Forecasting US inflation using dynamic general-to-specific model selection

George Bagdatoglou [$], Alexandros Kontonikas [+] and Mark E. Wohar [**]

Abstract

We forecast US inflation using a standard set of macroeconomic predictors and a dynamic model selection and averaging methodology that allows the forecasting model to change over time. Pseudo out-of-sample forecasts are generated from models identified from a multipath general-to-specific algorithm that is applied dynamically using rolling regressions. Our results indicate that the inflation forecasts that we obtain employing a short rolling window substantially outperform those from a well-established univariate benchmark, and contrary to previous evidence, are considerably robust to alternative forecast periods.

JEL: C22, C52, E31, E37.
Keywords: Inflation forecasting, dynamic general-to-specific.

[$] Deloitte Economic Consulting, London, email gbagdatoglou@deloitte.co.uk.
[+] University of Glasgow, Adam Smith Business School, email: alexandros.kontonikas@glasgow.ac.uk.
[**] University of Nebraska-Omaha, Department of Economics, email: mwohar@unomaha.edu.

1. Introduction

In a recent comprehensive study, Stock and Watson (2009) find that the success of Phillips curve forecasts of inflation is episodic. In line with previous work by Atkeson and Ohanian (2001), they show that since the mid-1980s there is little evidence that forecasts from a univariate benchmark model can be consistently improved upon. Nevertheless, they also find that there are periods in time, typically close to business cycle turning points (see also Stock and Watson, 2010), when economic fundamentals have useful predictive content even post the mid-1980s.

It is widely recognised that predictive failure is closely related to model instability, see e.g. Clements and Hendry (1998a,b) and Stock and Watson (1996). As Hendry and Clements (2003) point out, good forecasts are relying on the assumption that the model is a good representation of the economy, the structure of which remains relatively unchanged. More recently, Rossi (2012) provides a comprehensive review of issues related to forecasting in the presence of structural breaks. She stresses that predictive content instability implies that finding predictors that work well in one period is no guarantee of future success, and concludes that it is essential to improve methods to select good forecasting models in-sample.

Our paper's main contribution is related to this latter point. Specifically, we apply a general-to-specific (GETS) model selection algorithm similar in spirit to Autometrics (see Doornik, 2009), using rolling regressions, to select models in-sample that are subsequently employed for forecasting. At each rolling sub-sample, the GETS algorithm begins with a general model that includes all predictors, and applies model reduction on the basis of standard *t*- and *F*-tests considering, in principle, the whole model space, i.e. it is multipath in nature.[1] The combination of the GETS model selection, with rolling regressions, a scheme

[1] In single path model reduction algorithms such as the stepwise, model reduction is carried out by deleting one variable at a time, the most insignificant one, until all remaining variables are statistically significant, in which case the algorithm stops and a terminal model is reached. Our algorithm, searches multiple paths by considering

frequently employed when forecasting under structural breaks (see e.g. Pesaran and Timmerman, 2007), allows the predictive model to change over time. In other words, at each point of time, we take an agnostic view as to which fundamentals have predictive power, use the GETS algorithm to identify the fundamentals that are likely to contain useful predictive information, and employ these fundamentals to forecast out-of-sample. This approach allows not only the parameters of the forecasting model to change over time but also the actual predictors that enter the model to vary, hence allowing for a more severe form of model instability.[2]

Given the multipath nature of the GETS algorithm, multiple terminal models may be reached at any rolling sub-sample. In the case when more than one terminal model is reached, we employ two alternative approaches to obtain forecasts: first, we follow Doornik (2009) and use model-fit criteria in order to select a single terminal model; second, we combine forecasts from all terminal models, using various averaging methods, to further robustify our forecasts to potentially imprecise parameter estimates from individual models.[3]

Our approach is similar in spirit to Koop and Korobilis (2012) who implement the idea of dynamic model selection and averaging within a Bayesian framework and find that inflation forecasting performance can be substantially improved relative to univariate approaches. It is also related to Castle et al. (2012) who use a dynamic GETS approach via

---

alternative model reduction strategies. One of the paths that the GETS algorithm searches corresponds to the single stepwise path described above. An alternative path is searched by deleting instead of the most insignificant variable, the second, third, etc. most insignificant variable. This search can give rise to multiple terminal models.

[2] In the forecasting literature, model instability is typically considered with regards to model parameters and accounted for by utilising recursive or rolling estimation procedures.

[3] See Hendry and Clements (2002) for a discussion on how forecast pooling can generally improve forecast performance in the presence of model misspecification and structural change.

Autometrics to forecast US GDP and document benefits from multipath model selection.[4] Pesaran and Timmermann (1995) also apply a model selection approach based on model fit criteria to forecast US stock returns, and find significant variation in the predictive power of economic factors over time. Finally, Avramov (2002) finds that Bayesian model averaging generates lower forecast errors for US stock returns relative to model selection approaches that are based on model fit criteria.

We evaluate the out-of-sample forecast performance of the dynamic GETS procedure for US inflation using the Stock and Watson (2009) dataset. Stock and Watson (2009) examine the performance of 192 different forecasting procedures across 5 alternative measures of inflation, using 15 macroeconomic predictors, a long span of data and 6 alternative sub-samples. We build upon their extensive analysis to provide further insights on the ability of economic fundaments to forecast US inflation. Forecast performance is evaluated relative to the unobserved components stochastic volatility (UC-SV) model, the univariate model that is the most difficult to beat (see Stock and Watson, 2009). The principal result from our analysis is that the dynamic GETS methodology, combined with a short rolling window and a standard set of macroeconomic predictors of inflation, can substantially outperform the UC-SV benchmark across alternative forecast periods.

The remainder of the paper is structured as follows. Section 2 describes the dataset. Section 3 explains the econometric methodology. Section 4 presents and discusses the

[4] Banerjee and Marcellino (2006) also employ an automated model selection procedure using PcGets, Autometrics' predecessor, to forecast US inflation. Nevertheless, when pseudo out-of-sample forecasts are considered, the model selection algorithm is used to choose the optimum lag length in single-predictor dynamic models rather than to determine the best set of predictors in a multivariate context like in our case. Aron and Muellbauer (2012) use Autometrics to forecast US inflation but the model selection algorithm is applied only once, for the first estimation sample. They note that applying the model selection recursively would probably result in better forecast performance.

forecasting results. Section 5 provides evidence from robustness checks, and Section 6 concludes.

2. Data

The dataset is taken from Stock and Watson (2009), has quarterly frequency and covers the US over the period 1953.Q1-2008.Q1.[5] The data include 5 measures of inflation based upon: the CPI for all items (CPI-all), CPI excluding food and energy (CPI-core), personal consumption expenditure deflator (PCE-all), personal consumption expenditure deflator excluding food and energy (PCE-core), and the GDP deflator. 15 potential predictors of inflation are also included, mainly reflecting activity variables such as unemployment, GDP, and industrial production, but also other predictors proposed by economic theory (e.g., the term spread, a trade-weighted exchange rate index, etc.).[6] In line with the inflation variable used in Stock and Watson (2009), we calculate the annual (four-quarter) inflation at quarter *t*:

$$\pi_t^4 = 100 \ln(P_t / P_{t-4}) \tag{1}$$

where P_t is the relevant price index at quarter *t*.

3. Methodology

3.1 Model selection algorithm

The methodology is based on a GETS model selection algorithm, which is similar in spirit to Autometrics (Doornik, 2009), an automated model selection algorithm embedded in

[5] The data were obtained from M. Watson's website:http://www.princeton.edu/~mwatson/ddisk/bfed_Sept2008.zip.

[6] The full set of predictors is shown in Table A1 in the Appendix, together with the transformations that we apply so that the variables are stationary.

the OxMetrics econometrics software.[7] The starting point of the model selection process is the definition of a general unrestricted model (GUM), which should be formulated on the basis of theory, encompass competing models and provide sufficient information on the process that is being modelled (see Hendry and Krolzig, 2005; Doornik, 2009). Model reduction is carried out by removing statistically insignificant variables, and a terminal model is reached when all variables are statistically significant at a pre-specified level.

[FIGURE 1 HERE]

The algorithm considers, in principle, the whole model space, thus it is multipath in nature. In order to demonstrate how the algorithm works, consider for example that the GUM includes four explanatory variables (A, B, C and D) as shown in Figure 1. If all four variables are statistically significant the search terminates and the GUM is the terminal model. If, on the other hand, the GUM includes statistically insignificant variables, these are deleted one at a time based on their individual significance. If, for example, only variable A is insignificant, the GUM is reduced to BCD, which itself becomes the basis for another search. If all variables in the GUM are statistically insignificant, the algorithm removes each of them, one at the time, reaching four models: BCD, ACD, ABD and ABC. The reduction process is repeated at each of these four nodes.[8] If at each node all variables are insignificant, the algorithm will visit all 16 (=2^4) unique models represented by the solid dots in Figure 1.[9]

[7] Following PcGets (Hendry and Krolzig, 2001), Autometrics is the second generation automatic model selection algorithm in OxMetrics.

[8] For instance, if all three variables are insignificant at node BCD, the algorithm will consider three two-variable models: CD, BD and BC. If statistically insignificant variables are included in these two-variable models the search will continue. If, on the other hand, all variables at node ACD are significant, then ACD is a terminal model and no other variable combinations are searched in this branch.

[9] There are 15 unique models with at least one variable and one empty model omitted from Figure 1. Hollow dots represent duplicated models and can be ignored.

The variable combinations implied by the solid dots also reflect all possible terminal models the algorithm could reach.[10]

This multipath search offers two advantages over single path model selection procedures. First, given the multipath nature, it avoids the risk of getting trapped in the wrong search path (Doornik, 2009). In principle, this risk is higher the greater the intercorrelation of the explanatory variables. Second, it generates alternative valid reductions of the general model, and hence is more informative as to which models or variables have predictive power. This rich set of information can be utilised by means of model averaging to potentially improve forecast performance compared to single models.

In our setup, the GUM is given by the following equation and to avoid over-fitting, the 1% level of significance is utilised for model reduction:

$$\pi_{t+4}^{4} = \mu + \alpha\pi_{t}^{4} + \boldsymbol{\beta}\mathbf{X}_{t} + \gamma trend_{t+4} + v_{t+4} \qquad (2)$$

where μ is the constant; π_{t+4}^{4} is the four-quarters ahead annual inflation; $\mathbf{X}_{t}$ is a vector that contains the 15 potential predictors of inflation; and $trend_t$ is a linear trend. [11, 12]

With 17 independent variables in the GUM shown in Equation (2), there are up to (≈) 131,000 unique models, which make the search process computationally intensive. We adopt two strategies to move through the nodes efficiently. First, following Doornik (2009), we

[10] Note that it is difficult to assess the number of terminal models the algorithm is more likely to produce on the basis of the number of significant variables in the initial model. If just one variable is statistically significant, for instance, variable A, the algorithm could search all paths except the ones starting from BCD. In turn, the algorithm could identify several terminal models, if some of the insignificant variables turn significant as the model reduction progresses. But it could also lead to zero terminal models if A, the significant variable in the initial model, turns insignificant, and none of the other variables becomes significant at different stages of model reduction.

[11] The t-statistics are computed using Newey and West (1987) heteroskedasticity and autocorrelation consistent (HAC) standard errors based on a lag truncation of 3.

[12] Note that the constant is always included in the model(s).

consider deleting more than one variable at a time based on individual and joint significance levels. At each node, highly insignificant variables are grouped together, tested for joint significance and removed if they fail the test. There is obviously a trade-off between computational speed and the possibility of skipping unique terminal models. For the group deletion strategy, we define highly insignificant variables as those having *t*-statistics less than 0.5 (in absolute value), and the joint significance test is performed at the 1% level of significance.

The second strategy is based on the sign of the coefficients. We impose theory-consistent sign restrictions on the model space: if a variable is statistically significant but exhibits the 'wrong' sign, then it is deleted. Effectively, the sign restrictions impose priors on the model space to speed up the search process but also to improve the selection test power and ensure that the terminal model conforms to economic theory, at least in terms of coefficient signs (Hendry and Krolzig, 2005).[13] The group and sign deletion strategies are considered before the individual significance criterion, which is ignored if one or more variables are removed as a result of the aforementioned strategies.

The general-to-specific algorithm is applied dynamically (DGETS) across rolling windows of 20, 40 and 60 quarters always starting from the GUM shown in Equation (2). A smaller rolling window is expected to provide greater forecast gains when there are big and recurrent breaks (Pesaran and Timmermann, 2007).

3.2 Forecasting

[13] Theory-based restrictions have been used in studies of stock market predictability and have been shown to improve forecast performance (see e.g. Campbell and Thomson, 2008). The sign restrictions that we impose are shown in Table A1 in the Appendix. In line with economic intuition and theory they indicate a positive response of inflation to higher real economic activity, lower unemployment, higher long-term interest rates (relative to short-term rates), and exchange rate depreciation.

Pseudo out-of-sample forecasts are generated from the terminal model(s) that were identified using the automated model selection approach described in Section 3.1. Given the multipath nature of the algorithm, multiple terminal models may be reached at any rolling sub-sample.[14] In the case when more than one terminal model is reached, two methods are employed to obtain forecasts. First, following Doornik (2009), DGMS (dynamic GETS model selection) forecasts are produced by the terminal model that minimises either the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC). Second, DGMA (dynamic GETS model averaging) forecasts are generated by averaging forecasts from all terminal models. We use 5 alternative model averaging approaches: mean; trimmed mean (where the largest and smallest forecasts are excluded before calculating the mean forecast); median; weighted averaging based on BIC and AIC weights.[15]

Finally, we generate forecasts from a simple single-path stepwise model selection algorithm in order to compare its performance against the multipath approach. The stepwise algorithm reduces a general model by sequentially removing one variable at a time, the most insignificant one, until all remaining variables are statistically significant at a pre-specified level. The stepwise regression is also applied dynamically across each rolling sub-sample.

The DGETS and stepwise forecasting performance is compared to the UC-SV univariate benchmark model of Stock and Watson (2007), which is shown to typically outperform other statistical or fundamentals-based models in Stock and Watson (2009,

[14] For a summary of the number of terminal models reached and number of variables selected, see Table A2 in the Appendix.

[15] Following Garratt et al. (2003), approximate Bayesian model averaging involves the use of a weighted average of the forecasts, with weights defined by BIC and AIC. In an inflation forecasting exercise, Kapetanios et al. (2008) find that model averaging using BIC and AIC weights can be a powerful alternative to Bayesian model averaging and to factor models.

2010).[16] Finally, forecast accuracy is evaluated over the period 1968 Q1 to 2007 Q1 using the root mean squared forecast error (RMSE).

4. Forecasting results

Tables 1.1 to 1.5 summarise the pseudo out-of-sample forecasting performance of each forecasting procedure (DGMS, DGMA and stepwise) for five inflation series. RMSEs of 4-quarters ahead inflation forecasts, relative to the UC-SV benchmark, are reported for the five forecast periods that have been used by Stock and Watson (2009).[17] A value of less than 1 indicates that the DGETS method generates RMSEs that are lower relative to the benchmark. The most important message arising from our analysis is that DGETS-based forecast methods, combined with the use of a short rolling window and a standard set of macroeconomic predictors of inflation, outperform the benchmark univariate model. This finding is largely robust to the choice of inflation variable and forecasting period, and is in ample contrast with Stock and Watson (2009) who find only episodic success for fundamentals-based inflation forecasts. More specifically, the results can be summarised as follows:

First, there is evidence that the forecasting performance of models that include relevant economic variables as predictors of inflation, relative to the univariate UC-SV benchmark, depends on the length of the rolling window. Specifically, the shorter the window the larger the forecast accuracy of DGMS and DGMA. This pattern is consistent across forecasting periods and inflation measures. The deterioration of the forecast accuracy of these models when the estimation window lengthens indicates that there are significant regime

[16] The RMSEs of the UC-SV model for the five inflation series are taken directly from Tables 3.1 to 3.5 of Stock and Watson (2009).

[17] Stock and Watson (2009) also report results for an earlier sub-sample (1960-1967). Due to lack of data for the economic fundamentals, these refer only to univariate models.

changes that can be better captured by a shorter window.[18] In particular, using a rolling window of 20 quarters, DGMS and DGMA produce RMSEs that are almost invariably lower relative to the UC-SV model. Typically, it is easier to outperform the UC-SV model in the first two sub-samples (1968-1976 and 1977-1984), consistent with the results of Stock and Watson (2009). However, there are significant improvements in forecast accuracy associated with the DGETS methods even post mid-1980s. The notable exceptions involve CPI-all and PCE-all between 1993-2000.[19] The forecast performance of the DGMS and DGMA is more mixed when the rolling window is set to 40 quarters with the relative RMSEs being typically below 1 during the second (1977-1984) and last forecast period (2001-2007). Finally, when using a 60 quarters rolling window, the DGETS approaches only occasionally outperform the benchmark model, and often generate RMSEs that are considerably greater than those of UC-SV model.[20]

[18] See Pesaran and Timmerman (2007), among others, for theoretical work on the determination of the optimal observation window, which suggests that in the presence of structural change the optimal length of the window is weakly decreasing in the magnitude of the break.

[19] Stock and Watson (2009) also find that there are time periods, such as the mid-1990s, that the univariate forecast cannot be improved upon using fundamentals-based models. They point out that when the unemployment rate is close to the NAIRU, forecasts from Phillips curve models do not outperform the UC-SV.

[20] In their experiments, Stock and Watson (2009) use two moving window schemes, namely recursive and rolling of 40-quarters window size. Given the sensitivity of forecasting performance to the size of the rolling window documented in this paper, we repeat several of the Stock and Watson (2009) experiments using alternative window sizes. In particular, we generate forecasts from single-predictor models for each of the 15 macroeconomic factors using autoregressive distributed lag models (as described in their paper), and a simple model averaging by taking a simple mean of all single-predictor forecasts. The main message that arises from this exercise is that Stock and Watson's results do not seem to depend on the size of the rolling window. In other words, the relative success of the DGETS methods relative to the Stock and Watson fundamental-based models cannot be attributed to differences between the moving window schemes employed by the two papers. The results from this exercise are not reported here to save space but are available upon request.

[TABLES 1.1-1.5 HERE]

Second, DGMA forecasts tend to improve upon DGMS forecasts when the rolling window is set to 20 quarters. However, the relative improvement from model averaging decreases as the size of the rolling window increases, and disappears when the window increases to 60 quarters. This is shown more clearly in Table 2. For each inflation variable, and using the full forecasting sample, we calculate the RMSE of the DGMS forecasts relative to the DGMA-mean. At 20 quarters rolling window the improvement in forecast accuracy from model averaging is typically around 10%, while at 60 quarters, forecasts based on model averaging underperform the DGMS forecasts. Thus, model averaging is beneficial, relative to DGMS, only when the rolling window is relatively short. The forecast gains from pooling forecasts from multiple models has been extensively documented in the literature (see e.g. Hendry and Clements, 2002), but to our knowledge, the extent to which these gains depend on the rolling window has not been reported. Our results are novel in that respect. A smaller estimation window is likely to generate greater forecast error variance than a larger window if in both situations the model is well-specified. But a larger window will give rise to biased forecasts if the data generating process is subject to recurrent structural breaks. A potential explanation for our results is that model averaging might lead to greater forecast gains when applied on inefficient forecasts rather than forecasts generated from biased, subject to structural breaks, models.

[TABLE 2 HERE]

Third, BIC- and AIC-based DGMS approaches generate similar levels of forecast accuracy. BIC penalises models for the inclusion of irrelevant regressors more severely than AIC, however, the choice of the over-fitting penalty does not affect the out-of-sample performance of the DGETS methods. Fourth, there is not a particular model averaging approach that shows a systematic or substantial improvement over an alternative averaging method. Fifth, the simple stepwise model selection procedure does not exhibit much success.

It generally generates RMSEs that are somewhat lower than those of the UC-SV model prior to the mid-1980s but thereafter the stepwise approach tends to underperform the benchmark model. Interestingly, unlike the multipath DGETS approaches, the stepwise model selection method does not seem to consistently or substantially improve upon the UC-SV benchmark even when a short rolling window is used. On the contrary, the 20 quarters rolling window generates almost always inferior forecasts relative to 40 or 60 quarters window.

Finally, Figure 2 shows the inclusion frequency in the terminal models identified by the DGETS method for each of the 15 macroeconomic predictors. The plots refer to CPI-all with 20 quarters rolling window and indicate significant time variation in the forecasting model, with respect to the variables that enter it and the frequency of inclusion. Interestingly, the predictive information contained in the macroeconomic variables appears to be the lowest in mid to late 1990s, the period when the UC-SV forecasts are most difficult to beat, as discussed earlier.

A very prominent predictor, at least in terms of terminal model frequency inclusion, is the Chicago Fed National Activity Index (CFNAI). CFNAI is published by Federal Reserve Bank of Chicago, calculated as the principal component of 85 monthly indicators of economic activity, and its forecasting power has been demonstrated by Stock and Watson (1999) and Hansen (2006). Housing starts and capacity utilisation are two other proxies of economic activity that have good forecasting power, whereas GDP and industrial production only occasionally enter the terminal models. Strikingly, unemployment rate also shows very poor predictive power. It enters at least one terminal model only around 1970, 1985-1992 and 2005-2007. In general, this is consistent with previous studies that found significant instability in the Phillips curve relationship (see, e.g. Stock and Watson, 1996). Finally, the exchange rate and Treasury Bond–Treasury Bill spread appear much less important that the economic activity variables. Nonetheless, the exchange rate seems to have predictive content in early 1980s and late 1990s, two periods of protracted dollar appreciation, and the spread in the more recent years (2000-2007).

[FIGURE 2 HERE]

5. Robustness checks

This section investigates the sensitivity of the DGETS forecast performance across a variety of checks. In particular, we (i) remove the sign restrictions from the search space; (ii) vary the model reduction significance level from 1% to 0.5% and 5%; (iii) vary the group deletion *t*-statistic threshold absolute value from 0.5 to 0.7 and 0.3; (iv) consider a GUM specification that does not include a deterministic trend.[21] The results from these robustness checks, for the full sample, are reported in Table 3.[22] RMSEs relative to the benchmark forecasts, which are based upon DGETS models that use the settings discussed in Section 3.1, are shown; values greater than one indicate that the alternative settings deliver less accurate forecasts than the benchmark.

The most striking result that emerges is that the performance of the DGETS approaches deteriorates significantly when we remove the theory-based sign restrictions from the search space. This is most notable in the case of the DGMS forecasts where accuracy falls by 40% or more. The only exception is CPI-core where the impact of the sign restrictions is relatively small. The DGMA forecast accuracy is also adversely affected by the removal of the sign restrictions, but to a smaller extent as compared to DGMS. Thus, model averaging appears to be even more important when the sign restrictions are dropped, most likely

[21] We have also experimented with two additional simple forecasting models, the first involving the CFNAI Index, and the second one the first principal component (PC) of the 15 predictor variables. The results are available upon request and indicate that CFNAI- and PC-based forecasts only occasionally outperform those from the UC-SV model.

[22] Given the success of the 20 quarters rolling window size reported in the previous section, and to conserve space, we report results only for this window size.

reflecting greater forecast variance in individual model forecasts. All in all, these results imply that theory-based sign restrictions provide improvements in model selection power, which is translated into significantly improved forecast performance.

The performance of the DGETS model selection approaches is only marginally affected by changes in other settings. In general, the forecast accuracy drops mildly when we increase the model reduction significance level to 5% but remains unaffected when we decrease it to 0.5%. Altering the group deletion *t*-statistic threshold does not affect the results. Finally, excluding the trend from the GUM lowers forecast performance for two inflation measures and improves it for another two but in any case the changes in forecast accuracy are relatively small.

[TABLE 3 HERE]

6. Conclusions

In this paper we use a standard set of macroeconomic predictors and a dynamic model selection and averaging econometric methodology that allows the forecasting model of inflation to change over time. Identifying the specification of the forecasting models using a multipath DGETS model selection algorithm, pseudo out-of-sample forecasts are generated and their performance is compared with a well-established univariate benchmark, the UC-SV model of Stock and Watson (2009). Our results, across several inflation variables and forecasting periods, reveal that DGETS-based methods in association with a short rolling window lead to significant improvements in forecast performance.

References

Aron, J., Muellbauer, J. (2012). "New methods for forecasting inflation, applied to the US". *Oxford Bulletin of Economics and Statistics*, doi: 10.1111/j.1468-0084.2012.00728.x.

Avramov, D. (2002). "Stock return predictability and model uncertainty". *Journal of Financial Economics*, 64, 423–458.

Atkeson, A., Ohanian, L. (2001). "Are Phillips curves useful for forecasting inflation?". *Federal Reserve Bank of Minneapolis Quarterly Review*, 25, 2–11.

Banerjee, A., Marcellino, M. (2006). "Are there any reliable leading indicators for U.S. inflation and GDP growth?". *International Journal of Forecasting*, 22, 137–151.

Campbell, J., Thomson, S. (2008). "Predicting excess stock returns out of sample: can anything beat the historical average?". *Review of Financial Studies*, 21, 1509–1531.

Castle, J., Clements, M., Hendry, D. (2012). "Forecasting by factors, by variables, or both?". University of Oxford, Department of Economics, Discussion Paper Series, No. 600.

Clements, M., Hendry, D. (1998a). "Forecasting economic processes". *International Journal of Forecasting*, 14, 111–131.

Clements, M., Hendry, D. (1998b). "Intercept corrections and structural change". *Journal of Applied Econometrics*, 11, 475–494.

Doornik, J.A., (2009). "Autometrics". In J. L. Castle and N. Shephard (Eds), The methodology and practice of econometrics, 88–21. Oxford: Oxford University Press.

Garratt, A., Pesaran, M.H., Shin, Y. (2003). "Forecast uncertainties in macroeconometric modelling: An application to the U.K. economy". *Journal of the American Statistical Association*, 98, 829–838.

Hansen, P. (2006), "A Test for Superior Predictive Ability," *Journal of Business and Economic Statistics* 23(4), 365-380.

Hendry, D., Clements, M. (2002). "Pooling of forecasts". *Econometrics Journal*, 5, 1 – 26.

Hendry, D., Clements, M. (2003). “Economic forecasting: some lessons from recent research”. *Economic Modelling*, 20, 301–329.

Hendry, D., Krolzig, H. (2001). “Automatic econometric model selection”. London: Timberlake Consultants Press.

Hendry, D., Krolzig, H. (2005). “The properties of automatic Gets modelling”. *Economic Journal*, 115, 32–61.

Kapetanios, G., Labhard, V., Price, S. (2008). “Forecasting using Bayesian and information-theoretic model averaging: An application to U.K. inflation”. *Journal of Business and Economic Statistics*, 26, 33–41.

Koop, G., Korobilis, D. (2012). “Forecasting inflation using dynamic model averaging”. *International Economic Review*, 53, 867–886.

Newey, W., West, K., 1987. “A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix”. *Econometrica*, 55, 703–708.

Pesaran, H., Timmermann, A. (1995). “Predictability of stock returns: Robustness and economic significance”. *The Journal of Finance*, 50(4), 1201-1228.

Pesaran, M.H., Timmermann, A. (2007). “Selection of estimation window in the presence of breaks”. *Journal of Econometrics*, 137, 134–161.

Rossi, B. (2012). “Advances in Forecasting Under Instability”. In G. Elliott and A. Timmermann (Eds), Handbook of Economic Forecasting, Elsevier-North Holland.

Stock, J., Watson, M. (1996). “Evidence on structural instability in macroeconomic time series relations”. *Journal of Business and Economic Statistics*, 14, 11–30.

Stock, J., Watson, M. (2007). “Why has U.S. inflation become harder to forecast?”. *Journal of Money, Credit and Banking*, 39, 3–34.

Stock, J., Watson, M. (2009). “Phillips curve inflation forecasts”. In J. Fuhrer, Y. Kodrzycki, J. Little and G. Olivei (Eds), Understanding inflation and the implications for monetary policy, 101–186. Cambridge: MIT Press.

Stock, J., Watson, M. (2010). “Modeling inflation after the crisis”. NBER Working Paper, No. 16488.

Table 1.1
Root mean squared errors for 4-quarters ahead inflation forecasting models by sub-period, relative to the UC-SV model: CPI-all

Forecast Period	1968.Q1–1976.Q4	1977.Q1–1984.Q4	1985.Q1–1992.Q4	1993.Q1–2000.Q4	2001.Q1–2007.Q1
Number of observations	36	32	32	32	25
DGMS_BIC_R20	**0.71**	**0.61**	**0.87**	1.19	**0.93**
DGMS_AIC_R20	**0.72**	**0.63**	**0.88**	1.19	**0.88**
DGMA_MEAN_R20	**0.68**	**0.57**	**0.75**	1.06	**0.75**
DGMA_TMEAN_R20	**0.68**	**0.57**	**0.75**	1.06	**0.75**
DGMA_MEDIAN_R20	**0.68**	**0.58**	**0.75**	1.07	**0.75**
DGMA_MEANBIC_R20	**0.67**	**0.57**	**0.76**	1.07	**0.76**
DGMA_MEANAIC_R20	**0.68**	**0.57**	**0.76**	1.08	**0.76**
DGMS_BIC_R40	1.18	**0.59**	1.02	1.29	**0.78**
DGMS_AIC_R40	1.19	**0.58**	1.02	1.21	**0.78**
DGMA_MEAN_R40	1.02	**0.64**	**0.99**	1.20	**0.74**
DGMA_TMEAN_R40	1.05	**0.63**	**0.97**	1.19	**0.76**
DGMA_MEDIAN_R40	1.06	**0.63**	**0.99**	1.20	**0.76**
DGMA_MEANBIC_R40	1.02	**0.63**	**0.98**	1.21	**0.75**
DGMS_MEANAIC_R40	1.02	**0.63**	**0.99**	1.20	**0.74**
DGMS_BIC_R60	1.07	**0.76**	1.16	1.36	**0.98**
DGMS_AIC_R60	1.05	**0.76**	1.16	1.36	**0.97**
DGMA_MEAN_R60	1.09	**0.85**	1.15	1.28	1.00
DGMA_TMEAN_R60	1.10	**0.85**	1.14	1.32	1.00
DGMA_MEDIAN_R60	1.08	**0.86**	1.14	1.32	1.01
DGMA_MEANBIC_R60	1.09	**0.84**	1.14	1.28	**1.00**
DGMA_MEANAIC_R60	1.09	**0.84**	1.14	1.28	**1.00**
STEPWISE_R20	**0.95**	**0.96**	1.70	1.35	1.43
STEPWISE_R40	1.07	**0.57**	1.21	1.22	**0.77**
STEPWISE_R60	1.11	**0.65**	1.13	1.65	1.00

Note: This table reports root mean squared forecast errors (RMSEs) relative to the RMSEs of the unobserved components-stochastic volatility model (UC-SV), over the indicated sample period. The latter is taken from Table 3.1 in Stock and Watson (2009). Bold indicates a relative RMSE of less than 1. Blanks indicate insufficient data to compute forecasts over the indicated sample period. DGMS_BIC (AIC): forecasts based on the BIC (AIC) minimising terminal model identified through the dynamic general-to-specific (DGETS) algorithm. DGMA: forecasts based on averaging forecasts from terminal models identified through the DGETS algorithm. 5 alternative model averaging approaches are used: mean (DGMA_MEAN); trimmed mean (DGMA_TMEAN); median (DGMA_MEDIAN); and weighted averaging based on BIC (DGMA_MEANBIC) and AIC (DGMA_MEANAIC) weights. STEPWISE: forecasts based on model identified through dynamic single-path stepwise model selection. _R20, _R40, and _R60: indicate rolling estimation with a window size of 20, 40, and 60 quarters, respectively.

Table 1.2
Root mean squared errors for 4-quarters ahead inflation forecasting models by sub-period, relative to the UC-SV model: CPI-core

Forecast Period	1968.Q1–1976.Q4	1977.Q1–1984.Q4	1985.Q1–1992.Q4	1993.Q1–2000.Q4	2001.Q1–2007.Q1
Number of observations	36	32	32	32	25
DGMS_BIC_R20	**0.85**	**0.51**	1.01	**0.95**	**0.81**
DGMS_AIC_R20	**0.85**	**0.57**	1.04	**0.91**	**0.75**
DGMA_MEAN_R20	**0.76**	**0.55**	**0.86**	**0.89**	**0.67**
DGMA_TMEAN_R20	**0.77**	**0.56**	**0.87**	**0.90**	**0.66**
DGMA_MEDIAN_R20	**0.78**	**0.58**	**0.89**	**0.89**	**0.67**
DGMA_MEANBIC_R20	**0.77**	**0.53**	**0.90**	**0.86**	**0.65**
DGMA_MEANAIC_R20	**0.77**	**0.53**	**0.90**	**0.87**	**0.65**
DGMS_BIC_R40	**0.98**	**0.61**	1.73	1.34	**0.86**
DGMS_AIC_R40	**0.98**	**0.60**	1.87	1.36	**0.90**
DGMA_MEAN_R40	1.01	**0.61**	1.79	1.66	**0.63**
DGMA_TMEAN_R40	1.01	**0.62**	1.77	1.67	**0.67**
DGMA_MEDIAN_R40	1.01	**0.64**	1.75	1.66	**0.68**
DGMA_MEANBIC_R40	1.01	**0.62**	1.74	1.62	**0.65**
DGMS_MEANAIC_R40	1.01	**0.62**	1.73	1.61	**0.65**
DGMS_BIC_R60		**0.91**	2.37	1.40	**0.94**
DGMS_AIC_R60		**0.86**	2.37	1.48	**0.94**
DGMA_MEAN_R60		**0.82**	2.19	1.54	**0.93**
DGMA_TMEAN_R60		**0.84**	2.09	1.51	**0.94**
DGMA_MEDIAN_R60		**0.83**	2.09	1.47	**0.94**
DGMA_MEANBIC_R60		**0.82**	2.17	1.48	**0.93**
DGMA_MEANAIC_R60		**0.82**	2.17	1.48	**0.93**
STEPWISE_R20	1.07	**0.65**	1.45	2.51	1.04
STEPWISE_R40	**0.74**	**0.57**	1.52	1.45	**0.76**
STEPWISE_R60		**0.65**	1.83	1.26	**0.92**

Note: This table reports root mean squared forecast errors (RMSEs) relative to the RMSEs of the unobserved components-stochastic volatility model (UC-SV), over the indicated sample period. The latter is taken from Table 3.2 in Stock and Watson (2009). Bold indicates a relative RMSE of less than 1. Blanks indicate insufficient data to compute forecasts over the indicated sample period. DGMS_BIC (AIC): forecasts based on the BIC (AIC) minimising terminal model identified through the dynamic general-to-specific (DGETS) algorithm. DGMA: forecasts based on averaging forecasts from terminal models identified through the DGETS algorithm. 5 alternative model averaging approaches are used: mean (DGMA_MEAN); trimmed mean (DGMA_TMEAN); median (DGMA_MEDIAN); and weighted averaging based on BIC (DGMA_MEANBIC) and AIC (DGMA_MEANAIC) weights. STEPWISE: forecasts based on model identified through dynamic single-path stepwise model selection. _R20, _R40, and _R60: indicate rolling estimation with a window size of 20, 40, and 60 quarters, respectively.

Table 1.3
Root mean squared errors for 4-quarters ahead inflation forecasting models by sub-period, relative to the UC-SV model: PCE-all

Forecast Period	1968.Q1–1976.Q4	1977.Q1–1984.Q4	1985.Q1–1992.Q4	1993.Q1–2000.Q4	2001.Q1–2007.Q1
Number of observations	36	32	32	32	25
DGMS_BIC_R20	**0.63**	**0.92**	**0.90**	1.00	**0.99**
DGMS_AIC_R20	**0.63**	**0.91**	**0.90**	1.03	**0.99**
DGMA_MEAN_R20	**0.67**	**0.75**	**0.73**	**0.99**	**0.77**
DGMA_TMEAN_R20	**0.64**	**0.78**	**0.74**	1.00	**0.78**
DGMA_MEDIAN_R20	**0.66**	**0.86**	**0.76**	1.01	**0.80**
DGMA_MEANBIC_R20	**0.65**	**0.75**	**0.78**	**1.00**	**0.79**
DGMA_MEANAIC_R20	**0.66**	**0.75**	**0.78**	1.00	**0.79**
DGMS_BIC_R40		**0.90**	1.39	1.18	**0.90**
DGMS_AIC_R40		**0.85**	1.36	1.18	**0.90**
DGMA_MEAN_R40		**0.86**	1.21	1.20	**0.78**
DGMA_TMEAN_R40		**0.87**	1.18	1.17	**0.77**
DGMA_MEDIAN_R40		**0.87**	1.16	1.18	**0.77**
DGMA_MEANBIC_R40		**0.86**	1.22	1.18	**0.79**
DGMS_MEANAIC_R40		**0.86**	1.22	1.18	**0.79**
DGMS_BIC_R60		1.16	1.69	1.08	1.11
DGMS_AIC_R60		1.16	1.69	1.10	1.09
DGMA_MEAN_R60		1.27	1.75	1.09	1.22
DGMA_TMEAN_R60		1.28	1.76	1.07	1.23
DGMA_MEDIAN_R60		1.28	1.79	1.10	1.23
DGMA_MEANBIC_R60		1.26	1.74	1.08	1.21
DGMA_MEANAIC_R60		1.26	1.74	1.08	1.21
STEPWISE_R20	**0.91**	**0.91**	1.97	1.71	1.52
STEPWISE_R40		**0.93**	1.35	1.08	**0.78**
STEPWISE_R60		**0.97**	1.16	1.24	1.25

Note: This table reports root mean squared forecast errors (RMSEs) relative to the RMSEs of the unobserved components-stochastic volatility model (UC-SV), over the indicated sample period. The latter is taken from Table 3.3 in Stock and Watson (2009). Bold indicates a relative RMSE of less than 1. Blanks indicate insufficient data to compute forecasts over the indicated sample period. DGMS_BIC (AIC): forecasts based on the BIC (AIC) minimising terminal model identified through the dynamic general-to-specific (DGETS) algorithm. DGMA: forecasts based on averaging forecasts from terminal models identified through the DGETS algorithm. 5 alternative model averaging approaches are used: mean (DGMA_MEAN); trimmed mean (DGMA_TMEAN); median (DGMA_MEDIAN); and weighted averaging based on BIC (DGMA_MEANBIC) and AIC (DGMA_MEANAIC) weights. STEPWISE: forecasts based on model identified through dynamic single-path stepwise model selection. _R20, _R40, and _R60: indicate rolling estimation with a window size of 20, 40, and 60 quarters, respectively.

Table 1.4
Root mean squared errors for 4-quarters ahead inflation forecasting models by sub-period, relative to the UC-SV model: PCE-core

Forecast Period	1968.Q1–1976.Q4	1977.Q1–1984.Q4	1985.Q1–1992.Q4	1993.Q1–2000.Q4	2001.Q1–2007.Q1
Number of observations	36	32	32	32	25
DGMS_BIC_R20	**0.81**	**0.94**	**0.85**	**0.95**	**0.94**
DGMS_AIC_R20	**0.82**	**0.94**	**0.84**	**0.92**	**0.94**
DGMA_MEAN_R20	**0.80**	**0.83**	**0.78**	**0.85**	**0.81**
DGMA_TMEAN_R20	**0.82**	**0.83**	**0.77**	**0.86**	**0.84**
DGMA_MEDIAN_R20	**0.84**	**0.84**	**0.77**	**0.86**	**0.84**
DGMA_MEANBIC_R20	**0.81**	**0.83**	**0.82**	**0.83**	**0.82**
DGMA_MEANAIC_R20	**0.81**	**0.83**	**0.82**	**0.83**	**0.82**
DGMS_BIC_R40		1.05	1.22	1.35	1.06
DGMS_AIC_R40		1.07	1.16	1.37	1.06
DGMA_MEAN_R40		**0.96**	1.17	1.53	1.03
DGMA_TMEAN_R40		**0.97**	1.14	1.54	1.05
DGMA_MEDIAN_R40		**0.97**	1.13	1.54	1.03
DGMA_MEANBIC_R40		**0.95**	1.14	1.50	1.04
DGMS_MEANAIC_R40		**0.94**	1.14	1.49	1.04
DGMS_BIC_R60		1.02	1.23	**0.99**	1.54
DGMS_AIC_R60		1.01	1.24	1.04	1.54
DGMA_MEAN_R60		1.15	1.55	1.13	1.60
DGMA_TMEAN_R60		1.18	1.50	1.11	1.57
DGMA_MEDIAN_R60		1.17	1.42	1.11	1.56
DGMA_MEANBIC_R60		1.13	1.45	1.10	1.55
DGMA_MEANAIC_R60		1.13	1.44	1.09	1.55
STEPWISE_R20	**0.94**	1.16	1.93	2.24	2.01
STEPWISE_R40		**0.91**	**0.89**	1.00	1.04
STEPWISE_R60		**0.77**	**0.91**	1.04	1.31

Note: This table reports root mean squared forecast errors (RMSEs) relative to the RMSEs of the unobserved components-stochastic volatility model (UC-SV), over the indicated sample period. The latter is taken from Table 3.4 in Stock and Watson (2009). Bold indicates a relative RMSE of less than 1. Blanks indicate insufficient data to compute forecasts over the indicated sample period. DGMS_BIC (AIC): forecasts based on the BIC (AIC) minimising terminal model identified through the dynamic general-to-specific (DGETS) algorithm. DGMA: forecasts based on averaging forecasts from terminal models identified through the DGETS algorithm. 5 alternative model averaging approaches are used: mean (DGMA_MEAN); trimmed mean (DGMA_TMEAN); median (DGMA_MEDIAN); and weighted averaging based on BIC (DGMA_MEANBIC) and AIC (DGMA_MEANAIC) weights. STEPWISE: forecasts based on model identified through dynamic single-path stepwise model selection. _R20, _R40, and _R60: indicate rolling estimation with a window size of 20, 40, and 60 quarters, respectively.

Table 1.5

Root mean squared errors for 4-quarters ahead inflation forecasting models by sub-period, relative to the UC-SV model: GDP deflator

Forecast Period	1968.Q1–1976.Q4	1977.Q1–1984.Q4	1985.Q1–1992.Q4	1993.Q1–2000.Q4	2001.Q1–2007.Q1
Number of observations	36	32	32	32	25
DGMS_BIC_R20	**0.59**	**0.95**	**0.81**	**0.82**	**0.88**
DGMS_AIC_R20	**0.59**	1.01	**0.79**	**0.82**	**0.85**
DGMA_MEAN_R20	**0.61**	**0.89**	**0.75**	**0.78**	**0.74**
DGMA_TMEAN_R20	**0.62**	**0.93**	**0.76**	**0.79**	**0.75**
DGMA_MEDIAN_R20	**0.62**	**0.93**	**0.80**	**0.79**	**0.75**
DGMA_MEANBIC_R20	**0.58**	**0.89**	**0.77**	**0.79**	**0.72**
DGMA_MEANAIC_R20	**0.58**	**0.89**	**0.77**	**0.79**	**0.72**
DGMS_BIC_R40	**0.99**	**0.96**	1.34	1.32	**0.99**
DGMS_AIC_R40	**0.98**	**0.97**	1.39	1.32	**0.99**
DGMA_MEAN_R40	**0.96**	**0.92**	1.18	1.29	**0.77**
DGMA_TMEAN_R40	**0.99**	**0.93**	1.20	1.26	**0.77**
DGMA_MEDIAN_R40	**0.99**	**0.93**	1.22	1.26	**0.79**
DGMA_MEANBIC_R40	**0.96**	**0.92**	1.20	1.29	**0.76**
DGMS_MEANAIC_R40	**0.96**	**0.92**	1.20	1.28	**0.76**
DGMS_BIC_R60	**0.97**	1.15	**0.92**	1.23	**0.86**
DGMS_AIC_R60	**0.97**	1.04	**0.90**	1.23	**0.87**
DGMA_MEAN_R60	1.01	1.05	1.26	1.13	**0.99**
DGMA_TMEAN_R60	1.04	1.05	1.18	1.12	**0.99**
DGMA_MEDIAN_R60	1.04	1.05	1.15	1.11	1.01
DGMA_MEANBIC_R60	1.00	1.04	1.17	1.12	**0.98**
DGMA_MEANAIC_R60	1.00	1.04	1.16	1.12	**0.98**
STEPWISE_R20	**0.55**	1.11	1.77	1.17	1.22
STEPWISE_R40	**0.90**	**0.89**	1.24	1.30	1.14
STEPWISE_R60	**0.99**	**0.82**	1.27	1.24	1.05

Note: This table reports root mean squared forecast errors (RMSEs) relative to the RMSEs of the unobserved components-stochastic volatility model (UC-SV), over the indicated sample period. The latter is taken from Table 3.5 in Stock and Watson (2009). Bold indicates a relative RMSE of less than 1. Blanks indicate insufficient data to compute forecasts over the indicated sample period. DGMS_BIC (AIC): forecasts based on the BIC (AIC) minimising terminal model identified through the dynamic general-to-specific (DGETS) algorithm. DGMA: forecasts based on averaging forecasts from terminal models identified through the DGETS algorithm. 5 alternative model averaging approaches are used: mean (DGMA_MEAN); trimmed mean (DGMA_TMEAN); median (DGMA_MEDIAN); and weighted averaging based on BIC (DGMA_MEANBIC) and AIC (DGMA_MEANAIC) weights. STEPWISE: forecasts based on model identified through dynamic single-path stepwise model selection. _R20, _R40, and _R60: indicate rolling estimation with a window size of 20, 40, and 60 quarters, respectively.

Table 2
Root mean squared errors for 4-quarters ahead inflation forecasting models, full sample, DGMA_MEAN relative to DGMS

	CPI- all	CPI- core	PCE-all	PCE-core	GDP deflator
DGMS_BIC_R20	**0.91**	**0.90**	**0.88**	**0.92**	**0.97**
DGMS_AIC_R20	**0.90**	**0.87**	**0.88**	**0.92**	**0.94**
DGMS_BIC_R40	**0.93**	1.02	**0.89**	**0.95**	**0.95**
DGMS_AIC_R40	**0.93**	1.01	**0.92**	**0.95**	**0.95**
DGMS_BIC_R60	1.04	**0.97**	1.05	1.13	1.02
DGMS_AIC_R60	1.04	**0.99**	1.05	1.12	1.05

Note: This table reports root mean squared forecast errors (RMSEs) from model averaging relative to the RMSEs from non-model averaging. Bold indicates a relative RMSE of less than 1. DGMS_BIC (AIC): forecasts based on the BIC (AIC) minimising terminal model identified through the dynamic general-to-specific (DGETS) algorithm. DGMA_MEAN: forecasts based on the mean of forecasts from terminal models identified through the DGETS algorithm. _R20, _R40, and _R60: indicate rolling estimation with a window size of 20, 40, and 60 quarters, respectively.

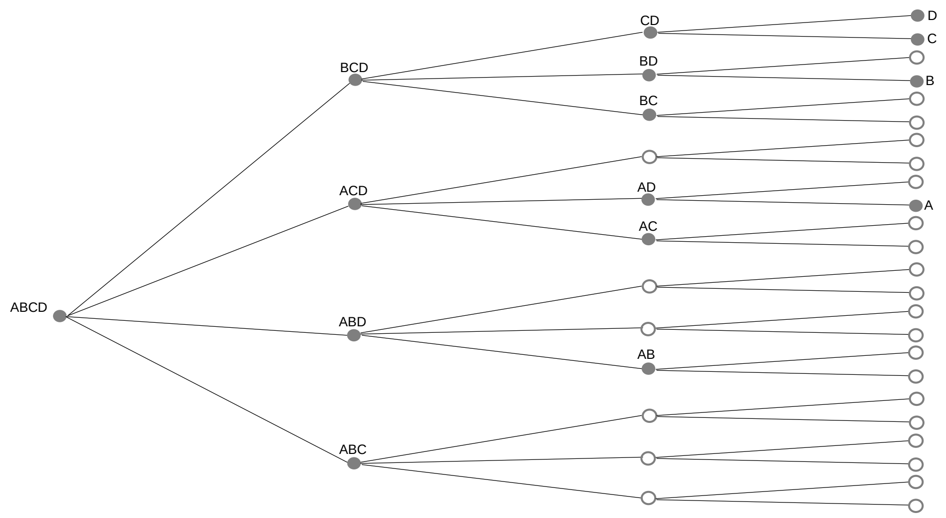
Table 3
Root mean squared errors for 4-quarters ahead inflation forecasting models, full sample, relative to benchmark DGMS and DGMA

	CPI- all	CPI- core	PCE-all	PCE-core	GDP deflator
No Sign Restrictions					
DGMS_BIC_R20	1.50	1.08	1.42	1.40	1.46
DGMS_AIC_R20	1.57	1.03	1.55	1.72	1.45
DGMA_MEAN_R20	1.33	**0.97**	1.19	1.11	1.04
DGMA_TMEAN_R20	1.33	**0.96**	1.20	1.09	1.01
DGMA_MEDIAN_R20	1.33	**0.96**	1.15	1.07	1.00
DGMA_MEANBIC_R20	1.34	1.00	1.20	1.14	1.08
DGMA_MEANAIC_R20	1.34	1.02	1.21	1.17	1.10
Model reduction sign. level = 0.5%					
DGMS_BIC_R20	1.00	1.00	**0.99**	1.00	1.00
DGMS_AIC_R20	1.00	1.00	1.00	1.00	1.00
DGMA_MEAN_R20	1.02	1.01	1.03	**0.99**	1.00
DGMA_TMEAN_R20	1.02	1.01	1.04	1.00	1.00
DGMA_MEDIAN_R20	1.01	**0.99**	1.04	1.01	1.00
DGMA_MEANBIC_R20	1.01	1.01	1.01	**0.99**	1.00
DGMA_MEANAIC_R20	1.01	1.01	1.01	**0.99**	1.00
Model reduction sign. level = 5%					
DGMS_BIC_R20	1.04	**0.99**	1.04	**0.99**	1.10
DGMS_AIC_R20	1.04	**0.99**	1.05	1.02	1.09
DGMA_MEAN_R20	1.05	**0.94**	1.12	1.02	1.10
DGMA_TMEAN_R20	1.06	**0.94**	1.13	1.02	1.07
DGMA_MEDIAN_R20	1.06	**0.92**	1.11	1.02	1.06
DGMA_MEANBIC_R20	1.06	**0.97**	1.11	1.01	1.10
DGMA_MEANAIC_R20	1.06	**0.97**	1.11	1.01	1.10
Group deletion abs. t-value = 0.7					
DGMS_BIC_R20	1.00	1.00	1.00	1.00	1.00
DGMS_AIC_R20	1.00	1.00	1.00	1.00	1.00
DGMA_MEAN_R20	1.01	1.01	1.00	1.00	**0.99**
DGMA_TMEAN_R20	1.01	1.01	1.00	1.00	**0.99**
DGMA_MEDIAN_R20	1.01	1.01	**0.99**	1.01	1.00
DGMA_MEANBIC_R20	1.00	1.00	1.00	1.00	1.00
DGMA_MEANAIC_R20	1.00	1.00	1.00	1.00	1.00
Group deletion abs. t-value = 0.3					
DGMS_BIC_R20	1.00	1.00	1.00	1.00	1.00
DGMS_AIC_R20	1.00	1.00	1.00	1.00	1.00
DGMA_MEAN_R20	1.00	1.00	1.00	**0.99**	1.00
DGMA_TMEAN_R20	1.01	1.00	1.00	1.00	1.00
DGMA_MEDIAN_R20	1.00	1.00	1.00	1.01	1.00
DGMA_MEANBIC_R20	1.00	1.00	1.00	**0.99**	1.00
DGMA_MEANAIC_R20	1.00	1.00	1.00	**0.99**	1.00

Table 3 ***Continued***					
	CPI- all	CPI- core	PCE-all	PCE-core	GDP deflator
No Trend					
DGMS_BIC_R20	1.02	**0.97**	1.01	1.00	1.04
DGMS_AIC_R20	1.00	**0.97**	1.01	**0.99**	1.00
DGMA_MEAN_R20	1.04	1.03	1.06	**0.97**	**0.95**
DGMA_TMEAN_R20	1.05	1.02	1.08	**0.98**	**0.93**
DGMA_MEDIAN_R20	1.05	1.00	1.04	**0.98**	**0.90**
DGMA_MEANBIC_R20	1.04	1.03	1.04	**0.97**	**0.97**
DGMA_MEANAIC_R20	1.03	1.03	1.04	**0.97**	**0.97**

Note: This table reports root mean squared forecast errors (RMSEs) relative to the RMSEs of the benchmark DGMA and DGMS models. The benchmark models utilise 20-quarters rolling window, sign restrictions, model reduction significance level of 1%, and group variable deletion *t*-statistic absolute value of 0.5. Bold indicates a relative RMSE of less than 1. *No Sign Restrictions*: indicates that no sign restrictions are imposed on the search space; *Model reduction sign. level*: is the significance level used for model reduction; *Group deletion abs. t-value:* is the *t*-statistic absolute value threshold used for the group deletion strategy; *No Trend:* indicates that there is no deterministic trend in the GUM specification. DGMS_BIC (AIC): forecasts based on the BIC (AIC) minimising terminal model identified through the dynamic general-to-specific (DGETS) algorithm. DGMA: forecasts based on averaging forecasts from terminal models identified through the DGETS algorithm. 5 alternative model averaging approaches are used: mean (DGMA_MEAN); trimmed mean (DGMA_TMEAN); median (DGMA_MEDIAN); and weighted averaging based on BIC (DGMA_MEANBIC) and AIC (DGMA_MEANAIC) weights. _R20 indicate rolling estimation with a window size of 20 quarters.

Figure 1: Multipath model space



Note: This figure has been reproduced from Doornik (2009). It shows all unique models starting from a general unrestricted model (GUM) with variables ABCD.

Figure 2: Predictors' frequency of inclusion in terminal models

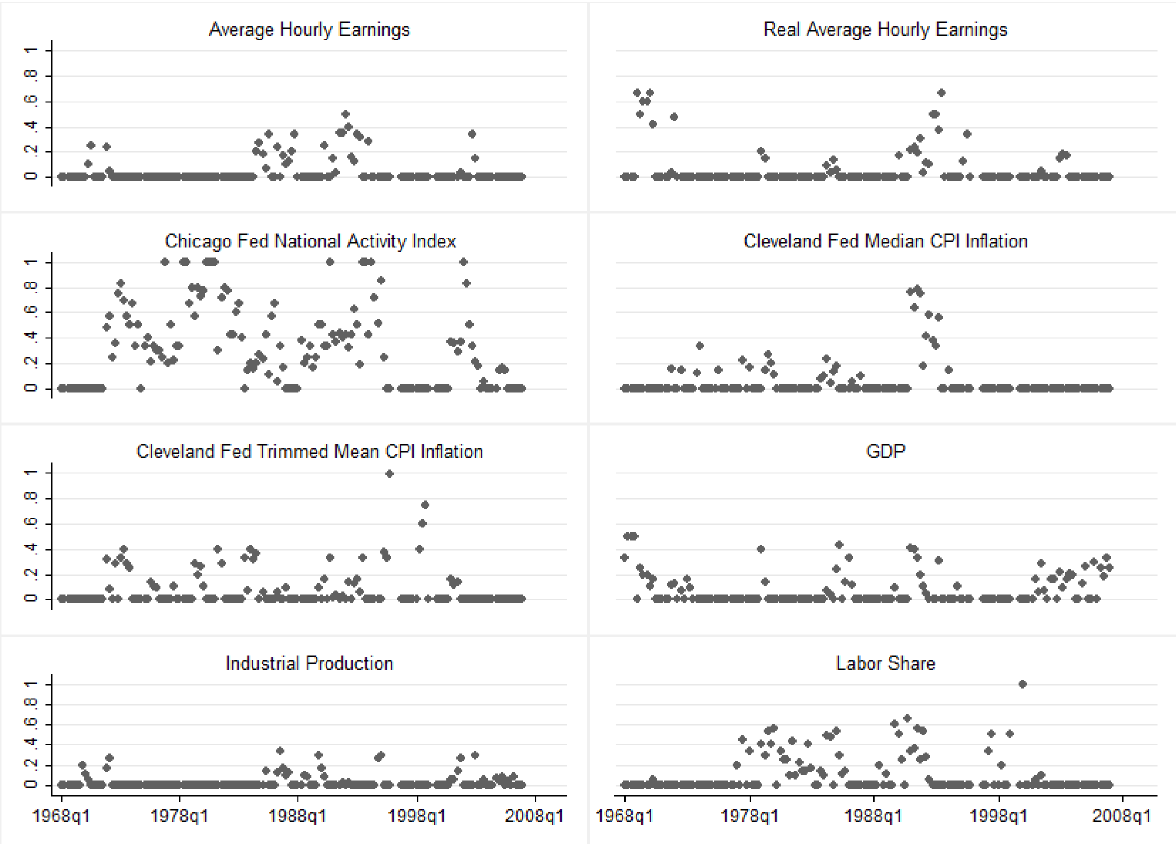
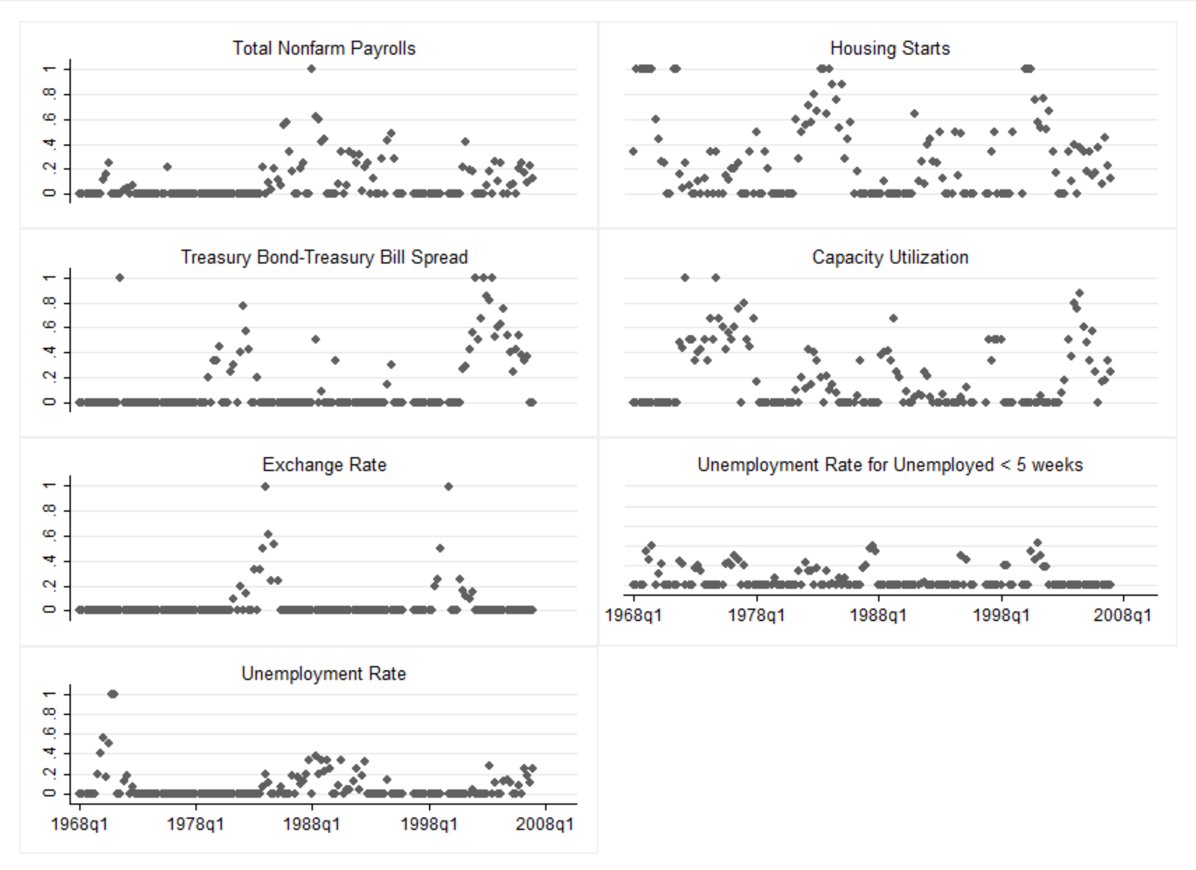


Figure 2 *Continued*: Predictors' frequency of inclusion in terminal models



Note: This figure plots the inclusion frequency (vertical axis) over time (horizontal axis) for each of the 15 predictors in the terminal models identified by the DGETS method with a rolling window of 20 quarters using the CPI-all inflation. A value of 1 for the inclusion frequency implies that the predictor enters all terminal models.

Appendix

Table A1: Inflation predictors and sign restrictions

Variable Name	Definition	Transformation	Sign Restriction
UNRATE	Civilian Unemployment Rate	First difference	-
GDPC96	Real Gross Domestic Product	Log first difference	+
INDPRO	Industrial Production Index	Log first difference	+
PAYEMS	Total Nonfarm Payrolls: All Employees	Log first difference	+
UR-5WK	Unemployment Rate for Unemployed < 5 weeks	First difference	-
TB_SP	1 Year Treasury Bond Rate minus 3 Months Treasury Bill Rate	First difference	+
PERMIT	New Private Housing Units Authorized by Building Permit	Log	+
AHETPI	Average Hourly Earnings: Total Private Industries	Log first difference	+
AHETPIR	Real Average Hourly Earnings: Total Private Industries	Log first difference	+
LS	Labor Share	Log first difference	+
TCU	Capacity Utilization: Total Industry	Level	+
CFNAI	Chicago Fed National Activity Index	Level	+
CPI_M	Cleveland Fed Median CPI Inflation	Level	+
CPI_TM	Cleveland Fed Trimmed Mean CPI Inflation	Level	+
TWEXMMTH	Trade Weighted Exchange Rate Index: Major Currencies	Log first difference	-

Note: In the final column we show the sign that we use for the sign restriction model reduction strategy. The exchange rate index is defined so that an increase is indicating an appreciation of the dollar. The source of the data is M. Watson's website: http://www.princeton.edu/~mwatson/ddisk/bfed_Sept2008.zip

Table A2: Terminal model summary

	CPI-all		CPI-core		PCE-all		PCE-core		GDP deflator	
Number of Terminal Models	Frequency	Average Number of Variables Selected	Frequency	Average Number of Variables Selected	Frequency	Average Number of Variables Selected	Frequency	Average Number of Variables Selected	Frequency	Average Number of Variables Selected
1	38	0.8	17	0.6	6	1.0	7	0.7	38	0.3
2	29	1.4	15	1.3	14	1.7	16	1.2	27	0.9
3	24	2.0	11	1.5	12	2.0	18	1.9	26	2.1
4	20	3.7	21	2.6	24	2.5	24	2.5	15	2.7
5	17	3.9	18	3.7	16	3.7	17	3.6	23	3.0
6	14	4.4	12	4.2	14	3.9	11	4.0	13	4.0
7	15	4.7	10	4.7	12	4.7	13	4.5	13	4.4
8	7	5.4	11	5.1	5	5.4	16	4.3	11	5.1
9	7	5.3	10	4.7	13	6.1	14	4.4	5	5.2
10	7	6.4	7	5.9	12	5.4	11	5.4	4	6.0
>10	29	7.8	40	7.0	38	7.8	19	6.2	34	6.7

Note: This table reports the frequency at which the DGETS algorithm finds 1, 2, 3, etc. terminal models together with the average number of variables identified as being statistically significant. Note that the same variable can be found to be significant in more than one terminal model. Average number of variables selected refers only to the 15 predictors, not counting lagged inflation and the trend term.