Forecasting Inflation

Jon Faust and Jonathan H. Wright*

First Version: May 23, 2011
This version: June 27, 2012

*Department of Economics, Johns Hopkins University, Baltimore MD 21218; faustj@jhu.edu and wrightj@jhu.edu. We are grateful to two anonymous referees, Todd Clark, Joseph Haubrich, Kirstin Hubrich, Sharon Kozicki, Evan Koenig, Lutz Kilian, Barbara Rossi, Ellis Tallman and Alan Timmermann for their very helpful comments on earlier drafts, and to Lu Xu for research assistance. All errors are our sole responsibility.

1 Introduction

Long-term nominal commitments such as labor contracts, mortgages and other debt, and price stickiness are widespread features of modern economies. In such a world, forecasting how the general price level will evolve over the life of a commitment is an essential part of private sector decision-making. The existence of long-term nominal obligations is also among the primary reasons economists generally believe that monetary policy is not neutral, at least over moderate horizons. While macroeconomists continue to debate whether these nonneutralities give rise to beneficially exploitable trade-offs for monetary policymakers, the recent New Keynesian formulation of optimal policy has raised the prominence of inflation forecasting in policymaking (Woodford (2003)).

Central banks aim to keep inflation stable, and perhaps also to keep output near an efficient level. With these objectives, the New Keynesian model makes explicit that optimal policy will depend on optimal forecasts (e.g., Svensson (2005)), and further that policy will be most effective when it is well understood by the general public. These results helped bolster a transparency revolution in central banking. A centerpiece of this revolution has been the practice of central banks announcing forecasts of inflation and other key variables. This practice was initiated by "inflation targeting" central banks who generally released forecasts in quarterly "Inflation Reports." The Fed (the U.S. central bank formally called the Federal Reserve System) lagged other central banks in joining this practice, but now publishes quarterly forecasts for inflation and other macroeconomic variables. The costs and benefits of transparency are widely debated, but the need for a central bank to be concerned with inflation forecasting is broadly agreed. In short, inflation forecasting is of great importance to households, businesses, and policymakers.

This chapter reviews the state of the art in inflation forecasting. A significant part of our review involves simply rounding up and summarizing the performance of the myriad forecasting models and methods that have been proposed. Now is a particularly good time for such a review, as there has been an explosion in the number and variety of methods in recent years. Along with a number of traditional time series models, a host of new methods for use with large numbers of predictors have recently come to the fore. There are also many newly popular forecast combination techniques. Financial markets have also provided new assets (indexed-linked bonds, inflation derivatives), the prices of which may contain clearer forward-looking information about inflation than was available from previously existing assets. Further, with the pathbreaking work of Smets and Wouters (2003, 2007), structural economic models have—for the first time since the 1970s—been put forward as serious forecasting models.

Finally, the creation of a number of vintage datasets in recent years makes possible quasi-realtime comparisons of proposed methods. By a quasi-realtime forecast we mean a model-based forecast for some point in time in the past based only on data that was available to forecasters at that time. These data allow us to ask whether a new method could, in principle, have been used to improve on the best forecasts that were actually made in real time.

One principle aim, then, in this chapter is to round up the traditional and newer methods and compare their strengths and weaknesses paying close attention where possible to data vintage issues. Many elements of this review are present in excellent forecasting papers in the recent literature. Few papers, however, do a comprehensive review, and the results in some cases seem to be sensitive to the particulars of the exercise. Thus, pulling things together on as consistent a basis as possible is of some value.

We go beyond mere review, however, following some themes that have emerged in several recent papers. Based on our reading of these themes, we argue that a more fundamental re-think of the topic of inflation forecasting is called for. One motivation for this re-think is that subjective inflation forecasts of inflation seem to outperform model-based forecasts in certain dimensions, often by a wide margin. This is in stark contrast to the results for, say, GDP forecasting where trivially simple time series models often perform similarly to the best subjective forecasts over any but the shortest horizons.

Thus, the chapter begins by attempting to isolate the apparent source of the advantage of subjective forecasts so that we can attempt to bring that information into the model-based forecasts. We note that the forecast path over, say, 8 quarters involves two boundary values: the path starts at a nowcast of where inflation is right now and it ends somewhere close to an estimate of what we will call the local mean inflation rate. There are good reasons reviewed below why the subjective forecasters outperform most conventional models regarding both boundary values. The advantage of the subjective methods historically seems to be largely due to the choice of boundary values.

If we accept this conclusion, it is natural to consider ways of giving the standard forecasting models the advantage of good boundary values. Given good boundaries could the models come closer to or surpass subjective forecasts? Our results here are basically negative. The methods that work well essentially choose some smooth path between the boundaries, and end up being about equivalent. One can do much worse by giving the model too much flexibility to fit sampling fluctuations. Every model constrained in a heavy-handed way performs about equally well given good boundary values. Less constrained methods often do worse, and never do appreciably better.

Overall, the big roundup of methods says that you should choose the two boundary values astutely and, given these values, effectively ignore most information that might lead you to deviate from a smooth path between the boundaries.

There is one important difference between inflation forecasting by a central bank and inflation forecasting by the public, which is worth stressing at the outset. Unlike the public, the central bank has control over monetary policy, which in turn affects inflation. So the central bank's inflation forecast depends on what they expect their own policy actions to be.

The central bank could integrate over this uncertainty—giving an unconditional forecast. Or the central bank could condition it's forecast on a particular path of interest rates—such as unchanged interest rates or a path implied by futures quotes. The Federal Reserve's Greenbook forecast and the Bank of England forecasts are both conditional forecasts. If one takes the conditioning seriously, this poses a substantial obstacle to assessing the quality of these forecasts (Faust and Wright (2008)). In this chapter, we nevertheless treat all forecasts as though they are unconditional forecasts.

The plan for the remainder of this chapter is as follows. Section 2 contains our review of forecasting methods, and our comparison of them on U.S. data. In section 3, we analyze forecasts of inflation extracted from financial market variables linked explicitly to inflation. These assets have not been traded long enough for inclusion in our broad review of forecasting methods, so we treat them separately. Section 4 discusses some other topics— the construction of inflation density forecasts, forecasting aggregates directly versus aggregating forecasts of components, and issues of forecasting core versus headline inflation. Section 5 gives a comparison of a few inflation forecasting methods for some international data. Section 6 concludes.

2 Approach for our General Review

In this section, we describe the set-up for our broad review of forecasting methods.

2.1 A triply Great Sample and its problems

As is generally the case in macroeconomics, the choice of sample period for our forecast comparison exercise is constrained by data availability. In practice, the available data do not allow us to start the forecast evaluation exercises before about 1980.[1] Thus, the feasible

[1] This is because of the forecasts that we wish to evaluate. Naturally some forecasts could be considered over much longer periods.

span for forecast evaluation is something like 1980 to the present. This includes about a 20-year period known as the Great Moderation, bookended by two short and distinctly immoderate subsamples: the end of the Great Inflation and the recent Great Recession. Thus, while all three parts of the sample are Great in some way, the sample as a whole presents some serious challenges.

It is well known that U.S. inflation was much more quiescent and harder to forecast during the Great Moderation period from, say, 1985 to 2007, than during the period before (e.g. Stock and Watson (2007)). Up until the financial crisis, it might have been fashionable to presume that the Great Moderation period would continue indefinitely, and thus focusing on the Great Moderation sample period might have seemed most relevant to the project of forecasting going forward.

Even during the Great Moderation, the "good luck or good policy" debate (e.g. Stock and Watson (2003a), Ahmed, Levin and Wilson (2004)) makes clear that the foundations for presuming that the Great Moderation would last forever were rather shaky. Of course, now we must add "bad policy" to the list of explanations for the extended buoyant period, and we have little firm basis for treating the Great Moderation sample as "the new normal."

At present, it is a question of fundamental importance whether "the new normal" will be a pattern more like the pre-Great Moderation period, the Great Moderation, or perhaps some third pattern such as the path followed by Japan during the "lost decade(s)."

One might make an argument that all three periods in our available sample should be pooled. After all, the ideal would be to have a forecasting model that performs well across the board and in all conditions. Of course, this may be an unattainable ideal. Our best macroeconomic theories are generally viewed as first order approximations around some steady-state. Similarly, our parsimonious (often linear) time series models are probably best viewed as local approximations to some more complicated process.

In this view, it would be natural, as a first goal, to choose forecasting models that perform well in something we might call "normal times," leaving periods far from the steady-state to be studied separately. While it may be hard to define "normal times," it is clear that the extreme periods that bookend our largest available sample are not normal times.

We take a pragmatic approach to this issue. We omit the unusual period of the early 1980s (which would be especially unrepresentative in including the end, but not the start, of the Great Inflation). However, we include data spanning the recent crisis, during which inflation behavior has not been as extreme (if we were modeling output growth or bank lending, our approach might be different). Thus, our baseline results for forecast evaluation are for the period 1985:Q1 to 2011:Q4. We also report separate results, excluding the recent crisis. While including the crisis raises root mean square prediction errors across the board, the basic conclusions about relative forecast accuracy that we emphasize are unaffected by inclusion or exclusion of this period.

2.2 Measures of inflation

We will focus on prediction of quarterly inflation as measured by the GDP deflator[2], the personal consumption expenditures (PCE) deflator, the CPI, and core CPI (CPI excluding food and energy). Inflation rates are computed as π_t =400 $\log(p_t/p_{t-1})$ where p_t is the underlying price index. CPI data are of course available at the monthly frequency, but our focus throughout this chapter is on quarterly data (using end-of-quarter CPI values). Figure 1 shows the evolution of these four inflation measures over the past half century. They tend to move together, but differences in composition and likely aggregation biases mean, however, that their short-run and even long-run behavior may differ.[3] Still, a slowly moving trend component—rising in the Great Inflation, and falling over subsequent decades—can

[2]GNP deflator prior to 1992.

[3]For example, CPI inflation tends to be about 0.3 percent per annum higher than PCE inflation, because the regular CPI index does not use chain weighting, and so has a well-known upward substitution bias.

clearly be seen for all four inflation measures. This slowly-varying trend is a recurrent theme of a large recent macroeconomic literature, much of which does not focus narrowly on the forecasting question. Many authors, including Sims (1993), Kozicki and Tinsley (2001, 2005), Gürkaynak, Sack and Swanson (2005), Cogley and Sargent (2005), Cogley and Sbordone (2008), de Graeve, Emiris and Wouters (2008), Cogley, Primiceri and Sargent (2010), Stock and Watson (2010), van Dijk, Koopman, van der Wel and Wright (2011), Gonzáles, Hubrich and Teräsvirta (2011), Clark (2011), Dotsey, Fujita and Stark (2011) and Wright (2012) have all emphasized the need to take account of slowly varying perceptions of long-run inflation objectives in forecasting inflation, understanding the term structure of interest rates, and/or modeling the relationship between inflation and economic slack.

In our forecast evaluations, forecast errors are calculated as actual minus forecast value, but for variables that are repeatedly and indefinitely revised with evolving definitions, an issue arises as to what to treat as the actual. Revisions to CPI and core CPI inflation are trivial; but revisions to the other inflation measures are large[4], and include benchmark revisions, which incorporate conceptual and definitional changes. It makes little sense to evaluate whether judgmental or time series models predict definitional changes, and the Greenbook, in particular, explicitly does not attempt to do so. Thus, we follow Tulip (2009) and Faust and Wright (2009) in measuring actual realized inflation by the data as recorded in the real-time dataset of the Federal Reserve Bank of Philadelphia two quarters after the quarter to which the data refer.

2.3 Metrics and inference

Our main results are for root mean square prediction errors. More specifically, we compute quasi-realtime recursive out-of-sample root mean square prediction errors (RMSPEs). Out-

[4]Croushore (2008) discusses data revisions to PCE inflation especially around 2003 when Federal Reserve officials were concerned about a very low level of inflation that subsequently got revised away.

of-sample accuracy is considered because parameter estimation error and structural breaks (as surveyed by Giacomini and Rossi (2012)) often mean that good in-sample fit fails to translate into out-of-sample forecasting performance.[5] We generally present RMSPEs relative to a benchmark (detailed later). The benchmark is in the denominator so that numbers less than one indicate that the alternative model outperforms the benchmark.

Assessing the statistical significance of any deviations in relative RMSPEs from unity raises some knotty econometric questions in the case where the forecasting models being compared are nested. Clark and McCracken (2009a, 2012) provide thorough discussions of the issues. One might think of the hypothesis as being that the smaller model is correctly specified in population. In this case, the two models are the same under the null. When viewed as a test of the null hypothesis that the small model is correct in population, the test of Diebold and Mariano (1995) has a nonstandard asymptotic distribution, and alternatives such as the test of Clark and West (2007) are often used instead, as they are close to being correctly sized.

The null hypothesis for the Clark and West (2007) test, however, is not that the two nested models have equal RMSPE in the current sample size, but rather that the small model is correctly specified and so that the two models have equal RMSPE in population. We instead prefer to think of the relevant hypothesis as being that the two models have equal *finite-sample* forecast accuracy.[6] Clark and McCracken (2012) find that comparing the Diebold and Mariano (1995) test statistic to standard normal critical values gives a test of the null hypothesis of equal finite-sample forecast accuracy (in both the nested and non-

[5]We emphasize that there is no such thing as a truly out-of-sample forecast evaluation, because the models that are considered in such an exercise are always the outcome of a data mining process conducted, if not by the individual researcher, then by the economics profession as a whole.

[6]To see the difference, consider the case where the restricted model has one highly informative predictor and the alternative model adds another 50 that add some very modest forecasting power. Owing to the effects of parameter estimation error, one would expect to find that the bigger model has substantially higher RMSPE in small sample sizes. The important distinction between population and finite-sample predictive accuracy was discussed in Inoue and Kilian (2004).

nested cases) that has size fairly close to the nominal level, provided that the standard errors use the rectangular window with lag truncation parameter equal to the forecast horizon, and the small-sample adjustment of Harvey, Leybourne, and Newbold (1997) is employed. This is the approach that we adopt throughout this chapter.

We would note two further caveats on our use of the Diebold and Mariano (1995) test. First, the results that we appeal to strictly apply to a comparison among primitive forecasting *models*. We will however be using them to compare forecast *methods* that each combine multiple models. Second, the presence of data revisions presents some potential to add to size distortions (Clark and McCracken (2009b, 2012)). However, for two of our inflation measures (CPI and core CPI), data revisions are trivial. As we find about as many rejections of the null of equal forecast accuracy for these inflation measures as for the others (PCE and GDP deflator inflation), we suspect that data revisions do not lead our tests of equal finite-sample accuracy seriously astray.

2.4 Forecasts

We focus on prediction of quarterly inflation rates made in the middle month of each quarter. We consider forecasts for the current quarter (horizon $h = 0$) and subsequent quarters. The first set of forecasts is made in February 1985; the final one is made in November 2011. For the most part (and unless explicitly noted otherwise) we perform a quasi-realtime forecasting exercise, using the vintage datasets from the database maintained by the Federal Reserve Bank of Philadelphia. We also make use of the vintage Greenbook databases used in Faust and Wright (2009). When using vintage data, our timing convention is based on the Federal Reserve Bank of Philadelphia's real-time dataset. Only data that were available in the middle of the second month of each quarter are included in forecasting. Since our forecasts are dated in the middle month of quarter, t, in all cases, the published inflation data go through the previous quarter $t - 1$.

We prefer to forecast single-quarter inflation rates rather than cumulative inflation rates over longer periods, because it makes it easier to see how the predictability varies with horizon. In contrast, inflation from the last quarter to, say, four quarters hence conflates short- and longer-term inflation predictions. We consider forecasts for the current quarter (horizon $h = 0$) and for the next 8 quarters ($h = 1, 2, \ldots, 8$).

2.5 A roundup of forecasting models

In this section, we describe the set of methods for forecasting inflation, π_t, that we shall evaluate. We consider a few models with purely stationary specifications for inflation. These models however imply, by construction, that the forecast of inflation converges to the unconditional mean as the horizon gets large. For example, long-horizon forecasts of inflation made in recent years using stationary models estimated on a period covering the Great Inflation have been over 4 percent. These seem unreasonable forecasts, and they result from ignoring the slowly-varying trend in inflation that is evident in Figure 1.

As a device to capture this varying local mean, we measure the trend level of inflation, τ_t, using the most recent five-to-ten-year-ahead inflation forecast from Blue Chip—Blue Chip has asked respondents to predict the average inflation levels from five to ten years' hence twice a year, since 1979.[7] Prior to 1979, we have no source of long-run survey inflation expectations, and so use exponential smoothing[8] of real-time inflation data instead, as a crude proxy. Then we define the inflation "gap" as $g_t = \pi_t - \tau_t$, and consider models in which g_t is treated as stationary, and for forecasting purposes, τ_t is assumed to follow a random walk. This idea of forecasting inflation in "gap" form around some slowly-varying local mean has

[7] The Blue Chip survey asks respondents for predictions of GDP deflator and CPI inflation only. We use the GDP deflator inflation projection as the trend measure for PCE inflation and use the CPI inflation projection as the trend measure for core CPI inflation.

[8] For any time series, $z(t)$, the exponentially smoothed counterpart, $z^{ES}(t)$, satisfies the recursion $z^{ES}(t) = \alpha z^{ES}(t-1) + (1-\alpha)z(t)$, where α is the smoothing parameter, set to 0.95 throughout this chapter.

been found to be quite successful (Kozicki and Tinsley (2001), Stock and Watson (2010), Cogley, Primiceri and Sargent (2010) and Clark (2011)). It controls for a low-frequency component that is evidently quite important in the behavior of inflation over the last few decades. Some analysts interpret the trend as representing agents' perceptions of the Fed's long-run inflation target, which in this view must have shifted over time, owing to changes in the Fed's preferences and also in its credibility. We also consider other non-stationary specifications for inflation and subjective forecasts. In all, we consider the following set of competing forecasting methods:

1. Direct forecast (Direct). For each horizon h, we run the regression $\pi_{t+h} = \rho_0 + \Sigma_{j=1}^{p}\rho_j\pi_{t-j} + \varepsilon_{t+h}$ and use this to obtain the forecast of $\pi_{T+h.}$

2. Recursive autoregression (RAR). We estimate $\pi_t = \rho_0 + \Sigma_{j=1}^{p}\rho_j\pi_{t-j} + \varepsilon_t$. The h-period forecast is constructed by recursively iterating the one-step forecast forward. If the AR model is correctly specified, then the AR forecast will asymptotically outperform the direct benchmark, but the direct forecast may be more robust to misspecification, as discussed by Marcellino, Stock and Watson (2006). Like the direct autoregression, this does not impose a unit root on the inflation process.

3. A Phillips-curve-motivated forecast (PC). The Phillips curve is the canonical economically motivated approach to forecast inflation (Phillips (1958), Gordon (1980, 1998), Brayton, Roberts and Williams (1999) Stock and Watson (1999, 2009) and many others). For each h, we estimate $\pi_{t+h} = \rho_0 + \Sigma_{j=1}^{p}\rho_j\pi_{t-j} + \lambda u_{t-1} + \varepsilon_t$, where u_{t-1} is the unemployment rate in quarter $t-1$, and use this to forecast π_{T+h}. Phillips curve forecasts are sometimes interpreted more broadly, replacing the unemployment rate with other economic activity measures, such as the output gap, industrial production growth, or marginal cost.

4. A random walk model. We consider two variants on this. The pure random walk model (RW) takes π_{T-1} as the forecast for π_{T+h}, $h = 0, \ldots, 5$. The closely-related forecast for

inflation considered by Atkeson and Ohanian (2001) (RW-AO) instead takes $\frac{1}{4}\Sigma_{j=1}^{4}\pi_{T-j}$ as the forecast for π_{T+h}.

5. An unobserved component stochastic volatility model (UCSV). The model is univariate: $\pi_t = \tau_t + \eta_t^T$ and $\tau_t = \tau_{t-1} + \eta_t^P$ where η_t^T is $iidN(0, \sigma_{T,t}^2)$, η_t^P is $iidN(0, \sigma_{P,t}^2)$, $\log(\sigma_{T,t}^2) = \log(\sigma_{T,t-1}^2) + \psi_{1,t}$, $\log(\sigma_{P,t}^2) = \log(\sigma_{P,t-1}^2) + \psi_{2,t}$ and $(\psi_{1,t}, \psi_{2,t})'$ is $iidN(0, I_2)$. The forecast of π_{T+h} is the filtered estimate of τ_T. If the variances of η_t^T and η_t^P were constant, then this would be an integrated moving average (IMA) model. Stock and Watson (2007) find that the UCSV model provides good forecasts for inflation.

6. The autoregression in gap form (AR-GAP). For each horizon h, we estimate the regression $g_{t+h} = \rho_0 + \Sigma_{j=1}^{p}\rho_j g_{t-j} + \varepsilon_{t+h}$. We then iterate this forward to provide a forecast of g_{T+h} and add τ_T back to the forecast to get the implied prediction of inflation, treating the trend as a random walk. Henceforth, all the time series predictions that we consider are in "gap" form—they yield a forecast of g_{T+h}, to which we add back the final observation on the trend to get the implied prediction of inflation.

7. The "fixed ρ" forecast. We assume that the inflation gap is an AR(1) with a fixed slope coefficient, ρ, which set to 0.46. This is in turn the slope coefficient from fitting an AR(1) to the 1985Q1 vintage of GDP deflator inflation from 1947Q2 to 1959Q4. Thus, the model is $g_t = \rho g_{t-1} + \varepsilon_t$, and absolutely no parameter estimation is involved. This can in turn be used to obtain a forecast of g_{T+h} and, adding τ_T back to the forecast, gives the implied prediction of inflation.

8. A Phillips curve forecast in "gap" form (PC-GAP). We apply the Phillips curve not to inflation, but to the inflation gap, g_t. For each h, we estimate $g_{t+h} = \rho_0 + \Sigma_{j=1}^{p}\rho_j g_{t-j} + \lambda u_{t-1} + \varepsilon_t$, where u_{t-1} is the unemployment rate in quarter $t-1$, and use this to forecast g_{T+h} and hence π_{T+h}. Phillips curves applied to the inflation gap have been considered by

Stock and Watson (2010) and Koenig and Atkinson (2012).

9. A Phillips curve forecast in "gap" form with a time-varying NAIRU (PCTVN-GAP). For each h, we estimate $g_{t+h} = \rho_0 + \Sigma_{j=1}^{p} \rho_j g_{t-j} + \lambda(u_{t-1} - u_{t-1}^*) + \varepsilon_t$, where u_t^* is an estimate of the NAIRU. We use the most recent five-to-ten-year-ahead Blue Chip survey forecast for the unemployment rate as the estimate of the time-varying NAIRU. This goes back to 1979—before this we use exponential smoothing of the real-time realized unemployment rate instead.

10. A term structure VAR based forecast (Term Structure VAR). Macro finance models aim to characterize the joint dynamics of Treasury yields and macroeconomic variables (Ang and Piazzesi (2003), Diebold, Rudebusch and Aruoba (2006)). A simple way of operationalizing this (following Diebold and Li (2006)) is to fit a Nelson-Siegel yield curve to the term structure of interest rates at the end of each quarter, specifying that the yield on a zero-coupon bond of maturity n is

$$y_t(n) = \beta_{1t} + \beta_{2t}(\frac{1 - e^{-\lambda n}}{\lambda n}) + \beta_{3t}(\frac{1 - e^{-\lambda n}}{\lambda n} - e^{-\lambda n}) \tag{1}$$

where λ is treated as fixed at 0.0609. The coefficients β_{1t}, β_{2t} and β_{3t} have interpretations as the level, slope and curvature of yields. The underlying zero-coupon yields for this exercise are from the dataset of Gürkaynak, Sack and Wright (2007). We can then fit a VAR(1) to β_{1t}, β_{2t}, β_{3t}, the inflation gap, and the unemployment rate. Vector autoregressions of this sort are familiar in the macro-finance term structure literature (see, for example, Joslin, Priebsch and Singleton (2010)).[9]

11. A forecast based on VAR with time-varying parameters (TVP-VAR). This is a VAR(2) in inflation, the unemployment rate and Treasury bill yields in which the intercept and slope

[9] Some authors impose no-arbitrage restrictions rather than estimating an unrestricted VAR, as we do here. Joslin, Le and Singleton (2012) however argue that the imposition of these no-arbitrage restrictions is empirically inconsequential.

coefficients are allowed to drift slowly over time, as in Primiceri (2005). The parameters follow random walks with stochastic volatility. This can be thought of as a multivariate generalization of the UCSV model.[10]

12. Equal-weighted averaging (EWA). This and the next two methods assessed are *large dataset methods*. We constructed a dataset of 77 predictors at the quarterly frequency, listed in Table 1. All of the series are available from 1960Q1 through to the end of the sample, and as such constitute a balanced panel. As is usual, the series were transformed such that the transformed series (levels, logs, log differences etc.) are arguably stationary. Unfortunately, unlike the rest of our forecasting exercise, these data are *not* real-time. Instead, a single recent vintage of data is used for the large dataset methods.[11] However, real-time forecasting exercises with large datasets have been considered (Bernanke and Boivin (2003), Faust and Wright (2009)), and those studies found that the relative performance of large dataset and simpler forecasting methods is not greatly affected by whether one uses real-time data or a single vintage of revised data. We first estimate and forecast using n simple models, each of the form $g_{t+h} = \rho_0 + \Sigma_{j=1}^{p} \rho_j g_{t-j} + \beta_i x_{i,t-1} + \varepsilon_{it+h}$ for $i = 1, \ldots n$ where $x_{i,t}$ is the value of the ith predictor in the large dataset at time t. Letting $\hat{g}^i_{T+h}$ be the forecast of g_{T+h} from the ith model, the EWA forecast of the inflation gap is $n^{-1}\Sigma_{i=1}^{n} \hat{g}^i_{T+h}$. This method was first proposed by Bates and Granger (1969) and its surprising empirical success is part of the folklore of forecasting. Stock and Watson (2003b) among others find continuing support for the folklore.

---

[10] Cogley and Sbordone (2008) estimate the parameters of a structural New-Keynesian Phillips curve by matching the coefficients of a reduced form VAR with time-varying parameters. Inflation forecasts from the New-Keynesian Phillips curve model are thus designed to be close to those from the VAR with time-varying parameters.

[11] The problem could be dealt with by using the vintage datasets of Faust and Wright (2009), but those datasets come from the Greenbook process and the forecast period would have to end outside the five-year embargo for those forecasts.

13. Bayesian model averaging (BMA). In this method, described in more detail by Wright (2009a), we assign a prior over the parameters of the n models used in EWA, just described; and a flat prior that each model is equally likely to be true. The prior for the model parameters follows Fernandez, Ley and Steel (2001). Write each model as $g_{t+h} = \lambda_i' w_{i,t} + \varepsilon_{it}$, where $\varepsilon_{it} \sim N(0, \sigma^2)$, let the prior for λ_i conditional on σ be $N(\bar{\lambda}, \phi(\sigma^2 \Sigma_{t=1}^T w_{i,t} w_{i,t}')^{-1})$ and the marginal prior for σ be proportional to $1/\sigma$. The models are then estimated and the forecast from each is evaluated at the posterior mean for the parameters. Finally, these n forecasts are then combined in a weighted average with weights determined by the posterior probability that each model is correct. The prior has a hyperparameter, ϕ, that determines how much the model weights are likely to vary from equal weighting. We set $\phi = 2$.

The theoretical justification of this method relies on strictly exogenous regressors and iid errors—assumptions that are patently false in our application. Earlier work (Koop and Potter (2003) and Wright (2009a)) shows that the method works well in cases like the one at hand, however, and we simply view BMA as a pragmatic shrinkage device.

14. Factor augmented vector autoregression (FAV). This uses the VAR $\xi_t = \mu_0 + \Sigma_{j=1}^p \mu_j \xi_{t-j} + \varepsilon_t$, where $\xi_t = (g_t, z_{1t}, z_{2t}, ...z_{mt})'$ and $\{z_{it}\}_{i=1}^m$ are the first m principal components of $\{x_{it}\}_{i=1}^n$, with the predictors first standardized to have mean zero and unit variance. The model can be estimated and iterated forward to provide a forecast of g_{T+h}. This method was proposed by Bernanke, Boivin and Eliasz (2005).

15. The Dynamic Stochastic General Equilibrium (DSGE) model of Smets and Wouters. Our choice of the Smets-Wouters (2007) model is due to its iconic nature, the existence of a body of prior results and due to the pragmatic fact that compiling a real-time dataset for the more elaborate models has not been done and would be very expensive. There are, of course, many versions of DSGE models and it might be nice to include more recent models and even the larger models in use at the Federal Reserve Board.

We base our DSGE forecasts on the real-time exercise conducted by Edge and Gürkaynak (2010). They created, and graciously made available to all, vintage datasets for the Smets-Wouters model for the period 1992 to 2009. We augment these vintage datasets with data from the vintage Greenbook databases used in Faust and Wright (2009), allowing us to extend the exercise back to 1985:Q1 as with our other models. The Smets-Wouters model gives forecasts for GDP deflator inflation alone.

In the baseline case, we follow Edge and Gürkaynak (2010) in using the Bayesian prior specified by Smets and Wouters to re-estimate the model by Markov Chain Monte Carlo for each vintage of the data. We take as forecasts the mean of the predictive density for inflation taking the full posterior distribution of the model parameters into account. We also investigated forming forecasts treating the parameters as fixed at the posterior mode, but found that this gave similar results.

The Smets-Wouters model is a stationary specification. However, the prior mean for the steady-state inflation rate is 2.5 percent, and so long-horizon forecasts for inflation from this model do not necessarily have to be close to the sample mean at the time that the forecast is being made (unlike for methods 1-3 above).

16. The Dynamic Stochastic General Equilibrium with shifting local mean (DSGE-GAP). It is hard to evaluate a model estimated by Bayesian methods on a quasi-realtime basis, as the choice of priors was inevitably influenced by the data observed at the time that the model is first proposed. As a crude device to mitigate this potential for "lookback" bias in forecasting with the Smets-Wouters model, we also consider a modification of this forecasting method in which the prior mean for the steady-state of inflation is set to our real-time measure of the local mean of inflation, τ_t.

17. Finally, we consider three fully real-time judgmental forecasts, each of which incorporates an immense range of information processed through an economics-influenced subjective filter:

(i) Blue Chip survey (BC). Blue Chip provides a forecast of both the GDP deflator and CPI. The AR-GAP benchmark forecast is affected by long-term Blue Chip survey predictions, but at least for the horizons for which they are available, one might just want to use the Blue Chip projections directly. Blue Chip forecasts are released at the start of each month. For each quarter, we take the second Blue Chip forecast, which is always released before the time at which our forecasts are being made (the middle of the middle month of the quarter).
(ii) Survey of Professional Forecasters (SPF). The SPF also provides quarterly GDP deflator and CPI forecasts. These are released at the start of the middle month of each quarter, again just before the time at which our forecasts are being made.
(iii) The Fed staff's Greenbook forecast. The Greenbook provides GDP deflator, CPI and CPI-Core forecasts. While the Greenbook forecast is informed by myriad small-scale and large-scale models, it is ultimately a judgmental forecast (Reifschneider, Stockton and Wilcox (1997)). The forecast is made once per FOMC meeting. To align these Greenbook forecasts with the rest of our exercise, we simply choose the Greenbook closest to, but before, the middle of the middle month of the quarter. The Greenbook forecast is conditioned on a particular path for the policy interest rate over the forecast horizon—which is not meant to be a prediction of that policy rate. That makes it a conditional forecast, as noted in the introduction. However, in evaluating the forecasts, we shall assess the Greenbook in exactly the same way as all the other forecasts, neglecting the effect of this conditioning.
The available evidence indicates that all these judgmental forecasts do remarkably well, generally dominating model-based forecasts (Ang, Bekaert and Wei (2007), Faust and Wright (2009), Croushore (2010))— even when the models are chosen *ex post* in light of known behavior of inflation in the forecast period.

In most proposed models, p lags of inflation (or the inflation gap) are included on the right-hand-side. We select p using the Bayes Information Criterion in the AR-GAP model, and use the *same* number of lags in all the other models that have p lags of inflation or the inflation gap (methods 1, 2, 3, 6, 8, 9, 12, 13 and 14 as listed above).[12]

The above set of models is a fairly comprehensive list of the sort of models that have appeared in the literature. But it is of course far from exhaustive. Other methods that have been proposed include threshold models (Dotsey, Fujita and Stark (2011)), LASSO methods that do prediction and variable selection jointly (Bai and Ng (2008)), bootstrap aggregation (Inoue and Kilian (2008)) and household survey expectations (Inoue, Kilian and Kiraz (2009)). There is moreover one additional natural family of forecasting approaches and that involves some sort of direct extraction of an inflation forecast from inflation-related financial market variables. These are treated separately in section 3 due to the limited available sample.

2.6 Results of the forecast comparison exercise

We can then evaluate the competing forecasts in terms of their real-time recursive out-of-sample RMSPEs. The results for all four inflation measures are shown in Table 2. All RMSPEs are reported relative to the benchmark of the "fixed ρ" forecast (an AR(1) in gap form with a fixed slope coefficient). A relative RMSPE below 1 means that the forecast is doing better than the benchmark. We pick this benchmark because it is very simple, yet is still amazingly hard to beat by much. We also assess the statistical significance of the deviations in relative RMSPEs from unity, using the test of Diebold and Mariano (1995), as discussed earlier.

The entries in Table 2 are mostly above 1, indicating that the AR in gap form with a fixed slope coefficient gives better out-of-sample forecasts than most alternatives.

Simple time series methods that treat inflation as a stationary process—the direct fore-

[12] There is some evidence that BIC does best for direct forecasts while AIC does best for iterated forecasts (Inoue and Kilian (2006), Marcellino, Stock and Watson (2006)).

cast, the autoregression and the Phillips Curve—do especially badly relative to the benchmark. The forecasts in gap form fare better. Still, most of the time, the "fixed ρ" benchmark does a bit better than the other forecasts in gap form, including the Phillips curve gap forecasts (PC-GAP and PCTVN-GAP). There are some cases in which other alternatives beat the benchmark, including the term structure VAR, or model averaging methods (equal-weighted or BMA), but the improvements are not great (at the very best about a 10 percent reduction in RMSPE), nor are these gains consistent across inflation measures or forecast horizons. Within the model averaging methods, Bayesian model averaging has a slight edge over equal-weighted model averaging in most, but not all, cases. Meanwhile, the forecasting performance of the factor augmented VAR is less good; the forecasting performance of factor-based models for inflation (and growth) seems to be fragile and dependent on the precise variables used in the large dataset (Faust and Wright (2009)). The Atkeson-Ohanian version of the random walk forecast and the UCSV and TVP-VAR forecasts, which are all nonstationary, generally do reasonably well, with performance comparable to the benchmark.

The DSGE model provides some of the better forecasts for GDP deflator inflation in Table 2. The observation that DSGE models that are competitive with alternatives has been made by Smets and Wouters (2003, 2007), Edge and Gürkaynak (2010), Edge, Kiley and Laforte (2010) and Kwon (2011) and this observation that DSGE models (incorporating certain frictions) can provide forecasts with reasonable accuracy has been greatly enhanced their appeal to central banks around the world. Still, even these forecasts have RMSPE that is higher than the simple "fixed ρ" benchmark at all horizons. The DSGE-GAP model does a bit better; it's forecast accuracy is roughly on a par with the simple benchmark.

2.7 Four Principles

We see four key principles emerging from our forecast comparison exercise:

2.7.1 Subjective forecasts do best

The very best forecasts in Table 2 are the subjective ones: Blue Chip, SPF and Greenbook. Indeed, these are the only forecasts that consistently significantly improve on our simple benchmark. The fact that purely subjective forecasts are in effect the frontier of our ability to forecast inflation has been found by a number of recent papers (Ang, Bekaert and Wei (2007), Faust and Wright (2009), Croushore (2010)). Perhaps it should not be too surprising—private sector and Fed forecasters have access to econometric models, but add expert judgment to these models. Relative to the benchmark, the subjective forecasts give reductions in RMSPE of up to 25 percent. This means that they are doing far better than direct, RAR and PC forecasts for inflation. Within the set of subjective forecasts, the Greenbook seems to have a small edge (consistent with Romer and Romer (2000)), although this result is known to be somewhat dependent on the sample period. Note that in this exercise, the Blue Chip and SPF forecasts span the financial crisis and the accompanying severe recession, while the Greenbook forecasts end in 2006. Atkeson and Ohanian (2001) found that the random walk forecast did better than the Greenbook in a particular sample with one observation per year from 1983 to 1995, but that result appears to be special to the precise sample period (Faust and Wright (2009)).

Thus our first principle is that purely judgmental forecasts of inflation are right at the frontier of our forecasting ability. This in turn has substantive implications for empirical work beyond just the narrow question of forecast accuracy. It suggests that a useful way of assessing models is by their ability to match survey measures of inflation expectations (e.g. Del Negro and Eusepi (2011)). And it implies that in estimating forward-looking macroeconomic models, it may be better to treat survey forecasts as direct measures of expected future inflation (e.g. Adam and Padula (2003)), instead of the more commonly-used method of replacing expected inflation with future realized values and then using instrumental variables

(Gali and Gertler (1999))[13].

The next three principles shed some additional light on the advantage of the subjective forecasts and on whether further modeling work can erode this advantage.

2.7.2 Good forecasts must account for a slowly varying local mean

All of the models that perform reasonably well have some method for taking account of a slowly evolving local mean for inflation. The models based on stationary specifications for inflation do consistently less well than models in gap form and three of the nonstationary models are among the best performers. During recent years, the stationary models have generated unreasonably high forecasts of inflation at longer horizons because inflation has been persistently below the full-sample average.

The evident desirability of using models that account for a slowly-varying trend naturally raises the question of the appropriate measurement of the trend component of inflation, τ_t. We have adopted the approach of using long-run survey expectations (Clark (2011), Kozicki and Tinsley (2012), Wright (2012)); but one might also measure the trend by exponential smoothing, or by measuring the permanent component from the UCSV model (following Stock and Watson (2010)). Any of these will capture the low-frequency shifts in inflation. Figure 2 plots three different real-time measures of the long-run trend inflation: the five-to-ten-year-ahead Blue Chip survey forecast, real-time exponentially smoothed inflation and the permanent component from the UCSV model. All share the same trend, but the survey forecasts declined more rapidly in the late 1980s and early 1990s than exponentially smoothed inflation. This suggests that in this case the subjective forecasters were quicker to realize the ongoing disinflation than one could have divined from the other methods, which essentially extrapolate recent inflation.[14]

[13] See Kleibergen and Mavroeidis (2009) and Rudd and Whelan (2007) for discussion of the econometric problems that can arise with the instrumental variables approach.

[14] Also, the trend from the UCSV model is a good bit less smooth than either the long-range Blue Chip

We re-did all our gap forecasts using exponential smoothing to measure trend inflation in place of the long-run survey value, but found that this typically made the inflation forecasts less accurate. This brief exploration by no means demonstrates that subjective forecasts of the local mean inherently dominate their model-based counterparts. However, if these low-frequency shifts mainly involve fundamental change in the economy or policymaking environment, there are reasons to believe that parsimonious econometric models based on standard determinants of inflation may have difficulty capturing all the relevant information. For example, a change in the long-run inflation target of the central bank might well be announced, but at the time of announcement this information would not be in the conventional determinants of inflation. This is one interpretation of the superior performance of the survey-based trend measure during the disinflation from the early 1980s peak.

Our second principle is that every inflation forecast for horizons longer than one or two quarters hence should involve some mechanism for capturing low-frequency local mean dynamics. Long-horizon survey forecasts seem to represent a good way of doing this.

2.7.3 The nowcast

For horizon zero, the nowcast for all three subjective methods have RMSPEs around 20 percent smaller than the very best of the model-based forecasts (which already use the subjective information on the local mean from long-horizon survey forecasts). Where does the nowcasting advantage come from? One thing is clear from our experience observing the nowcasting process at the Fed and from our discussions with professional forecasters: nowcasting and backcasting in practice involve a very different process from forward-looking forecasting. This is because the relevant information set regarding the current quarter is generally much different from that for future quarters. Often a large portion of the source data for inflation are available before the official data are released; thus, a nowcasting exercise can in part

survey forecast or exponentially smoothed inflation.

involve replicating the data construction agency. Further, often there will have been special events such as hurricanes, dock strikes, special sales programs by auto manufacturers, etc., in the current quarter that are known and that can be directly accounted for in some way in making the nowcast. A specific recent example is that when Vancouver hosted the Winter Olympics in 2010, there was an enormous spike in lodging prices in Vancouver, and the Bank of Canada took this special information into account in its nowcast for Canada-wide inflation.[15]

We do not intend to take a strong position on whether the nowcasting embedded in subjective forecasts could be rendered entirely systematic so that we could remove the word "subjective." Perhaps some of the recently proposed mixed-frequency nowcasting models that have been proposed (e.g. Giannone, Reichlin and Small (2008), Andreou, Ghysels and Kourtellos (2008), Banbura, Giannone, Modugno and Reichlin (2012)) could achieve this, or at least get close. On the other hand, we suspect that an econometric model rich enough to systematically take account of all possible special factors that might be present in a given quarter would be unwieldy and not cost-effective.

Part of the benefit of Greenbook and other judgmental forecasts at longer horizons flows from their advantage in measuring the current state of the economy (as suggested by Sims (2002), Giannone, Reichlin and Small (2008), Faust and Wright (2009)). Fortunately, as Faust and Wright (2009) argue, we can easily give any model forecast the benefit of a good nowcast by simply augmenting the predictor database for each model with a high quality nowcast—bring all predictor data series up to the current quarter based on a good nowcast and then compare all the models' forecasting ability, having given all an equal footing on the nowcast.

Our third principle is that good forecasts begin with high quality nowcasts.

[15] We are grateful to Sharon Kozicki for this example.

2.7.4 Heavy shrinkage in the use of information improves inflation forecasts

Table 2 includes the relative RMSPEs for GDP deflator and CPI inflation using an AR(1) model in gap form with a fixed slope coefficient (the "fixed ρ" method), but using the Blue Chip current-quarter nowcast as a jumping-off point. This is a trivial way to take account of the importance of the second and third principles. We use the Blue Chip projections for nowcasting, because they are available at the highest frequency, and are thus least likely to be stale, and also because they are available for the entire forecast period, whereas the Greenbook is not available for the last five years. Unfortunately, we do not have a Blue Chip nowcast for PCE deflator or core CPI inflation. So the "fixed ρ + nowcast" method is available only for GDP deflator and CPI inflation.

We can view this forecast path as having two boundaries—the forecast starts at the Blue Chip nowcast and converges at long horizons to the local mean estimated from Blue Chip. Our first two principles determine the boundary conditions, and the forecast path simply involves exponential decay from one boundary toward the other. This forecast turns out to do very well. At horizon 0 it is (by construction) the same as the Blue Chip forecast, which has excellent accuracy. But by getting the jumping-off point right, it also improves forecast accuracy at horizons 1 and 2, at least for GDP deflator inflation.

Indeed any of forecasting methods 1-10 above can be implemented using the current-quarter jumping-off point, based on Blue Chip nowcasts.[16] The DSGE forecasts can also be implemented using the current-quarter jumping-off point.[17] Table 3 reports the relative RMSPEs of these forecasts, all relative to the benchmark of the "fixed ρ" method starting

[16]We do not have nowcasts for all the predictors in the large dataset in this chapter, so it is not possible to give the large dataset methods the advantage of a good nowcast. This problem could be dealt with by using the vintage datasets used in Faust and Wright (2009), but those datasets come from the Greenbook process and the forecast period would have to end outside the 5 year embargo for Greenbook forecasts.

[17]We have nowcasts only for some of the variables that are used in the DSGE model. What we do is to use the DSGE model to construct current-quarter forecasts, and replace these with nowcasts, where those are available, before constructing forecasts at longer horizons.

from the current-quarter Blue Chip nowcast.

All of the models in gap form considered in Table 3 are effectively glide paths from the nowcast to the local mean. But the benchmark is different from the other forecasts in that it has only one parameter, and that parameter is fixed independently of the data used in the remainder of the forecasting exercise. Yet all the models in gap form, even given the nowcast, struggle to do better than this benchmark. Some alternative model-based forecasts do a bit better in some cases, but the improvements are small and inconsistent.[18] In other words, we simply ask the Blue Chip survey "Where are we now?" and "Where will we be in 5-10 years?" This is a very draconian form of shrinkage that uses no data and no subjective views to directly inform the answer to "where will be in 1 quarter, or 2, or . . . , 8?" Yet, the resulting forecast is close to the frontier of predictive performance.

Note that the UCSV forecast also does quite well relative to the alternatives, especially when it is given a good nowcast. Remember that the UCSV model is a univariate model involving only inflation. The UCSV forecast essentially involves taking current inflation, filtering out the bit that is taken to be purely transitory, and taking whatever is left as the forecast of inflation for all horizons. Viewed from the standpoint of the boundary conditions discussion, the forecast path for the UCSV model with nowcast starts with the Blue Chip nowcast and then for all other horizons jumps immediately to its estimate of the local mean. Thus, like the benchmark model, the UCSV makes no attempt to exploit *any* information regarding the path between the boundary conditions. The Atkeson-Ohanian random walk forecast can be thought of as doing a similar filtering exercise.

Our point is that these models seem to have a useful approach to measuring the local

[18] Table 3 also reports results for the subjective forecasts. The Blue Chip, SPF and Greenbook forecasts do better than the benchmark for GDP deflator inflation at one- to four-quarter-ahead horizons, by anything from 6 to 23 percent. Thus, for GDP deflator inflation, the benefit of subjective forecasts is not solely the result of getting the boundaries right. However, the benchmark and subjective forecasts have almost identical accuracy for CPI inflation.

mean and nowcasting, but they essentially eschew any other ambitions. One might have hoped that they would provide a starting point for finding even better models that make more savvy use of the myriad available data series in order to forecast how inflation will evolve between the boundary conditions, but Table 3 shows that this is broadly incorrect. Once one has a good estimate of the local mean and a good nowcast, it is difficult to find any way to constructively use additional information from a model.

Thus our fourth principle is that heavy-handedness helps (H^3). Draconian restrictions (or shrinkage, or use of very informative priors) are generally required to get models close to the frontier of inflation forecasting performance.

At first sight, it might seem surprising or even worrisome that we can do little better in forecasting inflation than choosing a fixed glide path that moves quite quickly from an initial condition toward the local mean. But, as observed by Edge and Gürkaynak (2010), if monetary policy is mainly directed toward offsetting deviations in inflation from a slowly-moving target, then deviations in inflation from that target should short-lived.

We now turn to some additional topics relating to our forecast comparison exercise.

2.8 Inflation forecasts and the financial crisis

Inflation was volatile during the recent financial crisis and its aftermath (as can be seen in Figure 1). As discussed earlier, this poses a dilemma of whether or not to include this period in a forecast evaluation exercise. In Table 4 we report the RMSPEs of all the forecasts for GDP deflator inflation, relative to the benchmark of a fixed ρ forecast in gap form, over a pre-crisis sample. Table 4 is therefore just like Table 2, except that only forecasts that were made for quarters 2007Q3 and earlier are included. As it turns out, perhaps surprisingly, the relative RMSPEs in the pre-crisis sample (Table 4) are very similar to those in the full sample (Table 2). Thus inclusion of the financial crisis does not materially change our conclusions as to the relative average accuracy of competing forecast methods for GDP deflator inflation.

The same is true for the other three inflation measures (pre-crisis results for these are not shown, to conserve space). It is also true for methods using a nowcast.

2.9 How different are the better model forecasts?

Table 5 shows the correlations of 9 good forecasts for GDP deflator inflation, at the four-quarter-ahead horizon. These forecasts are all roughly comparable in their *average* predictive performance. It might be that they are quite different forecasts and just so happen to give the same average forecast accuracy. Or, their comparable performance might owe to them being roughly the same forecast. Which it is matters a great deal—to the extent that they are fundamentally different forecasts, we might want to find some way of combining them to produce a more accurate prediction, or might be interested in characterizing the circumstances under which one forecast does particularly well. Consistent with Sims (2002), Table 5 shows that most of the forecasts are highly correlated with each other.

The exceptions to this result of high correlation are the DSGE forecast (which has little correlation with other alternatives over the full sample period) and the DSGE-GAP forecast (which has moderate correlation with other alternatives). Figure 3 plots the DSGE, DSGE-GAP and AR-GAP four-quarter-ahead forecasts for GDP deflator inflation, along with the subsequent actual realized values, shifted back four quarters, so that the forecasts and actuals would coincide if the forecast were perfect. The DSGE forecast bounces around 2.5 percent for the whole sample, which is the prior mean for steady-state inflation in this model. Meanwhile, the AR-GAP forecasts share the downward trend of long-run Blue Chip survey forecasts. Thus, the AR-GAP and DSGE forecasts have different trending behavior, and that's why they are effectively uncorrelated with each other in the full sample. The DSGE-GAP forecast, which has a prior mean centered around the most recent long-run Blue Chip survey forecasts, not surprisingly picks up some of the downward trend in these survey projections. It has consequently a higher correlation with the AR-GAP and other forecasts. Indeed, from the mid 1990s to the financial crisis, the DSGE, DSGE-GAP and AR-GAP forecasts do not just have comparable RMSPEs—they are strikingly similar forecasts.

The high correlation between forecasts means that while there could be some scope for improving predictive accuracy by forecast combination, it cannot be the magic bullet. Aiolfi, Capistran and Timmermann (2010) consider combinations of various survey and time series forecasts. They find that there can be some gains from such combinations, but survey forecasts are still hard to beat. To illustrate this point, we consider forecasts that put weight λ on the Blue Chip prediction and $1 - \lambda$ on some other forecast. Figure 4 plots the relative RMSPEs of these survey+time series forecasts against the weight λ (for GDP deflator inflation at various forecast horizons). In most cases, the optimal weight on the survey forecast turns out to be 1. The only exception to this is that at some horizons, even though the Blue Chip forecast is more accurate than the DSGE-GAP model, a combination can do a little better than even the Blue Chip projection.

2.10 A comment on the DSGE model forecasts

A long period of DSGE model refinement using the postwar sample preceded Smets and Wouters finding a particular DSGE model with forecasting performance on a par with standard benchmarks. This was a remarkable and important achievement. The results of this chapter suggest, however, that there is still more to understand about the DSGE model forecast accuracy. From the standpoint of the economics profession we may hope that the solid performance is due to the fact that the model correctly captures important economic relations. From a skeptical scientific perspective, however, this remains far from clear.

The results earlier in this section give us reason to believe that part of the advantage of the DSGE model stems from use of a prior for steady-state inflation specified in light of the full estimation sample—a prior that probably was at odds with what most agents actually expected at the time. More broadly, it's very hard to imagine a fully real-time assessment of the Smets-Wouters model. Unlike a simple time series model, it involves many specification decisions and choices about the prior. We would like to know how the model would have fared if *all* these choices had been made in real-time. But in the model that we have, these choices—with the possible exception of the steady-state inflation prior—are made in light of the full estimation sample.[19]

Further, a main result in our study is that very heavy-handed restrictions improve forecast accuracy over less restricted models. In short, we need models that keep the forecast from following the sample too closely. In our view it is an open question whether the DSGE model performance stems not from being "right" in some meaningful economic sense, but simply from being heavy-handed. Cynically, one can point out that fitting poorly is one way to avoid overfitting. More generously, the model may amount to a way to choose an arbitrary glide path between two end points. Any such glide path does pretty well, but doing well thereby need not be seen as much support for any particular economic story.

2.11 The Phillips Curve

Our results for the Phillips curve models are not very supportive, consistent with the available evidence on the forecasting performance of the Phillips curve, which is mixed and sensitive to the sample period (Brayton, Roberts and Williams (1999), Liu and Rudebusch (2010), Stock and Watson (2009)). There is plentiful evidence of a tradeoff between slack and inflation (e.g. Stock and Watson (2010) and Meier (2010)). Another example is that at the time of writing the euro-zone countries with the highest unemployment are exhibiting the most disinflation. Still, none of this translates into clear and consistent improvements in out-of-sample forecast accuracy. In our experience, defenders of the Phillips curve models often

[19] Tetlow and Ironside (2007) is a useful reminder of how revisions to model specification can be at least as important as revisions to data.

react to negative results by arguing that nonlinearities may render the model more useful for forecasting in some periods than in others.

Standard theory gives good reasons for possible nonlinearity in the Phillips curve, so these protests should be taken seriously. Filardo (1998) and Barnes and Olivei (2003) have posited a nonlinearity in the Phillips curve whereby it should be most useful for forecasting inflation when unemployment is cyclically high. Stock and Watson (2010) and Dotsey, Fujita and Stark (2011) find some support for this view. Ball, Mankiw and Romer (1988) have considered versions of the Phillips curve in which it flattens at low levels of inflation. That would suggest that Phillips curve should give larger improvements in forecast inflation at times when the inflation rate is relatively high.

To examine the forecasting performance of the Phillips curve over time in more detail, Figure 5 plots the PC-GAP and AR-GAP four-quarter-ahead forecasts for GDP deflator inflation, along with the subsequent actual realized values, again shifted back four quarters, so that the forecasts and actuals would coincide if the forecast were perfect. Indeed, it can be seen that there were periods in which inclusion of the AR-GAP turned out to be closer to the mark than the PC-GAP forecast. This is true in the mid to late 1990s, when the low level of unemployment led the PC-GAP forecast to project a rise in inflation that never occurred. It is also true at the end of the sample, where inflation came in higher than would be predicted by the PC-GAP forecast (see Ball and Mazumder (2011)). On the other hand, there are episodes where the PC-GAP forecast did well, including around the business cycle peak in 1990 and around the year 2000.

2.12 Conditional forecast comparisons

Claims that a particular forecast (such as the Phillips curve) may work better at some times than at others motivate us to consider conditional forecast comparisons. In the comparison between any two forecasts, one might ask the question of whether one forecast is better *conditional* on some other variable. In principle two forecasts could be equally good on average, but one forecast could be better than the other under certain conditions (Giacomini and White (2006)). The question can be posed in two different ways. We might ask if one forecast is better than another conditional on some variable that was known at the time that the forecast was made. Of course, if one forecast is better than another only conditional on some other observed variable, then this must indicate that both forecasting models are in some way misspecified and that some kind of nonlinearity is called for, but implementing this may be too ambitious. Alternatively, we might ask if one forecast is better than another conditional on some variable that is not yet known when the forecast is made. For example, we might ask if one forecast is better than another if the forecast turns out to refer to a period during a recession. Although this does not give a direct strategy for improving forecast accuracy, the question might still be of interest if the user of the forecast has a loss function that penalizes forecast misses at some times more than at others. For example, the central bank may be particularly averse to large forecast errors if they happen during recessions.

We consider both questions. Table 2 showed that over the last quarter century, it was hard to do much better than the "fixed ρ" forecast. But we can also compare the out-of-sample relative RMSPEs of the forecasts *conditional* on the forecast being made at a time of cyclically high inflation or cyclically high unemployment.

We define a period of cyclically high inflation as being one where the average of inflation over the previous four quarters is above the trend. Following Stock and Watson, we define the unemployment gap as the difference between the real-time unemployment rate and the real-time average unemployment rate over the previous 12 quarters. We then define a period of cyclically high unemployment as one in which this unemployment gap exceeds 0.5 percent. One might think of this as a proxy for NBER recession quarters, but it is a fully real-time measure, whereas the NBER peaks and troughs are called only well after the fact. Also, this criterion is a bit less stringent than the NBER recession dates. Over our sample, there are three periods of elevated unemployment using this definition: from 1990Q4 to 1993Q1, from 2001Q4 to 2003Q4 and from 2008Q3 to 2011Q1.

Tables 6 and 7 show the out-of-sample relative RMSPEs of GDP deflator inflation forecasts conditional on the forecast being made at times of cyclically high inflation/unemployment, respectively. As can be seen in Table 6, the ability of the Phillips curve, or other forecasting methods, to outperform the AR-GAP benchmark is about the same at times of cyclically high inflation as it is in the sample as a whole. The results in Table 7 are a little more encouraging. It seems to be slightly easier to beat the AR-GAP benchmark at times of high unemployment. In particular, Phillips curve forecasts, give small improvements in forecasting inflation. The term structure VAR also does better in this subsample than in the sample period as a whole. However, the improvements are not very big, and this exercise is based on quite a small sample size. In several cases the relative root mean square prediction errors are not significantly different from one. Although the evidence is tantalizing, and highly relevant to the current situation, it is hard to have great confidence that the PC-GAP will continue to deliver better inflation forecasts in future periods of cyclically high unemployment.

We also evaluated the RMSPEs of inflation forecasts conditional on the forecast referring to a period in the three years immediately after an NBER business cycle trough. In contrast to the results in Tables 6 and 7, this is an example of conditioning on a variable known only after the fact. Results are reported in Table 8. There is some slight evidence that the predictability of inflation is higher in the early stages of expansions than at other times. But the improvement in forecast accuracy is not great. We did the same exercise for the three years immediately before an NBER business cycle peak, and for NBER recessions. The results are not reported, but the "fixed ρ" benchmark was even harder to beat in these periods.

Overall, we have found only weak evidence for improved forecastability of inflation relative to the "fixed ρ" benchmark, even when we condition on other variables.

2.13 Time-varying predictability

Closely related to the idea of assessing the performance of a forecast relative to a benchmark conditional on some other variable, one might also ask if the forecast does better than a benchmark *over a certain period of time.* Giacomini and Rossi (2010) propose a "fluctuations test" which uses the test statistic of Diebold and Mariano (1995), but computed over m-year rolling windows. They derive the asymptotic distribution of the maximum value of this statistic, over all possible windows. The null hypothesis is that the forecast *never* beats the benchmark. Rossi and Sekhposyan (2010) apply this test to forecasting CPI inflation. They find that certain variables—such as industrial production and capacity utilization—had significant predictive power for future inflation in the early 1980s, but that it subsequently disappeared.

2.14 Alternative measures of forecast quality

In this chapter, we focus on assessing forecast quality by root mean square prediction error. Some researchers prefer instead to use more general loss functions (Patton and Timmermann (2007)), such as the asymmetric LINEX loss function. At least from the perspective of forecasting inflation in a central bank, or a similar public policy environment, we are a little skeptical that there would be a large and systematic preference for overpredicting inflation rather than underpredicting it (or vice versa). However, if one wishes to assess forecast quality using absolute prediction error, or an asymmetric loss function, that is of course possible too.

Another way of assessing forecast quality is to run a forecast efficiency regression; projecting the forecast errors onto variables that were known at the time that the forecast was being made, such as the forecast itself.[20] This is not a horse race between two forecasts. Rather it is a test of the null hypothesis that a particular forecast is giving the conditional expectation of the future value of the variable at the time that the projection is being made.[21] Clearly the null is rejected if the regression coefficients are significantly different from zero; in this case, there exists in principle some way of correcting the forecast to make it more accurate. Romer and Romer (2000) and Patton and Timmermann (2012) are among the authors who have applied tests of this sort to the Greenbook and other growth and inflation forecasts. Patton and Timmermann consider a variant of the test which evaluates the hypothesis of forecast efficiency at multiple horizons jointly (increasing power) and they find evidence against the efficiency of the Greenbook inflation forecasts. This implies that there is some scope to improve on Greenbook forecasts, despite their low root mean square prediction error.[22]

Patton and Timmermann (2012) also show that forecast rationality under quadratic loss implies certain bounds on the second moments of forecasts and forecast errors across horizons, and develop tests of these variance bounds. Again they find evidence against the rationality of Greenbook forecasts. For example, the variance of a rational forecast ought to be decreasing in the forecast horizon, but Patton and Timmermann (2012) document that the variance of Greenbook forecasts for GDP deflator inflation is actually *increasing* in the forecast horizon.

[20] A special case is the Mincer-Zarnowitz regression, which is most often written as a regression of the variable to be forecast on the forecast itself; efficiency requires the intercept and slope to be zero and one respectively. This is algebraically equivalent to requiring both intercept and slope coefficients to be zero in a regression of the forecast error on the forecast itself.

[21] If the forecaster has quadratic loss, the forecast efficiency regression tests whether this loss function is being minimized, but this does not apply to other loss functions (Elliott, Komunjer and Timmermann (2005)).

[22] It is important to bear in mind that even if such a test results in a rejection of the null hypothesis of forecast efficiency, this does not necessarily mean that some other forecast will be more accurate in terms of *out-of-sample* predictive ability. Croushore (2012) and Arai (2012) discuss using the Patton and Timmermann (2012) test to adjust Greenbook forecasts.

3 Market-Based Measures of the Inflation Outlook

For over a decade, the U.S. Treasury has issued Treasury inflation-protected securities (TIPS): debt securities for which the coupon and principal payments are indexed to the Consumer Price Index (CPI), in addition to conventional nominal bonds. Comparing the yields on these two types of Treasury bonds allows us to compute measures of *inflation compensation* or *breakeven inflation*, defined as the rate of inflation that would give an investors the same return at maturity on a nominal security and an indexed security[23]. They are often interpreted as *market-based measures of inflation expectations* and receive enormous attention from policymakers and in the press. The idea of being able to read inflation expectations directly out of market prices has long held an allure for economists and central bankers—indeed this was one of the motivations for issuing TIPS in the first place (Greenspan (1992), Campbell and Shiller (1996) and Bernanke and Woodford (1997)). We will argue that, while direct high-frequency market-based information is valuable, interpreting these spreads as pure measures of inflation expectations is wrong and potentially dangerous.

Figure 6 plots the five- and five-to-ten-year-ahead forward rates of inflation compensation from TIPS. They are quite volatile, especially during the acute phase of the recent financial crisis—if taken literally as inflation forecasts, they would lead policymakers to be in a constant state of panic, at some times about excessively high inflation, and at other times about excessively low inflation. This figure also shows the corresponding-maturity survey forecasts of inflation, from Blue Chip. These trended downwards during the 1990s, but have been very flat over the past decade. Unfortunately, it would take a very long sample to directly evaluate the RMSPEs of these different long-term inflation forecasts, which we clearly do not have.

[23]Smoothed yield curves have been fitted to both the nominal and TIPS coupon securities (Gürkaynak, Sack, and Wright (2007, 2010)) and can be used for computing inflation compensation (or forward inflation compensation) at different maturities.

However, it seems clear that forward inflation compensation from TIPS is too volatile to represent a rational forecast of the long-run expected level of inflation, or the implicit inflation target of the central bank. We can be more precise about this following a line of reasoning proposed by Gürkaynak, Sack and Wright (2010). If a five-to-ten-year forward rate of inflation compensation really is the rational expectation of inflation in the long-run, then it should be a martingale. Otherwise, the expectation of the long-run expectation of inflation tomorrow would differ from the long-run expectation of inflation today, which is impossible by the law of iterated expectations. And if forward inflation compensation is a martingale, then the volatility of k-period changes in forward inflation compensation must be k times the volatility of one-period changes. This can in turn be tested using the variance ratio test of Lo and MacKinlay (1988). Table 9 shows the standard deviation of one-day and one-, three- and six-month changes in five-to-ten-year forward inflation compensation, along with variance-ratio test statistics. Under the martingale hypothesis, the variance-ratio test statistics have a standard normal asymptotic distribution.[24] However, we see in Table 9 that the test rejects in the left tail, meaning that the volatility of longer-term changes in inflation compensation is too small relative to the volatility of daily changes for five-to-ten-year forward inflation compensation to be a martingale.

Thus, the spreads between nominal and index-linked debt embody inflation expectations, but their interpretation is evidently also complicated by inflation risk premia, and by the different liquidity of nominal and TIPS securities[25]. Normally, one would expect the inflation risk premium to drive TIPS yields down relative to their nominal counterparts, causing inflation compensation to widen. Meanwhile, one would expect the liquidity premium to drive the less liquid TIPS yields up relative to their nominal counterparts, causing inflation

[24] The test statistic is $z^*(q)$ in the notation of Lo and MacKinlay. This test allows for time-varying conditional heterokedasticity.

[25] The same conclusion is reached by Pflueger and Viceira (2011), who instead use predictive regressions to show that the excess returns on a long-nominal and short-TIPS portfolio are time-varying.

compensation to narrow. Both of these effects are presumably time-varying. As can be seen in Figure 6, five-year inflation compensation is typically below the survey expectation, perhaps because the liquidity premium is the dominant effect. On the other hand, five-to-ten-year forward inflation compensation is typically above the survey expectation. Perhaps this is because the liquidity premium is to some extent "differenced out" in the forward rate whereas investors are willing to pay a large risk premium to compensate for inflation risk over longer horizons. Term structure models can be used to attempt to decompose inflation compensation into inflation expectations, liquidity premia, and inflation risk premia (see, for example, D'Amico, Kim and Wei (2010) and Joyce, Lildholdt and Sorensen (2010)). The resulting inflation expectations are far more stable than raw inflation compensation, but the available sample size is too short to evaluate these as inflation forecasts.

3.1 New inflation derivatives

Recently some other alternative market-based inflation measures have developed. There is now an over-the-counter market in inflation swaps. These are contracts where one party agrees to pay an interest rate on a notional underlying principle that is fixed at the start of the contract, while the other party agrees to pay the realized inflation rate on that same notional principle. Only the net of the two amounts actually changes hands. Under risk-neutrality, the fixed rate should equal expected inflation over the life of the contract.

Figure 7 plots the ten-year inflation swaps rate along with the ten-year rate of TIPS inflation compensation. The two have generally moved together, with the swaps rate being slightly higher.[26] They diverged noticeably in late 2008, in the most severe part of the

[26]Fleckstein, Longstaff and Lustig (2010) discuss the fact that inflation swap rates are higher than the spread between nominal and TIPS bond yields. They interpret this as representing an anomaly in the pricing of TIPS. But as the inflation swaps market is quite small, it would seem to us more natural to view it as an anomaly in the pricing of inflation swaps. In a bit more detail, customers in the inflation swaps market almost uniformly want to buy insurance against inflation. Inflation swaps dealers must hedge this risk, and they do so using TIPS and nominal Treasuries. We interpret the relatively high inflation swap rates as the "fee" that dealers require to provide this insurance service.

financial crisis. Campbell, Shiller and Viceira (2009), Hu and Wohar (2009) suggest a rather technical explanation for this. Parties had lent money to Lehman in the collateralized repo market. Following the demise of Lehman, this collateral had to be sold. A good part of the collateral was in TIPS, which are comparatively illiquid. The prices of TIPS bonds fell and their yields rose, as the collateral had to be sold at firesale prices, even more than would be explained by the obviously disinflationary impact of the financial crisis and the resulting recession. The swaps market was apparently not affected in the same way, or at least not to the same extent. In any event, Figure 7 should cast considerable doubt on the idea of reading inflation expectations directly from either bond spreads or inflation derivatives. Haubrich, Pennacchi and Ritchken (2008) consider a term structure models using TIPS, inflation swaps, realized inflation data and surveys jointly to infer a measure of inflation expectations.

There are also short-horizon inflation swaps which provide information about investors' assessment of short-run inflation prospects, although again these are not pure inflation expectations. Figure 8 shows the time series of one-year-ahead inflation swap rates, the shortest maturity that is available. Subsequent realized CPI inflation is also shown in the figure, shifted back one year, so that the forecasts and actuals would coincide if the forecast were perfect. The inflation swap rates are available only back to 2005, and so it is still too soon to assess their performance as predictors of inflation. However, looking at Figure 8, we can say that while these short-term inflation swap rates may be telling us something about near-term inflation expectations, they appear to move almost in lockstep with *past* inflation.

Very recently, inflation options have started to be actively traded. These take the form of caps and floors. A simple cap is a contract which entitles the holder to receive a payment at maturity that is a fraction $\max((1+\pi)^n - (1+s)^n, 0)$ of a notional underlying principle, where π is the realized average inflation rate over the life of the contract, s is a strike price, and n is the life of the contract in years. The holder must pay an up-front fee for this contract. Inflation floors are similar, except that the holder receives a payoff when inflation turns out to be particularly low. The prices of these derivatives can be used to reverse-engineer investors' probability density for inflation, under the assumption that investors are risk-neutral. Kitsul and Wright (2012) find that during recent years these option-implied densities have implied non-negligible odds of both deflation and fairly high inflation (greater than 4 percent), even over five- and ten-year horizons. But of course investors are not risk-neutral, and they may be willing to pay a premium to hedge against the risks of both deflation and a sharp pickup in inflation. While inflation caps and floors do not provide physical density forecasts, except under risk neutrality, they do tell us something about the inflation concerns of some investors.[27]

4 Other Topics

4.1 Density Forecasts

Only the most foolishly overconfident forecasters claim to have perfect foresight. A point forecast of inflation without some measure of associated uncertainty is arguably of little value. Meanwhile, a density forecast gives a complete characterization of the beliefs of the forecaster about future inflation prospects. One particular special case of a density forecast that is of particular topical interest is estimating the probability of deflation, as policy makers are worried that this might result in a spiral in which falling prices push real interest rates up, depressing aggregate demand, and putting further downward pressure on prices. One can also construct measures of the risk of excessively low (or high) inflation, given a specification of the users' preferences (Kilian and Manganelli (2007, 2008)).

One way of constructing a density forecast is to take any point forecast and assume that the errors are homoskedastic. Assuming normality, one would take the point forecast

[27] TIPS actually contain an option-like feature. The principal repayment at maturity will be the *greater* of the nominal face value and the face value adjusted for inflation over the life of the security. This can be used to back out the implied probability of deflation (Wright (2009b), Christensen, Lopez and Rudebusch (2011)), although these calculation again assume risk-neutrality.

±1.96 standard deviations, or one could instead take the percentiles of the distributions of the historical errors and add those on to the point forecasts (Reifschneider and Williams (2000)). This latter approach is also taken in the density forecasts included in the Greenbook since 2004. The difficulty with an exercise like this is that it assumes that the errors are always drawn from the *same* distribution, which in turn renders the whole density forecasting problem a rather uninteresting of point forecasting. But the volatility of shocks to both output growth and inflation clearly vary over time. Indeed, the motivating example for Rob Engle's pathbreaking work on volatility clustering (Engle (1982)) was inflation, although these methods are now more widely applied to asset price returns. A density forecast ought therefore to aspire to be more informative than simply adding and subtracting constants from the point forecasts.

A number of such approaches have been proposed for constructing density forecasts for inflation. One is to use univariate or VAR forecasts, but with GARCH effects, stochastic volatility, or breaks in variance (Giordani and Söderlind (2003), Stock and Watson (2007), Groen, Paap and Ravazzolo (2009) or Clark (2011)) . Another is to use quantile regressions, in which the model specifies not the conditional mean of inflation, but rather some conditional quantiles (Manzan and Zerom (2009))—the different quantiles may exhibit different sensitivity to the predictors. And a third way is to use survey density forecasts, or the density forecasts provided by central banks. Since 1968, the Survey of Professional Forecasters has asked respondents to assign probabilities to inflation falling into a number of bins, which represents a simple density forecast, discussed and evaluated by Diebold, Tay and Wallis (1999). More recently, the ECB Survey of Professional Forecasters (ECB-SPF) has obtained density forecasts at short- and long-horizons for euro-zone inflation. The Bank of England has produced density forecast ("fan charts") in its quarterly Inflation reports since 1997 (Britton, Fisher and Whitley (1998), Clements (2004)). The Bank of England constructs these densities using a three parameter functional form for the density, and sets the parameters (effectively mean, volatility and skewness) judgmentally. Appropriately, the SPF, ECB-SPF and Bank of England inflation density forecasts all widened out substantially during the recent financial crisis.

As an illustration of questions that may be addressed with density forecasts, we use real-time estimation of the UCSV model to construct probabilities of average inflation (GDP deflator) over the next two years being above 4 percent (top panel) or below 0 percent (bottom panel) from 1985Q1 to 2011Q4. These probabilities are shown in Figure 9. The probability of a sustained period of excessively high inflation was elevated in the early part of the sample, but then declined and has remained near zero since then. The probability of sustained deflation remained below 10 percent throughout the sample, even during the financial crisis. That's because the variance of the permanent component of inflation in the UCSV model is estimated to be very small at the end of the sample period.

4.2 Forecasting Aggregates or Disaggregates?

A long-standing question is whether it is better to forecast inflation aggregrates directly, as we have done so far in this chapter, or to forecast the disaggregated inflation rates, and then aggregate these forecasts. Theoretically, if the data generating process is known, then aggregating disaggregate forecasts must be at least as good as constructing the aggregate forecast directly (Lütkepohl (1987)). But, when the data generating process has to be estimated, then it is possible for the direct aggregate forecast to be more accurate, because it entails the estimation of fewer parameters. If the disaggregates are all have similar dynamics, then we might expect forecasting aggregates to work best in small samples. On the other hand, if the disaggregates have very persistence properties, then we would expect aggregation of disaggregate forecasts to do better. In the end, it is an empirical question as to which of these two methods is more accurate. Hubrich (2005) compares these two approaches to forecasting euro area inflation, and finds that neither method necessarily works better. Bermingham and D'Agostino (2011) are more supportive of aggregating disaggregate forecasts.

In this chapter, we illustrate the trade-offs between forecasting inflation aggregates directly and forecasting the disaggregates, by considering the real-time forecasting of overall CPI inflation. One innovation relative to the existing work on comparing these two alternative forecasting strategies is that our forecasts are all in gap form, in keeping with a main theme of this chapter. The forecasts that we consider are as follows:

1. An AR-GAP forecast for overall CPI inflation. This method thus uses aggregates to forecast aggregates. The lag order is determined by the Bayes Information Criterion.

2. An AR-GAP forecast for overall CPI inflation using an AR(1) with a slope coefficient fixed at 0.46 (as in the "fixed ρ" forecasts discussed earlier).

3. As in method 1, except using a separate univariate autoregression for the food, energy and core (i.e. ex-food-and-energy) CPI inflation disaggregates. All disaggregates are expressed in gap form, relative to the overall CPI inflation trend. These give forecasts of the disaggregates, which can then be combined, using real-time CPI weights, to obtain an implied forecast for overall CPI inflation.

4. As in method 3, except imposing that the slope coefficients in the autoregressions for food and energy inflation are both equal to zero. This imposes that food and energy inflation have no persistence.

5. As in method 4, except imposing that core inflation is an AR(1) with a slope coefficient fixed at 0.46.

6. A VAR in the food, energy and core CPI inflation disaggregates. This VAR provides gives forecasts of disaggregates which can then be combined to obtain the implied forecast for overall inflation.

7. A VAR in overall CPI inflation, and food and energy disaggregates, used to forecast overall inflation. This provides a direct forecast of aggregate inflation, but uses the disaggregated variables to do so. The idea of projecting aggregates onto disaggregates was proposed by Hendry and Hubrich (2010), who found that it may be a promising direction, especially if the researcher combines the information in the disaggregates judiciously, perhaps using model selection procedures.

Table 10 shows the RMSPE of all of these forecasts. The VARs (methods 6 and 7) and the univariate autoregression for overall CPI inflation all have roughly comparable performance. Fitting univariate autoregressions to the three components separately (method 3) fares a bit better at short horizons. But our main finding is that heavy-handedness helps. The AR-GAP forecast for overall CPI inflation with a fixed slope coefficient does better than the unrestricted AR-GAP model. Within the disaggregate forecasts, restricting the slope coefficients for food and energy inflation to be zero improves forecast accuracy. Adding in the restriction that core inflation is an AR(1) with a fixed slope coefficient helps further, and in fact gives the best forecasts for overall CPI inflation.[28] The argument of Lütkepohl (1987) for aggregating disaggregate forecasts is right, but the benefits are easily swamped by parameter estimation error. It is by imposing parameter restrictions (that are obviously not literally correct) that we are able to get some gain out of a "bottom up" forecasting strategy.

Managing headline (i.e. total) inflation is presumably the appropriate ultimate objective of the central bank—people consume food and energy, and it seems hard to imagine a substantive rationale for why these should matter less than other elements of the consumption basket. Nonetheless, the Fed pays most attention to core inflation rather than headline inflation, arguing that food and energy components in headline inflation are overwhelmingly

[28]This is still true in a sample that ends before the financial crisis, but the gains from treating food and energy inflation as transitory are smaller in this earlier sample.

transitory. The point is discussed in Blinder (1997) who indeed defines a concept of core inflation as the part of inflation that has some persistence. Decomposing CPI into food, energy and all other items appears to be a useful way of splitting out more- and less-persistent inflation fluctuations.

4.3 Using Core Forecasts as Headline Forecasts

Indeed, if we treat the food and energy components of inflation as pure noise, with no pass-through into core prices, then any model estimation ought to be applied to core inflation even if prediction of total inflation is our end-objective. If food and energy are pure noise, then the population projection coefficients from regressions of core and total inflation onto explanatory variables are exactly the same, but the former are just estimated more precisely. As a result, even if we want to predict total inflation, we are better off forecasting core inflation and then using this as if it were a prediction of total inflation.[29]

Table 11 evaluates how this strategy works in practice. The table reports the out-of-sample RMSPEs of forecasts of core CPI inflation when used as predictions of subsequent total CPI inflation, relative to the benchmark of applying the same forecasting method directly to total CPI inflation. Most relative RMSPEs in Table 11 are slightly below 1, indicating that forecasting core inflation may indeed be appropriate, even if the end-goal is predicting total inflation.

4.4 Alternative Inflation Measures

CPI inflation excluding food and energy seems to be useful as a predictor of future total inflation. There are a number of alternative inflation indices that one might use for similar reasons. Bryan and Cecchetti (1994) advocated using median CPI inflation (across

[29]See Faust and Wright (2011) for further discussion of the idea that predictive regressions should be estimated removing any purely unforecastable components from the left-hand-side variable.

components) as a measure of core inflation.[30] Cutler (2001) and Bilke and Stracca (2007) have both constructed price indices that weight the components by their persistence. Bryan and Meyer (2010) decompose overall CPI inflation into the inflation associated with goods that have relatively sticky/flexible prices. They find a more pronounced tradeoff between slack and inflation when inflation is measured from the "flexible prices " CPI, which could potentially turn out to be useful for inflation forecasting.

5 International Inflation Forecasts

In this section, we briefly consider the forecasting of consumer price inflation in a few large foreign countries: Canada, Germany, Japan, and the United Kingdom. Analysis of the foreign experience is useful because it increases our effective sample size. In this respect, it is useful that Germany, Japan and the U.K. all had substantially different inflation experiences to the U.S. over the last half century. For example, the rise in inflation in the late 1970s was considerably more muted in Germany than in the U.S.. Also, inflation remained quite high in the UK into the early 1990s, well after the Great Inflation had subsided in most other developed countries. And, at the time of writing, Japan is experiencing a slow but sustained and pernicious deflation that has already lasted for more than a decade.

For the international forecast comparison, we use only *ex-post* revised data[31], and consider only a small subset of the methods that we applied to the main forecast comparison using U.S. data in section 2. For the foreign countries, we only consider the direct forecast, the two random walk forecasts (RW and RW-AO), the UCSV forecast, the Phillips curve (PC) forecast, the autoregression in gap form (AR-GAP), the "fixed ρ" forecast, and the

[30] Ball and Mazumder (2011) noted that median CPI inflation declined more distinctly than total inflation (or even than inflation excluding food and energy) during the Great Recession.

[31] Most countries outside the US report only inflation data without seasonal adjustment. The seasonal patterns in price data are however important. We therefore adjusted our inflation data for Canada, Germany, Japan and the United Kingdom using the X12 seasonal adjustment filter.

Phillips curve forecast in gap form (PC-GAP). For the last three of these forecasts, the inflation trend is measured from the most recent five-to-ten-year-ahead inflation forecast from Consensus Forecasts—a multi-country macroeconomic survey analogous to Blue Chip for the U.S..

The models are estimated on quarterly data from 1960Q1 through to 2011Q4. We consider recursive pseudo-out-of-sample forecasts for the current and subsequent quarters using all these methods, with the first forecast made in 1985Q1. We evaluate the competing forecasts in terms of their root mean square prediction errors (RMSPEs), relative to the benchmark of the "fixed ρ" forecast (the AR(1) in gap form with a fixed slope coefficient). The results are reported in Table 12.

The results are qualitatively similar to those that we earlier found for the U.S. in section 2. The "fixed ρ" benchmark—that is a glide-path from the most recent observation to an estimate of the local mean—is hard to beat by much. The relative RMSPEs are statistically significantly less than one in only a few cases. Even in these cases, they are around 0.9, corresponding to a 10 percent improvement relative to a very simple baseline. The pure random walk forecast does not do well: quarter-over-quarter inflation is too noisy to be a good estimate of the local mean. However, the other methods that allow for some nonstationarity in inflation—the "fixed ρ" benchmark forecasts, the Atkeson-Ohanian version of the random walk forecast, and the UCSV, AR-GAP and PC-GAP forecasts—are all about equally good. The forecasts that are based on treating inflation as a stationary process (direct and PC) do particularly badly at longer horizons, except perhaps for Germany. Germany has had more stable inflation dynamics over the last half century than the U.S., Canada, Japan or the U.K., and so it is perhaps not surprising that allowing for a time-varying local mean is less critical for Germany than it is for the other countries. The performance of the forecasts based on Phillips curve relationships is not terribly encouraging. For example, the PC-GAP forecast does worse than the "fixed ρ" benchmark for some country/horizon combinations, and never does much better.

The most natural interpretation for why the "fixed ρ" benchmark generally does so well is that central banks use monetary policy to push inflation back to target fairly quickly. If so, one would expect the benchmark to do less well in countries or at times where central banks have been less successful in this pursuit. It is noteworthy that for Japan, the AR-GAP, RW-AO and UCSV model give 5-16 percent reductions in RMSPE relative to the "fixed ρ" benchmark. Moreover, these improvements in forecast accuracy are statistically significant in some cases.[32] The fact that our benchmark seems a little easier to beat for Japan may be a reflection of the difficulty that the Bank of Japan has experienced in eliminating deflation.

Our results in Table 12 do not include judgmental forecasts of inflation, except of course that we use long-run surveys to measure the inflation trend.[33] Groen, Kapetanios and Price (2012) evaluated judgmental forecasts of inflation for the United Kingdom. They compared the Bank of England inflation forecasts with those from a range of econometric models. They found that the Bank of England forecast consistently did best, but that it did so mainly by incorporating judgment about the long-run value of inflation. This is again entirely consistent with our results for the U.S..

6 Conclusions

In this chapter, we have discussed a numbers of methods for forecasting inflation. We find that judgmental forecasts (private sector surveys and the Greenbook) are remarkably hard to beat. But we don't necessarily even need the whole term structure of judgmental

[32]The AR-GAP model has two estimated parameters: the intercept and the slope coefficient. Note that the estimate of the intercept coefficient is negative for Japan in recent years.

[33]Most central banks produce judgmental inflation forecasts. There are also several surveys for countries outside the U.S., such as the ECB-SPF. Consensus Forecasts is however the only survey that is of global scope, and it does not include forecasts of quarter-over-quarter inflation that we study in this chapter. For this reason, we don't include it in our forecast comparison.

inflation forecasts. If we just take the current-quarter and long-run survey forecasts, then a very simple glide path between these two boundary conditions—that doesn't involve any parameter estimation—turns out to be a surprisingly competitive benchmark.

In many forecasting contexts, very simple methods that limit or avoid parameter estimation turn out to work shockingly well. For example, Meese and Rogoff (1983) found that the driftless random walk is an excellent predictor of exchange rates. We find that in much the same way, extremely simple inflation forecasts—that however take account of nowcasting and secular changes in the local mean inflation rate—are just about the best that are available. If monetary policy mainly is directed toward smoothly eliminating deviations of inflation from some slowly-moving target, this result might be about what we should expect.

Table 1. Variables and transformations in our Large Dataset

Variable	Transform	Variable	Transform
Average Hourly Earnings: Construction	DLN	Payrolls: Goods-Producing	DLN
Average Hourly Earnings: Manufacturing	DLN	Payrolls: Government	DLN
Average Weekly Hours	Level	Payrolls: Information Services	DLN
Average Weekly Hours: Overtime	Level	Payrolls: Leisure	DLN
Civilian Employment	DLN	Payrolls: Natural Resources	DLN
Real Disposable Personal Income	DLN	Payrolls: Other Services	DLN
New Home Starts	Log	Payrolls: Professional	DLN
Housing Starts: 1-Unit Structures	Log	Payrolls: Retail Trade	DLN
Housing Starts: 5+Unit Structures	Log	Payrolls: Total Private Industries	DLN
Housing Starts in Midwest	Log	Payrolls: Trade, Transportation	DLN
Housing Starts in Northeast	Log	Payrolls: Wholesale Trade	DLN
Housing Starts in South	Log	Real GDP	DLN
Housing Starts in West	Log	Real Consumption	DLN
Industrial Production Index	DLN	Real Durables Consumption	DLN
IP: Business Equipment	DLN	Real Consumption (Services)	DLN
IP: Consumer Goods	DLN	Real Residential Investment	DLN
IP: Durable Consumer Goods	DLN	Real Nonresidential Investment	DLN
IP: Durable Materials	DLN	Real Government Spending	DLN
IP: Final Products (Market Group)	DLN	Real Exports	DLN
IP: Materials	DLN	Real Imports	DLN
IP: Nondurable Consumer Goods	DLN	Federal Funds Rate	FD
IP: Nondurable Materials	DLN	3 Month Treasury Bill Yield	FD
ISM Manufacturing: PMI Index	Level	1 Year Yield	FD
ISM: Employment Index	Level	3 Year Yield	FD
ISM: Inventories Index	Level	5 Year Yield	FD
ISM: New Orders Index	Level	10 Year Yield	FD
ISM: Production Index	Level	AAA Corporate Yield (Moody's)	FD
ISM: Prices Index	Level	BAA Corporate Yield (Moody's)	FD
ISM: Supplier Deliveries Index	Level	3 Month Bill/Fed Funds Spread	Level
Total Nonfarm Payrolls: All Employees	DLN	1 Year/3 Month Bill Spread	Level
Civilians Unemployed - 15 Weeks	DLN	3 Year/3 Month Bill Spread	Level
Civilians Unemployed for 15-26 Weeks	DLN	5 Year/3 Month Bill Spread	Level
Civilians Unemployed for 27 Weeks +	DLN	10 Year/3 Month Bill Spread	Level
Civilians Unemployed for 5-14 Weeks	DLN	AAA/10 Year Bill Spread	Level
Civilians Unemployed < 5 Weeks	DLN	BAA/AAA Spread	Level
Civilian Unemployment Rate	FD	Excess Stock Market Return	Level
Nonfarm Payrolls: Construction	DLN	SMB Fama French Factor	Level
Payrolls: Education	DLN	HML Fama French Factor	Level
Payrolls: Financial Activities	DLN		

Notes: Transformations are DLN: first difference of logs; FD: first differences.

Table 2: RMSPE of Selected Inflation Forecasts

Horizon	0	1	2	3	4	8
Panel A: GDP Deflator						
Direct	1.06**	1.00	0.96	1.04	1.09	1.34***
RAR	1.06**	1.02	1.01	1.17***	1.24***	1.53***
PC	1.07*	1.03	1.01	1.08	1.14*	1.41***
RW	1.19***	1.17**	1.09	1.04	1.06	1.25*
RW-AO	0.95	0.90*	0.91	0.94	0.96	1.05
UCSV	0.98	0.96	0.91	0.91	0.94	1.07
AR-GAP	1.03	0.97	0.95*	1.01	1.05	1.18***
PC-GAP	1.04	1.02	1.03	1.10*	1.17**	1.33***
PCTVN-GAP	1.04	1.02	1.03	1.10*	1.17**	1.30***
Term Structure VAR	1.07**	1.12**	1.16***	1.25***	1.32***	1.50***
TVP-VAR	0.99	0.94	0.95	0.94	1.00	1.21
EWA	1.02	0.94*	0.91**	0.97	1.01	1.15***
BMA	1.00	0.91**	0.89***	0.97	1.09	1.19**
FAVAR	1.02	1.03	1.07	1.06	1.13**	1.26***
DSGE	1.06	1.02	1.06	1.08	1.08	1.16
DSGE-GAP	1.02	0.95	0.97	0.98	0.97	1.05
BC	0.81***	0.85***	0.87***	0.90***	0.94**	
SPF	0.82***	0.84***	0.86***	0.88***	0.91**	
GB	0.84*	0.83**	0.82**	0.81**	0.82**	
Fixed ρ + nowcast	0.81***	0.93***	0.97**	1.00	1.00	1.00
Panel B: PCE Deflator						
Direct	1.13**	1.18**	1.22*	1.24**	1.16**	1.32***
RAR	1.13**	1.21**	1.18*	1.22***	1.17**	1.33***
PC	1.14**	1.21**	1.24*	1.27**	1.19**	1.33***
RW	1.23***	1.35**	1.32*	1.38**	1.28**	1.4*
RW-AO	1.10**	1.08	1.07	1.07	1.04	1.16*
UCSV	1.06*	1.07	1.06	1.09	1.06	1.17
AR-GAP	1.09**	1.12**	1.13*	1.14***	1.10***	1.23***
PC-GAP	1.09**	1.14**	1.16**	1.19***	1.18***	1.35***
PCTVN-GAP	1.10**	1.15**	1.17**	1.20***	1.18***	1.32***
Term Structure VAR	1.08***	1.11***	1.12***	1.18***	1.21***	1.38***
TVP-VAR	1.13**	1.18*	1.18*	1.14*	1.10	1.29*
EWA	1.08**	1.12**	1.13	1.13**	1.08***	1.22***
BMA	1.08**	1.12**	1.14	1.14***	1.12***	1.31***
FAVAR	1.04	1.13**	1.12**	1.15***	1.14***	1.25***

Table 2: RMSPE of Selected Inflation Forecasts, Continued

Horizon	0	1	2	3	4	8
Panel C: CPI						
Direct	1.02	1.10*	1.18**	1.23***	1.15***	1.23***
RAR	1.02	1.15*	1.11**	1.16***	1.13***	1.21***
PC	1.04	1.12*	1.19**	1.24***	1.16***	1.22***
RW	1.28**	1.40**	1.34**	1.46***	1.34**	1.51**
RW-AO	1.03	1.07*	1.07*	1.09**	1.07*	1.12*
UCSV	0.96	1.01	1.03	1.07**	1.03	1.10***
AR-GAP	1.00	1.06	1.10**	1.13***	1.11***	1.16***
PC-GAP	1.01	1.07*	1.12**	1.15***	1.15***	1.22***
PCTVN-GAP	1.02	1.09**	1.14***	1.16***	1.16***	1.22***
Term Structure VAR	1.00	1.06*	1.12***	1.15***	1.20***	1.23***
TVP-VAR	1.11	1.34	1.28	1.15**	1.40	1.48
EWA	0.99	1.06	1.10*	1.11***	1.09***	1.15***
BMA	1.02	1.09*	1.12*	1.12***	1.13***	1.19***
FAVAR	0.98	1.12**	1.12***	1.13***	1.14**	1.18***
BC	0.8**	0.98	1.00	0.97	0.99	
SPF	0.78**	0.97	1.00	0.99	0.99	
GB	0.82***	1.05	1.03	1.01	0.99	
Fixed ρ + nowcast	0.80**	1.00	1.01	1.00**	1.00	1.00***
Panel D: Core CPI						
DAR	1.05	1.06*	1.20***	1.39***	1.38***	1.76***
RAR	1.05	1.11***	1.54***	1.73***	1.71***	2.13***
PC	1.09*	1.09	1.22***	1.41***	1.40***	1.77***
RW	1.17***	1.09	1.06	1.17**	1.11	1.13
RW-AO	0.97	0.97	0.98	0.97	0.98	1.08
UCSV	0.98	0.95	0.94	0.99	0.98	1.04
AR-GAP	1.10**	1.14***	1.24***	1.42***	1.45***	1.69***
PC-GAP	1.17***	1.24***	1.35***	1.55***	1.57***	1.82***
PCTVN-GAP	1.19***	1.27***	1.38***	1.56***	1.60***	1.83***
Term Structure VAR	1.38***	1.60***	1.64***	1.69***	1.77***	1.95***
TVP-VAR	1.03	1.00	0.95	0.98	0.98	1.23
EWA	1.04	1.08**	1.16**	1.32***	1.35***	1.60***
BMA	1.05	1.11***	1.16**	1.38***	1.45***	1.65***
FAVAR	1.21***	1.35***	1.32***	1.40***	1.65***	1.77***
GB	0.95	0.91*	0.88*	0.89*	0.87*	

Notes: This table reports the pseudo-out-of-sample recursive RMSPE of alternative h-quarter-ahead forecasts of four inflation measures (quarter-over-quarter), all relative to the benchmark of an AR(1) in "gap" form with a fixed slope coefficient. All forecasts are fully real-time, except for the large dataset methods (EWA, BMA, FAVAR), which use revised data on the predictors. The sample consists of data as observed in 1985Q1-2011Q4, with data going back to 1960Q1 in all cases. Cases in which the relative root mean square prediction error is significantly different from one at the 10, 5 and 1 percent significance levels are denoted with one, two and three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Table 3: RMSPE of Selected Inflation Forecasts Using Blue Chip Nowcasts as Jumping-Off Point

Horizon	0	1	2	3	4	8
			Panel A: GDP Deflator			
Direct	1	0.98	0.95	1.02	1.07	1.32***
RAR	1	0.98	0.96	1.14***	1.22***	1.50***
PC	1	0.98	0.98	1.04	1.11*	1.39***
RW	1	0.93**	0.87**	0.92	0.94	1.06
RW-AO	1	0.95	0.88**	0.90	0.94	1.06
UCSV	1	0.95	0.89*	0.90*	0.92	1.04
AR-GAP	1	0.99	0.97	0.99	1.05**	1.18***
PC-GAP	1	1.01	1.02	1.07	1.14**	1.32***
PCTVN-GAP	1	1.01	1.02	1.07	1.15**	1.30***
Term Structure VAR	1	1.06*	1.11**	1.19***	1.27***	1.46***
TVP-VAR	1	0.95	0.89*	0.87**	0.92	1.09
DSGE	1	0.88**	0.92	1.01	1.02	1.13
DSGE-GAP	1	0.88***	0.89*	0.96	0.97	1.05
BC	1	0.91***	0.90***	0.91***	0.94**	
SPF	1.02	0.90***	0.88***	0.88***	0.91**	
GB	0.97	0.89*	0.84**	0.81**	0.82**	
			Panel B: CPI			
DAR	1	1.13**	1.06*	1.07*	1.06**	1.14***
RAR	1	1.13**	1.06**	1.09**	1.1***	1.19***
PC	1	1.14**	1.06*	1.08*	1.06*	1.13***
RW	1	1.08*	1.08*	1.09*	1.07*	1.07
RW-AO	1	1.09*	1.09**	1.09**	1.07*	1.15*
UCSV	1	1.01	1.02	1.04*	1.02	1.07*
AR-GAP	1	1.08**	1.05**	1.07**	1.07**	1.16***
PC-GAP	1	1.08**	1.07***	1.09***	1.11***	1.23***
PCTVN-GAP	1	1.09**	1.08***	1.09***	1.12***	1.23***
Term Structure VAR	1	1.04	1.10***	1.15***	1.18***	1.23***
TVP-VAR	1	1.15*	1.25*	1.16**	1.09*	1.28
BC	1	0.98	1.00	0.98	0.99	
SPF	0.97***	0.97	1.00	0.99	0.99	
GB	0.99	1.05	1.02	1.02	0.99	

Notes: As for Table 2, except that the Blue Chip current-quarter forecast is treated as the final observation for all forecasts except the SPF and the Greenbook. The table shows RMSPEs relative to the benchmark of an AR(1) in "gap" form with a fixed slope coefficient, using the Blue Chip nowcast. By construction, the relative RMSPEs for the current quarter are all equal to 1 (except for the SPF and the Greenbook). Blue Chip nowcasts are only available for GDP deflator and CPI inflation.

Table 4: RMSE of GDP Deflator Inflation Forecasts over the Pre-Crisis Period

Horizon	0	1	2	3	4	8
Direct	1.06*	0.98	0.93	0.97	1.03	1.27***
RAR	1.06*	1.01	0.98	1.13**	1.20**	1.49***
PC	1.07*	0.99	0.95	1.00	1.06	1.34***
RW	1.16**	1.14*	1.03	0.95	0.98	1.17
RW-AO	0.96	0.86**	0.86*	0.87*	0.90	1.00
UCSV	0.98	0.92	0.86*	0.84*	0.88	1.03
AR-GAP	1.03	0.95	0.93**	0.97	1.03	1.18***
PC-GAP	1.04	0.99	0.98	1.05	1.13*	1.32***
PCTVN-GAP	1.04	0.99	0.98	1.05	1.12*	1.30***
Term Structure VAR	1.06*	1.11**	1.16***	1.24***	1.31***	1.48***
TVP-VAR	1.02	0.95	0.89	0.86**	0.90	1.05
EWA	1.02	0.94*	0.91**	0.94*	1.00	1.15**
BMA	1.01	0.91*	0.87***	0.93*	1.02	1.12*
FAVAR	1.03	1.00	1.02	1.06	1.13*	1.28***
DSGE	1.07	1.04	1.07	1.07	1.06	1.13
DSGE-GAP	1.02	0.97	0.97	0.96	0.94	1.00
BC	0.86**	0.85***	0.87***	0.90**	0.93**	
SPF	0.85**	0.83***	0.84***	0.86***	0.89**	
GB	0.84*	0.83**	0.82**	0.80**	0.81**	
Fixed ρ + nowcast	0.86**	0.93***	0.98**	1.00	1.00	1.00

Note: As for Table 2, except that only forecasts made for 2007Q3 and earlier are included.

Table 5: Correlations among Selected Inflation Forecasts

	Fixed ρ	PC-GAP	VAR	EWA	BMA	BC	SPF	DSGE	DSGE-GAP
				Panel A: GDP Deflator					
Fixed ρ	1.00	0.78	0.80	0.89	0.82	0.96	0.94	-0.08	0.50
PC-GAP	0.78	1.00	0.96	0.90	0.89	0.84	0.84	0.22	0.62
VAR	0.80	0.96	1.00	0.85	0.85	0.85	0.85	0.12	0.55
EWA	0.89	0.90	0.85	1.00	0.93	0.91	0.90	0.18	0.66
BMA	0.82	0.89	0.85	0.93	1.00	0.84	0.84	0.17	0.63
BC	0.96	0.84	0.85	0.91	0.84	1.00	0.98	0.01	0.55
SPF	0.94	0.84	0.85	0.90	0.84	0.98	1.00	0.02	0.54
DSGE	-0.08	0.22	0.12	0.18	0.17	0.01	0.02	1.00	0.81
DSGE-GAP	0.50	0.62	0.55	0.66	0.63	0.55	0.54	0.81	1.00
				Panel B: PCE Deflator					
Fixed ρ	1.00	0.80	0.80	0.85	0.51				
PC-GAP	0.80	1.00	0.96	0.92	0.92				
VAR	0.80	0.96	1.00	0.84	0.84				
EWA	0.85	0.92	0.84	1.00	0.96				
BMA	0.81	0.92	0.84	0.96	1.00				
				Panel C: CPI					
Fixed ρ	1.00	0.79	0.72	0.83	0.73	0.96	0.94		
PC-GAP	0.79	1.00	0.87	0.92	0.90	0.83	0.84		
VAR	0.72	0.87	1.00	0.68	0.73	0.77	0.78		
EWA	0.83	0.92	0.68	1.00	0.92	0.82	0.83		
BMA	0.73	0.90	0.73	0.92	1.00	0.72	0.71		
BC	0.96	0.83	0.77	0.82	0.72	1.00	0.99		
SPF	0.94	0.84	0.78	0.83	0.71	0.99	1.00		
				Panel D: Core CPI					
Fixed ρ	1.00	0.88	0.80	0.97	0.94				
PC-GAP	0.88	1.00	0.97	0.95	0.91				
VAR	0.80	0.97	1.00	0.89	0.85				
EWA	0.97	0.95	0.89	1.00	0.98				
BMA	0.94	0.91	0.85	0.98	1.00				

Notes: This table shows the correlations among selected out-of-sample forecasts of each of the four inflation indicators considered in this chapter.

Table 6: RMSPE of GDP Deflator Inflation Forecasts
Forecasts Made During Periods of Elevated Inflation

Horizon	0	1	2	3	4	8
Direct	1.12**	1.13**	1.11	1.39**	1.38***	1.66***
RAR	1.12**	1.15**	1.16	1.42**	1.43***	1.74***
PC	1.15**	1.19**	1.17**	1.47***	1.48***	1.78***
RW	1.31***	1.34***	1.27	1.37***	1.29***	1.43***
RW-AO	0.92	1.01	1.01	1.10*	1.11***	1.21***
UCSV	0.94	1.05	1.03	1.12*	1.13***	1.24***
AR-GAP	1.05	1.06	1.03	1.20**	1.16***	1.22***
PC-GAP	1.08	1.12*	1.09*	1.27***	1.30***	1.43***
PCTVN-GAP	1.06	1.09	1.04	1.19***	1.20***	1.31***
Term Structure VAR	1.17***	1.19*	1.11	1.26***	1.33***	1.54***
TVP-VAR	0.91	1.00	1.03	1.01	1.08	1.23***
EWA	1.02	1.02	0.98	1.10***	1.05	1.12***
BMA	1.01	0.97	0.98	1.12***	0.94	1.00
FAVAR	0.92	1.04	1.00	0.88*	0.86*	1.06*
DSGE	1.07*	1.04	1.15	1.22	1.08*	1.26***
DSGE-GAP	1.06*	1.01	1.10	1.12	1.00	1.12*
BC	0.88	0.94	0.94	0.99	1.01	
SPF	0.89	0.91*	0.87**	0.96*	1.01	
GB	1.43**	0.87	0.98	1.14*	1.05	
Fixed ρ + nowcast	0.88	0.98	0.98	1.00	1.01***	1.00

Notes: As for Table 2, except that only forecasts made during periods of elevated inflation are considered.

Table 7: RMSPE of GDP Deflator Inflation Forecasts
Forecasts Made During Periods of Elevated Unemployment

Horizon	0	1	2	3	4	8
Direct	1.10**	1.03	1.03	1.10	1.16*	1.32***
RAR	1.10**	1.06	1.09	1.25**	1.31***	1.48***
PC	1.10**	1.07	1.08	1.12	1.17*	1.26***
RW	1.25***	1.28***	1.28*	1.07	1.19**	1.42*
RW-AO	0.94	0.98	1.04	1.05	1.03	1.08
UCSV	0.96	1.05	1.00	0.95	0.99	1.07
AR-GAP	1.07**	0.98	0.98	1.02	1.06**	1.19***
PC-GAP	1.05	1.01	1.05	1.08	1.11	1.12*
PCTVN-GAP	1.05	1.01	1.05	1.06	1.10	1.11
Term Structure VAR	1.04	0.95	1.02	1.04	1.12	1.23***
TVP-VAR	0.96	0.96	1.09	1.07	1.19	1.55
EWA	1.04	0.90	0.88*	0.92	0.99	1.13**
BMA	1.02	0.83**	0.82**	0.93	1.22	1.33
FAVAR	1.05	1.08	1.18	0.94	1.06	1.23**
DSGE	1.07	1.12	1.02	0.87	0.83	0.78
DSGE-GAP	1.05	1.05	0.96	0.84	0.79	0.93
BC	0.73**	0.78**	0.83*	0.86**	0.88*	
SPF	0.75**	0.76**	0.79**	0.82**	0.83*	
GB	0.61**	0.74*	0.81	0.77	0.70	
Fixed ρ + nowcast	0.73**	0.89*	0.96*	1.00	1.00	1.00

Notes: As for Table 2, except that only forecasts made during periods of elevated unemployment are considered.

Table 8: RMSPE of GDP Deflator Inflation Forecasts
Forecasts Made For Periods During the Early Stages of Expansions

Horizon	0	1	2	3	4	8
Direct	1.13**	1.06	0.99	1.17	1.24**	1.56***
RAR	1.13**	1.09	1.02	1.32***	1.37***	1.66***
PC	1.16**	1.10	1.07	1.22	1.30***	1.67***
RW	1.20**	1.29**	1.14	1.14	1.23*	1.45***
RW-AO	1.01	0.94	0.98	1.05	1.04	1.13
UCSV	1.02	1.00	0.94	0.95	1.02	1.16*
AR-GAP	1.09*	1.02	0.96	1.07	1.10**	1.24***
PC-GAP	1.11	1.04	1.05	1.15	1.22**	1.44***
PCTVN-GAP	1.13*	1.06	1.05	1.12	1.17	1.34***
Term Structure VAR	1.06	0.99	0.98	1.08	1.23**	1.55***
TVP-VAR	1.10	0.98	1.03	1.08	1.19	1.51*
EWA	1.08	0.97	0.88**	0.96	1.01	1.16***
BMA	1.02	0.85**	0.8**	0.96	1.21	1.28
FAVAR	1.24**	1.16	1.17	0.96	1.03	1.16***
DSGE	1.02	1.00	0.96	1.01	0.90	0.91
DSGE-GAP	1.00	0.97	0.91*	0.96	0.84	0.87
BC	0.90	0.81**	0.83*	0.85**	0.90*	
SPF	0.89	0.80**	0.79**	0.80**	0.86	
GB	0.84	0.91	0.82	0.80	0.74	
Fixed ρ + nowcast	0.90	0.90***	0.98**	1.00	1.00	1.00

Notes: As for Table 2, except that only forecasts made for one of the 12 quarters after business cycle troughs are considered.

Table 9: Volatility of changes in five-to-ten-year forward inflation compensation at selected horizons

Horizon	Standard Deviation (Basis Points)	Variance Ratio Statistic
One day	5.1	
One month	21.4	-1.38
Three months	27.8	-2.34**
Six months	33.9	-2.13**

Notes: This table shows the standard deviation of one-day and one-, three- and six-month changes in the five–to-ten-year forward rate of inflation compensation. They are computed assuming 22 days per month. The variance ratio statistic is the heteroskedasticity robust test statistic of Lo and MacKinlay (1988) and has a standard normal asymptotic distribution. The sample period is from the start of 1999 to the end of 2011. One, two and three asterisks denote significance at the 10, 5 and 1 percent significance levels respectively.

Table 10: Comparison of Aggregate and Disaggregate Forecasts of Total CPI

	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$	$h=8$
AR in aggregates	2.65	2.68	2.70	2.80	2.75	2.83
- Fixed ρ	2.66	2.53	2.44	2.48	2.48	2.43
Univariate ARs in disaggregates	2.53	2.56	2.59	2.70	2.74	2.78
- No persistence in food and energy	2.47	2.51	2.55	2.64	2.64	2.65
- Fixed slopes	2.44	2.44	2.45	2.47	2.48	2.43
VAR in disaggregates	2.71	2.80	2.65	2.70	2.70	2.73
VAR in aggregates and disaggregates	2.71	2.83	2.66	2.70	2.71	2.73

Notes: This table reports the pseudo-out-of-sample recursive RMSPE of alternative h-quarter-ahead forecasts of total CPI inflation (quarter-over-quarter; annualized percentage points). The sample consists of data vintages from 1985Q1 to 2011Q4, with the data going back to 1960Q1 in all cases. The alternative forecasting methods considered are (i) an AR fitted to total CPI inflation, (ii) an AR(1) fitted to total inflation imposing that the slope coefficient is 0.46, (iii) univariate autoregressions in core, food and energy inflation, with the resulting forecasts combined using real-time CPI weights, (iv) univariate autoregressions in core, food and energy inflation imposing that the slope coefficients on food and energy are zero, (v) univariate autoregressions in core, food and energy with fixed slope coefficients, imposing that the slope coefficients are 0.46, 0 and 0, respectively, (vi) a VAR in core, food and energy inflation, and (vii) a VAR in overall, food and energy inflation. In all cases, all inflation series are used in gap form, relative to the trend in overall CPI inflation.

Table 11: RMSPEs of Core CPI Forecasts Evaluated as Forecasts of Total CPI

	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$	$h=8$
Direct	0.90	0.89	0.88	0.88	0.95	0.93
RAR	0.90	0.86	0.97	0.96	0.98	0.99
PC	0.89	0.88	0.88	0.88	0.95	0.93
Random Walk	0.74	0.70	0.77	0.74	0.76	0.71
AR-GAP	0.94	0.94	0.96	0.97	1.01	0.99
PC-GAP	0.93	0.94	0.96	0.97	1.00	0.98
PCTVN-GAP	0.93	0.94	0.96	0.97	1.00	1.00
Term Structure VAR	0.96	1.01	1.02	1.01	1.01	0.98
EWA	0.93	0.93	0.94	0.96	1.00	0.99
BMA	0.91	0.89	0.92	0.94	0.96	0.96
FAVAR	0.93	0.92	0.98	0.98	0.97	1.00

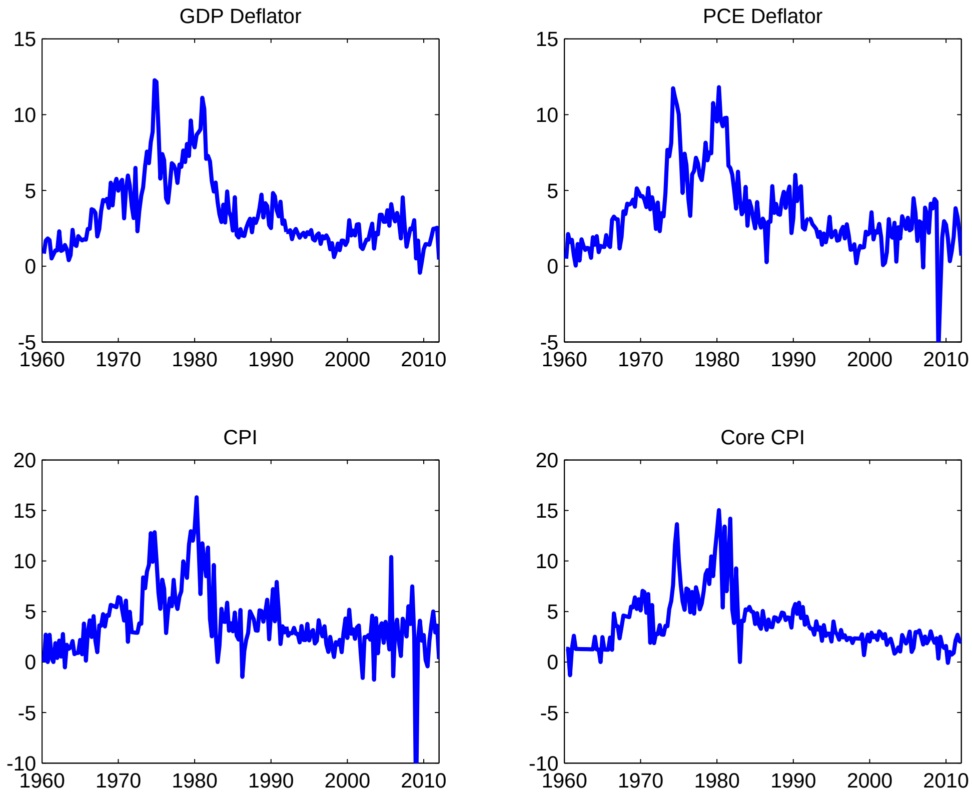
Notes: This table reports the pseudo-out-of-sample recursive RMSPE of h-quarter-ahead forecasts of core CPI inflation (quarter-over-quarter), evaluated as predictors of total CPI inflation, relative to the RMSPE of the corresponding direct forecasts of total CPI inflation. The sample consists of data vintages from 1985Q1 to 2011Q4, with the data going back to 1960Q1 in all cases.

Table 12: RMSPE of International Inflation Forecasts

Horizon	0	1	2	3	4	8
			Panel A: Canada			
Direct	1.04	1.11**	1.17**	1.10*	1.06	1.16**
RW	1.27***	1.36***	1.28**	1.39***	1.26**	1.22**
RW-AO	1.02	1.07	1.03	1.03	1.02	1.01
UCSV	0.97	1.04	1.05	1.05	1.01	0.99
PC	1.04	1.11*	1.16**	1.09	1.05	1.16***
AR-GAP	1.00	1.04	1.06	1.03	1.02	1.08**
PC-GAP	0.96	0.99	0.98	0.95	0.95	1.00
			Panel B: Germany			
Direct	1.01	0.95	0.98	1.00	0.99	1.12*
RW	1.15**	1.12*	1.07	1.22**	1.12	1.34**
RW-AO	1.01	0.96	0.96	1.02	1.04	1.16*
UCSV	0.98	0.96	0.96	1.01	0.97	1.13
PC	0.97	0.90*	0.92	0.96	0.94	1.05
AR-GAP	0.99	0.94	0.96	0.98	0.96	1.02
PC-GAP	1.01	0.94	0.95	1.00	0.98	1.14
			Panel C: Japan			
Direct	1.04	1.05	1.19***	1.42***	1.40***	1.55***
RW	1.15***	1.09	0.99	1.13	1.00	1.10
RW-AO	0.93*	0.89**	0.86*	0.88	0.85*	0.95
UCSV	0.91***	0.87**	0.85**	0.89	0.84*	0.91
PC	0.97	0.96	1.00	1.17	1.16	1.50*
AR-GAP	0.92***	0.87***	0.85***	0.88***	0.86***	0.87***
PC-GAP	0.97	0.92	0.92	1.06	1.00	1.22
			Panel D: United Kingdom			
Direct	1.04	1.07	1.10	1.24**	1.19	1.39**
RW	1.15*	1.12	1.04	1.16	1.14	1.18
RW-AO	0.97	0.91	0.88	0.90	0.93	1.03
UCSV	0.95	0.90	0.88	0.93	0.93	1.03
PC	1.14**	1.25**	1.48***	1.76***	1.78**	2.38***
AR-GAP	0.99	0.99	0.95	1.01	0.98	1.05*
PC-GAP	1.04	1.04	1.06	1.20*	1.19*	1.32***

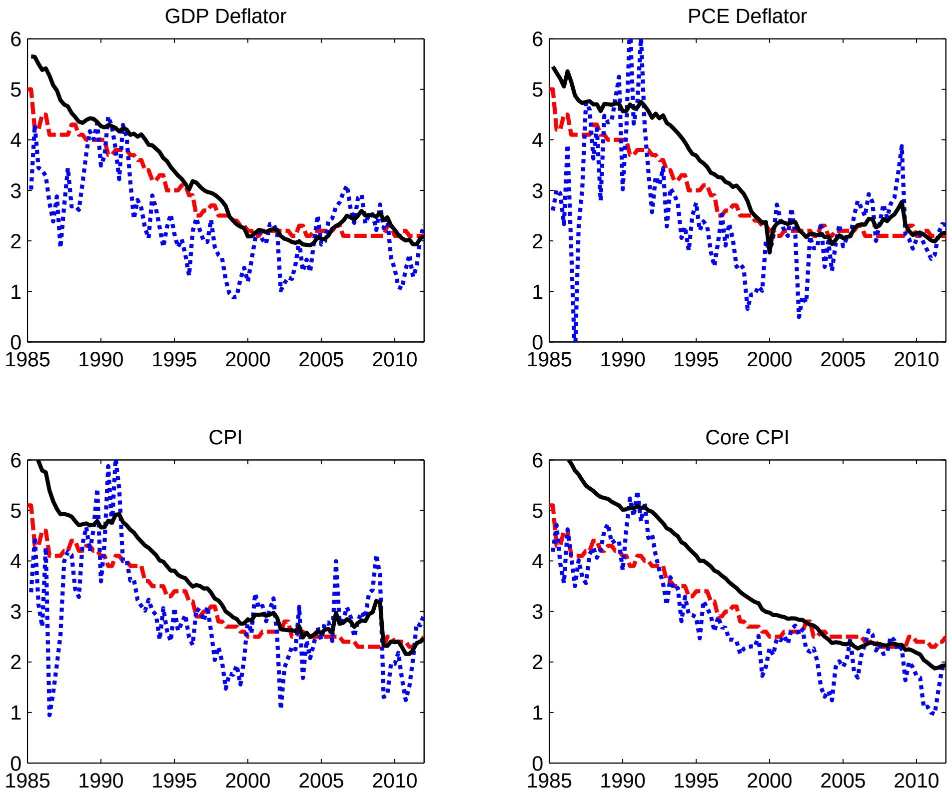
Notes: This table reports the pseudo-out-of-sample recursive RMSPE of alternative h-quarter-ahead forecasts of consumer price inflation (quarter-over-quarter) in four countries, all relative to the benchmark of an AR(1) in "gap" form with a fixed slope coefficient. Revised seasonally adjusted data were used. The sample is 1960Q1-2011Q4; the first pseudo-out-of-sample forecasts are made in 1985Q1, using data up to and including 1984Q4. Cases in which the relative root mean square prediction error is significantly different from one at the 10, 5 and 1 percent significance levels are denoted with one, two and three asterisks, respectively. These are based on the two-sided test of Diebold and Mariano (1995), implemented as described in the text.

Figure 1: Annualized Inflation Rates



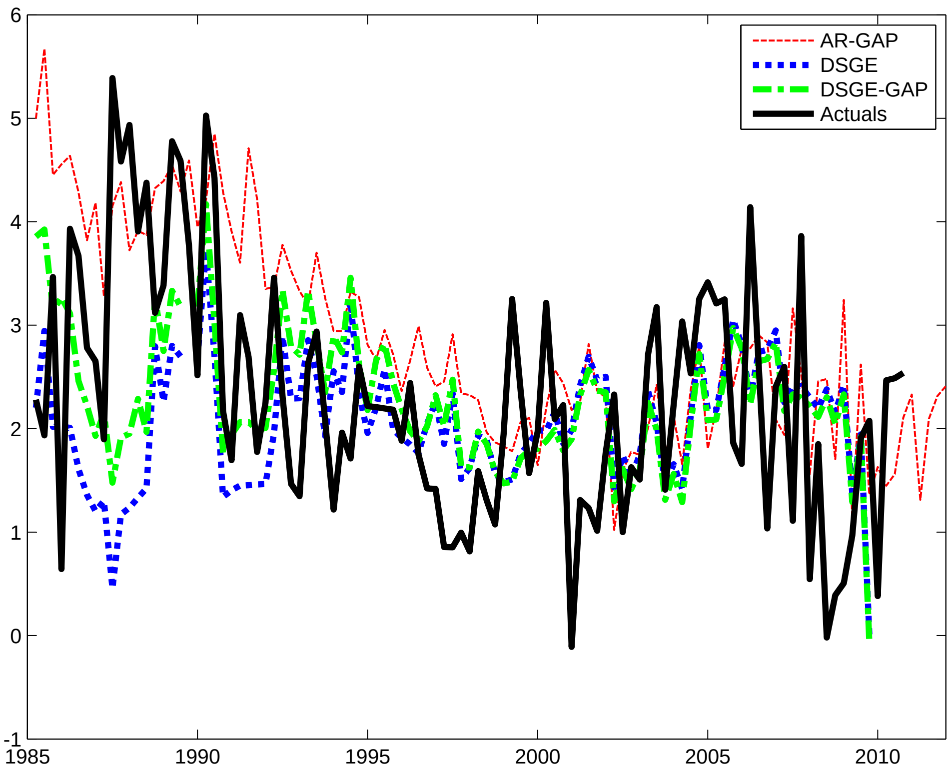
Notes: Quarter-over-quarter inflation rates corresponding to various price indexes (2011Q4 vintage data).

Figure 2: Alternative Measures of Trend Inflation



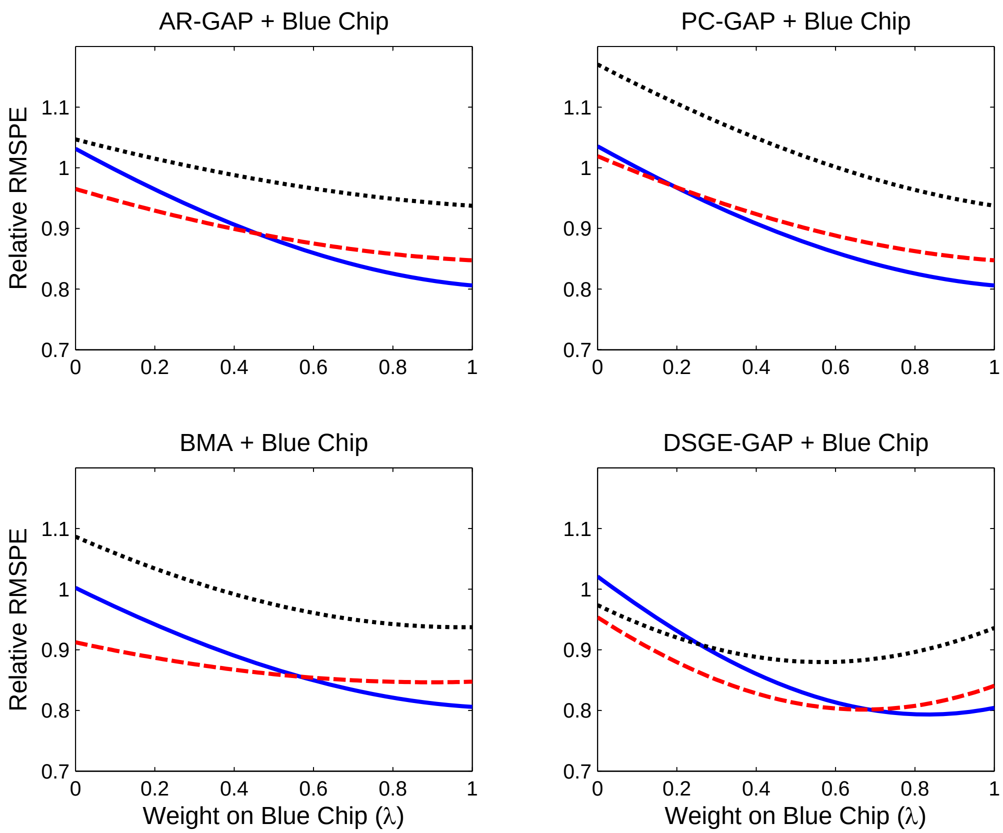
Note: Red dashes: Five-to-ten-year-ahead Blue Chip inflation projections (GDP deflator for the upper panels, CPI for the lower panels). Blue dots: the UCSV trend inflation series, i.e. the real-time filtered estimates of the final level of the permanent component of each of the four inflation measures. Black solid line: exponentially smoothed real-time inflation, using each of the four inflation measures, and a smoothing coefficient of 0.05.

Figure 3: Four-quarter-ahead GDP Deflator Inflation Forecasts and Realized Values



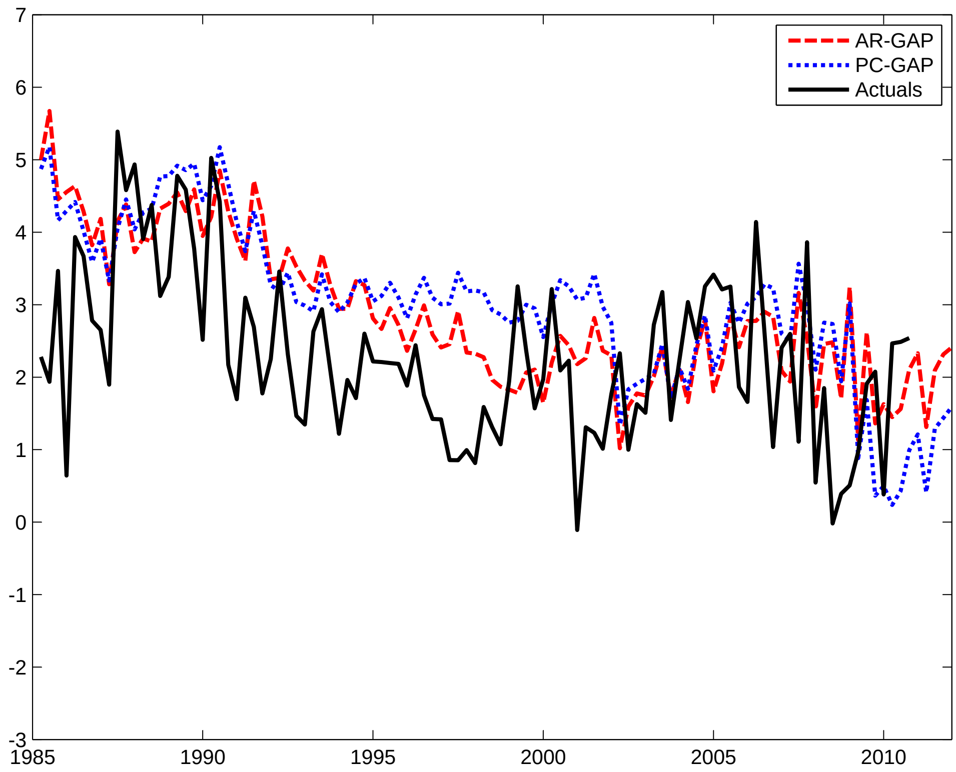
Notes: The forecasts extend through 2011Q4, but realized values are only available for forecasts made in 2010Q2 and earlier (given the convention of defining realized values as the data as observed in the middle of the second quarter after the quarter to which the data refer).

Figure 4: Relative RMSPEs of Combination Forecasts



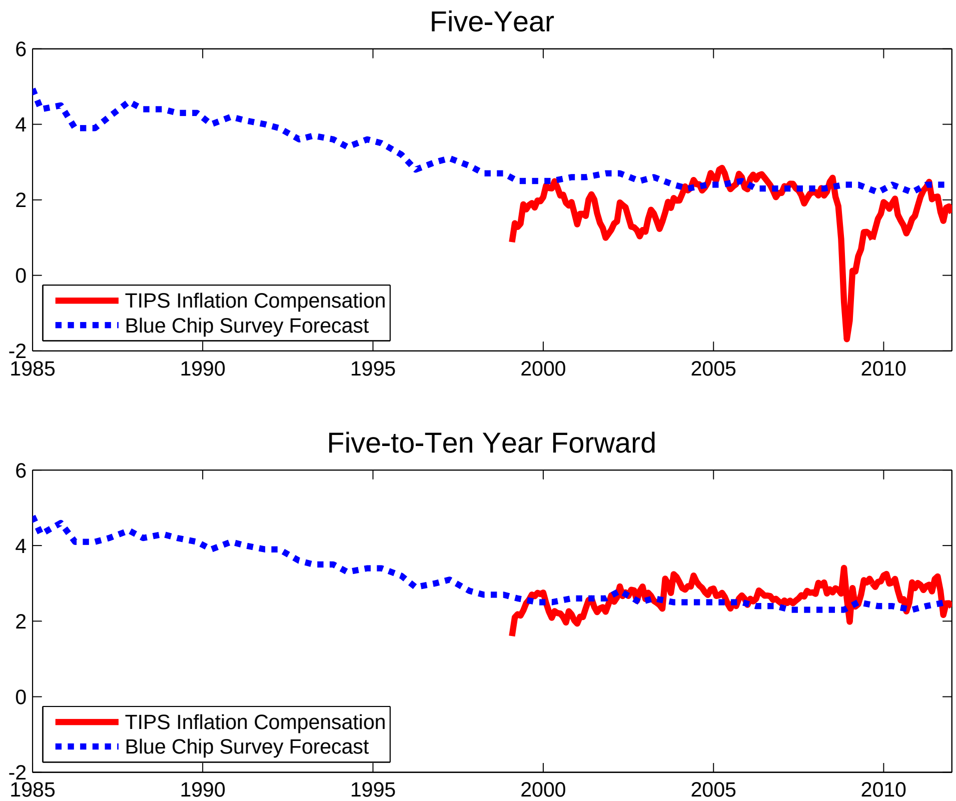
Notes: In each panel a combination forecast was formed that places weight λ on the Blue Chip forecast and $1-\lambda$ on some econometric forecast (real-time forecasts made 1985Q1-2011Q4). The blue solid line, red dashed line and black dotted line refer to current-quarter one-quarter-ahead and four-quarter-ahead forecasts, respectively.

Figure 5: Four-quarter-ahead GDP Deflator Inflation Forecasts and Realized Values



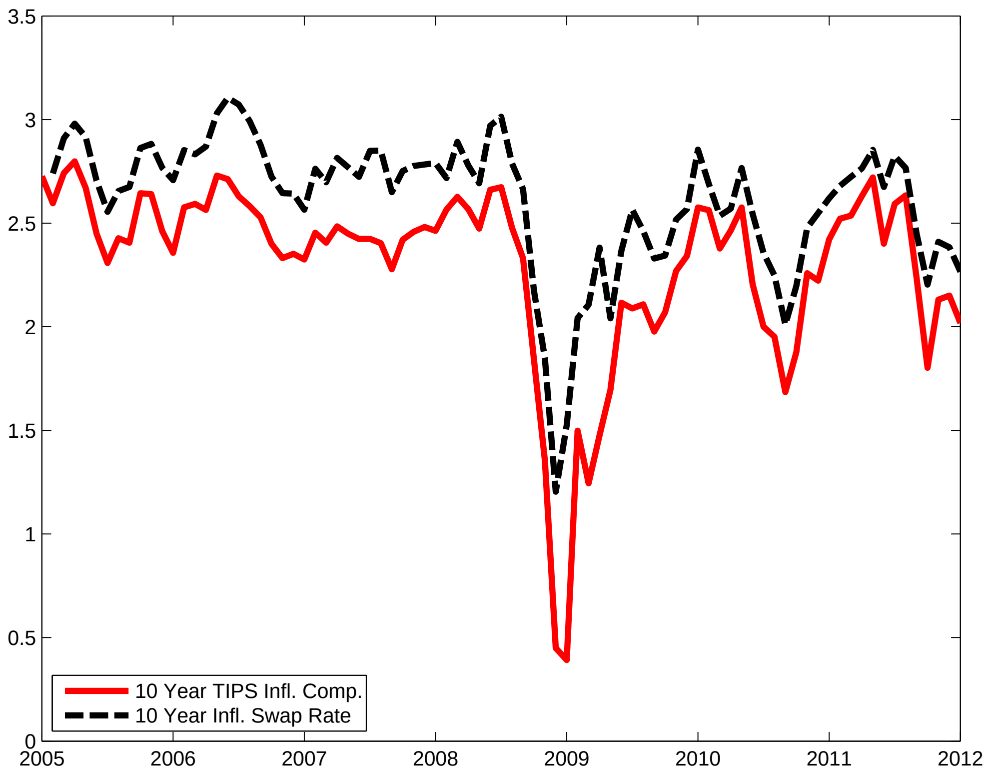
Notes: The forecasts extend through 2011Q4, but realized values are only available for forecasts made in 2010Q2 and earlier (given the convention of defining realized values as the data as observed in the middle of the second quarter after the quarter to which the data refer).

Figure 6: Inflation Compensation and Survey Forecasts



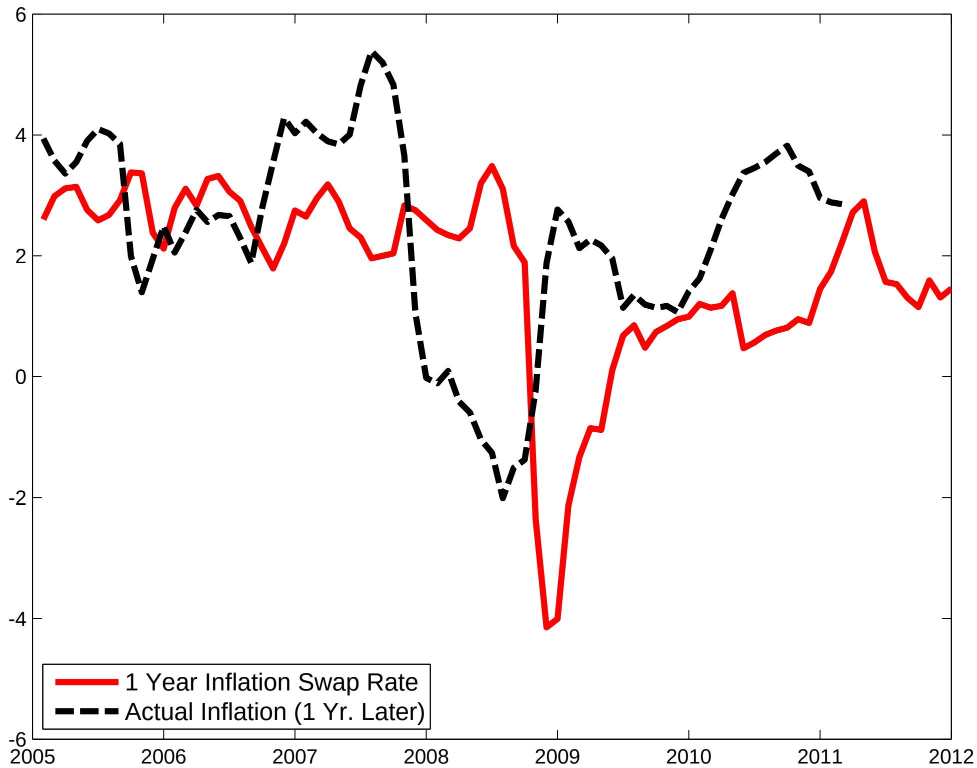
Notes: The panels of the figure show TIPS-based inflation compensation at five- and five-to-ten-year forward horizons and the corresponding maturity Blue Chip survey CPI forecasts. The TIPS-based inflation compensation series are from the dataset of Gürkaynak, Sack and Wright (2010): these data are available only back to 1999.

Figure 7: Ten-year TIPS-based Inflation Compensation and Inflation Swap Rates



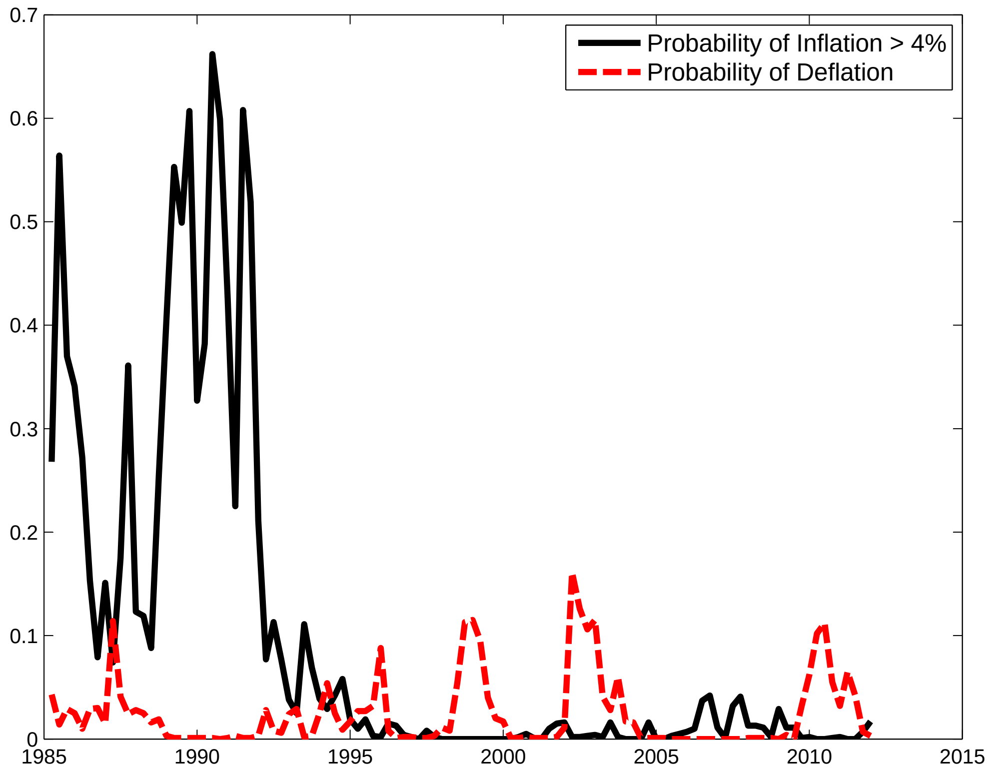
Notes: The TIPS-based inflation compensation series are from the dataset of Gürkaynak, Sack and Wright (2010). The inflation swap rates are from Bloomberg.

Figure 8: One-Year Inflation Swap Rates and Realized Inflation



Notes: The figure shows the one-year inflation swap rates, from Bloomberg. Subsequent realized values (of headline CPI inflation) are also shown: these are plotted against the month at which the forecast was made. Thus, perfect foresight forecasts would by construction line up perfectly with the subsequent realized values.

Figure 9: Probabilities of high and low inflation from the UCSV Model



Notes: The figure shows the probability of GDP deflator inflation being above 4 percent or below 0 percent on average over the subsequent two years. These are obtained from real-time estimation of the unobserved components stochastic volatility model of Stock and Watson (2007), applied to GDP deflator inflation.

References

[1] Adam, Klaus and Mario Padula (2003): Inflation Dynamics and Subjective Expectations in the United States, working paper.

[2] Ahmed, Shagil, Andrew T. Levin and Beth Anne Wilson (2004): Recent Improvements in US Macroeconomic Stability: Good Policy, Good Practices, or Good Luck?, *Review of Economics and Statistics*, 86, pp.824-832.

[3] Andreou, Elena, Eric Ghysels, and Andros Kourtellos (2008): Should macroeconomic forecasters look at daily financial data?, working paper.

[4] Aiolfi, Marco, Carlos Capistran and Allan Timmermann (2010): Forecast Combinations, in "Oxford Handbook of Economic Forecasting," Michael P. Clements and David F. Hendry (eds.), Oxford University Press, Oxford.

[5] Ang, Andrew, Geert Bekaert and Min Wei (2007): Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better?, *Journal of Monetary Economics*, 54, pp.1163-1212.

[6] Ang, Andrew and Monika Piazzesi (2003): A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables, *Journal of Monetary Economics*, 50, pp.745-787.

[7] Arai, Natsuki (2011): Using Forecast Evaluation to Improve Accuracy of the Greenbook Forecast, working paper.

[8] Atkeson, Andrew and Lee E. Ohanian (2001): Are Phillips Curves Useful for Forecasting Inflation?, *Federal Reserve Bank of Minneapolis Quarterly Review*, 25, pp.2-11.

[9] Bai, Jushan and Serena Ng (2008): Forecasting Economic Time Series Using Targeted Predictors, *Journal of Econometrics*, 146, pp.304-317.

[10] Ball, Laurence, N. Gregory Mankiw, and David Romer (1988): The New Keynesian Economics and the Output-Inflation Trade-Off, *Brookings Papers on Economic Activity*, 1, pp.1-65.

[11] Ball, Laurence and Sandeep Mazumder (2011): Inflation Dynamics and the Great Recession, *Brookings Papers on Economic Activity*, 1, pp.337-387.

[12] Banbura, Marta, Domenico Giannone, Michele Modugno and Lucrezia Reichlin (2012): Nowcasting, forthcoming in "Handbook of Economic Forecasting," Volume 2, Allan Timmermann and Graham Elliott (eds.), Elsevier, Amsterdam.

[13] Barnes, Michelle L. and Giovanni P. Olivei (2003): Inside and Outside Bounds: Threshold Estimates of the Phillips Curve, *New England Economic Review*, pp.1-18.

[14] Bates, John M. and Clive W.J. Granger (1969): The Combination of Forecasts, *Operations Research Quarterly*, 20, pp.451-468.

[15] Bermingham, Colin and Antonello D'Agostino (2011): Understanding and Forecasting Aggregate and Disaggregate Price Dynamics, , ECB Working Paper 1365.

[16] Bernanke, Ben S. and Jean Boivin (2003): Monetary Policy in a Data-Rich Environment, *Journal of Monetary Economics*, 50, pp.525-546.

[17] Bernanke, Ben S., Jean Boivin and Piotr Eliasz (2005): Measuring Monetary Policy: A Factor Augmented Vector Autoregressive (FAVAR) Approach, *Quarterly Journal of Economics*, 120, pp.387-422.

[18] Bernanke, Ben S. and Michael Woodford (1997): Inflation Forecasts and Monetary Policy, *Journal of Money, Credit and Banking*, 29, pp.653-684.

[19] Bilke, Laurent and Livio Stracca (2007): A Persistence-Weighted Measure of Core Inflation in the Euro Area, *Economic Modelling*, 24, pp. 1032 - 1047.

[20] Blinder, Alan S. (1997): Commentary on "Alternative Strategies for Aggregating Prices", *Federal Reserve Bank of Saint Louis Review*, May/June 1997, pp.139-142.

[21] Brayton, Flint, John. M. Roberts, and John C. Williams (1999). What's Happened to the Phillips Curve? FEDS Discussion Paper 99-49.

[22] Britton, Erik, Paul Fisher and John Whitley (1998): The Inflation Report projections: Understanding the Fan Chart, Bank of England Quarterly Bulletin 38, pp.30-37.

[23] Bryan, Michael F. and Stephen G. Cecchetti (1994): Measuring Core Inflation in "Monetary Policy" N. Gregory Mankiw (ed.), Chicago University Press, Chicago.

[24] Bryan, Michael F. and Brent Meyer (2010): Are Some Prices in the CPI More Forward Looking Than Others? We Think So, Federal Reserve Bank of Atlanta Economic Commentary, 2010-2.

[25] Campbell, John Y. and Robert J. Shiller (1996): A Scorecard for Indexed Government Debt, National Bureau of Economic Research Macroeconomics Annual, MIT Press.

[26] Campbell, John Y., Robert J. Shiller and Luis M. Viceira (2009): Understanding Inflation-Indexed Bond Markets, Brookings Papers on Economic Activity, 1, pp.79–120.

[27] Christensen, Jens H.E., Jose A. Lopez and Glenn D. Rudebusch (2011): Extracting Deflation Probability Forecasts from Treasury Yields, working paper 2011-10.

[28] Clark, Todd E. (2011): Real-Time Density Forecasts from VARs with Stochastic Volatility, *Journal of Business and Economic Statistics*, 29, pp.327-341.

[29] Clark, Todd E. and Taeyoung Doh (2011): A Bayesian Evaluation of Alternative Models of Trend Inflation, working paper.

[30] Clark, Todd E. and Michael W. McCracken (2009a): Nested Forecast Model Comparisons: A New Approach to Testing Equal Accuracy, Federal Reserve Bank of Kansas City working paper 2009-11.

[31] Clark, Todd E. and Michael W. McCracken (2009b): Tests of Equal Predictive Ability with Real-Time Data, *Journal of Business and Economic Statistics*, 27, pp.441-454.

[32] Clark, Todd E. and Michael W. McCracken (2012): Advances in Forecast Evaluation, forthcoming in "Handbook of Economic Forecasting," Volume 2, Allan Timmermann and Graham Elliott (eds.), Elsevier, Amsterdam.

[33] Clark, Todd E. and Kenneth D. West (2007): Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics*, 138, pp.291-311.

[34] Clements, Michael P. (2004): Evaluating the Bank of England Density Forecasts of Inflation, *Economic Journal*, 114, pp.855-877.

[35] Cogley, Timothy, Giorgio Primiceri, and Thomas J. Sargent (2010): Inflation-Gap Persistence in the U.S., *American Economic Journal: Macroeconomics* 2, 43-69.

[36] Cogley, Timothy and Thomas J. Sargent (2005): Drifts and Volatilities: Monetary Policies and Outcomes in the Post World War II U.S., *Review of Economic Dynamics*, 8, pp.262-302.

[37] Cogley, Timothy, and Argia M. Sbordone (2008): Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve, *American Economic Review*, 98: 2101–26.

[38] Croushore, Dean (2008): Revisions to PCE Inflation Measures: Implications for Monetary Policy, working paper.

[39] Croushore, Dean (2010): An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data, *BE Journal of Macroeconomics, Contributions* 10.

[40] Croushore, Dean (2012): Comment on "Understanding Inflation-Indexed Bond Markets", *Journal of Business and Economic Statistics*, 30, pp.17-20.

[41] Cutler, Joanne (2001): A New Measure of Core Inflation in the UK, working paper.

[42] D'Amico, Stefania, Don H. Kim and Min Wei (2010): Tips from TIPS: the informational content of Treasury Inflation-Protected Security prices, Finance and Economics Discussion Series working paper, 2010-19.

[43] De Graeve, Ferris, Marina Emiris and Rafael Wouters (2009): A Structural Decomposition of the U.S. Yield Curve, *Journal of Monetary Economics* 56, pp.545-559.

[44] Del Negro, Marco and Stefano Eusepi (2011): Fitting Observed Inflation Expectations, *Journal of Economic Dynamics and Control*, 35, pp.2105-2131.

[45] Diebold, Francis X. and Canlin Li, (2006): Forecasting the Term Structure of Government Bond Yields, *Journal of Econometrics*, 130, pp.337-364.

[46] Diebold, Francis X. and Roberto S. Mariano (1995): Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, pp.253-263

[47] Diebold, Francis X., Glenn D. Rudebusch and S. Boragan Aruoba (2006): The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach, *Journal of Econometrics*, 131, pp.309-338.

[48] Diebold, Francis X, Anthony S. Tay and Kenneth F. Wallis (1999): Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters, in "Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger", Robert F. Engle and Halbert White (eds.), Oxford University Press, Oxford.

[49] Van Dijk, Dick, Siem Jan Koopman, Michel van der Wel and Jonathan H. Wright (2011), Forecasting Interest Rates with Shifting Endpoints, working paper.

[50] Dotsey, Michael, Shigeru Fujita and Tom Stark (2011): Do Phillips Curves Conditionally Help to Forecast Inflation, working paper.

[51] Edge, Rochelle and Refet S. Gürkaynak (2010): How Useful are Estimated DSGE Model Forecasts for Central Bankers, *Brookings Papers on Economic Activity*, 2, pp.209-244.

[52] Edge, Rochelle, Michael T. Kiley and Jean-Philippe Laforte (2010): A Comparison of Forecast Performance Between Federal Reserve Staff Forecast, Simple Reduced-Form Models, and a DSGE Model, *Journal of Applied Econometrics*, 25, pp.720-754.

[53] Engle, Robert F. (1982): Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 50, pp.987-1007.

[54] Erceg, Christopher and Andrew T. Levin (2003): Imperfect Credibility and Inflation Persistence, *Journal of Monetary Economics*, 50, pp.915-944.

[55] Faust, Jon and Jonathan H. Wright (2008): Efficient forecast tests for conditional policy forecasts, *Journal of Econometrics*, 146, pp.293-303.

[56] Faust, Jon and Jonathan H. Wright (2009): Comparing Greenbook and Reduced Form Forecasts Using a Large Real-time Dataset, *Journal of Business and Economic Statistics*, 27, pp.468-479.

[57] Faust, Jon and Jonathan H. Wright (2011): Efficient Prediction of Excess Returns, *Review of Economics and Statistics*, 93, pp.647-659.

[58] Fernandez, Carmen, Eduardo Ley and Mark F.J. Steel (2001): Model Uncertainty in Cross-Country Growth Regressions, *Journal of Applied Econometrics*, 16, pp.563-576.

[59] Filardo, Andrew J. (1998): New Evidence on the Output Cost of Fighting Inflation, *Federal Reserve Bank of Kansas City Economic Review* (3rd quarter), pp.33-61.

[60] Fleckenstein, Matthias, Francis A. Longstaff and Hanno Lustig (2010): Why does the Treasury Issue TIPS? The TIPS-Treasury Bond Puzzle, NBER working paper 16358.

[61] Gali, Jordi and Mark Gertler (1999): Inflation Dynamics: A Structural Econometric Analysis, *Journal of Monetary Economics*, 44, pp.195-222.

[62] Giacomini, Raffaella and Halbert White (2006): Tests of Conditional Predictive Ability, *Econometrica*, 74, pp.1545-1578.

[63] Giacomini, Raffaella and Barbara Rossi (2010): Forecast Comparisons in Unstable Environments, *Journal of Applied Econometrics*, 25, pp.595-620.

[64] Giacomini, Raffaella and Barbara Rossi (2012): Advances in Forecasting under Model Instability, forthcoming in "Handbook of Economic Forecasting," Volume 2, Allan Timmermann and Graham Elliott (eds.), Elsevier, Amsterdam.

[65] Giannone, Domenico, Lucrezia Reichlin and David Small (2008): Nowcasting GDP: The real time Informational Content of Macroeconomic Data Releases, *Journal of Monetary Economics*, 55, pp.665-676.

[66] Gioradani, Paolo and Paul Söderlind (2003): Inflation Forecast Uncertainty, *European Economic Review*, 47, pp.1037-1059.

[67] Gonzáles, Andrés, Kirstin Hubrich and Timo Teräsvirta (2011): Forecasting Inflation with Gradual Regime Shifts and Exogenous Information, ECB Working Paper 1363.

[68] Gordon, Robert J. (1982): Inflation, Flexible Exchange Rates, and the Natural Rate of Unemployment, in "Workers, Jobs and Inflation", Martin N. Baily (ed). The Brookings Institution, Washington.

[69] Gordon, Robert J. (1998): Foundations of the Goldilocks Economy: Supply Shocks and the Time-Varying NAIRU, *Brookings Papers on Economic Activity*, 2, pp.297-333.

[70] Greenspan, Alan (1992): Statement to the Congress, Federal Reserve Bulletin, August 1992, pp.603-607.

[71] Groen, Jan J.J., George Kapetanios and Simon Price (2009): A Real Time Evaluation of Bank of England Forecasts of Inflation and Growth, *International Journal of Forecasting*, 25, pp.74-80.

[72] Groen, Jan J.J., Richard Paap and Francesco Ravazzolo (2009): Real-Time Inflation Forecasting in a Changing World, Federal Reserve Bank of New York Staff Reports, 388.

[73] Gürkaynak, Refet S., Brian P. Sack and Eric T. Swanson (2005): The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models, *American Economic Review*, 95, pp.425-436.

[74] Gürkaynak, Refet S., Brian P. Sack and Jonathan H. Wright (2007): The U.S. Treasury Yield Curve: 1961 to the Present, *Journal of Monetary Economics*, 54, pp.2291-2304.

[75] Gürkaynak, Refet S., Brian P. Sack and Jonathan H. Wright (2010): The TIPS Yield Curve and Inflation Compensation, *American Economic Journal: Macroeconomics*, 2, pp.70-92.

[76] Harvey, David I., Stephen J. Leybourne and Paul Newbold (1997): Testing the Equality of Prediction Mean Squared Errors, *International Journal of Forecasting*, 13, pp.281-291.

[77] Haubrich, Joseph G., George G. Pennacchi and Peter H. Ritchken (2008): Estimating Real and Nominal Term Structures using Treasury Yields, Inflation, Inflation Forecasts, and Inflation Swap Rates, Federal Reserve Bank of Cleveland working paper 2008-10.

[78] Hendry, David and Kirstin Hubrich (2011): Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate, *Journal of Business and Economic Statistics*, 29, pp.216-227.

[79] Hu, Gang, and Mihir Worah (2009): Why TIPS Real Yields moved significantly higher after the Lehman Bankruptcy, PIMCO, Newport Beach, California.

[80] Hubrich, Kirstin (2005): Forecasting Euro Area Inflation: Does Aggregating Forecasts by HICP component Improve Forecast Accuracy?, *International Journal of Forecasting*, 21, pp.119-136.

[81] Inoue, Atsushi and Lutz Kilian (2004): In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?, *Econometric Reviews*, 23, pp.371-402.

[82] Inoue, Atsushi and Lutz Kilian (2006): On the Selection of Forecasting Models, *Journal of Econometrics*, 130, pp.273-306.

[83] Inoue, Atsushi and Lutz Kilian (2008): How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation, *Journal of the American Statistical Association*, 103, pp.511-522.

[84] Inoue, Atsushi, Fatma B. Kiraz and Lutz Kilian (2009): Do Actions Speak Louder than Words? Household Expectations of Inflation Based on Micro Consumption Data, *Journal of Money, Credit and Banking*, 41, pp.1331-1363.

[85] Joslin, Scott, Marcel Priesbsch and Kenneth J. Singleton (2010): Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks, working paper.

[86] Joslin, Scott, Anh Le and Kenneth J. Singleton (2012): Why Gaussian Macro-Finance Term Structure Models are (Nearly) Unconstrained Factor-VARs, *Journal of Financial Economics*, forthcoming.

[87] Joyce, Michael, Peter M. Lildholdt and Steffen Sorensen (2010): Extracting Inflation Expectations and Inflation Risk Premia from the Term Structure: A Joint Model of the UK Nominal and Real Yield Curves, *Journal of Banking and Finance*, 34, pp.281–94.

[88] Kilian, Lutz and Simone Manganelli (2007): Quantifying the Risk of Deflation, *Journal of Money, Credit and Banking*, 39, pp.561-590.

[89] Kilian, Lutz and Simone Manganelli (2008): The Central Banker as a Risk Manager: Estimating the Federal Reserve's Preferences under Greenspan, *Journal of Money, Credit and Banking*, 40, pp.1103-1129.

[90] Kitsul, Yuriy and Jonathan H. Wright (2012): The Economics of Options-Implied Inflation Probability Density Functions, working paper.

[91] Kleibergen, Frank and Sophocles Mavroeidis (2009): Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve, *Journal of Business and Economic Statistics*, 27, pp.293-311.

[92] Koenig, Evan F. and Tyler Atkinson (2012): Inflation, Slack and Fed Credibility, Dallas Fed Staff Paper 16.

[93] Koop, Gary and Simon Potter (2003): Forecasting in Large Macroeconomic Panels Using Bayesian Model Averaging, Federal Reserve Bank of New York Staff Report 163.

[94] Kozicki, Sharon and Peter A. Tinsley (2001): Shifting Endpoints in the Term Structure of Interest Rates, *Journal of Monetary Economics*, 47, pp.613-652.

[95] Kozicki, Sharon and Peter A. Tinsley (2005): What do You Expect? Imperfect Policy Credibility and Tests of the Expectations Hypothesis, *Journal of Monetary Economics*, 52, pp.421-447.

[96] Kozicki, Sharon and Peter A. Tinsley (2012): Effective Use of Survey Information in Estimating the Evolution of Expected Inflation, *Journal of Money, Credit and Banking*, 44, pp.145-169.

[97] Kwon, So Jung (2011): Can Judgmental Forecasts help DSGE Forecasting?, working paper.

[98] Lo, Andrew W. and A.Craig Mackinlay (1988): Stock Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test, *Review of Financial Studies*, 1, pp. 41-66.

[99] Liu, Zheng and Glenn D. Rudebusch (2010): Inflation: Mind the Gap, Federal Reserve Bank of San Francisco Economic Letter 2010-02.

[100] Lütkepohl, Helmut (1987). Forecasting Aggregated Vector ARMA Processes, Springer Verlag, Berlin.

[101] Manzan, Sebastiano and Dawit Zerom (2009): Are Macroeconomic Variables Useful for Forecasting the Distribution of U.S. Inflation?, working paper.

[102] Marcellino, Massimo, James H. Stock and Mark W. Watson (2006): A Comparison of Direct and Iterated Multistep AR methods for Forecasting Macroeconomic Time Series, *Journal of Econometrics*, 135, pp.499-526.

[103] Meese, Richard A. and Kenneth Rogoff (1983): Empirical Exchange Rate Models of the Seventies? Do they fit out of sample?, *Journal of International Economics*, 14, pp.3-24.

[104] Meier, André (2010) Still Minding the Gap—Inflation Dynamics during Episodes of Persistent Large Output Gaps, IMF Working Paper 10-189.

[105] Patton, Andrew J. and Allan Timmermann (2007): Properties of Optimal Forecasts Under Asymmetric Loss and Nonlinearity, *Journal of Econometrics*, 140, pp.884-918.

[106] Patton, Andrew J. and Allan Timmermann (2012): Forecast Rationality Tests Based on Multi-horizon Bounds, *Journal of Business and Economic Statistics*, 30, pp.1-17.

[107] Pflueger, Carolin E. and Luis M. Viceira (2011): Inflation-Indexed Bonds and the Expectations Hypothesis, Annual Review of Financial Economics, 3, pp.139-158.

[108] Phillips, A. William (1958): The Relationship between Unemployment and the Rate of Change of Money Wages in the United Kingdom 1861-1957, *Economica*, 25, pp.283–299.

[109] Primiceri, Giorgio E. (2005): Time Varying Structural Vector Autoregressions and Monetary Policy, *Review of Economic Studies*, 72, pp.821-852.

[110] Reifschneider, David L., David J. Stockton and David W. Wilcox (1997): Econometric Models and the Monetary Policy Process, *Carnegie Rochester Series on Public Policy*, 47, pp.1-37.

[111] Reifschneider, David L., and John C. Williams (2000): Three Lessons for Monetary Policy in a Low-Inflation Era, *Journal of Money, Credit and Banking*, 32, pp.936-966.

[112] Romer, Christina D. and David H. Romer (2000): Federal Reserve Information and the Behavior of Interest Rates, *American Economic Review*, 90, pp.429-457.

[113] Rossi, Barbara and Tatevik Sekhposyan (2010): Have Economic Models' Forecasting Performance for US Output Growth and Inflation Changed Over Time, and When?, International Journal of Forecasting, 26, pp.808-835.

[114] Rudd, Jeremy and Karl Whelan (2005): New Tests of the New-Keynesian Phillips Curve, *Journal of Monetary Economics*, 52, pp.1167-1181.

[115] Sims, Christopher A. (1993): A Nine-Variable Probabilistic Macroeconomic Forecasting Model in "NBER Studies in Business: Business Cycles, Indicators, and Forecasting," Volume 28, James H. Stock and Mark Watson (eds.), National Bureau of Economic Research, Cambridge.

[116] Sims, Christopher A. (2002): The Role of Models and Probabilities in the Monetary Policy Process, *Brookings Papers on Economic Activity*, 2, pp.1-40.

[117] Smets, Frank and Raf Wouters (2003): An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area, *Journal of the European Economic Association*, 1, pp.1123-1175.

[118] Smets, Frank and Raf Wouters (2007): Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach, *American Economic Review*, 97, pp.586-607.

[119] Stock, James H. and Mark W. Watson (1999): Forecasting Inflation, *Journal of Monetary Economics*, 44, pp.293-335.

[120] Stock, James H. and Mark W. Watson (2003a:): Has the Business Cycle Changed? Evidence and Explanations, Federal Reserve Bank of Kansas City Symposium: Monetary Policy and Uncertainty, Jackson Hole, Wyoming.

[121] Stock, James H. and Mark W. Watson (2003b): Forecasting Output and Inflation: The Role of Asset Prices, *Journal of Economic Literature*, 41, pp.788-829.

[122] Stock, James H. and Mark W. Watson (2007): Has Inflation Become Harder to Forecast?, *Journal of Money, Credit and Banking*, 39, pp.3-34.

[123] Stock, James H. and Mark W. Watson (2009): Phillips Curve Inflation Forecasts, in "Understanding Inflation and the Implications for Monetary Policy", Jeffrey Fuhrer, Yolanda Kodrzycki, Jane Little, and Giovanni Olivei (eds), MIT Press, Cambridge.

[124] Stock, James H. and Mark W. Watson (2010): Modeling Inflation After the Crisis, in *Macroeconomic Challenges: the Decade Ahead*, Federal Reserve Bank of Kansas City Economic Policy Symposium.

[125] Svensson, Lars E.O. (2005): Monetary Policy with Judgment: Forecast Targeting, *International Journal of Central Banking*, 1, pp.1-54.

[126] Tetlow, Robert J., and Brian Ironside (2007). Real-Time Model Uncertainty in the United States: The Fed, 1996-2003, *Journal of Money, Credit and Banking*, 39, pp.1533-1849.

[127] Tulip, Peter (2009): Has the Economy Become More Predictable? Changes in Greenbook Forecast Accuracy, *Journal of Money, Credit and Banking*, 39, pp.1217-1231.

[128] Woodford, Michael (2003): Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton University Press, Princeton, New Jersey.

[129] Wright, Jonathan H. (2009a): Forecasting U.S. Inflation by Bayesian Model Averaging, *Journal of Forecasting*, 28, pp.131-144.

[130] Wright, Jonathan H. (2009b): Comment on "Understanding Inflation-Indexed Bond Markets", *Brookings Papers on Economic Activity*, 1, pp.126–38.

[131] Wright, Jonathan H. (2012): Evaluating Real-Time VAR Forecasts with an Informative democratic Prior, *Journal of Applied Econometrics*, forthcoming.