DISCUSSION PAPER SERIES

No. 7877

NEW METHODS FOR FORECASTING INFLATION, APPLIED TO THE US.

Janine Aron and John Muellbauer

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP7877.asp**

ISSN 0265-8003

NEW METHODS FOR FORECASTING INFLATION, APPLIED TO THE US.

Janine Aron, University of Oxford
John Muellbauer, Nuffield College, University of Oxford

Discussion Paper No. 7877
June 2010

Centre for Economic Policy Research
53–56 Gt Sutton St, London EC1V 0DG, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 7877

June 2010

ABSTRACT

New methods for forecasting inflation, applied to the US.*

Models for the twelve-month-ahead US rate of inflation, measured by the chain weighted consumer expenditure deflator, are estimated for 1974-99 and subsequent pseudo out-of-sample forecasting performance is examined. Alternative forecasting approaches for different information sets are compared with benchmark univariate autoregressive models, and substantial out-performance is demonstrated. Three key ingredients to the out-performance are: including equilibrium correction terms in relative prices; introducing non-linearities to proxy state dependence in the inflation process; and replacing the information criterion, commonly used in VARs to select lag length, with a 'parsimonious longer lags' (PLL) parameterisation. Forecast pooling or averaging also improves forecast performance.

JEL Classification: C22, C51, C52, C53, E31, E37 and E52
Keywords: error correction models, evaluating forecasts, model selection and multivariate time series

Janine Aron
Centre for the Study of African Economics (CSAE)
University of Oxford, St. Cross Building
Manor Road
Oxford
OX1 3UQ
Email: janine.aron@economics.ox.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=131312

John Muellbauer
Nuffield College
Oxford
OX1 1NF
Email: john.muellbauer@nuffield.ox.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=100344

* The authors acknowledge funding support from the Economic and Social Research Council, U.K. (grant RES-000-22-2066). Comments by Chris Carroll, Kevin Lansing, Roland Meeks, John Williams and other seminar participants at the San Francisco Federal Reserve are gratefully acknowledged. An earlier version of this work appeared in a discussion paper at the Department of Economics, Oxford, in late 2008.

Submitted 21 May 2010

I. Introduction

Stabilising inflation is a key objective of monetary policy in the USA, and a large subset of OECD and a few emerging market countries now target inflation as a primary objective of policy. Since monetary policy is based on the likely path of inflation, it is important that central banks have a reliable forecasting framework to avoid costly policy errors. Forecasting inflation is notoriously difficult, however. Clements and Hendry (1998, 2002) have long argued that structural breaks are the chief cause of forecast failure. Indeed, there have been large structural shifts in the world economy, for instance in. trade and financial globalization, and in individual economies, such as a decline in trade union power and the increasing economic weight of services. Monetary policy itself has generally shifted to a greater focus on inflation. In addition, energy and food price shocks can be both sizeable and largely unpredictable, with the speed of price changes tending to rise with larger shocks.

It is not surprising, therefore, that most forecasting models used by central banks put a large weight on *recent* inflation. This approach tracks inflation quite well, *except at turning points*, because the models miss underlying or long-term influences. In practice, central banks do not rely on a single econometric model to guide policy making. Most central banks augment the main model inflation forecasts by examining trends in individual price components and specific sectoral information, for example on the expiry of gas supply contracts, which might give clues to future changes. Outside the US Federal Reserve, the New Keynesian Phillips Curve (NKPC) has been the dominant paradigm for modelling inflation among macro-economists and central bankers.[1]

The current, mainstream version of the NKPC, proposed by Calvo (1983), claims micro-foundations for sticky prices in the adjustment process. It assumes every firm has the same probability of changing its prices: thus one that has just adapted its prices, has the same probability of adjusting prices again as a firm that has not changed for a long time. Furthermore, this probability is constant over time – independent of the state of the economy. This model implies that the rate of inflation is largely determined by the expected rate of inflation (with a coefficient close to 1), and by the output gap (or the price level relative to unit labour costs, in some versions). The implication for forecasting inflation is that any information relevant for forecasting the output gap or unit labour costs should help forecast inflation. There are open economy variants of the NKPC in which import prices also enter e.g. Batini *et al.* (2005). A hybrid form of the NKPC adds lagged inflation with the sum of expected and lagged inflation close to 1, known as the "accelerationist" restriction.

One problem for the NKPC is inflation heterogeneity. Figure 1 illustrates this with annual inflation rates for non-durable goods, durable goods and services in the US (see also Peach *et al*., 2004). Durables goods have experienced negative inflation for the last decade, while non-durable goods only occasionally experienced price falls over a 12-month period. Prices of services have not fallen in any 12-month period in the last four decades. Food and oil products, prominent among non-durables tend to have highly flexible and volatile prices, while some service sector prices are set in annual contracts.

[1] The NKPC is a modern version of the expectations-augmented Phillips curve, i.e. a relationship between inflation, the output gap or other cyclical demand measure, and expected inflation. The classic reference is Woodford (2003), though Clarida *et al.* (1999) did much to popularise the NKPC. Roberts (1995) discusses its earlier variants.

Moreover, rents, which are a key service sector component, adjust slowly to house prices and interest rates, which are variables outside the NKPC framework. Empirical tests against more general models than closed or open economy versions of the NKPC usually find that the model restrictions are rejected.[2] There is also empirical evidence against the rational expectations hypothesis embodied in the NKPC, using inflation forecasts from surveys of households, Forsells and Kenny (2002).

Substantial research has contrasted alternative formulations of forecasting models, evaluated through their ability to forecast out-of-sample. Prominent work by Stock and Watson (2003) concludes that it is very hard to improve on simple univariate autoregressive models, i.e. where inflation, say over the next year, is forecast only by inflation in the recent past. In Stock and Watson (2007), these authors argue that a univariate model with gradually evolving parameters is the best forecasting model of all. Unfortunately, however, such models, being formulated in differences, cannot capture turning points such as occurred in 2008, when the cost of policy errors tends to be at its maximum.

Our paper takes an eclectic view of the economic drivers of the 12-month-ahead chained consumer expenditure deflator (*PC*) in the US, the price index preferred by the Federal Reserve[3], and attempts to throw light on the factors likely to enhance forecasting performance. We disentangle six major factors that influence forecasting performance and compare the results using benchmark models from the literature.

---

[2] Examples are Bardsen *et al.* (2004), Boug *et al.* (2006), Mavroides (2005) and Rudd and Whelan (2007).

[3] The consumer expenditure deflator PC is a chain-weighted index defined by $\Delta \log PC_t = \sum_i w_{it} \Delta \log PC_{it}$, where W_{it} is the aggregate expenditure share of the ith good or service in month t and PC_{it} is its price. The *PC* is preferred to the *CPI* since the latter underwent a major shift in methodology in 1983, when the treatment of homeowners' costs shifted from a mortgage cost basis to an imputed rent basis.

The first factor concerns whether to work with first differences or levels of non-stationary data. The former could benefit from greater robustness to structural breaks, see Hendry and Clements (2003), while the latter potentially exploits cointegration properties of non-stationary data.

The second factor is whether other parameter restrictions can improve forecasting performance. Vector Auto Regression *(*VAR) models aim to preserve generality by not imposing an *a priori* structure on models, (Sims, 1980), but suffer from the 'curse of dimensionality', as increases in lag lengths or in the number of variables covered rapidly raise the number of parameters to be estimated. In practice, their gain in generality comes at the cost of restricting the number of variables and lag lengths that can be considered. One way of achieving a better trade-off between these objectives is to impose other restrictions such as 'parsimonious longer lags' (PLL) used in this paper. PLL here takes the following form: for variables in differences, full generality is permitted at short lag lengths; for lags at three months or longer, these appear as the three month change or Δ_3, and as Δ_6 if six months or longer and Δ_{12} if twelve months or longer. Compared to unrestricted lags up to 23 months, 24 parameters are thus replaced by six. Further, for variables in levels, three-month or twelve-month moving averages are tested as alternatives to the monthly level. Formulating the 'general unrestricted model' (GUM) in this way offers potential benefits in enabling longer lags to play a role, and permitting smoother responses to shocks. Minimal sign restrictions are also examined, for example, requiring equilibrium correction terms to have signs consistent with such correction, and ensuring that output gaps or their proxies have a positive effect on future inflation.

The third factor considered concerns state dependence, for example whether the speed of price adjustment has fallen with a lower inflation volatility environment. There is a large literature on inflation persistence and price stickiness, recently intensively studied at the micro as well as macro level by the *Inflation Persistence Network* set up by the ECB and main central banks from round the world (see Angeloni *et al.*, 2006; Altissimo *et al.* 2006; and Alvarez *et al.*, 2006). A key issue concerns whether the probability of price changes is state dependent, as argued by Sheshinski and Weiss (1977), or whether the popular Calvo model, the work-horse of modern monetary economics, applies. Reis (2006) supports state dependence: in his sticky information model, producers re-optimise more frequently when cost changes are more volatile, suggesting a higher speed of adjustment in high inflation periods such as the 1970s. One plausible implication of these ideas is that there could be non-linearities in the inflation process so that high current or recent rates of inflation are associated with disproportionately higher future inflation. This question is explored by testing for such non-linearities and checking their contribution to forecast performance.

A fourth question concerns the role of long-run trends. For instance, the causes of the reduction in inflation and in its volatility since the early 1980s, an aspect of the 'great moderation' (Bernanke, 2004), remains controversial. Some have suggested that improved monetary policy from around 1980 is an important explanation. Others have emphasised the globalization of trade and finance, changes in technology, the decline of trade unions, and the rise of the "Asian tigers". Yet others claim that the greater stability of the last two decades has been largely a matter of luck. Although it is always difficult to identify unambiguously the role of trending variables, the contributions to forecasting performance of three trending variables are examined: a measure of trade openness, the share of Asian producers in global exports of manufactures, and union density.

The last two factors are whether forecast performance is improved by model averaging and by model selection using Autometrics (Doornik, 2009). Model averaging or pooling takes an average of two or more forecasts from different models instead of any one model (Bates and Granger, 1969; Hoeting *et al.*, 1999; Hendry and Clements, 2002; and Stock and Watson, 2004). Automatic model selection applies a sequence of tests to reduce the general unrestricted model to a parsimonious model. If the restrictions consistent with data in the estimation period continue to hold into the forecast period, one might expect that less noisy parameter estimates could improve the forecasts.

II. The modelling framework

In this section, we forecast *PC* directly using various information sets. The dependent variable is the h-step-ahead rate of aggregate inflation in single equation equilibrium correction models or simpler variants of these, where h is 12 months. Multi-step models for inflation forecasting have been popularised by Stock and Watson (1999, 2003). Methodologically, multi-step models can be regarded as single equation, reduced-forms of the related VAR system. Some research suggests that where VAR models suffer from specification errors such as omitted moving average error components or certain kinds of structural breaks, single-equation, multi-step models can sometimes provide more robust forecasts (Weiss, 1991; Clements and Hendry, 1996, 1998).

The forecasting performance of three different methodologies is contrasted in terms of reported root mean square forecast errors. Each methodology is applied to the same range of information sets, and in all, twelve information sets are considered. The first methodology uses an unrestricted standard AR(k+1) or VAR(k+1) specification for inflation rates, with lag length selected by an information criterion.[4] This model is contrasted with a restricted model applying the parsimonious and longer lag (PLL) structure discussed above.[5] The third methodology applies automatic model selection to the PLL models. We then check whether averaging forecasts confers any advantages in forecasting.

The simplest information set is a univariate model in differences only for the aggregate price index, PC. This is progressively extended by adding a time trend and/or other trending variables such as union density or the Asian share of world exports,, and changes in the unemployment rate. The information set is further enriched by adding log changes in the producer price index, labour earnings, import prices, the real exchange rate and oil prices. Later information sets include equilibrium correction terms, and finally, the role of non-linearities is tested for. The models are now discussed in further detail.

---

[4] It is standard in the VAR and the multi-step forecasting literature based on VARs to use the Akaike or Schwarz information criterion (AIC or BIC) to choose the maximum lag length of the model. The same is done here.

[5] Let $\Delta_j \log PC_t$ denote $\log PC_t - \log PC_{t-j}$. Thus an AR(k+1) or VAR(k+1) includes $\Delta \log PC_{t-i}$ for i=0, k, where the forecast variable is for t+h, with h greater than zero. Our preferred parsimonious longer lag (PLL) alternative for monthly data includes $\Delta \log PC_{t-i}$ for i=0 to 2, $\Delta_3 \log PC_{t-3}$, $\Delta_6 \log PC_{t-6}$, and $\Delta_{12} \log PC_{t-12}$. As noted above, this is a parsimonious way of allowing longer lags to enter, using six parameters to summarise 24 lags (from 0 to 23).

The specific modelling framework and sign priors

The single equation, reduced-form of the related VAR(k) system, without here imposing the parsimonious longer lag structure (i.e. Δ_3, Δ_6 and Δ_{12} restrictions), and which includes long-run equilibrium correction model (ECM) terms, is as follows:

$$
\begin{aligned}
\Delta_h \log PC_{t+h} = \alpha\ &+ \eta(\log PPI_t - \log PC_t) + \sum_{j=0}^{k} \eta_j \Delta \log PPI_{t-j} \\
&+ \theta(\log ULC_t - \log PC_t) + \sum_{j=0}^{k} \theta_j \Delta \log EARN_{t-j} \\
&+ \lambda(\log PIMP_t - \log PC_t) + \sum_{j=0}^{k} \lambda_j \Delta \log PIMP_{t-j} \\
&+ \varphi(\log OTHERP_t - \log PC_t) + \sum_{j=0}^{k} \varphi_j \Delta \log OTHERP_{t-j} \\
&+ \sum_{l=1}^{n} \beta_l X_{l,t} + \sum_{l=1}^{n} \sum_{j=0}^{k} \beta_{l,j} \Delta X_{l,t-j} \\
&+ \sum_{j=0}^{k} \omega_j \Delta \log PC_{t-j} + trends_t + \varepsilon_t \qquad (1)
\end{aligned}
$$

In the dependent variable, for h=12, PC_{t+12} is the twelve-month-ahead value of the price index. For the independent variables, *PPI* is a producer price index, *ULC* is unit labour costs, *PIMP* is an index of import prices (excluding petroleum products), and *OTHERP* includes other prices such as foreign prices relative to domestic prices (e.g. represented by the real effective exchange rate, *REER*), oil prices (*POIL*) and house prices (*HP*). Since unit labour costs are not available monthly, *ULC* is the 3-month moving average of interpolated quarterly data, lagged two months given lags in the availability of the quarterly data in real time. For the dynamics, changes in log *ULC* are replaced by changes in log *EARN*, earnings per person hour in the private sector, which are monthly. The variables in X include the unemployment rate (*UNR*) or other measures of the output gap. The trends could include variables such as trade union density (*UNDENS)*, trade openness, or the share of Asian 'tigers' in world exports of manufacturing (*ASIAEXP*). The stochastic error term, ε_t ,is almost certainly positively auto-correlated given the overlapping nature of the dependent variable.

The first four lines of the equation capture both the dynamics and "equilibrium correction" mechanisms for four or more types of prices. Long run homogeneity is imposed through the ECMs. The long-run solution for log *PC* is then a weighted average (weights adding to 1), of log *PPI*, log *ULC*, log *PIMP*, and log *OTHERP*,[6] with the X_l and trends as potential shift factors in the relationship.

Equation (1) does not explicitly include the non-linearities discussed above. As discussed there, if the frequency of price change is state dependent, as in sticky information models such as that of Reis (2006), high recent rates of inflation could be associated with more rapid pass-through of inflation shocks. This might suggest that the parameters in equation (1), which incorporate the speed of adjustment, should vary with recent inflation experience. However, this would produce a complex model, non-linear in both variables and parameters. A simpler model that captures some of the same ideas and possible asymmetries in price adjustment would include additive terms of a non-linear transformation of recent changes in log prices. To be specific, for a basic specification, the residual could be extracted from the regression of the $(\Delta_6 \log PC)^2$ on a constant and on $\Delta_6 \log PC$ itself. This residual has the virtue of being orthogonal to the log change, so

[6] A proviso is that the log level of the real exchange rate, defined as log domestic prices + log nominal trade weighted exchange rate – log trade weighted foreign price indices, is defined using the domestic consumer price index (CPI) rather than the expenditure deflator, PC.

avoiding collinearity. The residual and its six-month lag capture the non-linearities from recent inflation experience. The choice of six month lags at t and at t-6 encapsulates the idea of ‘recent changes’ while preserving parsimony.

Sign priors for the main regressors are now discussed, and serve as a guide for selecting parsimonious models. The equilibrium correction terms defined in equation (1) should all have positive coefficients, except for the log of the real exchange rate index. This should have a negative sign since the higher is the US Dollar relative to other currencies, and the lower are foreign prices, the cheaper are imports and the more difficult it is for domestic price setters to push through price increases. The proxies for excess demand or the output gap should have a positive coefficient. Thus, the unemployment rate should have a negative coefficient. When its level is excluded, leading changes in the unemployment rate should have negative coefficients.

Sign priors for potential X variables are less clear cut. For example, real interest rates might be expected to have a negative effect on inflation, but this may already be reflected in excess demand proxies. There is also the ‘cost channel’ of Barth and Ramey (2001) which argues for a positive coefficient on current interest rates. For example, mortgage interest rates feed into rents which are an important part of the consumer price deflator thus raising inflation. Furthermore, if the Federal Reserve has information about future inflation not reflected elsewhere in the model, it may raise interest rates to head off higher inflation.

Similar ambiguities surround the yield spread, defined as the yield on longer-dated Treasury bonds, e.g. three or ten years, minus the short rate, e.g. the three-month T-bill rate. This could have an expectations interpretation through term-structure theory. A higher yield spread suggests that the market expects rising future short rates relative to current rates, in line with raised inflation expectations, see Fama (1990). The empirical evidence in Kozicki (1997) and De Bondt and Bange (1992), suggests that the yield spread is poor predictor for inflation and interest rates over one or two year horizons. Indeed, the term structure interpretation seems to rest on the idea that the private sector has superior information to the central bank. However, there is evidence that the Fed's Green Book forecasts are superior to those of the private sector, see Romer and Romer (2000). Also, the consensus is that consumer survey- based measures of inflation expectations are not rational, and tend to lag inflation, see Roberts (1997). In this case, a fall in the yield spread indicates a tightening of monetary policy as the Fed's anticipates higher inflation in advance of the market, and a positive coefficient on the yield spread could then be implied in equation (1). Given the ambiguities, and more importantly, probable shifts in monetary policy, interest rates and spreads are omitted.

For the trends that could be included, the decline in union density should be associated with lower inflation - hence a positive coefficient, while increased trade openness and a rising share of Asian 'tigers' in world exports of manufactures would both be expected to have a negative effect on inflation. However, these are all I(2) variables[7] so that the risk of 'spurious regression' is considerable when most of the other variables in equation (1) are I(1) or I(0). Furthermore, the last two show acceleration over time and could potentially increase substantially further, though it seems implausible that US inflation would trend downwards limitlessly. Union density, on the other hand, effectively moves in an ogive from a high plateau to a low plateau, and can be thought of

[7] Data on union density and the Asian export share are interpolations of annual data so the tests are performed at an annual frequency.

as a smooth version of a step dummy. *A priori*, union density seems the most likely candidate of the 'trend' variables for obtaining sensible results.

By contrast with the level variables, no priors are imposed on price dynamics. Negative coefficients on lagged cost changes have several possible interpretations. For example, when an ECM is present in the equation, they can indicate longer lags in the reaction of the consumer price index to costs (i.e. that the ECM should be lagged as appropriate). They can indicate the lagged effects of a policy response to inflation, or the negative effect on excess demand of lagged cost changes, for example in the price of oil. They may also reflect a temporary squeeze in profit margins, followed by an inflationary restoration of margins.

Automatic selection with Autometrics

A range of models of different levels of generality is examined, as outlined above, using the Schwarz or Bayesian information criterion (BIC) to select the maximum lag length, and imposing sign priors in those models that include ECM terms. A check is carried on whether more extreme reductions towards parsimony are helpful for forecasting, using automatic model selection in the form of Autometrics, Doornik (2009).

Autometrics is an objective and easily reproducible tool, not affected by the subjective choices of the modeller. Any other investigator with the same data and the same specification of the 'general unrestricted model' (GUM), will then make the same model selection, given the chosen settings in Autometrics. This software examines a full set of general to simple reduction paths to select a parsimonious form of the GUM to satisfy a set of test criteria. The test criteria include tests for normality, heteroscedasticty, ARCH residuals, residual autocorrelation, parameter stability in the form of a Chow test, and the RESET test. There is also the option of automatically dummying out large outliers. In our context, the overlapping nature of the dependent variable means that residuals will be auto-correlated and so the corresponding tests, including portmanteau tests, are switched off. Further, outliers can easily arise, especially over six or twelve-month horizons because of unpredictable changes in energy and other commodity prices. Heteroscedasticity could therefore be endemic: the corresponding tests are therefore switched off, but we use heteroscedasticity and autocorrelation corrected (HAC) t-ratios and F-tests for model selection.

Model selection given a GUM involves two potential errors: omitting relevant variables and including irrelevant variables. Heavy protection against the latter error results in smaller models being chosen. The default setting is chosen, limiting the probability to 0.05 of including an irrelevant variable, see Doornik (2009) for further details.

Forecasting

The alternative models are evaluated in terms of their pseudo out-of-sample forecasting performance based on root mean squared forecast error (RMSFE). Model selection takes place for monthly data running from January 1974 to December 1999. The reason for the January 1974 start (with a twelve-month-ahead inflation rate at January 1975) is to avoid some of the most volatile period of the first oil price shock, the collapse of Bretton

Woods and the confusion of the Nixon price and wage controls. Forecasts are then run up to December 2007. For example, for the twelve-month-ahead inflation model, a form of equation (1) or its simpler variants, is run for data on the regressors from 1974(1) to 1998(12), meaning that the dependent variable runs to 1999(12). From this model, the twelve-month inflation rate for 2000(1) is forecast using data on the regressors up to 1999(1). Adding one month of data at a time, the forecasting equation and the forecasts are recursively updated, concluding with the 2006(12) forecast for the 2007(12) twelve-month inflation rate. This generates 84 out-of-sample forecast residuals.

The second oil shock in 1979-80, coinciding with large policy shocks, results in large outliers in all models. For all the twelve-month-ahead models therefore, dummies are included for the largest of these in the last month of 1978 and the first five months of 1979 (reflecting price shocks and dramatic shifts in monetary policy in 1979-80). All twelve-month-ahead models also incorporate dummies for the first three months of 1974, reflecting the aftermath of the removal of the Nixon price controls, see Frye and Gordon (1980) and Campbell and Duca (2006). Two dummies for hurricane Katerina[8] are also included in all models.

[8] Hurricane Katerina caused fuel prices to spike in September and October 2005. We define an impulse dummy =1 in 2005(9). The twelve-month lead of the twelve-month change in this dummy, and its lag, would then capture the inflation effect seen in September and October 2005 if correctly anticipated in the twelve-month-ahead forecast made in September and October 2004. In recursive out-of-sample forecasting, we have no estimate of the coefficient on these dummies in 2004, so the surprise inflation of September-October 2005 shows up in the forecast errors. The advantage of including the dummies is that the dropping out of these effects twelve months later *is* captured in the forecasting model, mimicking real-time forecasting. Further, the precision of the parameter estimates of the models should improve by the inclusion of these dummies.

III. Results for the aggregate PC with data from 1974.

The definitions of the variables used in the study, and means and standard deviations of levels and changes are given in Table 1. This table also indicates the order of integration of the data: changes in log prices are I(0), while log levels are mostly I(1). The aggregate PC price index, first differences and twelfth differences (all in logarithms) are shown in Figure 2. The trending variables and the unemployment rate are shown in Figure 3. Note that trade openness and the Asian tigers' export share show the acceleration mentioned above.

The forecasting performance of the various methodologies is contrasted in Table 2 with reported root mean squared forecast errors (RMSFE). The methodologies are listed as column headings in Table 2. Each methodology is applied to a range of information sets, listed below Table 2, *Information Sets 1 to 7.*

The discussion of the empirical results begins with 'naïve' autoregressive models in *Information Set 1*. Standard methodology is to use the Akaike or Schwarz information criterion (BIC) to choose the lag length for the AR(k+1) process. We regress $\Delta_{12}\log PC_{t+12}$ on a constant, the above dummies, and $\Delta\log PC_{t-i}$, for i up to 23 months. Using BIC in this rudimentary form of model selection, we find k=5. This corresponds to the standard bench-mark model used in the forecasting literature. The recursive out-of-sample RMSFE for 2000(1) to 2006(12) is 0.0095 shown in row 1, column 1 of Table 2. This contrasts with 0.0088 for the unrestricted AR(24) model (not shown), suggesting that there is information in the longer inflation lags beyond k=5, which the information criterion does not pick up. However, when we adopt the PLL approach allowing longer lags to play a role: we regress $\Delta_{12}\log PC_{t+12}$ on a constant, the above dummies, and $\Delta\log PC_t$, $\Delta\log PC_{t-1}$, $\Delta\log PC_{t-2}$, $\Delta_6\log PC_{t-6}$, and $\Delta_{12}\log PC_{t-12}$.[9] The BIC criterion suggests dropping only the last term. This results in an RMSFE of 0.0087, shown in row 1, column 2 of Table 2. We adopt this univariate PLL model as the naïve reference model for comparison with more sophisticated models, and also to check whether model averaging or 'pooling' forecasts is helpful for forecasting performance, as often reported in the literature.

Automatic model selection applied to the univariate PLL model suggests a three parameter model.[10] The RMSFE is also 0.0087, shown in row 1, column 6. We thus have prima facie evidence, albeit in a simple context, that the main gain in RMSFE relative to the AR(k+1) benchmark with k=5, comes from our PLL parameterisation (which also uses 5 parameters for lags but uses lags up to 12 months).

It is important to note that throughout this exercise, automatic model selection is applied only once, for the estimation sample. An alternative procedure would have been to apply automatic model selection recursively, every time the data set expands by one month. This would be likely to lead to switches in the choice of parsimonious models to better reflect the evolving information content in the data, probably resulting in better forecast performance.

The information set is then sequentially expanded and the exercise repeated. *Information Set 2* of Table 2 adds a linear trend to the univariate autoregression: the RMSFE for the AR(5) model and our PLL variant are substantially worse, see row 2,

[9] We also check an equivalent parameterisation, $\Delta\log PC_t$, $\Delta\log PC_{t-1}$, $\Delta_3\log PC_t$, $\Delta_6\log PC_t$, $\Delta_{12}\log PC_t$ and $\Delta_{12}\log PC_{t-12}$ which can result in the choice of even more parsimonious models when model selection is introduced.

[10] Retaining only $\Delta\log PC_t$, $\Delta_6\log PC_t$ and $\Delta_{12}\log PC_t$ in the PLL parameterisation of the previous footnote, and four parameters in the alternative form of PLL, but with the same RMSFE implications.

columns 1 and 4. This corresponds to the intuition that the trend may be no friend for the forecaster - unbounded linear trends have implausible long-run implications. However, model averaging with the naïve reference model, see row 2, columns 3 and 5, restores performance close to that found when the linear trend is omitted. Here, automatic model selection helps forecast performance because it excludes the trend. Adding union density and changes in the unemployment rate, in place of a linear trend, in *Information Set 3* tends to improve on the naïve reference model, particularly for the pooled forecast. In Information Set 3A, we add the Asian export share, with little effect on the pooled forecasts, though the AR(k+1) and PLL based forecasts are a little worse. Replacing the Asian export share by US trade openness similarly does not lead to forecast gains effects (not shown).

The next extension of the data set is a major one: *Information Set 4* introduces log changes in producer prices, hourly earnings, import prices, house prices, oil prices and the real exchange rate. This is the single equation reduced form of a VAR in these variables, plus changes in the unemployment rate. However, we retain union density as an additional regressor. The BIC criterion now selects k=2 and the RMSFE is shown in column 1. Clearly, the six times k extra parameters weigh heavily in BIC, relatively to the improvement in fit they bring. The RMSFE is a little better than for the simple AR(6) specification and the pooled forecasts are also notably better than in row 1. However, our PLL parameterisation now brings a substantial improvement both for the standard and the pooled forecast, columns 4 and 5. Automatic model selection is unhelpful here, though forecast pooling again protects performance.

Adding the Asian export share in *Information Set 4A*, brings a severe deterioration in forecast performance, though forecast pooling again protects against the worst consequences. Clearly and not surprisingly, the unlimited nature of the non-linear trend represented by the Asian export share is a danger for forecasting. Similar findings apply for US trade openness (not shown). Automatic model selection provides no protection against these dangers, as columns 6 and 7 for *Information Set 4A* reveal.

Information Set 5 adds the ECM terms to *Information Set 4* (which includes union density, but not the Asian export share). Several ECM terms have the wrong sign (oil and import prices) and are excluded. Leading changes in the unemployment rate have positive (though not very significant) coefficients in the PLL specification in column 4 and are omitted, leaving $\Delta_6 UNR_{t-6}$ and $\Delta_{12} UNR_{t-12}$ as the included terms. Forecast performance deteriorates somewhat relative to *Information Set 4* though is still substantially better than for the *Information Sets 1 to 3* (including the naïve reference model). Automatic model selection has relatively neutral consequences here. On the face of it, the trending nature of the ECM terms looks problematic, in that there may be structural breaks in their trends. Adding the level of the unemployment rate to the *Information Set 5A* brings marginal improvements, most evident in the pooled forecasts.

The next extension to the information set brings substantial improvements: *Information Set 6* adds the non-linearity in recent inflation, SQRES, i.e. the residual from the squared value of $\Delta_6 \log PC_t$ as defined above, and its six month lag. This improves the RMSFEs quite strikingly, particularly for the PLL version, and as ever, for its pooled forecast version. Automatic model selection improves marginally in its pooled forecast version, see column 7 and column 5. Clearly, the non-linearities, which are highly significant over the estimation period, have important information content. Extending the *Information Set 6A* to include the level of the unemployment rate is relatively neutral for the AR(k) and general PLL versions of the model, but causes some deterioration when automatic model selection is applied

In *Information Set 6B*, the sign restrictions on the ECMs in oil and import prices and leading changes in the unemployment rate are relaxed: the effects are somewhat mixed. For the PLL specification and its model selected version, the standard forecasts are worse. The pooled forecasts for the general PLL model improve a little while those for the model selected version deteriorate notably. On balance, this suggests that sign restrictions are mildly helpful.

Finally, consider an extension of the PLL parameterisation incorporating some moving average formulations. This takes twelve-month moving averages of the log real exchange rate and the ECM terms are defined by [ma12log PC_t – log PC_t] . This formulation was inspired by noting the pronounced negative Δ_3, Δ_6 and Δ_{12} effects corresponding to most ECM terms, consistent with a backward shift in the average lag of the ECM term. For changes in the unemployment rate and the log real exchange rate, we use the PLL methodology (in the variant described in footnote 8), with three-month moving averages: $\Delta_3\text{ma3}X_t$, $\Delta_6\text{ma3}X_t$, $\Delta_{12}\text{ma3}X_t$, and $\Delta_{12}\text{ma3}X_{t-12}$. Effectively this averages monthly data to a quarterly frequency, and saves two parameters for each variable. With the sign restrictions again imposed, and union density and the squared terms included, we now achieve an RMSFE of 0.0056, the best of the models considered so far. The pooled forecast version now shows a slight deterioration with an RMSFE of 0.0057. This model also has the best BIC for the 1974-99 sample of all models considered. Automatic model selection here causes some deterioration in forecast performance, though forecast pooling offers substantial protection.

To summarise, the general points that emerge are as follows: first, our parsimonious longer lag (PLL) specification outperforms the AR(k+1) or VAR(k+1) alternative, where BIC is used to select the lag length k, in almost all comparisons. This suggests that PLL offers a powerful practical tool for overcoming the curse of dimensionality suffered by VAR models in modelling and forecasting contexts. Secondly, in almost all comparisons, model averaging or pooling of the more sophisticated models with a naïve reference model outperforms the forecasts from the more sophisticated models. Unsurprisingly, the worse the performance of the sophisticated model, the more dramatic is the benefit of pooling. In the only case where pooling did not help, the cost was small: raising RMSFE from 0.0056 to 0.0057. This suggests that the robustness benefits from the type of pooling carried out here, averaging the naïve autoregressive benchmark model, known for its robustness under structural breaks with models bringing in long-run information, are considerable.

Thirdly, there is useful information in lagged changes in log PPI, log earnings, log import prices, log oil prices, log house prices and the log real exchange rate for forecasting inflation. Fourthly, there is powerful evidence that non-linearities in the inflation process improve forecasting performance, when combined with equilibrium correction specifications of key drivers. Finally, automatic model selection confined to the estimation period has mixed outcomes for forecast performance: the benefits tend to be small, but more often there is a slight deterioration in forecast performance, and occasionally a more noticeable deterioration. In other words, basic GUM design appears to be more important than the use of model selection, once the GUM has been chosen. However, it seems likely that recursive automatic selection, where model selection occurs every time the sample is extended by one observation, would yield better results.

The selection of more parsimonious models with Autometrics has two potential advantages if there is little structural change. Parameter estimation uncertainty is reduced in more parsimonious representations consistent with the underlying data generating process. Lower estimation uncertainty should lower the forecast error variance. Further, the selected models pass several specification tests over the estimation period, of which the Chow test for parameter stability is the most important. This reduces the risk of selecting well-fitting but unstable models. However, if there is a structural break in the forecast period, a more parsimonious model may forecast less well since the more general model has more variables whose parameters can shift as the sample is updated recursively, to respond to the structural break. It is also possible that because of data collinearities in the estimation period, automatic model selection can exclude relevant variables. If adding further observations resolves such collinearities, a more general model may forecast better. Structural breaks and unresolved collinearities are plausible explanations for the tendency reported in Table 2 for the GUMs often to forecast better than the more parsimonious models selected on the basis of data ending in 1999.

IV. State dependence or a structural break?

For estimation over 1974-98, the previous section presented powerful evidence that the quadratic term SQRES (and its lag) is not only highly significant but brings sharp improvements in subsequent pseudo out-of-sample forecast performance. The quadratic term is a crude proxy for state dependence, with only a simple additive effect on the forecast twelve-month inflation rate. More complex models could allow several interaction effects, for example with the speeds of adjustment to the key drivers. The prospect of identifying such models is uncertain, to say the least.

An alternative approach is to throw away the data for the high inflation period before 1983, and estimate forecasting models on data from 1983. This allows for the possibility of a more general structural break in the early 1980s, perhaps because of a shift in monetary policy. The problem is that the low-variance sample for 1983 to 1998 is likely to be too short for the development of robust models. Thus, we undertake a more restrictive forecasting exercise using a longer estimation period of 1983 to 2007. The twelve-month-ahead inflation rate then spans 1984(1) to 2008(12). For this sample, the quadratic term SQRES and its lag are insignificant. The forecasting performance for 2009(1) to 2009(12) is compared for this short sample model, and the model estimated for 1974 to 2007 that included (highly significant) squared terms. Each is evaluated against the yardstick of the univariate autoregressive models. Forecasts for 2009 made in 2008 take one of the most challenging episodes in US inflation history of the last 35 years, given the rapid transition from 12-month inflation of 4.4% in July 2008 to mild 12-month deflation by May 2009.[11]

Table 3 shows the RMSFEs for forecasts of the twelve-month inflation rate in 2009(1)-2009(12) for the naïve models based on *Information Set* 1, for models based on *Information Set 7* and for pooled models of the two (described above). In each case,

[11] The price level peaked in September 2008, reached a trough in December 2008 and did not regain its previous peak until December 2009.

models are estimated for 1974-2008 and 1983-2008 samples. The RMSEs for all models are worse for this estimation period, than for the 1998-2006 forecasting exercise in Section 4. As before, the pooled models perform far better than either naïve model, with RMSEs around 75% of the naïve yardsticks. Further, the shorter sample model, 1983-2008, does somewhat better than the longer sample model, 1974-2008, both for the naïve and for the pooled forecasts. Given the extreme nature of the turning point in inflation, it is perhaps not surprising that the sophisticated models, i.e. not averaged with the naïve models that automatically build in inflation persistence, perform best of all for this period. The out-performance of the shorter sample model is consistent with a more general structural break in the inflation process pre-1983 than can be represented by the simple SQRES proxy for state dependence.

The forecast of 2009(12) based on 2008(12) data has the special distinction of being a *real time* forecast, that was based on data published for December 2008 in the second half of January 2009 without any knowledge of 2009 data. This can be verified by consulting a non-technical summary and application of these methods to examining the inflation outlook in January 2009 (Aron and Muellbauer, 2009). This concluded that, with model averaging, the point forecast for the 2009(12) annual inflation rate would be 0.008 based on the short sample estimates, and 0.015 based on the sample back to 1974. On early April 2010 vintage data, the actual annual inflation rate at 2009(12) was 0.021 Given forecast RMSEs of around 0.007 based on findings reported in section 3, these forecasts look like moderate success.

V. Conclusions

This study set out to disentangle six key factors that influence forecasting performance, using monthly inflation models estimated for the US personal consumption expenditure deflator (*PC*), the price index preferred by the Federal Reserve.

The first three factors concern the forecasting performance of different methodologies: (i) unrestricted standard AR(k+1) or VAR(k+1) specifications for inflation rates, with lag length selected by an information criterion, versus restricted models, applying the parsimonious longer lag (PLL) structure and other restrictions discussed above; (ii) the use of automatic model selection applied to the PLL models; and (iii) whether averaging forecasts confers any advantages in forecasting.

The remaining three factors concern the possible advantages of including in the models: (iv) long-run trends e.g. in union density, openness or the Asian share of world exports; (v) non-linearities in the inflation process, when high current or recent rates of inflation are associated with disproportionately higher future inflation; and (vi) long-run information in equilibrium correction terms, as against using models specified only in differences for non-stationary data.

The first three methodologies were applied to twelve different information sets. Twelve-month ahead inflation models were estimated over 1974-1998 and inflation was then forecast twelve months ahead, up to the end of 2007, updating the estimated model by one month every time the forecasts moved one month forward (i.e. recursive forecasting). This process attempts to replicate what forecasters might have done in real time. Since data revisions for most of the variables tend to be relatively small, the results should approximate genuine real time forecasting.

The literature on inflation forecasting agrees that simple autoregressive models are hard to beat. However, this paper demonstrates substantial out-performance against simple benchmark models with a variety of information sets. First, a key innovation is to use parsimonious longer lags (PLL) in place of the standard AR or VAR approach with a lag length of k+1 months, where k is chosen by the use of an information criterion. Section 3 demonstrated that in almost every information set, the PLL specification produces better forecasts than the AR(k+1) or VAR(k+1) approach. This suggests that standard VAR methods tend to omit relevant longer lags and that PLL offers a powerful practical tool for overcoming the curse of dimensionality suffered by VAR models in modelling and forecasting contexts.

Second, further evidence is provided for the often-reported finding in the forecasting literature that forecast pooling or averaging improves forecast performance. In particular, the simple averaging of a naïve forecast based on univariate data and a more sophisticated model in almost all comparisons outperforms both the sophisticated and the naïve model. The worse the performance of the sophisticated model, the more dramatic is the benefit of pooling. The robustness benefits of pooling the naïve autoregressive benchmark model, known for its robustness under structural breaks, with models, bringing in long-run information, are considerable.

Third, the paper has demonstrated the usefulness of information on oil prices, producer price indices, hourly earnings, the real exchange rate, import prices, house prices, and changes in the unemployment rate for forecasting PC inflation, both in models specified in first differences, and in more sophisticated models with error correction terms. Fourth, the combination of equilibrium correction terms, which bring gradual adjustment of relative prices into the inflation process, and non-linearities, was shown to contribute importantly to out-performance relative to the benchmark model. It is possible that these factors account for important parts of the drift in the univariate inflation process for the US, reported by Stock and Watson (2006).

Fifth, some of the potential pitfalls of forecasting are illustrated by incorporating a linear trend and, even worse, the export share of Asian 'tiger' economies, in some of the models. Unbounded trends have implausible implications for inflation rates in the long run, and examples of poor pseudo out-of-sample forecasting performance are reported when such trends are included in equations for log *PC*. However, the inclusion of trade union density, which has declined towards a plausible lower bound, mostly improves forecast performance, particularly in a multivariate context.

Finally, another issue concerning forecasting methodology explored at both the aggregate and the sectoral level is whether model selection improves performance. Automatic model selection with Autometrics (Doornik, 2009), was used to select parsimonious models from general unrestricted models (GUMs). The evidence proves to be somewhat mixed: the gains from selection are usually small, but in several instances the deterioration of forecast performance is notable. It is likely that less restricted models are better able to handle structural breaks and the resolution of previous multi-collinearities. However, the test posed in this paper for the automatic selection method is tougher than real time forecasters actually face. In the paper, model selection takes place just once – for the 1974 to 1998 sample. In practice, real time forecasters could reselect the parsimonious model on a rolling monthly basis, allowing the model to react far better to structural breaks and other new information on the parameters. Thus, it is plausible that regular reselection would result in better performance.

Our work also throws modest light on the causes of the reduction since the early 1980s in inflation and in its volatility (the ‘great moderation’, Bernanke, 2004). Our key findings are two: the first is that the decline in union density has a significant negative effect on inflation in most specifications. The second is consistent with an element of the ‘luck’ hypothesis: evidence of non-linearities in the inflation process suggests anything that reduces inflation will tend to keep inflation down, and recent lower inflation volatility is associated with lower inflation to come.

Forecast errors are clearly associated with unpredictable shocks to oil and raw food prices. To illustrate the importance of the equilibrium correction terms, we find using *Information Set 7* (Table 3, column 4) that the long-run solution is approximately of the form:[12]

$\log PC = \text{constant} + 0.0168\ UNDENS + 0.57 \log ULC + 0.31 \log HP + 0.12 \log FP$,

where *FP* measures foreign consumer prices and *UNDENS* is union density. By coincidence, the weight on housing rents in the CPI-U (consumer price index for urban households), including imputed rent and lodging away from home, has been 0.32 in recent years, though the housing rents weight is somewhat lower in the PCE-deflator. House prices are strongly correlated with prices of commercial property which ultimately feed into business costs, including commercial rents. Hence the weight of about 31% in the long run solution for *PC* seems reasonable. Broadly similar weights are found in a

[12] The sample is from 1974(1) to 2007(12) (i.e. using 2008(12) data on the 12-month inflation rate) and the coefficients are: 0.33 (17.4) for log unit labour costs, 0.18 (26.8) for log house prices, – 0.07 (-18.5) for the log real exchange rate, and 0.0098 (22.2) for trade union density, where t-ratios are given in parenthesis.

model using *Information Set 7*, forecasting the Federal Reserve's measure of *core* inflation twelve months ahead. This is another indication of robustness, with food and energy prices removed.

Our equations suggest that the inflation outlook in the US is likely to remain relatively subdued for several years from 2010. The lags from unit labour costs and house prices to PCE inflation are long. With unit labour costs showing little risk of major rises and falls in house prices still to feed through, potential rises in oil and other commodity prices will be offset by these important moderating forces. Our models also throw important light on the monetary transmission mechanism linking interest rates to consumer price inflation via the exchange rate, house prices and unemployment. The house price channel of monetary transmission is missing in standard New Keynesian models.

References

Altissimo, Filippo, Laurent Bilke, Andrew Levin, Thomas Mathä and Benoit Mojon. 2006. "Sectoral and Aggregate Inflation Dynamics in the Euro Area." *Journal of the European Economic Association* 4(2-3): 585-593.

Aron, J. and J. Muellbauer. 2009. "US Deflation? New Methods of Forecasting Consumer Prices." *Economic Outlook* 33 (1): 33-45, January.

Álvarez, Luis, Emmanuel Dhyne, Marco Hoeberichts, Claudia Kwapil, Hervé Le Bihan, Patrick Lünnemann, Fernando Martins, Roberto Sabbatini, Harald Stahl, Philip Vermeulen and Jouko Vilmunen. 2006. "Sticky Prices in the Euro Area: A Summary of New Micro-evidence." *Journal of the European Economic Association* 4(2-3): 575-584.

Angeloni, I., Aucremanne, L., Ehrmann, M., Galí, J., Levin, A. and F. Smets. 2006. "New Evidence on Inflation Persistence and Price Stickiness in the Euro Area: Implications for Macro Modelling." *Journal of the European Economic Association.* 4(2-3), 562-574.

Bårdsen, Gunnar, Eilev S. Jansen, Ragnar Nymoen. 2004. "Econometric Evaluation of the New Keynesian Phillips Curve." *Oxford Bulletin of Economics & Statistics* 66(1): 671-686.

Barth, M. and V. Ramey. 2001. “The Cost Channel of Monetary Transmission.” *NBER Macroeconomics Annual 16*: 199-240.

Bates, J. M. C. W. J. Granger. 1969. “The Combination of Forecasts.” *Operations Research* 20(4) (Dec., 1969): 451-468

Batini, N., B. Jackson and S. Nickell. 2005. “An Open Economy New Keynesian Phillips Curve for the UK.” *Journal of Monetary Economics* 52: 1061-1071.

Bernanke, Ben S. 2004. "The Great Moderation", speech delivered at the meetings of the Eastern Economic Association, Washington, D.C., February 20.

Boug, Pål , Ådne Cappelen and Anders Rygh Swensen. 2006. "The New Keynesian Phillips Curve for a Small Open Economy." Discussion Papers 460, Research Department of Statistics Norway.

Calvo, Guillermo A. 1983. “Staggered Pricing in a Utility Maximizing Framework.” *Journal of Monetary Economics* 12: 383-398.

Campbell, Carl and John Duca. 2006. "The Impact of Evolving Labor Practices and Demographics on U.S. Inflation and Unemployment." ms. Dallas Federal Reserve, December.

Clarida, R., Galí, J. and Gertler, M. 1999. “The science of monetary policy: a New Keynesian perspective.” *Journal of Economic Literature* 37: 1661–1707.

Clements, M.P. and D.F. Hendry. 2002. “Modelling methodology and forecast failure.” *Econometrics Journal* 5: 319–344.

Clements, M. P. and D. F. Hendry. 1998. *Forecasting Economic Time Series.* Cambridge: Cambridge University Press.

Clements, Michael P. and David F. Hendry. 1996. “Multi-step estimation for forecasting.” *Oxford Bulletin of Economics and Statistics* 58: 657-684.

De Bondt, Werner F. M. and Mary M. Bange. 1992. “Inflation Forecast Errors and Time Variation in Term Premia.” *The Journal of Financial and Quantitative Analysis* 27 (4): 479-496.

Doornik, J.A. 2009. “Autometrics.” in J. L. Castle and N. Shephard (eds*.), Festschrift in Honour of David F. Hendry*, Oxford University Press.

Fama, Eugene. 1990. “Term-structure forecasts of interest rates, inflation and real returns.” *Journal of Monetary Economics* 25: 59-76.

Forsells, Magnus and Kenny, Geoff. 2002. “The Rationality of Consumers' Inflation Expectations: Survey-based Evidence for the Euro Area.” ECB Working Paper No. 163 (August).

Frye, J. and R. Gordon. 1980. “The variance and acceleration of inflation in the 1970s: alternative exploratory models and methods.” NBER working paper 0551.

Hendry, D.F. M.P. Clements. 2002 “Pooling of Forecasts.” *Econometrics Journal* 5: 1-26.

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

Hoeting, Jennifer A. David Madigan, Adrian E. Raftery, Chris T. Volinsky. 1999. “Bayesian Model Averaging: A Tutorial.” *Statistical Science* 14 (4) (Nov., 1999): 382-401.

Kozicki, Sharon. 1997. “Predicting real growth and inflation with the yield spread.” Federal Reserve Bank of Kansas City - Economic Review, Fourth Quarter.

Mavroeidis, S. 2005. “Identification Issues in Forward-looking Models Estimated by GMM, with an Application to the Phillips Curve.” *Journal of Money, Credit and Banking* 37: 421-448.

Peach, R. W., R. Rich, and A. Antoniades. 2004. “The historical and recent behaviour of goods and services inflation.” *Economic Policy Review*, Federal Reserve Bank of New York, December, 19-31.

Reis, R. 2006. “Inattentive Producers.” *Review of Economic Studies* 73: 793–821.

Rich, R. and C. Steindel. 2005. “A Review of Core Inflation and an Evaluation of Its Measures.” Federal Reserve Bank of New York Staff Reports, Staff Report no. 236, December.

Roberts, John M. 1995. "New Keynesian Economics and the Phillips Curve." *Journal of Money, Credit and Banking* 27: 975–84

Roberts, John M. 1997. “Is Inflation Sticky?” *Journal of Monetary Economics* 39(2): 173-196.

Romer C.D. and D.H. Romer. 2000. “Federal Reserve information and the behavior of interest rates.” *American Economic Review* 90: 429–457.

Rudd, J. and K. Whelan, 2007. "Modeling Inflation Dynamics: A Critical Review of Recent Research," *Journal of Money, Credit and Banking* 39(s1): 155-170.

Sheshinski, E. and Y. Weiss. 1977. “Inflation and costs of price adjustment.” *Review of Economic Studies* 44: 287-303.

Sims, C. 1980. “Macroeconomics and reality.” *Econometrica* 48: 1-48.

Stock, J.H., and M.W. Watson. 2007. “Why Has U.S. Inflation Become Harder to Forecast?” *Journal of Money, Credit, and Banking* 39, 3-34.

Stock, J. H. and M. W. Watson. 2006. “Why Has US Inflation Become Harder to Forecast?” NBER working paper, w12324, June.

Stock JH, Watson MW. 2004. “Combination forecasts of output growth in a seven country dataset.” *Journal of Forecasting* 23: 204–430.

Stock, J. H. and M. W. Watson. 2003. "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature* 41(3): 788-829, September.

Stock, J. H. and M. W. Watson. 1999. “Forecasting Inflation.” *Journal of Monetary Economics* 44 (2): 293–335.

Watson, Mark W. 2003. “Macroeconomic Forecasting Using Many Predictors.” In Dewatripont, M., L. Hansen, and S. Turnovski (eds), *Advances in Econometrics: Theory and Applications*, Eighth World Congress of the Econometric Society, Vol. III, 87-115.

Weiss, A. A. 1991. “Multi-step estimation and forecasting in dynamic models.” *Journal of Econometrics* 48: 135-149.

Woodford, Michael. 2003. *Interest and Prices*. Princeton, NJ: Princeton University Press.

TABLE 1: *Statistics and Variable Definitions: 1974:1-2007:12*

Variable	*Definition of variable*	*Mean*	*Std Dev.*	*I(1)†*	*I(2)‡*
log PC	Log of consumer expenditure deflator	4.29	0.354	-3.69*	-5.90**
log POIL	Log of West Texas Intermediate crude oil price	3.12	0.423	-1.71	-16.9**
log REER	Log of the real effective exchange rate (uses consumer price indices for trading partners to deflate nominal effective exchange rate)	4.55	0.098	-1.96	-14.8**
log PIMP	Log of non-petroleum import prices	4.56	0.173	-3.08	-4.75**
log HP	Log of median US house prices	11.6	0.515	-2.35	-21.0**
log ULC	Log of unit labour costs in non-agric. private sector (monthly interpolation of quarterly data)	4.46	0.303	-3.55*	-5.37**
log HEPRIV	Log of hourly earnings in non-agric. private sector	2.27	0.370	-3.94*	-4.81**
log PPIALL	Log of wholesale price for total domestic goods	4.66	0.280	-3.64*	-8.33**
UNDENS	Union density, 12-month moving average	13.8	4.99	NA	NA
ASIAEXP	Share in World manufactured exports of Asian economies (excl. Japan, Israel), 12-month moving average of annual interpolated data	5.27	2.69	NA	NA
ROPENMA12	Conventional trade policy measure in real terms: ratio of real exports plus real imports to real GDP, 12-month moving average of monthly data	134	42.9	NA	NA
UNR	Unemployment rate for those aged 25 or over (%)	4.75	1.12	-3.43*	-7.18**

Source: Data from BEA (US), BLS (US), BIS, BP, IFS (International Monetary Fund), US Census Bureau, UN Monthly Digest, FRED and Dallas Federal Reserve Bank. Statistics are reported to three significant figures.

Notes:

†For a variable X, the augmented Dickey-Fuller (1981) statistic is the t ratio on π from the regression: $\Delta X_t = \pi X_{t-1} + \Sigma_{i=1,k} \theta_i \Delta X_{t-i} + \psi_0 + \psi_1 t + \varepsilon_t$, where k is the number of lags on the dependent variable, ψ_0 is a constant term, and t is a trend. The kth-order augmented Dickey-Fuller statistic is reported, where k is the last significant lag of the 3 lags employed. The trend is included if significant. For null order I(2), ΔX replaces X in the equation above. Critical values are obtained from MacKinnon (1991). Asterisks * and ** denote rejection at 5% and 1% critical values. Stationarity tests are performed for the variables in levels before time-transformation.

‡For several price level variables the test statistics are based on the inclusion of a time trend, though this is not strictly significant. If the trend is excluded, the test would suggest they are I(0) but with an implausibly low speed of adjustment (less than 1%).

TABLE 2: *Root mean square forecast errors (12 months ahead) for the aggregate price index, PC, comparing information sets and models.*

*Information sets**	*AR(k+1) or VAR(k+1) choosing k with BIC†*	*Univariate AR(k+1)) imposing PLL‡ (reference model)*	*Averaging using PLL: average of col.1 + col.2 forecasts*	*AR(k+1) or VAR(k+1) imposing PLL*	*Averaging plus PLL: average of col.2 + col.4 Forecasts*	*AR(k+1) or VAR(k+1) imposing PLL plus model selection*	*Averaging plus PLL plus model selection: average of col.2 + col.6 forecasts*
	1	*2*	*3*	*4*	*5*	*6*	*7*
Info Set 1	0.0095	0.0087	0.0091	0.0087	0.0087	0.0087	0.0087
Info Set 2	0.0109		0.0087	0.0107	0.0088	0.0087	0.0087
Info Set 3	0.0094		0.0084	0.0097	0.0085	0.0099	0.0086
Info Set 3A	0.0097		0.0084	0.0099	0.0085	0.0099	0.0085
Info Set 4	0.0089		0.0080	0.0078	0.0077	0.0085	0.0079
Info Set 4A	0.0122		0.0089	0.0113	0.0087	0.0149	0.0099
Info Set 5	0.0106		0.0084	0.0099	0.0080	0.0097	0.0081
Info Set 5A	0.0098		0.0082	0.0094	0.0079	0.0101	0.0080
Info Set 6	0.0093		0.0074	0.0070	0.0063	0.0074	0.0062
Info Set 6A	0.0088		0.0073	0.0070	0.0064	0.0094	0.0071
Info Set 6B	0.0086		0.0072	0.0079	0.0061	0.0094	0.0069
Info Set 7	0.0108		0.0078	0.0056	0.0057	0.0083	0.0061

Notes:

†*AR(k)* is an unrestricted autoregressive model (constant and lags) in the relevant information set. An information criterion is used to select k.

‡The parsimonious longer lags (PLL model) for variable Z was defined above by ΔZ_t , ΔZ_{t-1}, ΔZ_{t-2}, $\Delta_3 Z_{t-3}$, $\Delta_6 Z_{t-6}$, and $\Delta_{12} Z_{t-12}$.

*Information sets are (variable definitions are in Table 1):

Info Set 1: *Differences only in univariate model*: constant and lags in Δlog *PC,* plus 1974 and 1979 dummies (oil shocks, monetary policy regime change)
Info Set 2: *Differences only plus trend:* Set 1 plus linear trend
Info Set 3: *Differences only plus trends:* Set 1 plus *UNDENS* and changes in unemployment rate, Δ*UNR*
Info Set 3A: *Differences only plus trends:* Set 3 plus *ASIAEXP*
Info Set 4: *Differences only plus trends:* Set 3 plus changes in log *PPI*, log *EARN*, log *REER*, log *PIMP*, log *HP* and log *PC*
Info Set 4A: *Differences only plus trends:* Set 4 plus *ASIAEXP*
Info Set 5: *ECM model plus trends:* Set 4 plus ECM terms and level log *REER* (with sign restrictions)
Info Set 5A: *ECM model plus trends:* Set 5 plus unemployment rate level, *UNR*
Info Set 6: *ECM model plus trends plus non-linear terms:* Set 5 plus squared terms
Info Set 6A: *ECM model plus trends plus non-linear terms:* Set 6 plus *UNR*
Info Set 6B: *ECM model plus trends plus non-linear terms:* Set 6 without sign restrictions
Info Set 7: *ECM model plus trends plus non-linear terms:* Set 6 using MA12 in ECM terms and level log *REER* and MA3 in Δ log *REER* and Δ*UNR* terms (with sign restrictions)

TABLE 3:

Root mean square forecast errors (12 months ahead) for the aggregate price index, PC, comparing information sets and models.

*Information sets**	*Info Set 1 1974-2007*	*Info Set 7 1974-2007*	*Pooled 1974-2007*	*Info Set 1 1983-2007*	*Info Set 7 1983-2007*	*Pooled 1983-2007*
	1	*2*	*3*	*4*	*5*	*6*
RMSFE	0.0367	0.0209	0.0263	0.0303	0.0180	0.0235

Notes:

*Information sets are (variable definitions are in Table 1):

Info Set 1: *Differences only in univariate model*: constant and lags in Δlog *PC,* plus 1974 and 1979 dummies (oil shocks, monetary policy regime change)

Info Set 7: *ECM model plus trends plus non-linear terms:* Set 6 using MA12 in ECM terms and level log *REER* and MA3 in Δ log *REER* and Δ*UNR* terms (with sign restrictions); Set 6 is specified below Table 2.

Figure 1. Annual inflation rates for all goods and services, durables, non-durables and services

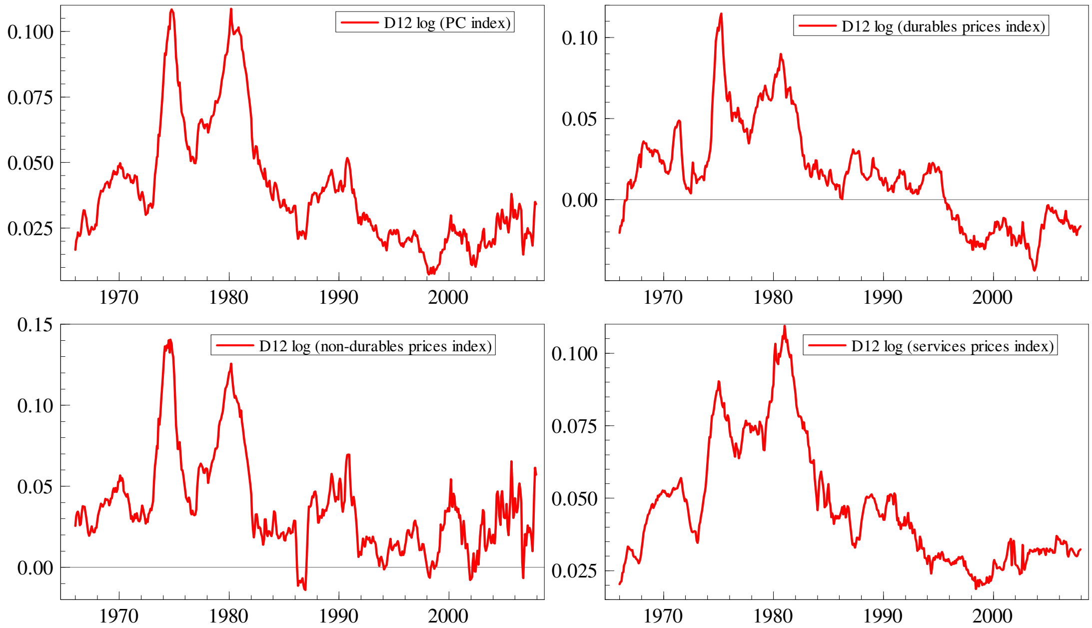
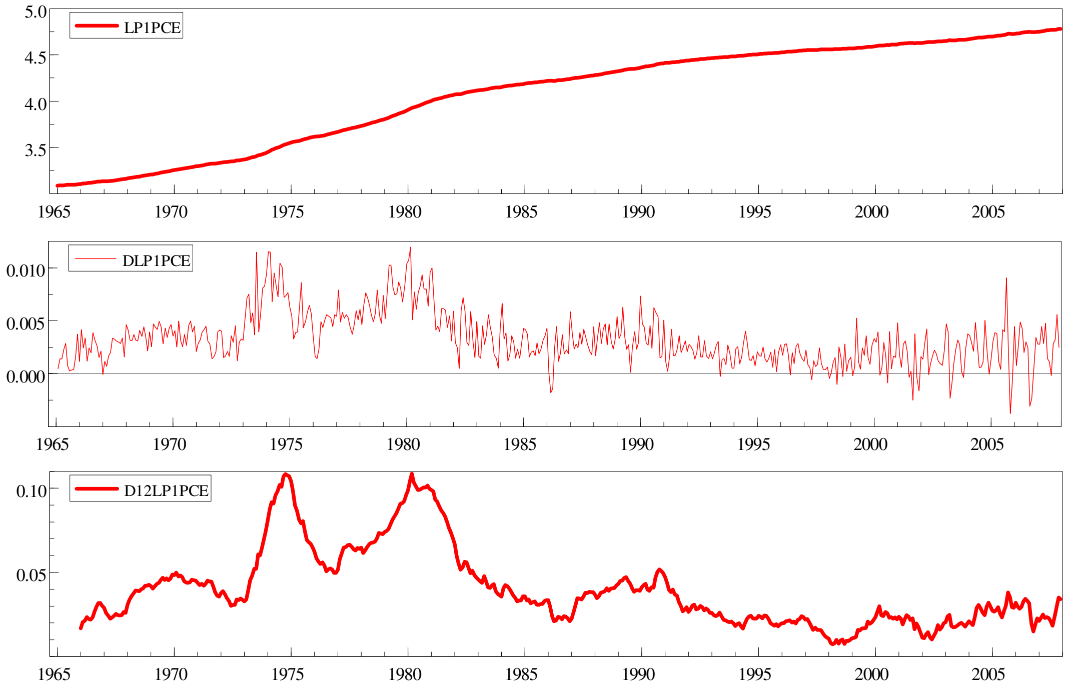
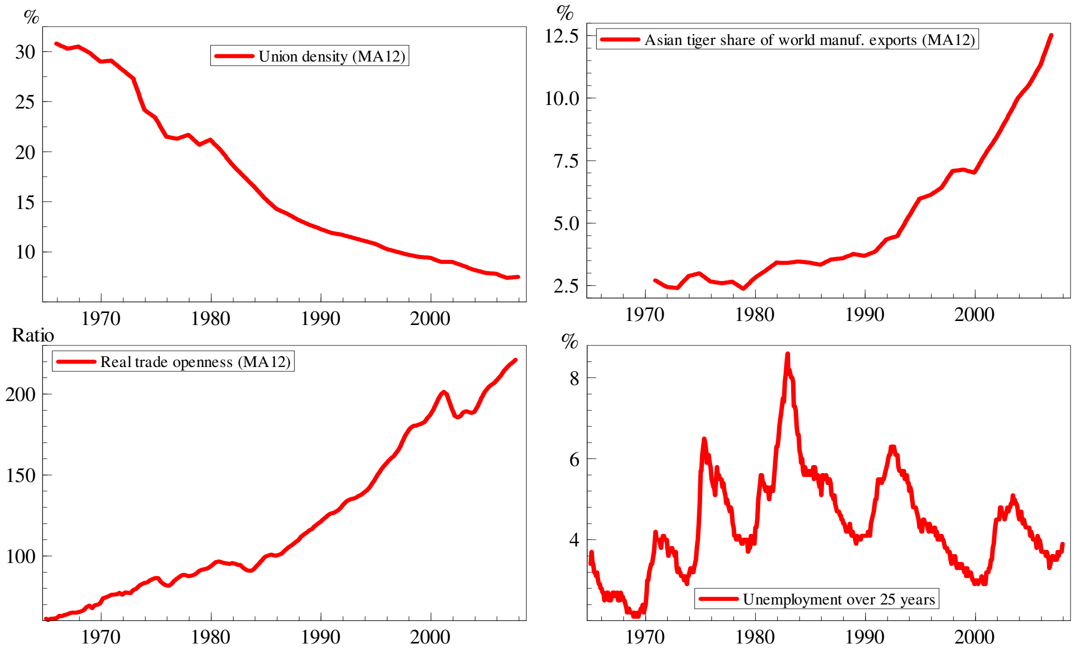


Figure 2. The aggregate PC price index, first differences and twelfth differences (all in logarithms)



Note: Definitions are in Table 1.

Figure 3. Union density, the export share of Asian 'tigers', real trade openness and unemployment



Note: Definitions are in Table 1.