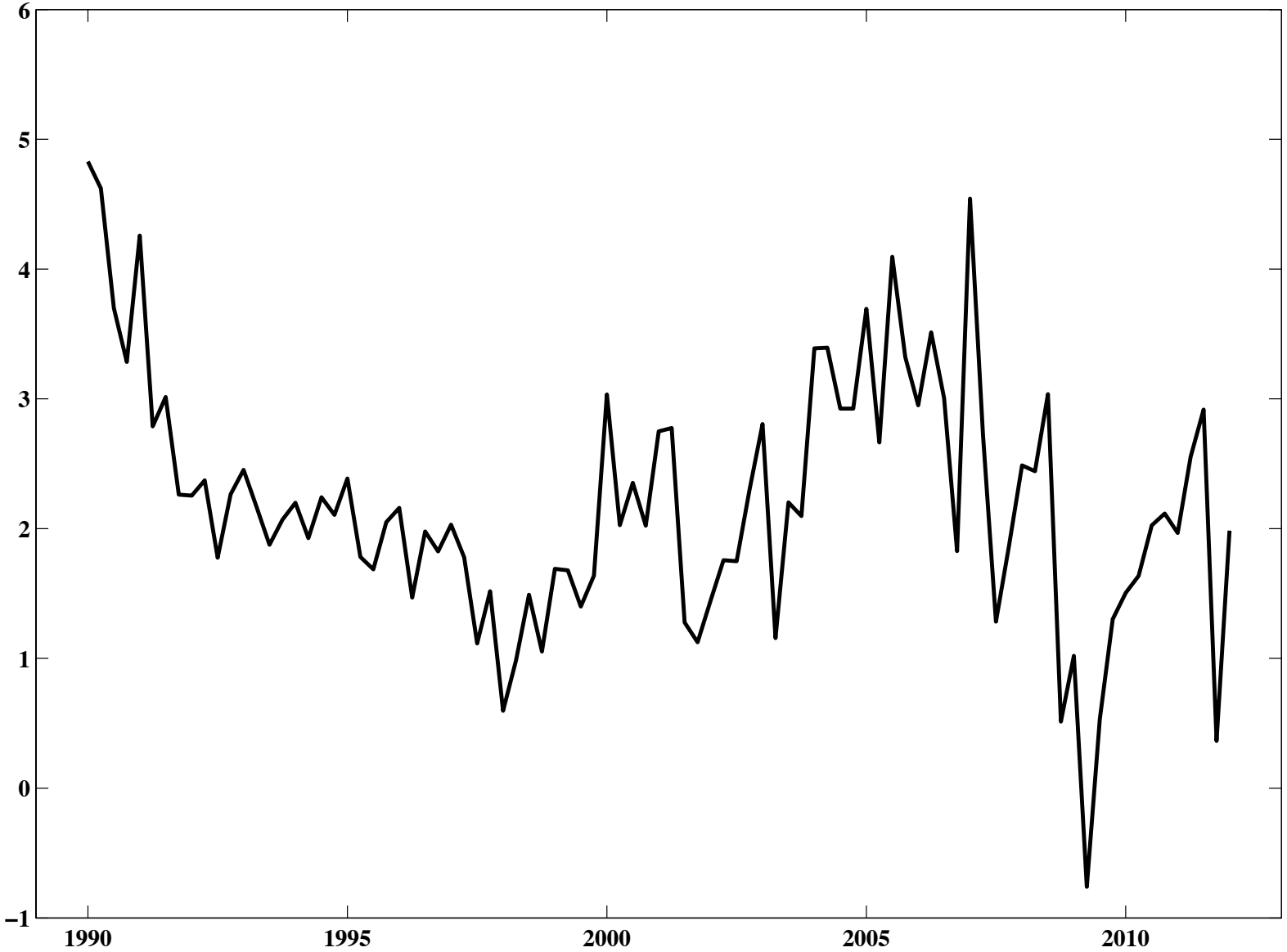


**Figure 2: Output growth (annualized, %)**

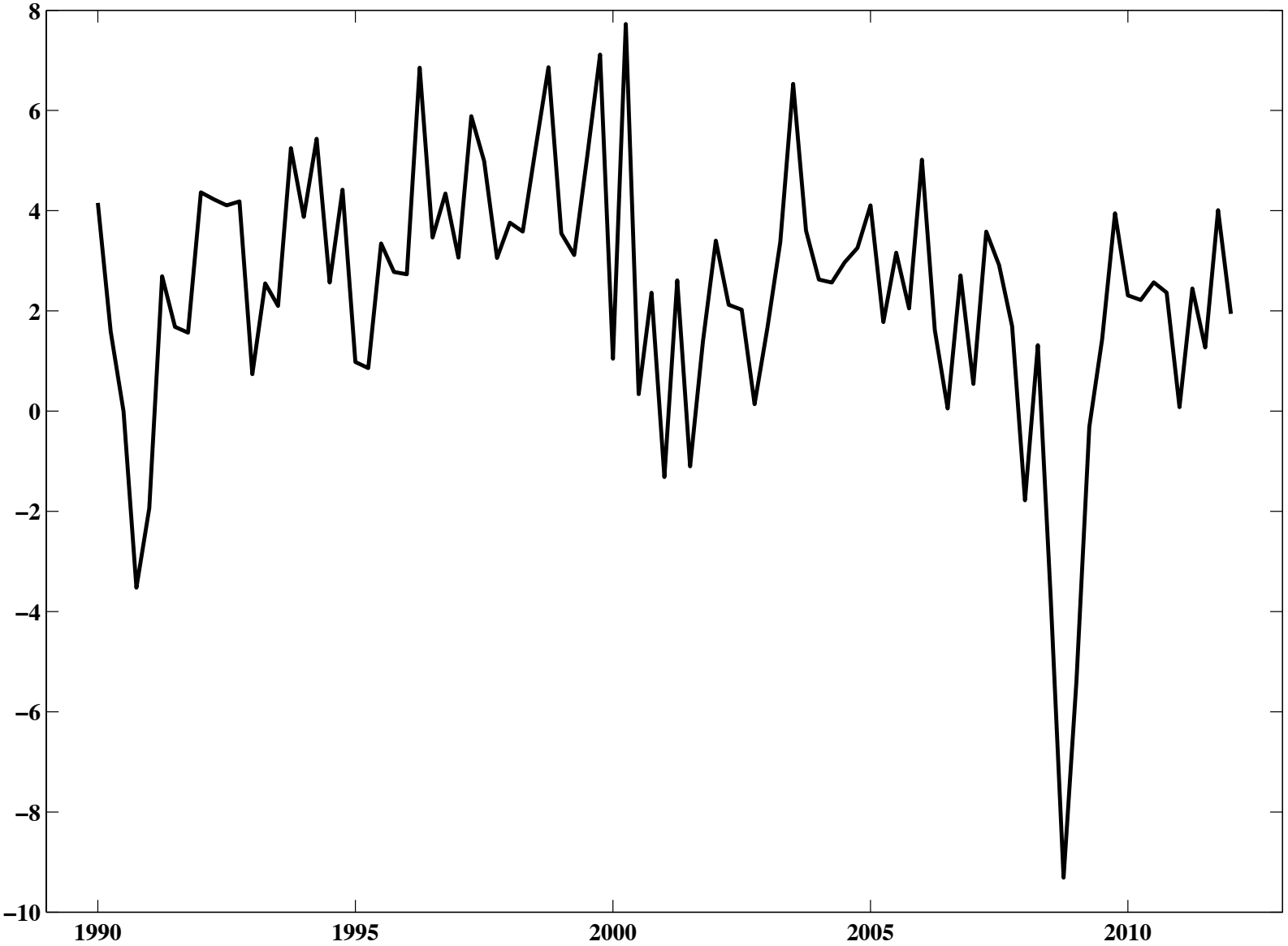


**Figure 3: Total Economic Loss, $\pi < 1.0$ event**

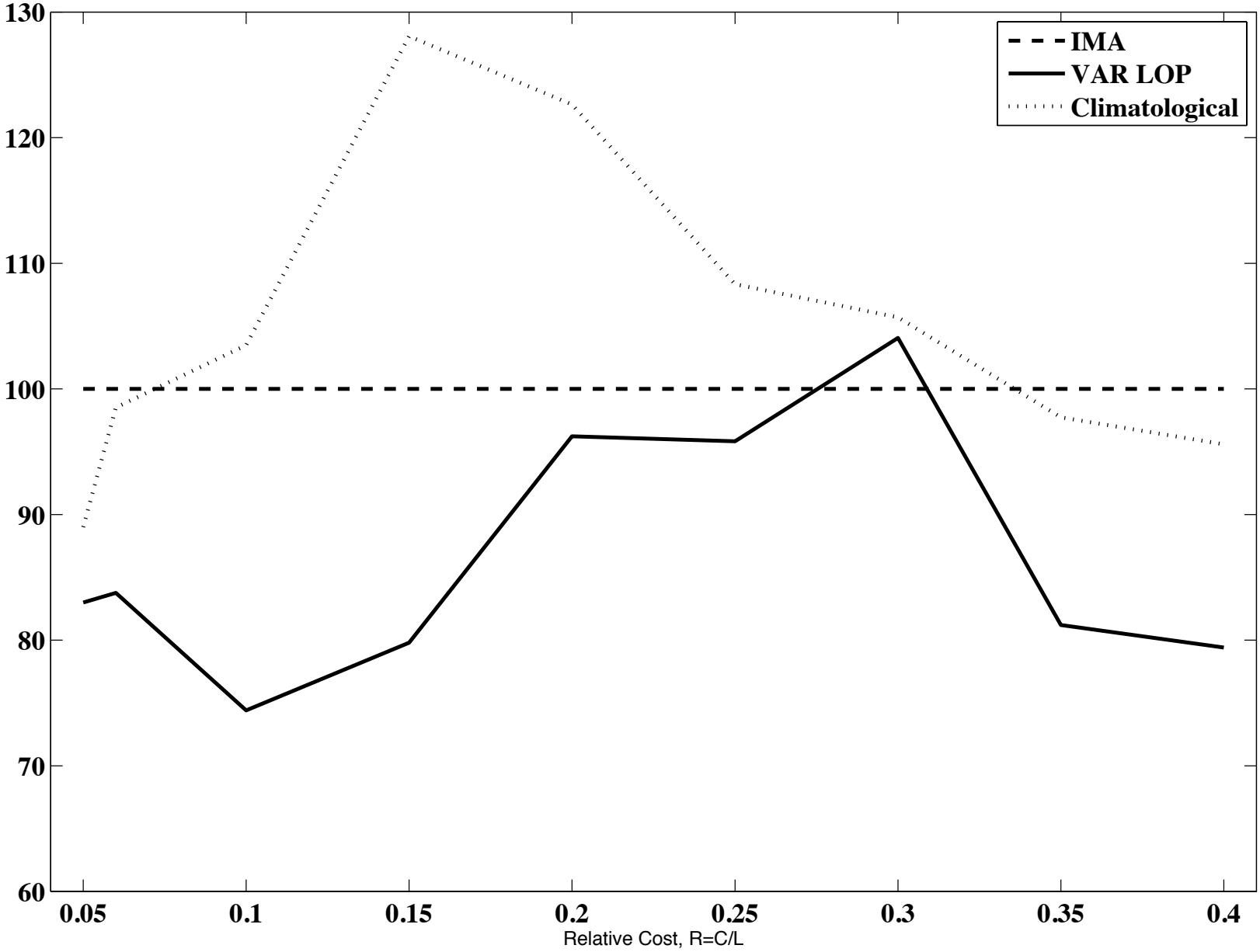


**Figure 4: Total Economic Loss, $\pi < 1.25$ event**

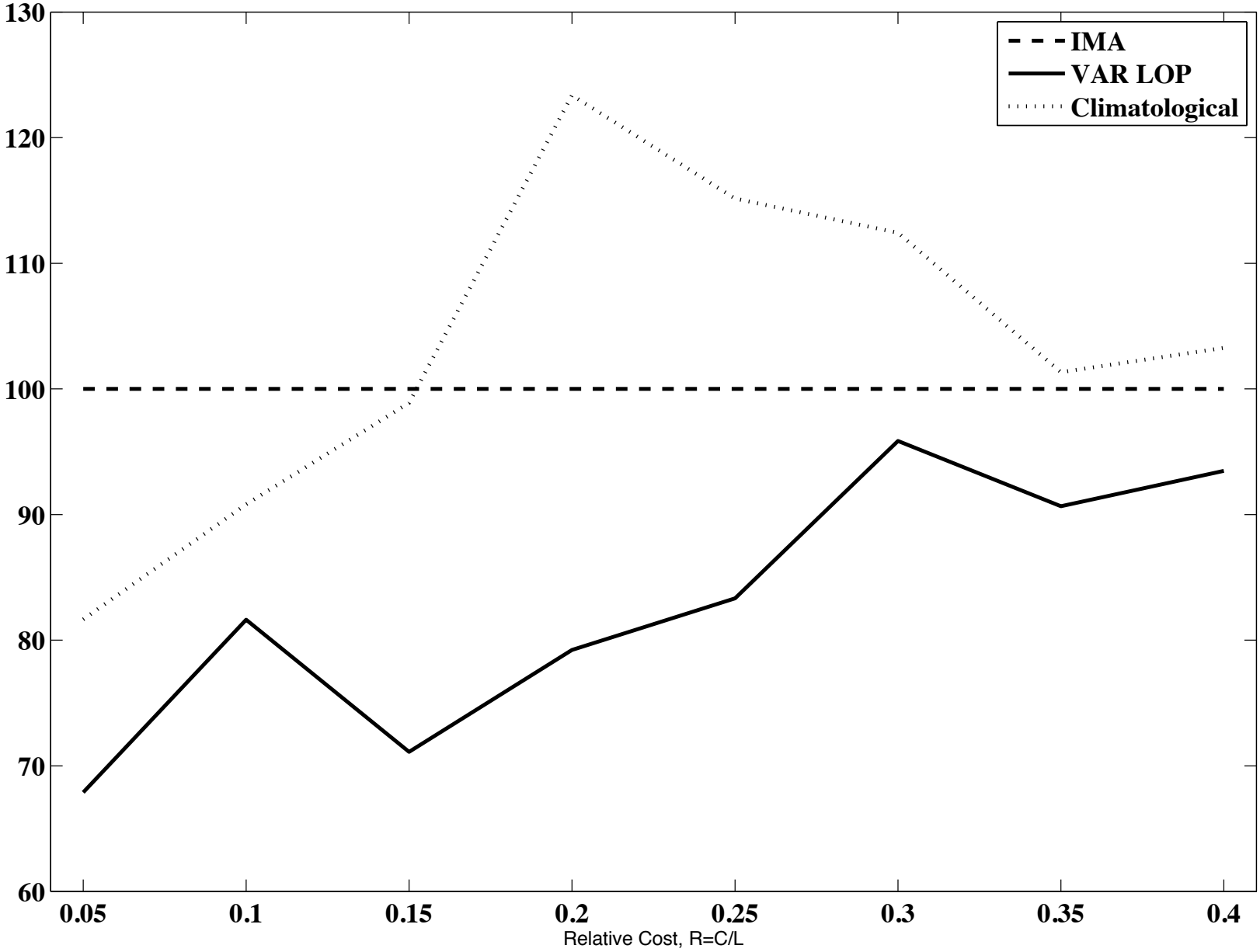
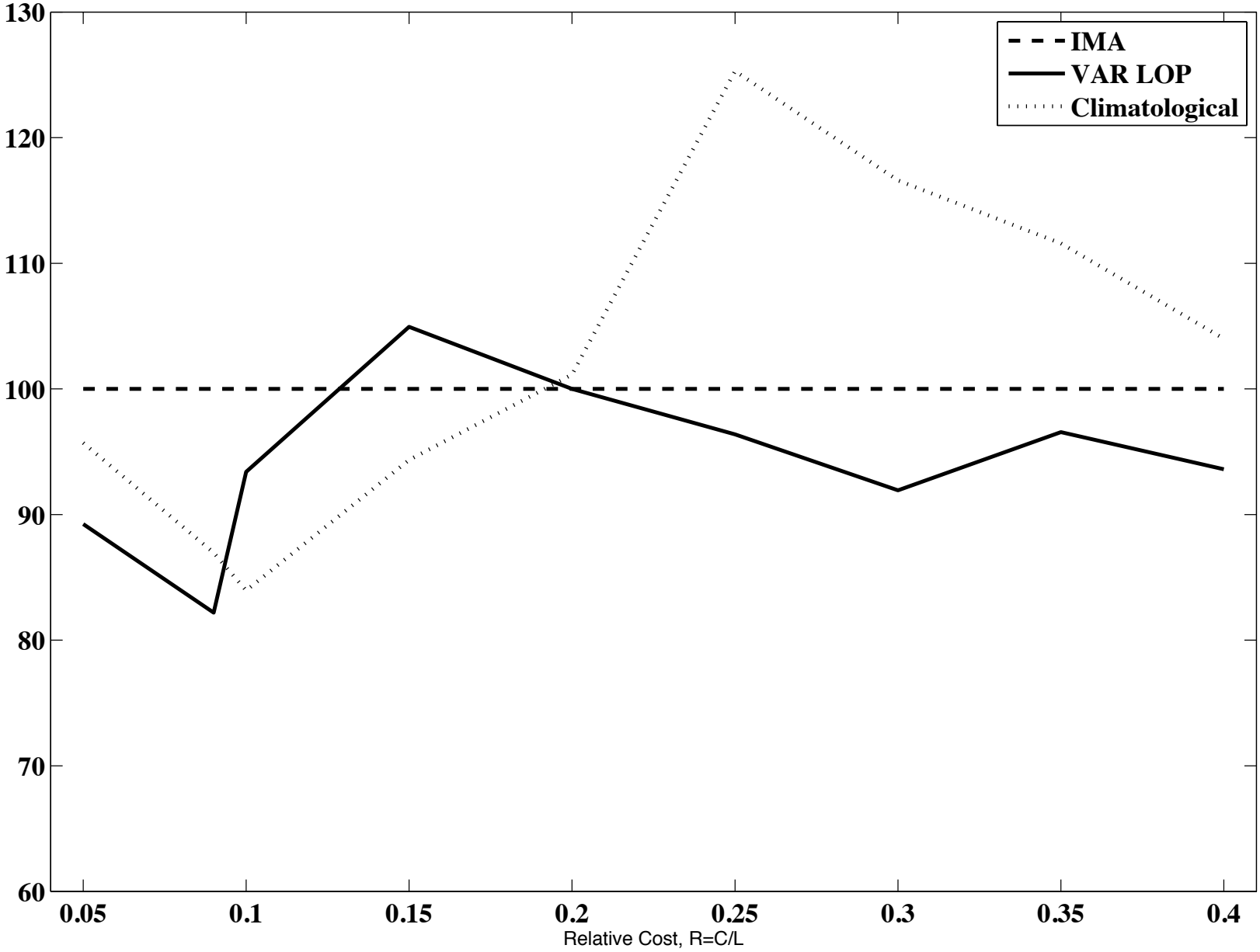


**Figure 5: Total Economic Loss, $\pi < 1.5$ event**



## Figure 6: Pr(π<1.0)

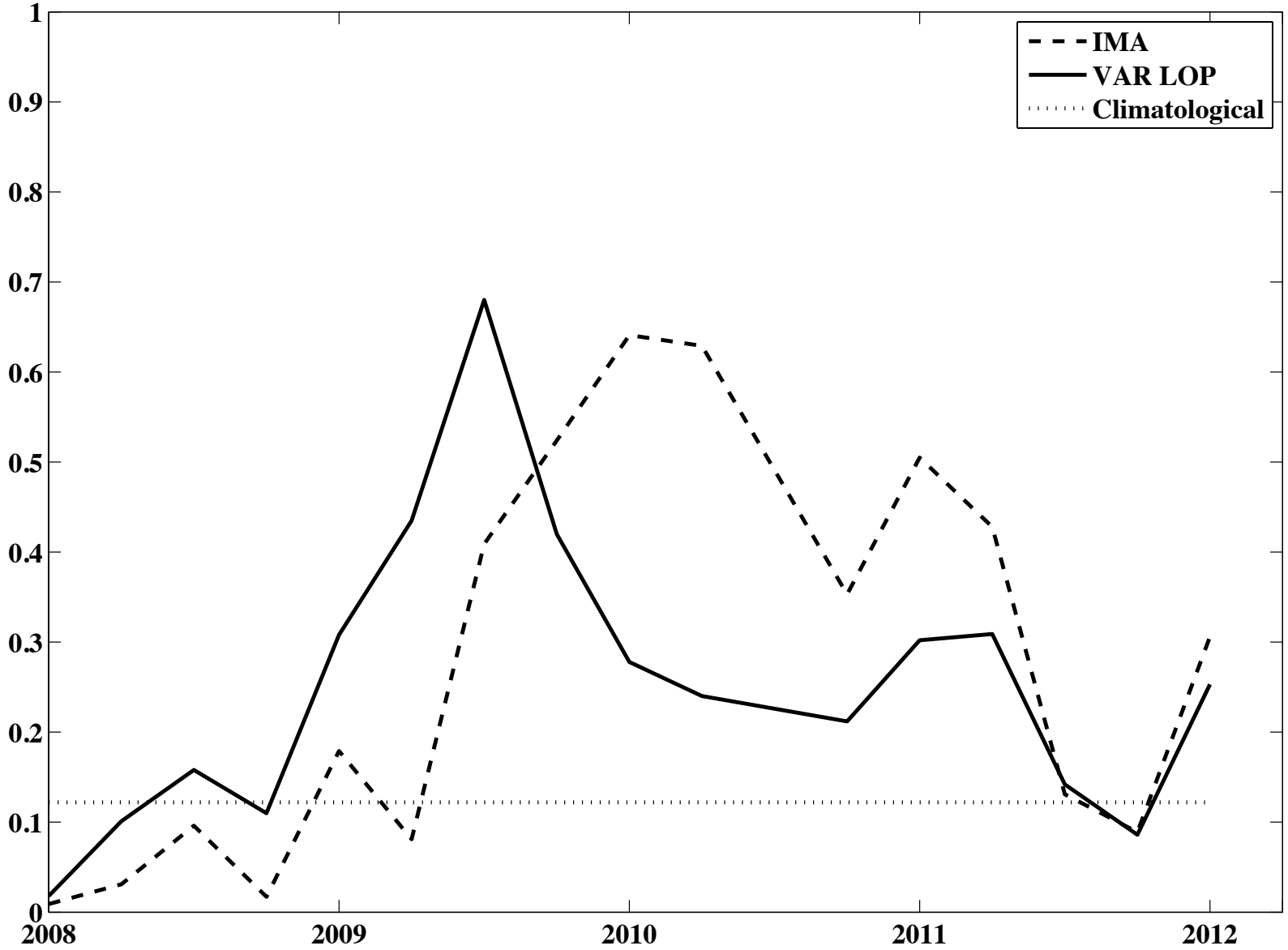
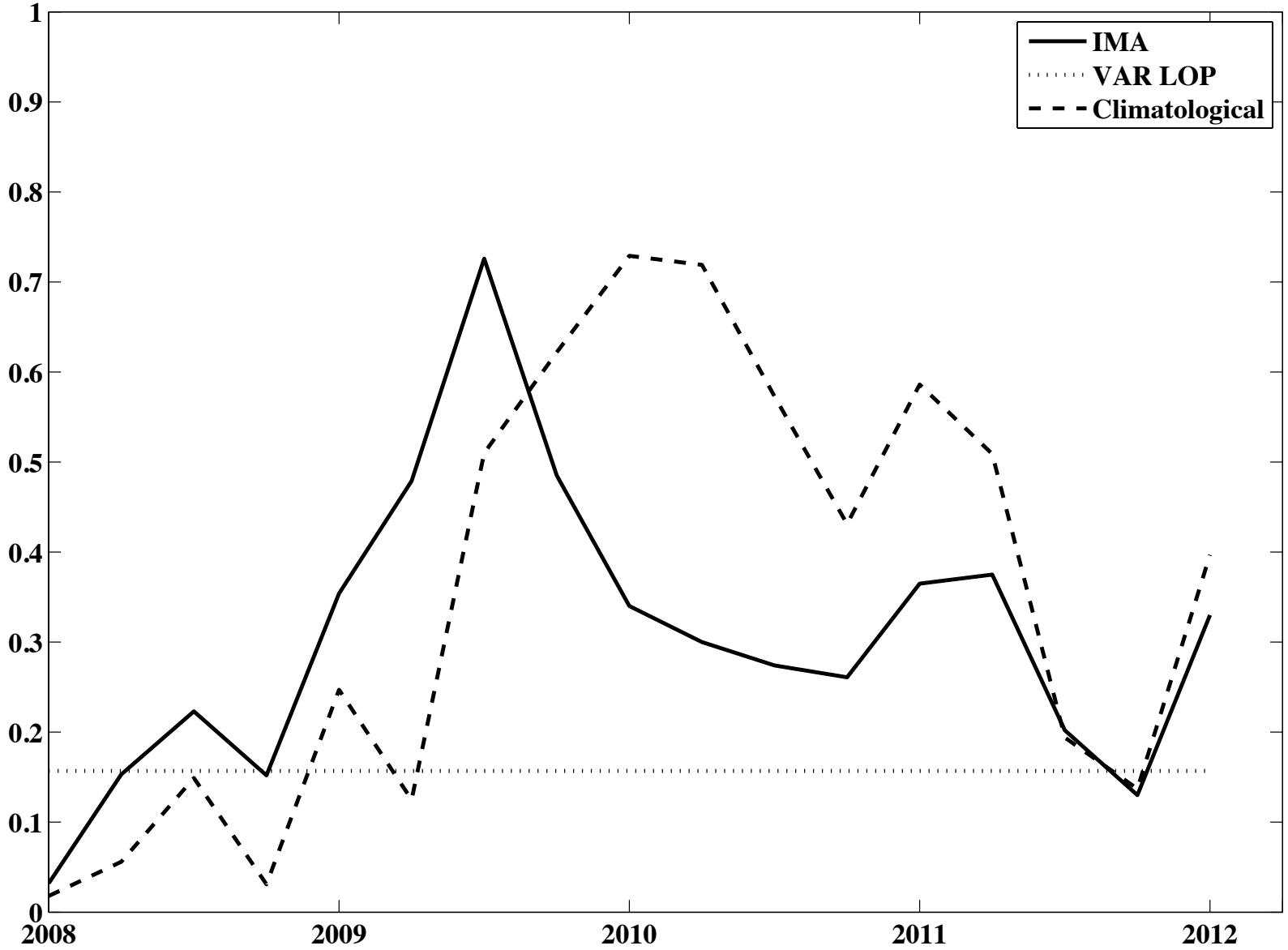


**Figure 7: Pr(π<1.25)**



## Figure 8: Pr(π<1.5)