UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

High-Dimensional Forecasting with Known Knowns and Known Unknowns

M. Hashem Pesaran

Ron P. Smith

Abstract

Forecasts play a central role in decision making under uncertainty. After a brief review of the general issues, this paper considers ways of using high-dimensional data in forecasting. We consider selecting variables from a known active set, known knowns, using Lasso and OCMT, and approximating unobserved latent factors, known unknowns, by various means. This combines both sparse and dense approaches to forecasting. We demonstrate the various issues involved in variable selection in a high-dimensional setting with an application to forecasting UK inflation at different horizons over the period 2020q1-2023q1. This application shows both the power of parsimonious models and the importance of allowing for global variables.

Replication files: https://www.econ.cam.ac.uk/people-files/emeritus/mhp1/fp24/Replication_files_for_Pesaran_Smith_NIER_4_April_2024.zip

Reference Details

CWPE	2406
Published	6 February 2024
Revised	21 March 2024
Key Words	Forecasting, high-dimensional data, Lasso, OCMT, latent factors, principal components
JEL Codes	C53, C55, E37, E52
Website	www.econ.cam.ac.uk/cwpe

High-dimensional forecasting with known knowns and known unknowns*

M. Hashem Pesaran
University of Southern California, and Trinity College, Cambridge

Ron P. Smith
Birkbeck, University of London

March 21, 2024

Abstract

Forecasts play a central role in decision making under uncertainty. After a brief review of the general issues, this paper considers ways of using high-dimensional data in forecasting. We consider selecting variables from a known active set, known knowns, using Lasso and OCMT, and approximating unobserved latent factors, known unknowns, by various means. This combines both sparse and dense approaches to forecasting. We demonstrate the various issues involved in variable selection in a high-dimensional setting with an application to forecasting UK inflation at different horizons over the period 2020q1-2023q1. This application shows both the power of parsimonious models and the importance of allowing for global variables.

JEL Classifications: C53, C55, E37, E52

Key Words: Forecasting, high-dimensional data, Lasso, OCMT, latent factors, principal components

*This paper developed from Pesaran's Deane-Stone Lecture at the National Institute of Economic and Social Research, London. We greatly benefitted from comments when earlier versions of this paper were presented in 2023 at NIESR on 21 June, the International Association of Applied Econometrics Annual Conference in Oslo, June 27-30, 2023, Bayes Business School, City University, 22 November 2023, and at Economics Department-Wide Seminar, Emory University, 16 February, 2024. We are also grateful for comments from Alex Chudik, Anthony Garratt, George Kapetanios, Essie Maasoumi, Alessio Sancetta, Mahrad Sharifvaghefi, Allan Timmermann and Stephen Wright. We particularly thank Hayun Song, our research assistant, for his work in coding, empirical implementation, and the tabulation of results, along with his invaluable research assistance.

"All the business of war, indeed all the business of life, is endeavour to find out what you don't know by what you do; that's what I called '*guess what was the other side of the hill*' "

Duke of Wellington

1 Introduction

Forecasts play a central role in decision making under uncertainty. Good forecasts are those that lead to good decisions, in the sense that the expected payoff to the decision maker using the forecast is greater than it would be otherwise.[1] In the case of inflation forecasts, which we consider below, the Bank of England makes forecasts to help it set monetary policy to keep inflation within a target range.[2] The payoff is the variation of inflation around target. However, it is not clear how one would quantify the contribution of the forecast to the payoff in terms of a specific central bank loss function.

Since forecasts are designed to inform decisions, they are inherently linked to policy making. However, there is an issue as to whether one should use the same model for both forecasting and setting the policy instruments. Different questions require different types of model to answer them. A policy model might be quite large, while a forecasting model might be quite small. There is also an issue of how transparent the model should be. It may be difficult to interpret why a machine learning statistical model makes the predictions it does and this can be a major disadvantage when policy requires communication of a persuasive narrative.

In recent years forecasting has been influenced by the increasing availability of high-dimensional data, improvements in computational power and advances in econometrics and machine learning techniques. In some areas, such as meteorology, this has resulted in improved forecasts, increasing the number of hours ahead for which accurate predictions can be made. The improved forecasts lead to better decision making as people

[1] The linkage between forecasting and decision making is discussed in Granger and Pesaran (2000a,b), who argue in favour of a closer link between the decision and forecast evaluation problems. Pesaran and Skouras (2004) provide a more general survey of decision theoretic approaches to forecast evaluation.

[2] The letter of 26 June 2023 by Huw Pill, Bank chief economist to the chair of the House of Commons Treasury Committee sets out his assessment of the role played by the forecasts in the policy process.

change their behaviour in response to the predictions and the effect of such responses on mortality from heat and cold is examined in Shrader, Bakkensen, and Lemoine (2023). Despite advances in data, computation and technique, the improvement in accuracy of weather forecasts has not been matched by economic forecasts. This is a cause for concern, since as emphasised in the classic analysis of Whittle (1983), prediction and control are inherently linked and decisions over such elements of economic management as monetary policy are dependent on a view of the future.

Macroeconomic forecasting is challenging because lags in responses to policies or shocks are long and variable and the economic system is responsive, events prompting changes in the structure of the economy. Forecasting tends to be relatively successful during normal times, but in times of crises and change, in the face of large shocks or structural changes, when accurate predictions are most needed, forecasters tend to fail. For instance, inflation was 5.4% in December 2021. This was the last figure they had when, in February 2022, the Bank forecast that inflation would peak at 7.1% in 2022q2, and fall to 5.5% in 2023q1, and be back within target at 2.6% in 2024q1. The 2023q1 actual was 10.2%, almost 5 percentage points higher than forecast. This burst in inflation was a global phenomenon and other central banks made similar errors.

Economic forecasters may use purely statistical models or more structural economic models, which include the policy variables and important economic linkages. We will call these more structural models 'policy models', given the way structural has a number of interpretations. The statistical models will typically be conditional on information available at the time of the forecast, which may be inaccurate: knowing where one is at the time of forecast, nowcasting, is an important element. Policy models may also be conditional on assumed future values. The Bank of England makes forecasts conditional on market expectations of future interest rates, assumptions about future energy prices and government announcements about future fiscal policy as well as other measures. Wrong assumptions about those future values may cause problems as it did for the Bank in August 2022, when it anticipated a rise in energy prices but not the government response to them, so over-estimated inflation. There is also a policy issue as to whether fiscal and monetary policy should be determined independently by different institutions or jointly.

If both statistical and policy models are used, there is an issue as to how to integrate them. Forecast averaging has been widely shown to improve forecast performance, but forming averages on many variables may lack coherence and consistency. In September 2018, the Bank of England Independent Evaluation Office, IEO reported back to Court of the Bank on the implementation and impact of the 2015 IEO review of forecasting performance. IEO reported that some 'non-structural' models had been introduced as a source of challenge, and outputs were routinely shown to the Monetary Policy Committee as a way of cross checking the main forecast. While, some but not all members found them helpful, but there was no desire to develop more models of this sort and internally they had not been integrated into the forecast process as a source of challenge.

Forecasts by central banks fulfill multiple purposes including as a means of communication to influence expectations in the wider economy. This also makes it difficult to choose a loss function to evaluate forecasts. For instance, Bank of England forecasts for inflation at a 2 year horizon are always close to the target of 2 per cent. Even were the Bank to think it unlikely that it could get back to target within 2 years, it might feel that its credibility might be damaged were it to admit that. An institutional issue is who "owns" the forecast. The Bank of England forecast is the responsibility of the 9 member Monetary Policy Committee (MPC); other central banks have different systems. For instance the US Federal Reserve has a staff forecast not necessarily endorsed by the decision makers.

There is an issue about the optimal amount of information to use: both with respect to breadth, how many variables, and length, how long a run of data. With respect to breadth, in principle, one should use information on as many variables as possible and not just for the country being forecast, since in a networked world foreign variables contain information. This is the information that is used in the Global VAR (GVAR) whose use is surveyed in Chudik and Pesaran (2016). While the use of many variables might imply a large model, in practice quite parsimonious small models tend to be difficult to beat in forecasting competitions.

With respect to length, in a May 2023 hearing, the Chair of the House of Commons Treasury Select Committee asked the chief economist of the Bank of England, Huw Pill: "Are you saying that, despite the Bank of England having been in existence for over 300 years, you look at only the last 30 years when you think about what the risks are to inflation?". Pill emphasised the importance of the policy regime, which had been different in the past 30 years of inflation targeting than in earlier high inflation periods. The 30 years up to 2019 had also been different in terms of the absence of large real shocks, like Covid-19 and the effects of the Russian invasion of Ukraine.

Whether it is statistical or policy, the model will typically be supplemented by a judgemental input, justified by the argument that the forecaster has a larger information set than the model. In evidence to the Treasury Select Committee in September 2023, Sir Jon Cunliffe said: "We start with the model. All models are caricatures of real life. There is a suite of models; that is the starting point. But then the MPC itself puts judgments that change the model, and we have made some quite big judgments in the past about inflation persistence and the like. Finally, when we have the best collective view of the committee, which is our judgment on top of the model, the model keeps us honest. It ensures that there is a general equilibrium and we cannot just move things around."

In short, macroeconomic forecasting faces important challenges. It depends on how forecasts are announced and used in the decision making process. To deal with a constantly changing economic environment, forecasts must continually adapt to new data sets, statistical techniques and theory-based economic insights, knowing that there are still key variables that might have been left out, either due to difficulties in measurement, oversight or ignorance. Forecasters must answer a range of difficult questions. What sample periods and potential variables to consider? How to decide which variables to use for forecasting, and whether to use the same sample periods for variable selection and for forecasting? Should one use ensemble forecasting from forecasts obtained either from different models or from the same model estimated over different sample sizes or with different degrees of down-weighting? One must only be humbled by the sheer extent of the uncertainty that these choices entail. It is within this wider context that this paper tries to formalize some elements of the problem of forecasting with high-dimensional data and illustrates the various issues involved with an application to forecasting UK inflation.

The rest of this paper is organized as follows. Section 2 sets out the high-dimensional forecasting framework we will be considering. Section 3 considers "known knowns", selecting relevant variables from a known active set. Section 4 considers "known unknowns" where there are known to be unobserved latent variables. Section 5 presents the empirical application on forecasting UK inflation. Section 6 contains some concluding comments.

2 The high-dimensional forecasting problem

Suppose the aim is to forecast a scalar target variable, denoted by y_{T+h}, at time T, for the future dates, $T+h$, $h=1,2,...,H$. Given the historical observations the optimal forecast of the target variable, y_{T+h}, depends on how the forecasts are used, namely the underlying decision problem. In practice, specifying loss functions associated with decision problems is hard, hence the tendency to fall back on mean squared error loss. Under this loss function the optimal forecasts are given by conditional expectations, $E\left(y_{T+h}\left|\mathcal{I}_T\right.\right)$, where $\mathcal{I}_T$ is the set of available information, and expectations are formed with respect to the joint probability distribution of the target variable and the set of potential predictors under consideration. But when the number of potential predictors, say K, is large even this result is too general to be of much use in practice.

The high-dimensional nature of the forecasting problem also presents a challenge of its own when we come to multi-step ahead forecasting when forecasts of the target variable are required for different horizons, $h=1,2,...,H$. Many decision problems require having forecasts many periods ahead, months, years and even decades ahead. Monetary policy is often conducted over the business cycle, at least 2-3 years ahead of the policy formulation. Climate change policy requires forecasts over many decades ahead. In interpreting Pharaoh's dreams, Joseph considered a two-period decision problem whereby seven years of plenty are predicted to be followed by seven years of drought. Multi-horizon forecasting is relatively straightforward when the number of potential predictors is small and a complete system of equations, such as a vector autoregression (VAR), can be used to generate forecasts for different horizons from the same forecasting model in an iterative manner. Such an **iterated** approach is not feasible, and might not even be desirable, when the number of potential predictors is too large, since future forecasts of predictors are also needed to generate forecasts of y_{T+h} for $h \geq 2$. This is why in high-dimensional set ups multi-period ahead forecasts are typically formed using different models for different horizons. This is known as the

direct approach and avoids the need for forward iteration by directly regressing the target variable y_{t+h} on the predictors at time t, thus possibly ending up with different models and/or estimates for each h.[3]

To be more specific, ignoring intercepts and factors which we introduce below, suppose y_t is the first element of the high-dimensional vector $\boldsymbol{w}_t$, assumed to follow the first order VAR model,

$$\boldsymbol{w}_t = \boldsymbol{\Phi}\boldsymbol{w}_{t-1} + \mathbf{u}_t. \tag{1}$$

Higher order VARs can be written as first order VARs using the companion form. The error vector, $\mathbf{u}_t$, satisfies the orthogonality condition $E\left(\mathbf{u}_t \left|\mathcal{I}_{t-1}\right.\right) = \mathbf{0}$, where $\mathcal{I}_{t-1} = (\boldsymbol{w}_{t-1}, \boldsymbol{w}_{t-2}, ...)$. Then

$$\boldsymbol{w}_{T+h} = \boldsymbol{\Phi}^h \boldsymbol{w}_T + \mathbf{u}_{h,T+h}, \tag{2}$$

where, except for $h = 1$, the overlapping observations cause the error in (2) to have the moving average structure of order $h - 1$:

$$\mathbf{u}_{h,t+h} = \mathbf{u}_{t+h} + \boldsymbol{\Phi}\mathbf{u}_{t+h-1} + \boldsymbol{\Phi}^2\mathbf{u}_{t+h-2} + ... + \boldsymbol{\Phi}^{h-1}\mathbf{u}_{t+1}.$$

Under the VAR specification $E\left(\mathbf{u}_{h,T+h} \left|\mathcal{I}_T\right.\right) = 0$, for $h = 1, 2, ...$ and the optimal (in the mean squared error sense) h-step ahead forecast of $\boldsymbol{w}_{T+h}$ is $E\left(\boldsymbol{w}_{T+h} \left|\mathcal{I}_T\right.\right) = \boldsymbol{\Phi}^h\boldsymbol{w}_T$. But given that in most forecasting applications the dimension of $\boldsymbol{w}_t$ is large, it is not feasible to estimate $\boldsymbol{\Phi}$ directly without imposing strong sparsity restrictions. Instead we take the target variable, y_{T+h}, to be the first element of $\boldsymbol{w}_{T+h}$ and consider the direct regression

$$y_{t+h} = \boldsymbol{\phi}_h'\boldsymbol{w}_t + u_{h,t+h}\ ,$$

where $\boldsymbol{\phi}_h'$ is the first row of $\boldsymbol{\Phi}^h$, and $u_{h,t+h}$ is the first element of $\mathbf{u}_{h,t+h}$. We still face a high-dimensional problem since there are a large number of potential covariates in $\boldsymbol{w}_t$. We consider the implementation of the direct approach under two scenarios concerning the potential predictors. First, when it is known that the target variable y_{t+h} is a sparse linear function of a large set of observed variables $\boldsymbol{x}_t$ (a subset of $\boldsymbol{w}_t$) known as the 'active set'. The model is sparse in the sense that y_{t+h} depends on a

[3] Marcellino, Stock, and Watson (2006) discuss the pros and cons of iterated and direct approaches to forecasting when K is small, and the target variable and the predictors can be jointly modelled as low-dimensional VARs or VARMAs. It is shown that if the underlying VAR model is correctly specified then iterated forecasts, being coherent, are preferred to direct forecasts. However, under misspecification direct forecasts could perform better. Pesaran, Pick, and Timmermann (2011) reconsider the comparison of iterated and direct forecasts to factor augmented VARs.

small number of covariates, that are *known* to be a subset of the much larger active set. The machine learning literature focuses on this case, which we refer to as the case of "known knowns". Second, when y_{t+h} could also depend on a few latent (unobserved) factors, $\mathbf{f}_t$, not directly included in the active set, which we call the case of "known unknowns".

Specifically, we suppose that for each h, y_{t+h} can be approximated by the following linear model, where the predictors are also elements of $\boldsymbol{w}_t$ in the high-dimensional VAR (1),

$$y_{t+h} = c_h + \mathbf{a}_h' \mathbf{z}_t + \sum_{j=1}^{K} \beta_{jh} I(j \in DGP) x_{jt} + \psi_h' \mathbf{f}_t + u_{h,t+h} \, , \tag{3}$$

for $t = 1, 2, ..., T-h$, where c_h is the intercept, $\mathbf{z}_t$ is a vector of small number, p, of pre-selected covariates included across all horizons h. Obvious examples, include lagged values of the target variable $(y_t, y_{t-1}, ...)$. Other variables can also be included in $\mathbf{z}_t$ on the basis of *a priori* theory or strong beliefs. The third component of y_{t+h} specifies the subset of variables in the active set $\boldsymbol{x}_{Kt} = (x_{1t}, x_{2t}, ..., x_{Kt})'$. $I(j \in DGP)$ is an indicator variable which takes the value of unity if x_{jt} is included in the data generating process (DGP) for y_{t+h} and zero otherwise. It is only if $I(j \in DGP) = 1$ that β_{jh} will be identified. We discuss ways to determine the selection indicator $I(j \in DGP)$ below. The number of variables included in the DGP is given by $k = \sum_{j=1}^{K} I(j \in DGP)$, which is supposed to be small and fixed as T (and possibly K) become large. This assumption imposes sparsity on the relationship between the target and the variables in the active set. In addition, we allow for a small number of latent factors, $\mathbf{f}_t$, that represent other variables influencing y_{t+h} that are not observed directly, but known to be present - the known unknowns.

Giannone, Lenza, and Primiceri (2021) contrast **sparse methods**, that select a few variables from the active set as predictors, such as Lasso and OCMT discussed below, and **dense methods**, that select all the variables in the active set but attach small weights to many of them, such as principal components (PC), ridge regression and other shrinkage techniques. Rather than having to choose between sparse and dense predictors, we consider approaches that combine the two. We apply sparse selection methods to the variables in the active set, and use dense shrinkage methods to approximate $\mathbf{f}_t$ from a wider set of variables with $\boldsymbol{x}_t$ included as a subset. We first consider the selection problem, known knowns, where we know the active set of potential covariates, and we then consider known unknowns where there are unobserved factors. The elastic net regression of Zou and Hastie (2005) discussed below also combines sparse and dense techniques.

Throughout, we shall assume that the errors, $u_{h,t+h}$, in (3) satisfy the orthogonality condition $E\left(u_{h,t+h}\ |\mathbf{z}_t, \boldsymbol{x}_t, \mathbf{f}_t\right) = 0$, for $h \geq 1$. In the context of the high-dimensional VAR model discussed above, this orthogonality condition holds so long as the underlying errors, $\mathbf{u}_t$, are serially uncorrelated. This is so despite the fact that due to the use of overlapping observations $u_{h,t+h}$ will be serially correlated when $h > 1$. This is an important consideration when high-dimensional techniques are applied to select predictors for multi-step ahead forecasting; an issue to which we will return.

3 Known knowns

In the case of known knowns, forecasts are obtained assuming that y_{t+h} is a linear function of $\boldsymbol{x}_t$

$$y_{t+h} = c_h + \mathbf{a}_h' \boldsymbol{z}_t + \sum_{j=1}^{K} \beta_{hj} x_{jt} + u_{h,t+h}, \text{ for } t = 1, 2, ..., T, \tag{4}$$

subject to some penalty condition on $\{\beta_{hj}\}$. Some of the covariates, x_{jt}, could be transformations of other covariates, such as interaction terms. It is assumed that the model is correctly specified, in the sense that, apart from $\boldsymbol{z}_t$, the variables that drive y_{t+h} are all included in the active set, $\mathbf{x}_{Kt}$.

Penalized regressions estimate $\boldsymbol{\beta}$ by solving the following optimization problem:

$$\min_{\mathbf{a}_h, \beta_h} \left\{ \sum_{t=1}^{T} \left(y_{t+h} - \mathbf{a}_h' \boldsymbol{z}_t - \boldsymbol{\beta}_h' \mathbf{x}_{Kt}\right)^2 + \lambda_{hT} \sum_{i=1}^{K} \left[(1-\alpha)|\beta_{hi}| + \alpha \beta_{hi}^2\right] \right\},$$

where $\boldsymbol{\beta}_h = (\beta_{h1}, \beta_{h2}, ..., \beta_{hK})'$, for given values of the "tuning" parameters λ and α. When $\alpha = 1$, we have ridge regression. When $\alpha = 0$, $\lambda_{hT} \neq 0$, we have the **Lasso** regression, which is better suited for variable selection. When $\lambda_{hT} \neq 0$ and $\alpha \neq 0$ we have the Zou and Hastie (2005) **elastic net** regression, which also mixes sparse and dense approaches.

Many standard forecasting techniques result from the particular choice of the penalty function. Shrinkage estimators such as ridge or some Bayesian forecasts can be derived using the ℓ_2 norm $\sum_{i=1}^{K} \beta_{hi}^2 < C_h < \infty$. Lasso (least absolute shrinkage and selection operator) follows when the ℓ_1 norm is used $\sum_{i=1}^{K} |\beta_{hi}| < C_h < \infty$. The difference is shown in Figure 1 below in Tibshirani (1996) where the ℓ_1 norm yields corner solutions with many of the coefficients, β_{hj}, estimated to be zero. In contrast, the use of ℓ_2 norm yields non-zero estimates for all the coefficients with many very close to zero.

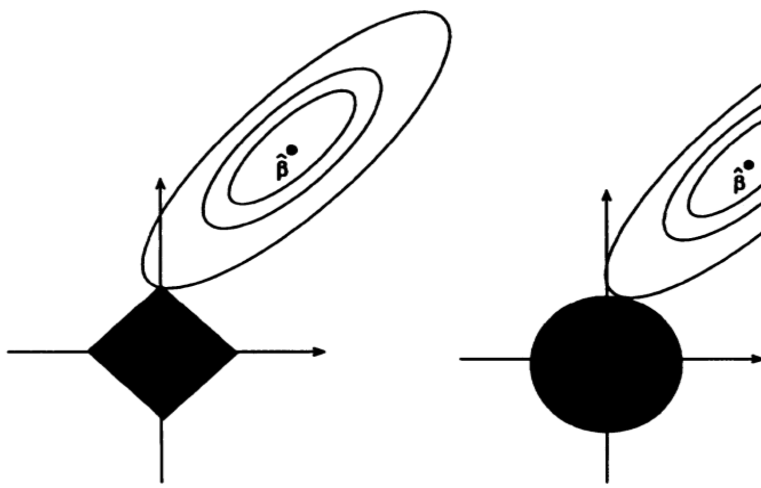

Figure 1: Estimation for the Lasso (left) and ridge (right) regression

There are also a large number of variants of Lasso, including adaptive Lasso, group Lasso, double Lasso, fused Lasso and prior Lasso. We will focus on Lasso itself, which we use in our empirical application and which we will compare to OCMT as an alternative procedure which is based on inferential rather than penalized procedures.

3.1 Lasso

In this paper we focus on Lasso, but acknowledge that are many variations on Lasso such as: adaptive Lasso, group Lasso, fused Lasso, and prior Lasso. Lasso estimates $\boldsymbol{\beta}_h$ by solving the following optimization problem:

$$\min_{\boldsymbol{\beta}_h} \left\{ \sum_{t=1}^{T} \left(y_{t+h} - c_h - \boldsymbol{\beta}_h' \boldsymbol{x}_t \right)^2 + \lambda_{hT} \sum_{i=1}^{K} |\beta_{hi}| \right\} , \tag{5}$$

where $\boldsymbol{\beta}_h = (\beta_{h1}, \beta_{h2}, ..., \beta_{hK})'$, $\boldsymbol{x}_{Kt} = (x_{1t}, x_{2t}, ..., x_{Kt})'$ for a given choice of the "tuning" parameter, λ_{hT}. The variable selection consistency of Lasso has been investigated by Zhao and Yu (2006), Meinshausen and Bühlmann (2006) and more recently by Lahiri (2021). The key condition is the so-called "Irrepresentable Condition, IRC" that places restrictions on the magnitudes of the correlations between the signals ($\mathbf{X}_{1h}$, standardized) and the rest of the covariates ($\mathbf{X}_{2h}$, standardized), taken as given (deterministic). The IRC is:

$$\text{IRC: } \left\| \left(T^{-1}\mathbf{X}'_{2h}\mathbf{X}_{1h}\right)\left(T^{-1}\mathbf{X}'_{1h}\mathbf{X}_{1h}\right)^{-1} sign(\boldsymbol{\beta}_h^0) \right\|_\infty < 1, \tag{6}$$

where $\boldsymbol{\beta}_h^0 = (\beta_{1h}^0, \beta_{2h}^0, ..., \beta_{k_h h}^0)'$ denotes the vector of true signal coefficients.[4] The IRC condition is met for pure noise variables, but need not hold for proxy variables, noise variables that are correlated with the true signals.

To appreciate the significance of the IRC, suppose the DGP contains x_{1t} and x_{2t} and the rest of the covariates in the active set are $x_{3t}, x_{4t}, ..., x_{Kt}$. Denote the sample correlation coefficient between x_{1t} and x_{2t} by $\hat{\rho}$ ($\hat{\rho}^2 < 1$) and the sample correlation coefficient of x_{1t} and x_{2t} with the rest of the covariates in the active set by $\hat{\rho}_{1s}, \hat{\rho}_{2s}$, for $s = 3, 4, ..., K$. Then, dropping the subscript h, the IRC for the s^{th} covariate is given by

$$\left| \left[sign\left(\beta_{01}\right), sign\left(\beta_{02}\right)\right]' \begin{pmatrix} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_{1s} \\ \hat{\rho}_{2s} \end{pmatrix} \right| < 1$$

which yields

$$\left| sign\left(\beta_{01}\right)\left(\hat{\rho}_{1s} - \hat{\rho}\hat{\rho}_{2s}\right) + sign\left(\beta_{02}\right)\left(\hat{\rho}_{2s} - \hat{\rho}\hat{\rho}_{1s}\right) \right| < 1 - \hat{\rho}^2$$

for $s = 3, 4, ..., K$. In this example there are two cases to consider: A: $sign\left(\beta_{02}\right) = sign\left(\beta_{01}\right)$; and B: $sign\left(\beta_{02}\right) = -sign\left(\beta_{01}\right)$. For case A $\sup_s \left|\hat{\rho}_{1s} + \hat{\rho}_{2s}\right| < 1 + \hat{\rho}$, and for case B $\sup_s \left|\hat{\rho}_{1s} - \hat{\rho}_{2s}\right| < 1 - \hat{\rho}$. Since the signs of the coefficients are unknown, for all possible values of $\hat{\rho}$, $\hat{\rho}_{1s}$ and $\hat{\rho}_{2s}$, we can ensure the IRC condition is met if $\left|\hat{\rho}\right| + \sup_s \left|\hat{\rho}_{1s}\right| + \sup_s \left|\hat{\rho}_{2s}\right| < 1$. This example shows the importance of the correlations between the true covariates in the DGP as well as between the true covariates and the other members of the active set that do not belong to the DGP. The IRC is quite a stringent condition and it is not just when one has proxies in the active set that are highly correlated with the true covariates that Lasso will tend to choose too many variables. In practice one cannot check the IRC condition since one does not know

[4] The number of signals, k_h, could vary with h.

which variables are the true signals.

In addition to the IRC it is also required that

$$\text{MinC: } \min_{j=1,2,...,k} \left|\beta_{jh}^0\right| > (2T)^{-1}\lambda_{hT} \left|\left(T^{-1}\mathbf{X}_{1h}'\mathbf{X}_{1h}\right)^{-1} sign(\beta_h^0)\right|_j,$$

$$\text{Penalty Condition : } T^{-1}\lambda_{hT} = o(1).$$

The penalty condition, which follows from MinC, says that the penalty has to rise with T, but not too fast and not too slowly. The expansion rate of λ_{hT} depends on the magnitude and the sign of β_{jh}^0, and the correlations of signals with the proxy variables. Lahiri (2021) shows that the penalty condition can be relaxed to $\lim_{T\to\infty} T^{-1}\lambda_{hT} < \liminf_{T\to\infty} d_{hT}$, where

$$d_{hT} = 2\min_j \left|\beta_{jh}^0\right| / \left|\left(T^{-1}\mathbf{X}_{1h}'\mathbf{X}_{1h}\right)^{-1} sign(\beta_h^0)\right|_j.$$

The above conditions do not restrict the choice of λ_{hT} very much, hence the recourse to cross-validation to determine it. In practice, λ_{hT} is calibrated using M-fold cross-validation techniques. The observations, $t = 1, 2, ..., T$, are partitioned into M disjoint subsets (folds), of size approximately $m = T/M$. Then $M - 1$ subsets are used for training and one for evaluation. This is repeated with each fold being used in turn for evaluation. M is typically set to 5 or 10. Cross-validation methods are often justified in machine learning literature under strong assumptions, such as independence and parameter stability across the sub-samples used in cross-validation. These assumptions are rarely met in the case of economic time series data, an issue that is discussed further in the context of the empirical example in Section 5.

3.2 OCMT

The need for cross-validation is avoided in the procedure proposed by Chudik, Kapetanios, and Pesaran (2018), (CKP). This is the "**O**ne **C**ovariate at a time **M**ultiple **T**esting" (**OCMT**) procedure, where covariates are selected **one at a time**, using $t-$statistic for testing the significance of the variables in the active set, **individually.**[5] Ideas from the multiple testing literature are used to control the false discovery rate, and ensure the selected covariates encompass the true covariates (signals) with probability tending to unity, under certain regularity conditions. Like Lasso, OCMT has

[5]Since t-ratios are invariant to scale, no pre-standardization of the covariates in the active set is required. This is in contrast to Lasso which is typically implemented after in-sample standardization of the covariates.

no difficulty in dealing with (pure) noise variables, and is very effective at eliminating them. Also, like Lasso, it requires some *min* condition such as $\left|\beta_{jh}^{0}\right| >> \sqrt{\frac{k\log(K)}{T}}$, for $j = 1, 2, ..., k$.[6] But because it considers a single variable at a time, OCMT does not require the IRC condition to hold and is not affected by the correlation between the members of the DGP as Lasso is. Instead it requires the number of proxies signals, say k_T^*, to rise no faster than $\sqrt{T}$. Chudik, Pesaran, and Sharifvaghefi (2023), discussed below, is primarily concerned with parameter instability, but Section 4 of that paper has a detailed comparison of the assumptions required for Lasso and OCMT under parameter stability.

OCMT's condition on k_T^* has been recently relaxed by Sharifvaghefi (2023) who allows $k_T^* \to \infty$, with T. He considers the following DGP

$$y_{t+h} = c_h + \mathbf{a}_h' \boldsymbol{z}_t + \sum_{j=1}^{K} \beta_{jh} I(j \in DGP) x_{jt} + u_{h,t+h} \ , \tag{7}$$

where as before $\boldsymbol{z}_t$ is a known vector of pre-selected variables, and it is assumed that the k signals are contained in the known active set $\mathcal{S}_{K,t} = \{x_{jt}, j = 1, 2, ..., K\}$. Note that for now the DGP in (7) does not include the additional latent factors, $\mathbf{f}_t$, introduced in (3). Without loss of generality, consider the extreme case where there are no noise variables and **all** proxy or pseudo signal variables (x_{jt}, for $j = k+1, k+2, ..., K$) are correlated with the signals, $\mathbf{x}_{1t} = (x_{1t}, x_{2t}, ..., x_{kt})'$. In this case k_T^* rises with K and OCMT is no longer applicable. However, in this case because of the correlation with the proxies, the signals, $\mathbf{x}_{1t}$, become latent factors for the proxy variables and we have

$$x_{jt} = \phi_{j0} + \sum_{i=1}^{k} \phi_{ji} x_{it} + \varepsilon_{jt} = \phi_{j0} + \phi_j' \mathbf{x}_{1t} + \varepsilon_{jt},$$

for $j = k+1, k+2, ..., K$. Although the identity of these common factors are unknown, because we do not know the true signals, they can be approximated by the principal components of the variables in the active set.

Specifically, following Sharifvaghefi (2023), denote the latent factors that result in non-zero correlations between the noise variables in the active set and the signals by $\varkappa_t$ and consider the factor model

$$x_{jt} = \boldsymbol{\kappa}_j' \varkappa_t + v_{jt}, \text{ for } j = 1, 2, ..., K \ , \tag{8}$$

[6] To simplify the exposition we have dropped explicit reference to the forecast horizon, h. But in practice, and as we shall see from the empirical applications below, the number and identity of selected signals could differ with h.

where $\boldsymbol{\kappa}_j$, for $j = 1, 2, ..., K$ are the factor loadings and the errors, v_{jt}, are weakly cross-correlated and distributed independently of the factors and their loadings. Under (8), the DGP, (7), can be written equivalently as

$$y_{t+h} = c_h + \mathbf{a}_h' \boldsymbol{z}_t + \mathbf{b}_h' \varkappa_t + \sum_{j=1}^{K} \beta_{jh} I(j \in DGP) v_{jt} + u_{h,t+h} \ , \tag{9}$$

where $\mathbf{b}_h = \sum_{j=1}^{K} I(j \in DGP)\beta_{jh}\boldsymbol{\kappa}_j$. When $\varkappa_t$ and v_{jt} are known, the problem reduces to selecting v_{jt} from $\mathcal{S}_{K,t}^{v} = \{v_{jt}, j = 1, 2, ..., K\}$, conditional on $\mathbf{z}_t$ and $\varkappa_t$. Sharifvaghefi shows that the OCMT selection can be carried out using the principal component estimators of $\varkappa_t$ and $\mathbf{v}_{jt}$ - denoted by $\hat{\varkappa}_t$ and $\hat{\mathbf{v}}_{jt}$, if both K and T are large. He labels this procedure as generalized OCMT (GOCMT). Note that at the moment it is assumed that $\varkappa_t$ does not **directly** affect y_{t+h}, it only enters through the x_{jt}. It represents the signals, the common factors correlated with the proxies, and provides a way of filtering the correlations in the first step.

3.3 GOCMT

The GOCMT procedure simply augments the OCMT regressions with the PCs, $\hat{\varkappa}_t$, and considers the statistical significance of $\hat{\mathbf{v}}_{jt}$ for each j, *one at the time*. Lasso-factor models have also been considered by Fan, Ke, and Wang (2020) and Hansen and Liao (2019). In practice, since $\mathbf{x}_j = \hat{\boldsymbol{\Xi}}\hat{\psi}_j + \hat{\mathbf{v}}_j$, where $\hat{\boldsymbol{\Xi}} = (\hat{\varkappa}_1, \hat{\varkappa}_2, ..., \hat{\varkappa}_T)'$, then $\mathbf{M}_{\hat{\boldsymbol{\Xi}}}\mathbf{x}_j = \mathbf{M}_{\hat{\boldsymbol{\Xi}}}\hat{\mathbf{v}}_j$, where $\mathbf{M}_{\hat{\boldsymbol{\Xi}}} = \mathbf{I}_T - \hat{\boldsymbol{\Xi}}\left(\hat{\boldsymbol{\Xi}}'\hat{\boldsymbol{\Xi}}\right)^{-1}\hat{\boldsymbol{\Xi}}'$, and GOCMT reduces to OCMT when $\mathbf{z}_t$ is augmented with $\hat{\varkappa}_t$, where the statistical significance of x_{jt} as a predictor of y_{t+h} is evaluated for each j, one at a time. Like OCMT, GOCMT allows for the multiple testing nature of the procedure (K separate tests - with K large) by increasing the level of significance with K. The number of PCs, $dim(\hat{\varkappa}_t)$, can be determined using one of the criteria suggested in the factor literature.

In the first stage, K **separate** OLS regressions are computed, where the variables in the active set are entered one at a time:

$$y_{t+h} = c_h + \mathbf{a}_h' \boldsymbol{z}_t + \mathbf{b}_h' \hat{\varkappa}_t + \phi_{jh} x_{jt} + e_{j,h,t+h},\ t = 1, 2, ..., T, \text{ for } j = 1, 2, ..., K, \tag{10}$$

Denote the t-ratio of ϕ_{jh} by $t_{\hat{\phi}_{j,(1)}}$. Then variable j is selected if

$$\widehat{\mathcal{J}}_{j,(1)} = I\left[\left|t_{\hat{\phi}_{j,(1)}}\right| > c_p(K, \delta)\right], \text{ for } j = 1, 2, ..., K \ , \tag{11}$$

where $c_p(K, \delta)$ is a critical value function given by

$$c_p(K, \delta) = \Phi^{-1}\left(1 - \frac{p}{2K^\delta}\right), \tag{12}$$

p is the nominal size (usually set to 5%), $\Phi^{-1}(\cdot)$ is the inverse of a standard normal distribution function and δ is a fixed constant set in the interval $[1, 1.5]$. In the second step a multivariate regression of y_{t+h} on $\boldsymbol{z}_t$ and all the selected regressors is considered for inference and forecasting. Serial correlation will arise with OCMT when selection is based on one variable at the time, and the omitted variables are mixing (serially correlated). CKP discuss this in section C of the online theory supplement to their paper and suggest using a more conservative (higher) critical value – namely using $\delta = 1.5$ rather than $\delta = 1.5$.

When the covariates are **not** highly correlated, OCMT applies irrespective of whether K is small or large relative to T, so long as $T = \ominus(K^c)$, for some finite $c > 0$. But to allow for highly correlated covariates, GOCMT requires K to be sufficiently large to enable the identification of the latent factor, $\varkappa_t$. In cases where K is not that large, it might be a good idea to augment the active set for the target variable, y_{t+h}, $\mathcal{S}_{K,t} = \{x_{jt},\ j = 1, 2, ..., K\}$, with covariates for other variables determined simultaneously with y_{t+h}, ending up with $\overline{\mathcal{K}} > K$ covariates for identification of $\varkappa_t$. GOCMT does not impose any restriction on the correlations between the variables other than that they cannot be perfectly collinear.

3.4 High-dimensional variable selection in presence of parameter instability

OCMT has also been recently generalized by Chudik, Pesaran, and Sharifvaghefi (2023) to deal with parameter instability. Under parameter instability OCMT correctly selects the covariates with non-zero average (over time) effects, using the full sample. However the adverse effects of changing parameters on the forecast may mean that while the full sample is the best to use for selection, it need not be the best to use for estimating the forecasting model. Instead, it may be better to use shorter windows or weight the observations in the light of the evidence on break points and break sizes.

Determining the appropriate window or weighting for the observations before estimation is a difficult problem and no fully satisfactory procedure seems to be available.

It is common in finance to use rolling windows of 60 or 120 months, but one problem with shorter windows is that if you have periods of instability interspersed with periods of stability, like the Great Moderation, estimates using a short window from the stable period may understate the degree of uncertainty. This happened during the financial crisis when the short windows used for estimation did not reflect past turbulence. Similarly, the Bank of England estimating their models using the low inflation regime of the past 30 years discounted the evidence from the high inflation regime of the 1970s and 1980s.

While identifying the date of a break might not be difficult, identifying the size of the break may be problematic if the break point is quite recent. If there is a short time since the break, there is little data on which to estimate the post-break coefficient with any degree of precision. If there is a long time since the break, then using post break data is sensible. Pesaran, Pick, and Pranovich (2013) examine optimal forecasts in the presence of continuous and discrete structural breaks. These present quite different sorts of challenges. With continuous breaks the parameters change often by small amounts. With discrete breaks the parameters change rarely but by large amounts. They propose weighting observations to obtain optimal forecasts in the MSFE sense and derive optimal weights for one step ahead forecasts for the two types of break. Under continuous breaks, their approach largely recovers exponential smoothing weights. Under discrete breaks between two regimes the optimal weights follow a step function that allocates constant weights within regimes but different weights in different regimes. In practice, the time and size of the break is uncertain and they investigate robust optimal weights. Averaging forecasts with different weighting schemes, for instance with exponential smoothing parameters between 0.96 and 0.99, may also be a way to produce more robust forecasts.

4 Known unknowns

So far we have considered techniques (penalized regressions and OCMT) that assume y_{t+h} depends on $\mathbf{z}_t$ and a subset of a set of covariates - the active set - which is assumed **known**. In contrast, shrinkage type techniques such as PCs, (implicitly) assume that

y_{t+h} depends on $\mathbf{z}_t$ and the $m \times 1$ vector of **unknown** factors $\mathbf{f}_t$

$$y_{t+h} = c_h + \mathbf{a}_h' \boldsymbol{z}_t + \theta_h' \mathbf{f}_t + u_{h,t+h}.$$

This is a simple example of techniques that in our terminology can be viewed as belonging to a class of forecasting models based on **known unknowns**. The uncertainty about $\mathbf{f}_t$ is resolved assuming it can be identified from a **known** active set, such as $\mathcal{S}_{K,t} = \{x_{jt}, j = 1, 2, ..., K\}$. Individual covariates in $\mathcal{S}_{K,t}$ are not considered for selection (although a few could be pre-selected and included in $\boldsymbol{z}_t$). To forecast y_{t+h} one still requires to forecast the PCs and to allow for the uncertainty regarding $m = \dim(\mathbf{f}_t)$.

Factor augmented VARs (FAVAR), initially proposed by (Bernanke, Boivin, and Eliasz, 2005, BBE), augment the standard VAR models with a set of unobserved common factors. In the context of our set up, FAVAR can be viewed as a generalized version of (3), where $\boldsymbol{y}_{t+h}$ is a vector and $\boldsymbol{z}_t = \left\{\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, ..., \boldsymbol{y}_{t-p}\right\}$. BBE argue that small VARs gave implausible impulse response functions, such as the "price puzzle", which were interpreted as reflecting omitted variables. One response was to add variables and use larger VARs, but this route rapidly runs out of degrees of freedom, since Central Bankers monitor hundreds of variables. The FAVAR was presented as a solution to this problem. Big Bayesian VARs are an alternative solution.

The assumptions that underlie both penalized regression and PC shrinkage are rather strong. The former assumes that $\mathbf{f}_t$ can affect y_{t+h} only indirectly through x_{jt}, $j = 1, 2, ..., K$, and the latter does not allow for individual variable selection. Suppose that $\mathbf{f}_t$ also enters (7) then the model can be written as (3) above, repeated here for convenience:

$$y_{t+h} = c_h + \mathbf{a}_h' \boldsymbol{z}_t + \sum_{j=1}^{K} \beta_{jh} I(j \in DGP) x_{jt} + \psi_h' \mathbf{f}_t + u_{h,t+h}.$$

The forecasting problem now involves both selection and shrinkage. The $\mathbf{f}_t$ can be identified by, for instance, the PCs of the augmented active set x_{jt}, $j = 1, 2..., K, K + 1, ..., \overline{\mathcal{K}}$ which can be wider than the active set of covariates used to predict y_t. There are various other ways that the unobserved $\mathbf{f}_t$ could be estimated, but we use PCs as an example, since they are widely used.

The latent factors are unlikely to be only specific to the target variable under consideration. Observed global factors, such as oil and raw material prices or inflation and output growth of major countries such as US can be included in the active set.

The main issue is how to deal with global factors, such as technology, political change and so on that are unobserved and tend to affect many countries in the world economy. Call this vector of global factors $\mathbf{g}_t$. A natural extension is to introduce forecast equations for other countries (entities) who have close trading relationships with UK and use penalized panel regressions, where the panel dimension allows identification of the known unknowns.

More specifically, suppose there are N other units (countries) that are affected by observed country specific covariates, $\mathbf{z}_{it}$, $i = 1, 2, ..., N$, and $x_{i,jt}$ for $j = 1, 2, ..., K_i$, plus domestic latent factors $\mathbf{f}_{it}$, and global latent factors, $\mathbf{g}_t$. The forecasting equations are now generalized as

$$\begin{aligned} y_{i,t+h} = c_h + \mathbf{a}'_{ih}\boldsymbol{z}_{it} + \sum_{j=1}^{K_i} \beta_{ijh} I(j \in DGP_i) x_{i,jt} + \\ \theta'_{ih}\mathbf{f}_{it} + \psi'_{ih}\mathbf{g}_t + u_{i,h,t+h}, \end{aligned}$$

where $k_i = \sum_{j=1}^{K_i} I(j \in DGP_i)$ is finite as $K_i \to \infty$, for $i = 0, 1, 2, ..., N$. For the country-specific covariates we postulate that there is an augmented active set

$$x_{i,jt} = \gamma'_{ij}\mathbf{f}_{it} + v_{i,jt},\ j = 1, 2, ..., K_i,\ K_{i+1}, ..., \overline{\mathcal{K}}_i,$$

where $\mathbf{f}_{it}$ are latent factors. The global factors are then identified as the common components of the country-specific factors, namely

$$\mathbf{f}_{it} = \boldsymbol{\Psi}_i\mathbf{g}_t + \boldsymbol{\xi}_{it}\ ,$$

for $i = 1, 2, .., N$, with N large.

Variable selection for the target variable (say UK inflation) can now proceed by applying GOCMT, with the UK model augmented with UK-specific PCs, $\hat{\mathbf{f}}_{it}$ as well as the PC estimator of the global factor, $\mathbf{g}_t$, that drives the country specific factors. This can be extracted from $\hat{\mathbf{f}}_{it}$ as PCs of the country-specific PCs. In addition to common factor dependence, countries are also linked through trade and other more local features (culture, language). Such "network" effects can be captured by using "starred" variables, to use the GVAR terminology. A simple example would be (for $i = 0, 1, ..., N$)

$$\begin{aligned} y_{i,t+h} = c_{ih} + \delta_{ih}y^*_{it} + \mathbf{a}'_{ih}\boldsymbol{z}_{it} + \sum_{j=1}^{K_i} \beta_{ijh} I(j \in DGP_i) x_{i,jt} + \\ \theta'_{ih}\mathbf{f}_{it} + \psi'_{ih}\mathbf{g}_t + u_{i,h,t+h}\ , \end{aligned} \tag{13}$$

where $i = 0$ represents UK, and $y^*_{it} = \sum_{j=1}^{N} w_{ij}y_{jt}$, w_{ij} (trade weights) measures the relative importance of country j in determination of country i^{th} target variable. Similarly, $\boldsymbol{z}_{it}^{*} = \sum_{j=1}^{N} w_{ij}^{*} \boldsymbol{z}_{jt}$ can also be added to the model if deemed necessary.

The network effects can be included either as an element of $\boldsymbol{z}_{it}$ or could be made subject to variable selection. The problem becomes much more complicated if we try to relate $y_{i,t+h}$ simultaneously to $y_{i,t+h}^{*}$. Further, for forecasting, following Chudik, Grossman, and Pesaran (2016), one might also need to augment the UK regressions with time series, forecasting models for the common factors.

Equation (13) allows for a number of different approaches to dimension reduction. As has been pointed out by Wainwright (2019): "Much of high-dimensional statistics involves constructing models of high-dimensional phenomena that involve some implicit form of low-dimensional structure, and then studying the statistical and computational gains afforded by exploiting this structure". Shrinkage methods, like PCs, assume a low dimensional factor structure. The two selection procedures that we have considered, Lasso and OCMT, exploit different aspects of the low-dimensional sparsity structure assumed for the underlying data generating process. Lasso restricts the magnitude of the correlations within and between the signals and the noise variables. OCMT limits the rate at which the number of proxy variables rises with the sample size. GOCMT relaxes this restriction by filtering out the effects of latent factors that bind the proxies to the true signals before implementing the OCMT procedure.

5 Forecasting UK inflation

5.1 Introduction

We apply the procedures proposed above to the problem of forecasting quarterly UK inflation at horizons $h = 1, 2$ and 4. The target variable is the headline rate, average annual UK inflation, which is also forecast by the Bank of England. It is labelled DPUK4, defined as $\pi_{t+h} = 100 \times \log(p_{t+h}/p_{t+h-4})$, where p_t is the UK consumer price index taken from the IMF International Financial Statistics. Forecasting annual rates of inflation at quarterly frequencies is subject to the overlapping observations problem when $h > 1$, and it is important that the pre-selected variables in $\boldsymbol{z}_t$, or the variables include in the active set $\mathcal{S}_{K,t} = \{x_{jt}, j = 1, 2, ..., K\}$ are all pre-determined (known) at time t. Furthermore, as discussed earlier, the variables selected for forecasting inflation at different horizons need not be the same, and the selected variables are also likely to change over time.

Since we have emphasised the importance of international network effects, we need to use a quarterly dataset that includes a large number of countries to estimate global factors and to allow the construction of the y_{it}^* variables that appear in (13). The Global VAR (GVAR) data set provides such a source. The publicly available data set compiled by Mohaddes and Raissi (2024) covers 1979q1-2023q3. We are very grateful to them for extending the data. While the latest GVAR dataset goes up to 2023q3 we only had access to data till 2023q1 when we started the forecasting exercise, the results of which are reported in this paper, but the data that we used matches that GVAR 2023 vintage which was released in January 2024.[7]

The data base includes quarterly macroeconomic data for 6 variables (log real GDP, y, the rate of inflation, dp, short-term interest rate, r, long-term interest rate, lr, the log deflated exchange rate, ep, and log real equity prices, eq), for 33 economies as well as data on commodity prices (oil prices, poil, agricultural raw material, pmat, and metals prices, pmetal). These 33 countries cover more than 90% of world GDP. The GVAR data was supplemented with other specific UK data on money, wages, employment and vacancies, in the construction of the active set discussed below.

In the light of the argument in Chudik, Pesaran, and Sharifvaghefi (2023), we use the full sample beginning in 1979q1 for variable selection. There are arguments for down-weighting earlier data for estimation when there have been structural changes, as discussed by Pesaran, Pick, and Pranovich (2013). However, the full sample was used both for variable selection and estimation of the forecasting model in order to allow evidence from the earlier higher inflation regime to inform both aspects.

Two sets of variables are considered for inclusion in $\boldsymbol{z}_t$. The first set, which we label AR2, includes lags of the target variable π_t, π_{t-1} (or equivalently π_t and $\Delta\pi_t$). Given the importance we attach to global variables and network effects, the second set, which we label ARX2, also includes π_t^*, π_{t-1}^* (or equivalently π_t^* and $\Delta\pi_t^*$) where π_t^* is a measure of UK specific foreign inflation constructed using UK trade weights with

[7] GVAR Data 1979q1-2023q3 (2023 Vintage) is available at https://www.mohaddes.org/gvar, further material on the GVAR is provided at https://sites.google.com/site/gvarmodelling/gvar-toolbox

the other countries.[8]

If there is a global factor in inflation, the inflation rates of different countries will be highly correlated and tend to move together. Figure 2 demonstrates that this is in fact the case. It plots the inflation rates for 19 countries over the period 1979-2022. It is clear that they do move together, reflecting a strong common factor. The dispersion is somewhat greater in the high inflation 1980s. At times individual countries break away from the herd with idiosyncratic bursts of inflation, like New Zealand in the mid 1980s. But it is striking that inflation in every country increases from 2020.

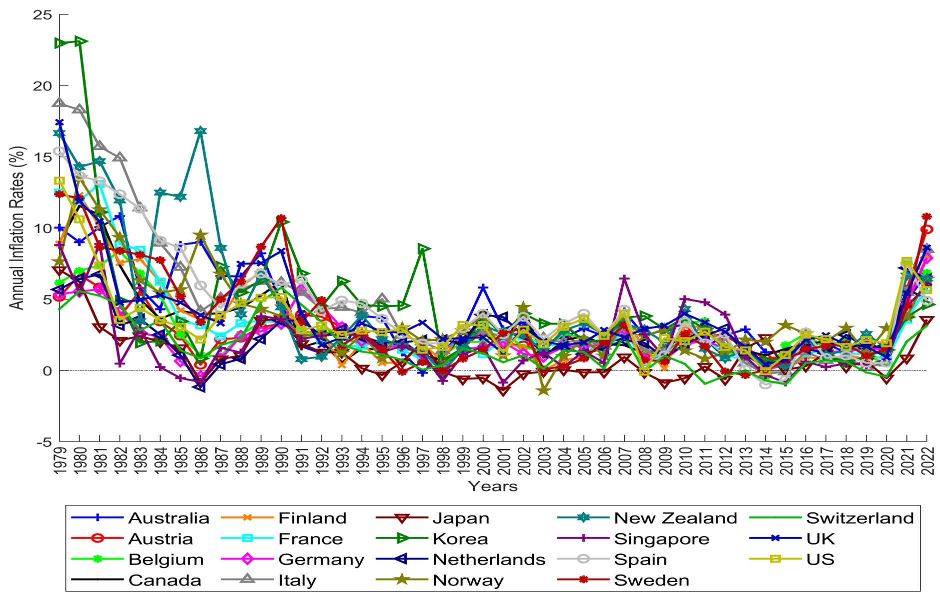

Figure 2: Inflation Across Advanced Economies

To demonstrate the importance of the global factor for the UK, Figure 3 plots π_t, and π_t^*, UK inflation and UK specific foreign inflation. The two series move together, and from the mid 1990s they are very close. This indicates that not only is one unlikely

[8]Specifically, $\pi_t^* = \sum_j w_j \pi_{jt}$, where π_{jt} is the inflation rate in country j and w_j is the trade weight of country j with UK.

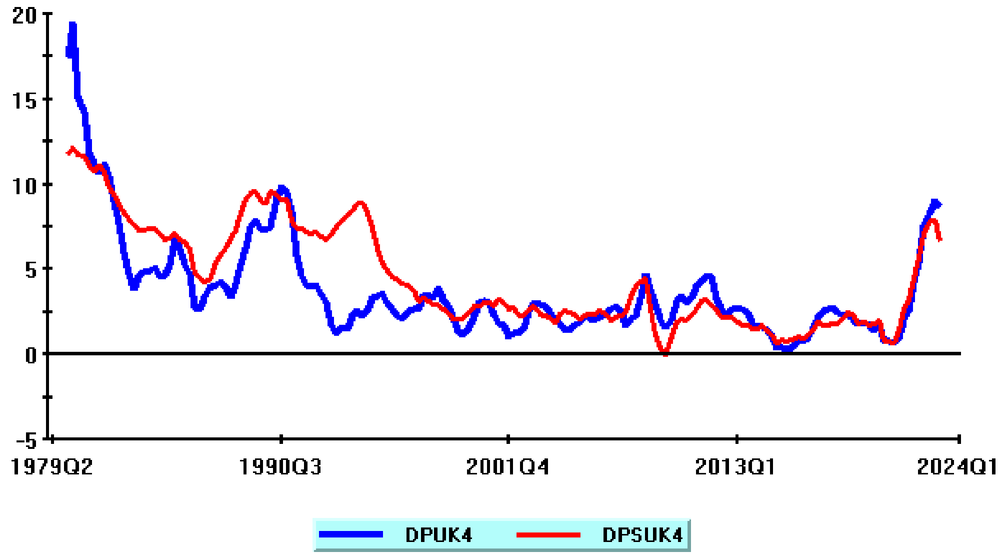

Figure 3: UK and UK-specific global inflation

to be able to explain UK inflation just by UK variables but that there are good reasons to include this UK specific measure of foreign inflation in one of our specifications for $\boldsymbol{z}_t$. The GVAR estimates also indicate the higher sensitivity of the UK to foreign variables than the US or euro area. This is not surprising, they are larger, less open, economies.

5.2 Active set

We now turn to the choice of the members of the active set, $\mathcal{S}_{K,t} = \{x_{jt}, j = 1, 2, ..., K\}$, some of which may be included in $\boldsymbol{z}_t$. While our focus is on forecasting not on building a coherent economic model, our choice of the covariates in the active set is motivated by the large Phillips curve literature, which suggests important roles for demand, supply and expectations variables. The demand and supply variables in the active set are both domestic and foreign from both product and labour markets. Expectations are captured by financial variables. Interaction terms were not included, but the non-linearities that have been investigated in the literature may be picked up by the latent foreign variables. These are represented by UK specific measures of foreign inflation and output that could be viewed as estimates of $\mathbf{g}_t$ that are tailored to the UK in relation to her trading partners.

Accordingly, we consider 26 covariates (x_{jt}) listed in Table 1, and their changes $\Delta x_{jt} = x_{jt} - x_{j,t-1}$, giving an active set with $K = 52$ variables to select from. Whereas in a regression including current and lagged values of a regressor (say x_t and x_{t-1}) is equivalent to including current and change (namely x_t and Δx_t), in selection the two specifications can result in different outcomes. Including Δx_t is better since, as compared to x_{t-1}, it is less correlated with the level of the other variables in the active set. The 26 included covariates are measured as four-quarter rates of change, changes or averages to match the definition of the target variable. The rates of change are per cent per annum.

UK goods market demand indicators: rate of change of output, two measures of the output gap: log output minus either a $P = 8$ or $P = 12$ quarter moving average of log output, $Gap(y_t, P) = y_t - P^{-1} \sum_{p=1}^{P} y_{t-p}$;

UK labour market demand indicators: rate of change of UK employment, vacancies, and average weekly earnings and the change in unemployment;

UK financial indicators: annual averages of UK short and long interest rates, the rate of change of money, UK M4, and of UK real equity prices;

Global cost pressures on the UK: rate of change of the price of oil, metals, materials, UK import prices and deflated dollar exchange rate;

Foreign demand and supply variables: UK specific global measures, foreign inflation, rate of change of foreign output and two measures of the foreign output gap: log foreign output minus either an 8 or 12 quarter moving average of log foreign output. In addition, large country variables were added: annual average of US short and long interest rates, rates of change of US output and prices, and of Chinese output.

Table 1: List of covariates for UK inflation forecasting

Variable	Description
DPUK4	4 quarter UK rate of inflation,
DYUK4	4 quarter rate of change of UK real GDP
GAPUK8	UK log real GDP relative to its 8 quarter moving average
GAPUK12	UK log real GDP relative to its 12 quarter moving average
DEMUK4	4 quarter rate of change of UK employment
DVUK4	4 quarter rate of change of UK vacancies
DUUK	4 quarter change in UK unemployment
DWUK4	4 quarter rate of change of average weekly earnings,
RUK4	4 quarter average UK short interest rate
LRUK4	4 quarter average UK long interest rate
DMUK4	4 quarter rate of change of UK M4 money
DEQUK4	4 quarter rate of change of UK real equity prices
DPOIL4	4 quarter rate of change of oil prices
DPMAT4	4 quarter rate of change of material prices
DPMETAL4	4 quarter rate of change of metal prices
DPMUK4	4 quarter rate of change of UK import prices
DEPUK4	4 quarter rate of change of UK deflated dollar exchange rate
DPSUK4	UK specific measure of 4 quarter foreign inflation
DYSUK4	4 quarter rate of change of UK specific foreign real GDP
RUS4	4 quarter average US short interest rate
LRUS4	4 quarter average US long interest rate
DYCHINA4	4 quarter rate of change of Chinese real GDP
GAPSUK8	UK specific log foreign real GDP relative to 8 quarter moving average
GAPSUK12	UK specific log foreign real GDP relative to 12 quarter moving average
DYUS4	4 quarter rate of change of US real GDP
DPUS4	4 quarter US rate of inflation

5.3 Variable selection

5.3.1 Variable selection procedures

We consider Lasso, Lasso conditional on $\boldsymbol{z}_t$, and GOCMT conditional on $\boldsymbol{z}_t$. With Lasso, the variables are standardized in-sample before implementing variable selection. The Lasso penalty parameter, λ_T, is estimated using 10 fold cross-validation (CV), across subsets of the observations. As noted above, the assumptions needed for standard CV procedures, for instance those used in the program cv.glmnet, are not appropriate for time series. Time series show features such as persistence and changing variance that are incompatible with those assumptions. In the standard procedure the CV subsets (folds) are typically chosen randomly. This is appropriate if the observations are independent draws from a common distribution, but this is not the case with time series. Since order matters in time series, we retain the time order of the data within each subset. See Bergmeir, Hyndman, and Koo (2018) who provide Monte Carlo evidence on various procedures suggested for the case of serially correlated data. We use all the data, and do not leave gaps between subsets. In addition, the standard procedure chooses the $\hat{\lambda}_{hT}$ that minimises the pooled MSE over the ten subsets. But when variances differ substantially over subsets pooling is not appropriate, instead we follow Chudik, Kapetanios, and Pesaran (2018, CKP), and use the average of the $\hat{\lambda}_{hT}$ chosen in each subset. Full details are provided in the online simulation appendix to CKP (2018).

As well as standard Lasso, for consistency with OCMT, we also generated Lasso forecasts conditional on $\boldsymbol{z}_t$ by including a pre-selected set of variables $\boldsymbol{z}_t$ in the optimization problem (5). This generalized Lasso procedure solves the following optimization problem:

$$\min_{\mathbf{a}_h,\beta_h}\left\{\sum_{t=1}^{T}\left(y_{t+h}-c_h-\mathbf{a}_h'\boldsymbol{z}_t-\boldsymbol{\beta}_h'\boldsymbol{x}_t\right)^2+\lambda_T\sum_{i=1}^{K}|\beta_{hi}|\right\}\ , \tag{14}$$

where the penalty is applied only to the variables in the active set, $\boldsymbol{x}_t$, and not to the pre-selected variables, $\boldsymbol{z}_t$. The above optimization problem can be solved in two stages. In the first stage the common effects of $\boldsymbol{z}_t$ are filtered out by regressing y_{t+h} and $\boldsymbol{x}_t$ on the pre-selected variables $\boldsymbol{z}_t$ and saving the residuals $e_{y.z}$ and $e_{xj.z}$, $j=1,2,..,K$. In the second stage Lasso is applied to these residuals. A proof that this two-step procedure solves the constrained minimization problem in (14) is provided by Sharifvaghefi and reproduced in the Appendix.

In the OCMT critical value function, $c_p(K,\delta)=\Phi^{-1}\left(1-\frac{p}{2K^\delta}\right)$, we set $p=0.05$ and $\delta=1$. With $K=52$ this means that we only retain variables with t-ratios (in absolute value) exceeding $c_{0.05}(52,1)=3.3$. To allow for possible serieal correlation, we also experimented with setting $\delta = 1.5$, which yields, $c_{0.05}(52, 1.5) = 3.82$. The results were reasonably robust and we focus on the baseline choice of $\delta = 1$, also recommended by CKP.

We implement Lasso and OCMT conditional on two pre-selected sets of variables, either an AR2 written as level and change $\mathbf{z}_t = (\pi_t,\ \Delta\pi_t)'$ or given the role of foreign inflation, shown above, the AR2 augmented by the level and change of the UK specific measure of foreign inflation, denoted ARX, $\mathbf{z}_t = (\pi_t,\ \Delta\pi_t, \pi_t^*, \Delta\pi_t^*)'$. As noted above, for selection including current and change is better than including current and lag. For comparative purposes we also generated forecasts with the pre-selected variables only, namely the AR2 forecasts generated from the regressions

$$AR2: \qquad \pi_{t+h} = c_h + a_1\pi_t + a_{2h}\Delta\pi_t + u_{h,t+h},$$

and the ARX forecasts generated from

$$ARX: \qquad \pi_{t+h} = c_h + a_{1h}\pi_t + a_{2h}\Delta\pi_t + a_{3h}\pi_t^* + a_{4h}\Delta\pi_t^* + u_{h,t+h}.$$

Variable selection is carried out recursively, for each forecast horizon h separately, using an expanding windows approach. All data samples start in $1979q2$ and end in the quarter that forecasts are made. To forecast the average inflation over the four quarters to $2020q1$ using a forecast horizon of $h = 4$, the sample used for selection and estimation ends in $2019q1$. The end of the sample is then moved to $2019q2$ to forecast the average inflation over the four quarters to $2020q2$, and so on. Similarly, to forecast the average inflation over the four quarters to $2020q1$ using $h = 2$, the sample ends in $2019q3$, and using $h = 1$ the sample ends in $2019q4$. These sequences continue one quarter at a time until the models are selected and estimated to forecast inflation over the four quarters to $2023q1$. Thus for $h = 4$, there are 17 samples used for variable selection, while for $h = 2$ and $h = 1$ there are 15 and 14 such variable selection samples. This process of recursive model selection and estimation means that the variables selected can change from quarter to quarter, and for each forecast horizon, h.

Section S-3 of the online supplement list the variables selected by each of the procedures, for each quarter and each forecast horizon. The main features are summarised here.

5.3.2 Number of variables selected

Table 2 gives the minimum, maximum, and average number of variables selected for the 3 forecast horizons and 5 variable selection procedures. Except for AR2-OCMT, at $h = 4$, OCMT chooses fewer variables than Lasso. Lasso conditional on the pre-selected variables selects a larger number of variables in total than the standard Lasso without conditioning. Conditioning on pre-selected variables is much more important for OCMT as compared to Lasso. This finding is in line with the theoretical results obtained by Sharifvaghefi (2023) who establishes the importance of conditioning on the latent factors when applied to an active set with highly correlated covariates. The number of variables Lasso selects falls with the forecast horizon,[9] whilst the number of variables selected by OCMT rises with the horizon. These results show that Lasso and OCMT could select very different models for forecasting.

Table 2: Number of variables selected by Lasso and OCMT including preselected

	Forecast horizon, h, in quarters										
	Total number of pre-selected and selected variables										
	$h = 1$				$h = 2$				$h = 4$		
	Min	Max	Mean	..	Min	Max	Mean	..	Min	Max	Mean
Lasso	7	12	8.1		5	9	6.1		3	6	5.2
AR2-Lasso	5	11	8.2		9	16	13.5		8	11	9.5
AR2-OCMT	2	3	2.2		4	5	4.5		5	14	6.2
ARX-Lasso	8	16	12.4		12	19	16.3		2	15	9.9
ARX-OCMT	4	4	4		5	5	5		5	8	5.8

Note: The reported results are based on 14, 15 and 17 variable selection samples for 1, 2 and 4 quarter ahead models, respectively. The AR2 and ARX components include 2 and 4 pre-selected variables, respectively.

As expected, the number of variables selected by Lasso correlates with the estimates of the penalty parameter $\hat{\lambda}_{hT}$, computed by cross-validation. These values are summarized in Table 3. For all three Lasso applications the mean of the estimated penalty parameter increases with the forecast horizon, though more slowly for the

[9]This is possibly because, as h increases the β_h^0 get smaller in (6), the IRC is more likely to be satisfied and Lasso is less likely to falsely select additional variables.

specifications that include pre-selected variables. For Lasso (without pre-selection) the number of variables selected falls with the forecast horizon because of the increasing penalty parameter. This is not as clear cut for the specifications including pre-selected variables.

Table 3: Estimates of the Lasso penalty parameter computed by 10-fold cross-validation procedure.

	Forecast horizon, h, in quarters										
	$h=1$				$h=2$				$h=4$		
	Min	Max	Mean	..	Min	Max	Mean	..	Min	Max	Mean
Lasso	0.07	0.12	0.11		0.22	0.31	0.26		0.33	0.47	0.41
AR2-Lasso	0.08	0.10	0.09		0.09	0.14	0.11		0.18	0.27	0.22
ARX-Lasso	0.04	0.08	0.06		0.07	0.12	0.09		0.15	0.33	0.22

Note: The reported estimates are based on Lasso penalty estimates (obtained from 10-fold cross-validation) for 14, 15 and 17 variable selection samples for 1, 2, and 4 quarter ahead models, respectively.

5.3.3 OCMT: selected variables by horizon

OCMT selects only a few variables in addition to the pre-selected UK and UK-specific foreign inflation (π_t, $\Delta\pi_t$,π_t^*, and $\Delta\pi_t^*$).[10] For $h = 1$, the variables selected are given in sub-section S-3.1.1 of the online supplement. In addition to the 2 pre-selected variables ($\pi_t, \Delta\pi_t$), AR2-OCMT selects the rate of change of wages (DWUK4) for samples ending in 2021q2, 2021q4 and 2022q1, and no other variables. ARX-OCMT does not select any additional variables for any of the 14 variable selection samples!

For $h = 2$, the variables selected are given in subsection S-3.1.2 of the online supplement. AR2-OCMT selects the rates of change of money (DMUK4) and exchange rate (DEPUK4) from samples ending in 2019q3 to 2020q3, then the rate of change of wages is added till the sample ending in 2022q2, from then the rates of change of money (DMUK4) and wages (DWUK4) are selected. ARX-OCMT selects just the rate

[10]The OCMT results reported here are based on the critical value function given by (12) with $\delta = 1$. Using the larger value of $\delta = 1.5$ reduces the number of selected variables for a few of sample periods, but the outcomes are generally robust to the choice on δ on the interval $[1, 1.5]$. The selection results for OCMT with $\delta = 1.5$ are reported in the online supplement.

of change of money (DMUK4) as an additional variable in every sample for $h = 2$.

For $h = 4$, the variables selected are given in subsection S-3.1.3 of the online supplement. AR2-OCMT chooses the same 3 extra variables - the rate of change of money (DMUK4) and exchange rate (DEPUK4) as well as the UK-specific measure of foreign inflation (DPSUK4, π_t^*) - for every sample ending from 2019q1 to 2021q1. Then, in the sample ending in 2021q2, AR2-OCMT chooses 12 extra variables. The number of variables selected then falls to 7 in 2021q3, 6 in 2021q4, 5 in 2022q1 and 4 in 2022q2 − 2022q4. These 4 are the rate of change of money (DMUK4), of material prices (DPMAT4), of wages (DWUK4), and π_t^* (DPSUK4). The number of variables selected falls to 3 in the sample ending in 2023q1 when the foreign inflation measure is no longer selected. ARX-OCMT chooses the rate of change of employment (DEMUK4) and of money for samples ending in 2019q1 − 2021q4, then adds material prices in 2022q1, and selects just the rate of change of money for the last four samples.

5.3.4 Lasso: selected variables by horizon

Lasso selections for each sample and horizon are given in the online supplement, subsection S-3.2. Lasso tends to select more variables than OCMT so we give less detail. Table 4 lists the variables chosen by standard Lasso at each horizon and the number of times they were chosen out of the maximum number of possible samples: 14, for $h = 1$, 15 for $h = 2$, and 17 for $h = 4$. UK inflation, π_t (DPUK4) is always chosen in every sample at every horizon as is the UK measure of foreign inflation, π_t^* (DPSUK4). The change in UK inflation, $\Delta\pi_t$, (DDPUK4) is chosen in every sample in the case of models for $h = 1$, and $h = 2$, but never for $h = 4$. The change in foreign inflation $\Delta\pi_t^*$ (DDPSUK4) is chosen in every sample at $h = 1$, in 3 samples at $h = 2$ but never at $h = 4$. Thus Lasso provides considerable support for the choice of pre-selected variables in $\boldsymbol{z}_t$ that include foreign inflation as well as the two lagged inflation variables.

Apart from these variables, the rate of change of wages and of money figure strongly when using Lasso. The rate of change of wages (DWUK4) is chosen in all the samples for $h = 1$ and $h = 2$ and 14 of the 17 samples for $h = 4$. The rate of change of money (DMUK) is chosen in 10 of the 14 samples for $h = 1$, and in every sample for $h = 2$ and $h = 4$. Money and wages are also chosen by OCMT but the rate of change of the exchange rate selected by OCMT is never chosen by Lasso.

When $h = 1$ Lasso also always selects two other variables, namely the change in long interest rates (DLRUK4), and import price inflation (DPMUK).

Table 4: The number of times covariates from the active set are selected by Lasso at different forecast horizons

	Horizon			Selected covariates
	$h=1$	$h=2$	$h=4$	
DPUK4	14	15	17	4 quarter UK rate of inflation
DPSUK4	14	15	17	UK specific measure of 4 quarter foreign inflation
DWUK4	14	15	14	4 quarter rate of change of average weekly earnings
DDPUK4	14	15	0	Change in 4 quarter UK rate of inflation
DPMUK	14	8	0	4 quarter rate of change of UK import prices
DLRUK4	14	4	0	Change in 4 quarter average UK long interest rate
DDPSUK4	14	3	0	Change in UK specific measure of 4 quarter foreign inflation
DMUK	10	15	17	4 quarter rate of change of UK M4 money
DRUK4	2	0	0	Change in 4 quarter average UK short interest rate
DEMUK4	1	0	6	4 quarter rate of change of UK employment
DDPOIL4	1	0	0	Change in 4 quarter rate of change of the oil price
DDEQUK4	1	0	0	Change in 4 quarter rate of change of UK equity prices
DDYSUK4	1	0	0	Change in 4 quarter rate of change of UK foreign real GDP
DGAPSUK12	1	0	0	UK log foreign real GDP relative to 8 quarter moving average
DPUS4	0	2	2	4 quarter US rate of inflation

Note: The number of variable selection samples are 14, 15 and 17 for $h = 1, 2,$ and 4 quarter ahead models, respectively.

5.4 Forecasts

The point forecasts of inflation for $h = 1, 2$ and $h = 4$ for the various selection procedures are summarized in Section S-4 of the online supplement. For each forecast horizon we have 13 forecasts and their realizations for the quarters $2020q1$ to $2023q1$ inclusive. These are summarised in Table 5 below. We use root mean square forecast error (RMSFE) as our forecast evaluation criterion. Since 13 forecast errors represent a very short evaluation sample with considerable serial correlation, testing for the significance of the loss differences using the Diebold and Mariano (1995) test would not be reliable, and is not pursued here.

5.4.1 One quarter ahead forecasts

Figure 4 gives plots of actual inflation and forecasts one quarter ahead. Section S-4.1 of the online supplement gives the point forecasts. For $h = 1$, ARX has the lowest RMSFE, the π_t^* and $\Delta\pi_t^*$ improve forecast performance relative to the AR2. AR2-OCMT adds wage growth in three periods. Lasso suffers from choosing too many variables relative to OCMT. The forecasts are very similar, except Lasso predicted a large drop in 2020q3 with a subsequent rebound. This results from selecting an output gap measure, when UK output dropped sharply in 2020q2. This sharp drop and rebound was also a feature of Lasso forecasts at other horizons. The Bank of England over-estimated inflation in 2022q4, correctly anticipating higher energy prices but not anticipating the government energy price guarantees.

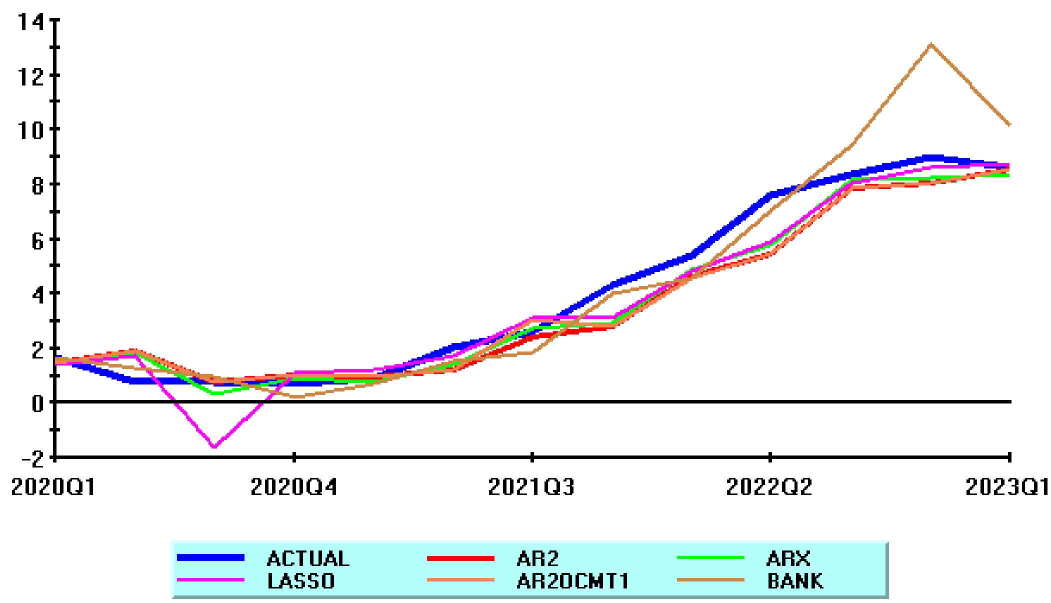

Figure 4: One quarter ahead forecasts.

5.4.2 Two quarter ahead forecasts

Figure 5 gives plots of actual inflation and forecasts two quarters ahead. Section S-4.2 of the online supplement gives the values. For $h = 2$, ARX again has the lowest RMSFE. ARX-OCMT selects money growth in every period. Lasso selects between 5 and 9 variables.

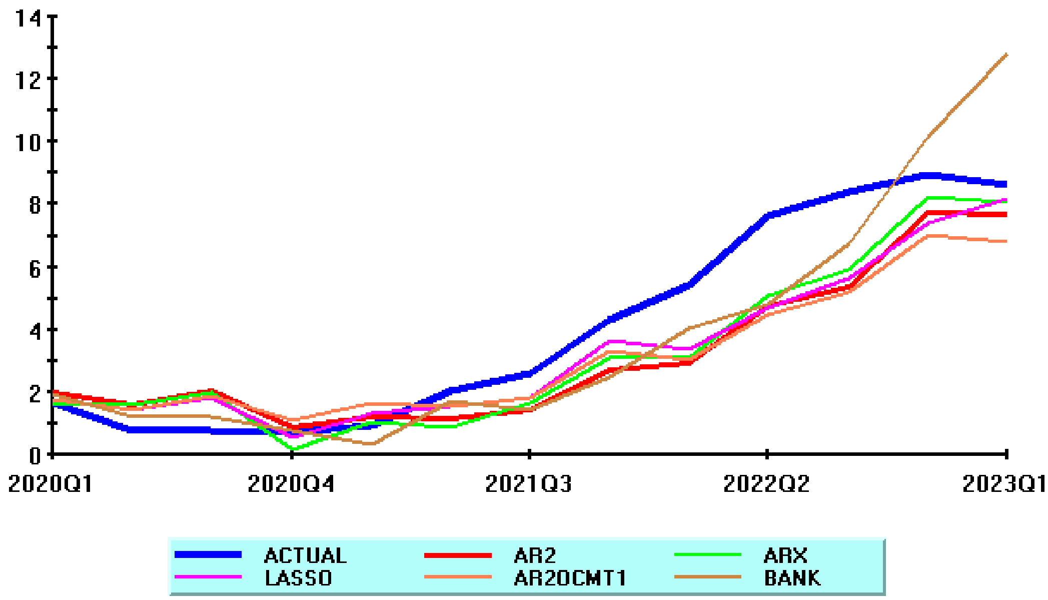

Figure 5: Plot of forecasts two quarters ahead

5.4.3 Four quarter ahead forecasts

Figure 6 gives plots of actual inflation and forecasts four quarters ahead. Section S-4.3 of the online supplement gives the values. The case of $h = 4$ is the only one where the ARX does not have the lowest RMSFE. The lowest RMSFE is obtained by AR2-OCMT. It does well by having a very high inflation forecast in 2022Q2. This corresponds to the selection of 12 extra variables in the sample ending in 2021q2. It then rejoins the pack in 2022q3.

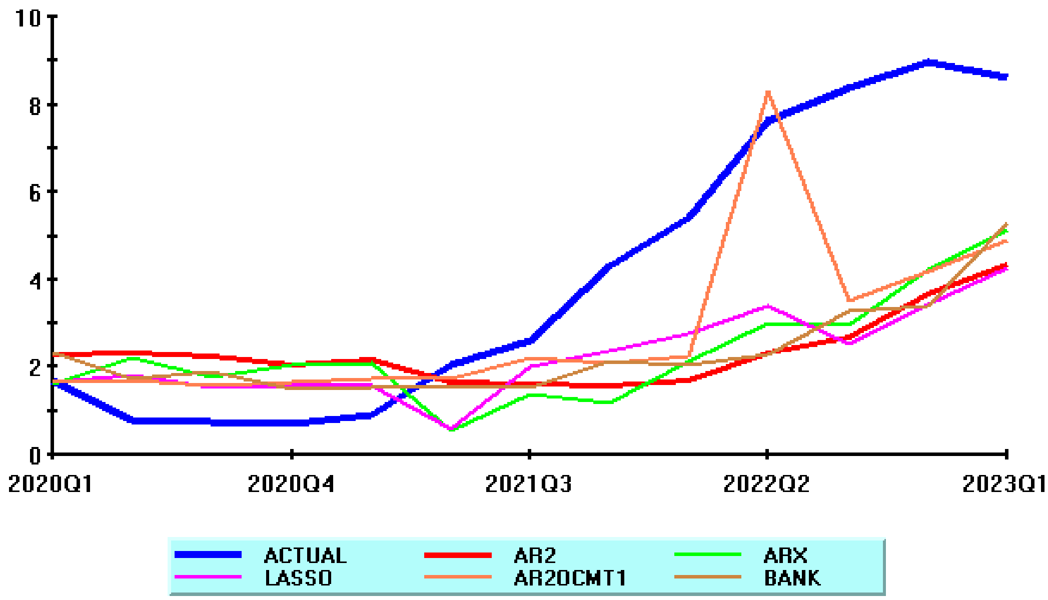

Figure 6: Plot of forecasts four quarters ahead

5.4.4 Summary

Table 5 brings together the RMSFE for each of the selection methods at the different horizons. Both the variable selection and forecasting exercises highlight the importance of taking account of persistence and foreign inflation for UK inflation forecasting. Lasso selects π_t and π_t^* in all three forecast horizon models. ARX which includes UK and foreign inflation as pre-selected variables tends to perform best in forecasting but in the present application the OCMT component does not seem to add much once the pre-selected variables are included. However, Lasso performs rather poorly when it is conditioned on the preselected variables.

Table 5: Root mean square forecast errors by forecast horizons, h=1,2 and 4 over the period 2020q1-2023q1

Forecast Source	**h=1**	**h=2**	**h=4**
AR2	0.9141	1.6039	3.2524
ARX	**0.7884**	**1.3813**	2.9883
Lasso	0.9696	1.4109	3.0131
AR2–Lasso	0.9617	2.8719	4.3440
ARX–Lasso	0.8750	2.9231	4.5021
AR2–OCMT	0.9233	1.6800	**2.4643**
ARX–OCMT	**0.7884**	1.4217	3.2470
Bank of England	1.3288	1.6959	3.0042

Note: The RMSFE figures are taken from online supplement Tables S-4.1, S-4.2, and S-4.3. The least value for RMSFE for each forecast horizon is shown in bold.

5.5 Contemporaneous drivers

Our forecasts of π_{t+h} are based on variables observed at time t, and do not depend on any conditioning. But, as noted above with respect to the Bank of England, it is common to condition on contemporaneous values of variables which are considered as *proximate* causes of the variable to be forecast. Even if such causal variables can be identified, however, it does not mean that they help with forecasting - often such causal variables are themselves difficult to forecast. The Bank of England over-estimated inflation in 2022q4, correctly anticipating higher energy prices but not anticipating the government energy price guarantees. This illustrates the dangers of conditioning on variables that cannot be forecast. Understanding does not necessarily translate into better forecasts. For example, knowing the causes of earthquakes does not necessarily help in predicting them in a timely manner.

This point can be illustrated by including contemporaneous changes in oil prices, in the UK inflation equation (over the period pre Covid-19 and the full sample) for the case $h = 1$. For both samples $\Delta poil_{t+1}$ are highly statistically significant, but their lagged values $\Delta poil_t$ are not.

Table 6: Contemporaneous and lagged effects of oil price changes on UK inflation (π_{t+1})

Covariates	**1979q2–2019q4**		**1979q2–2022q4**	
π_t	0.936	0.936	0.957	0.962
	(13.40)	(12.97)	(13.70)	(13.23)
π_{t-1}	–0.134	–0.136	–0.145	–0.151
	(–2.05)	(–2.02)	(–2.20)	(–2.21)
π_t^*	0.480	0.512	0.600	0.524
	(4.14)	(3.31)	(5.50)	(4.39)
π_{t-1}^*	–0.381	–0.316	–0.504	–0.432
	(–3.28)	(–2.52)	(–4.64)	(–3.61)
$\Delta poil_{t+1}$	**0.013**	·	**0.012**	·
	(3.40)	·	**(3.91)**	·
$\Delta poil_t$	·	0.006	·	0.005
	·	(1.47)	·	(1.37)
$\bar{R}^2$	0.953	0.950	0.952	0.948
SER	0.612	0.631	0.626	0.651

6 Conclusion

High dimensional data are not a panacea; the data must have some predictive content which might come from spatial or temporal sequential patterns. Forecasting is particularly challenging either if there are unknown unknowns (factors that are not even thought about) or if there are known factors that are falsely believed to be important. When there are new global factors, like Covid-19, or when a relevant variable has shown little variation over the sample period, forecasting their effect is going to be problematic.

Many forecasting problems require a hierarchical structure where latent factors at local and global levels are explicitly taken into account. This is particularly relevant for macro forecasting in an increasingly inter-connected world. It is important that we allow for global factors in national forecasting exercises - and GVAR was an attempt in this direction.

A number of key methodological issues were illustrated with a simple approach to forecasting UK inflation which has become a topic of public discussion. This example showed both the power of parsimonious models and the importance of global factors. There remain many challenges. How to allow for regime change and parameter instability in the case of high-dimensional data analysis? How to choose data samples? Our recent research suggests that it is best to use long time series samples for variable selections, but consider carefully what sample to use for forecasting. Given a set of selected variables, parameter estimation can be based on different window sizes, or down-weighting. Should we use ensemble or forecast averaging? Forecast averaging will only work if the covariates used to forecast the target variable are driven by strong common factors, otherwise one will be averaging over noise.

There are some more general lessons. Econometric and statistical models must not become a straightjacket. Forecasters should be open minded about factors not included in their model, and acknowledge that forecasts are likely to be wrong if unexpected shocks hit.

References

Bergmeir, C., R. J. Hyndman, and B. Koo (2018): "A note on the validity of cross-validation for evaluating autoregressive time series prediction," *Computational Statistics & Data Analysis*, 120, 70–83.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005): "Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach," *The Quarterly journal of economics*, 120, 387–422.

Chudik, A., V. Grossman, and M. H. Pesaran (2016): "A multi-country approach to forecasting output growth using PMIs," *Journal of Econometrics*, 192, 349–365.

Chudik, A., G. Kapetanios, and M. H. Pesaran (2018): "A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models," *Econometrica*, 86, 1479–1512.

Chudik, A., and M. H. Pesaran (2016): "Theory and practice of GVAR modelling," *Journal of Economic Surveys*, 30, 165–197.

Chudik, A., M. H. Pesaran, and M. Sharifvaghefi (2023): "Variable selection in high dimensional linear regressions with parameter instability," *arXiv:2312.15494 [econ.EM] https://arxiv.org/abs/2312.15494*.

Diebold, F. X., and R. S. Mariano (1995): "Comparing predictive accu racy," *Journal of Business and Economic Statistics*, 13, 253–263.

Fan, J., Y. Ke, and K. Wang (2020): "Factor-adjusted regularized model selection," *Journal of Econometrics*, 216, 71–85.

Giannone, D., M. Lenza, and G. E. Primiceri (2021): "Economic predictions with big data: The illusion of sparsity," *Econometrica*, 89, 2409–2437.

Granger, C. W., and M. H. Pesaran (2000a): "Economic and statistical measures of forecast accuracy," *Journal of Forecasting*, 19, 537–560.

Granger, C. W. J., and M. H. Pesaran (2000b): "A decision theoretic approach to forecast evaluation," in *Statistics and Finance: An Interface*, pp. 261–278. World Scientific.

Hansen, C., and Y. Liao (2019): "The factor-lasso and k-step bootstrap approach for inference in high-dimensional economic applications," *Econometric Theory*, 35, 465–509.

Lahiri, S. N. (2021): "Necessary and sufficient conditions for variable selection consistency of the LASSO in high dimensions," *The Annals of Statistics*, 49, 820 – 844.

Marcellino, M., J. H. Stock, and M. W. Watson (2006): "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," *Journal of econometrics*, 135, 499–526.

Meinshausen, N., and P. Bühlmann (2006): "Variable selection and high-dimensional graphs with the lasso," *Annals of Statistics*, 34, 1436–1462.

Mohaddes, K., and M. Raissi (2024): "Compilation, Revision and Updating of the Global VAR (GVAR) Database, 1979Q2-2023Q3," Mendeley Data, V1.

Pesaran, M. H., A. Pick, and M. Pranovich (2013): "Optimal forecasts in the presence of structural breaks," *Journal of Econometrics*, 177, 134–152.

Pesaran, M. H., A. Pick, and A. Timmermann (2011): "Variable selection, estimation and inference for multi-period forecasting problems," *Journal of Econometrics*, 164, 173–187.

Pesaran, M. H., and S. Skouras (2004): "Decision-based methods for forecast evaluation," in *A Companion to Economic Forecasting*, ed. by D. F. H. Michael P. Clements, chap. 11, pp. 241–267. Wiley Online Library.

Sharifvaghefi, M. (2023): "Variable selection in linear regressions with many highly correlated covariates," *Available at SSRN: https://ssrn.com/abstract=4159979*.

Shrader, J. G., L. Bakkensen, and D. Lemoine (2023): "Fatal Errors: The Mortality Value of Accurate Weather Forecasts," *Working Paper Series, National Bureau of Economic Research, Number 31361*.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58, 267–288.

Wainwright, M. J. (2019): *High-dimensional Statistics: A Non-asymptotic Viewpoint*. Cambridge University Press, Cambridge.

Whittle, P. (1983): *Prediction and Regulation by Linear Least-Square Methods*. University of Minnesota Press.

Zhao, P., and B. Yu (2006): "On model selection consistency of Lasso," *The Journal of Machine Learning Research*, 7, 2541–2563.

Zou, H., and T. Hastie (2005): "Regularization and variable selection via the elastic net,," *Journal of the Royal Statistical Society B,*, 67, 301–320.

Appendix

The Lasso procedure with a set of preselected variables[11]

Let $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$ be the vector of observations for the target variable. Suppose we have a vector of pre-selected covariates denoted by $\mathbf{z}_t = (z_{1t}, z_{2t}, \cdots, z_{mt})'$. Additionally, there is a vector of covariates denoted by $\mathbf{x}_t = (x_{1t}, x_{2t}, \cdots, x_{nt})'$, from which we aim to select the relevant ones for the target variable using the Lasso procedure. We can further stack the observations for $\mathbf{z}_t$ and $\mathbf{x}_t$ in matrices $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_T)'$ and $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T)'$, respectively. For a given value of the tuning parameter, λ, the Lasso problem can be written as:

$$\left(\hat{\boldsymbol{\delta}}(\lambda), \hat{\boldsymbol{\beta}}(\lambda)\right)' = \text{argmin}_{\mathbf{b}_z, \mathbf{b}} \left\{(\mathbf{y} - \mathbf{Z}\mathbf{b}_z - \mathbf{X}\mathbf{b})' (\mathbf{y} - \mathbf{Z}\mathbf{b}_z - \mathbf{X}\mathbf{b}) + \lambda \|\mathbf{b}\|_1\right\}.$$

Partition $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ where $\mathbf{X}_1$ is the matrix of covariates with the corresponding vector of estimated coefficients, $\hat{\boldsymbol{\beta}}_1(\lambda)$, different from zero and $\mathbf{X}_2$ is the matrix of covariates with the corresponding vector of estimated coefficients, $\hat{\boldsymbol{\beta}}_2(\lambda)$, equal to zero. So, $\mathbf{X}\hat{\boldsymbol{\beta}}(\lambda) = \mathbf{X}_1\hat{\boldsymbol{\beta}}_1(\lambda)$. By the first order conditions we have:

$$\mathbf{X}_1' \left(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\delta}}(\lambda) - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1(\lambda)\right) - \lambda \text{sign}\left(\hat{\boldsymbol{\beta}}_1(\lambda)\right) = 0, \tag{A.15}$$

$$\mathbf{Z}' \left(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\delta}}(\lambda) - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1(\lambda)\right) = 0. \tag{A.16}$$

and

$$-\lambda\mathbf{1} \leq \mathbf{X}_2' \left(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\delta}}(\lambda) - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1(\lambda)\right) \leq \lambda\mathbf{1}, \tag{A.17}$$

Where $\mathbf{1}$ represents a vector of ones. We can further conclude from Equation (A.16) that:

$$\hat{\boldsymbol{\delta}}(\lambda) = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \left(\mathbf{y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1(\lambda)\right). \tag{A.18}$$

By substituting $\hat{\boldsymbol{\delta}}(\lambda)$ from (A.18) into (A.15), we have

$$\mathbf{X}_1' \left(\mathbf{y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1(\lambda) - \mathbf{Z}\,(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \left(\mathbf{y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1(\lambda)\right)\right) - \lambda \text{sign}\left(\hat{\boldsymbol{\beta}}_1(\lambda)\right) = 0.$$

[11] We are grateful to Dr. Mahrad Sharifvaghefi for providing the proofs in this appendix.

We can further write this as:

$$\mathbf{X}_1' \left(\mathbf{I} - \mathbf{Z} \left(\mathbf{Z}'\mathbf{Z} \right)^{-1} \mathbf{Z}' \right) \left(\mathbf{y} - \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1(\lambda) \right) - \lambda \text{sign} \left(\hat{\boldsymbol{\beta}}_1(\lambda) \right) = 0.$$

Therefore,

$$\tilde{\mathbf{X}}_1' \left(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}_1 \hat{\boldsymbol{\beta}}_1(\lambda) \right) - \lambda \text{sign} \left(\hat{\boldsymbol{\beta}}_1(\lambda) \right) = 0, \tag{A.19}$$

where $\tilde{\mathbf{X}}_1 = \mathbf{M}_Z \mathbf{X}_1$, $\tilde{\mathbf{y}} = \mathbf{M}_Z \mathbf{y}$ and $\mathbf{M}_Z = \mathbf{I} - \mathbf{Z} \left(\mathbf{Z}'\mathbf{Z} \right)^{-1} \mathbf{Z}'$.

Similarly, by substituting $\hat{\boldsymbol{\delta}}(\lambda)$ from (A.18) into (A.17), we have

$$-\lambda \mathbf{1} \leq \tilde{\mathbf{X}}_2' \left(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}_1 \hat{\boldsymbol{\beta}}_1(\lambda) \right) \leq \lambda \mathbf{1}. \tag{A.20}$$

Note that (A.19) and (A.20) are the first order conditions of the following Lasso problem:

$$\hat{\boldsymbol{\beta}}(\lambda) = \text{argmin}_{\mathbf{b}} \left\{ \left(\tilde{\mathbf{y}} - \tilde{\mathbf{X}} \mathbf{b} \right)' \left(\tilde{\mathbf{y}} - \tilde{\mathbf{X}} \mathbf{b} \right) + \lambda ||\mathbf{b}||_1 \right\}. \tag{A.21}$$

Therefore, we can first obtain the estimator of the vector coefficients for $\mathbf{X}$, $\hat{\boldsymbol{\beta}}(\lambda)$, by solving the Lasso problem given by (A.21) and then estimate the vector of coefficients for $\mathbf{Z}$, $\hat{\boldsymbol{\delta}}(\lambda)$, by using Equation (A.18).

Online Supplement

to

High-dimensional forecasting with known knowns and known unknowns

by

M. Hashem Pesaran

University of Southern California, USA and Trinity College, Cambridge, UK

Ron P. Smith

Birkbeck, University of London, United Kingdom

March 21, 2024

S-1 Introduction

This online supplement presents the details of the methods used to forecast UK inflation, the variables selected and the results obtained. For this purpose the framework set out in the paper can be summarised as:

$$\pi_{t+h} = c_h + \mathbf{a}_h' \mathbf{z}_t + \sum_{j=1}^{K} \beta_{jh} I(j \in DGP) x_{jt} + u_{h,t+h}. \tag{S.1}$$

The target variable is average annual UK inflation, labelled DPUK4, defined as $\pi_{t+h} = 100 \times \log(p_{t+h}/p_{t+h-4})$, where p_t is the UK consumer price index taken from the IMF International Financial Statistics.[12] $\mathbf{z}_t$ is the vector of pre-selected variables which we set to $(\pi_t, \Delta\pi_t)$ or ($\pi_t, \Delta\pi_t$,π_t^*, and $\Delta\pi_t^*$). The active set, x_{jt}, $j = 1, 2, ..., K$, is the list of variables from which selection is made.

The active set consists of the 26 variables detailed in Table 1 of the paper plus their first differences (in total $K = 52$). As noted in the paper, whereas including a current and lagged value is equivalent to including a current value and its change in estimation, it is not in the selection. We denote $\mathcal{A}_m^h$ as the active set corresponding to model m among the 9 models that are examined. For each horizon, selection proceeds recursively using an expanding window starting in 1979q2 initially ending in 2019q1, for $h = 4$, 2019q3 for $h = 2$, and 2019q4 for $h = 4$. The window then expands one quarter at a time to 2023q1. We end up with 13 point forecasts for the quarters $2020q1 - 2023q1$. Since the model is re-estimated for each forecast period and horizon, the variables selected may change by quarter and horizon.

S-2 Models

The first two models set all the $\beta_{jh} = 0$ and differ in what is included in $\mathbf{z}_t$.

The $AR2$ model is:

$$\pi_{t+h} = c_h + a_{1h}\pi_t + a_{2h}\Delta\pi_t + u_{h,t+h}.$$

[12] In terms of the GVAR data, which report $dp_t = \log(p_t/p_{t-1})$ where p_t is the CPI, π_t is defined by $\pi_t = 100 \times (dp_t + dp_{t-1} + dp_{t-2} + dp_{t-3})$.

The ARX adds UK-specific foreign inflation, π_t^* and its change:

$$\pi_{t+h} = c_h + a_{1h}\pi_t + a_{2h}\Delta\pi_t + a_{3h}\pi_t^* + a_{4h}\Delta\pi_t^* + u_{h,t+h}$$

The next model sets $\mathbf{a}_h' = 0$ and just selects using Lasso, which performs a variable selection from the whole active set of 52 variables inclusive of $\pi_t, \Delta\pi_t, \pi_t^*$, and $\Delta\pi_t^*$. All covariates are standardized before solving the following minimization problem:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{|\mathcal{A}^h_{\text{Lasso}}|}} \frac{1}{2T}\sum_{t=1}^{T}\left(\pi_{t+h} - c_h - \sum_{j\in\mathcal{A}^h_{\text{Lasso}}} \tilde{x}_{tj}\beta_j\right)^2 + \lambda \sum_{j\in\mathcal{A}^h_{\text{Lasso}}} \|\beta_j\|_1 ,$$

where $\tilde{x}_{tj} = (x_{jt} - \bar{x}_{jT})/s_{jT}$, $\bar{x}_{jT} = T^{-1}\sum_{t=1}^{T} x_{jt}$ and $s_{jT} = T^{-1}\sum_{t=1}^{T}(x_{jt} - \bar{x}_{jT})$. $|\mathcal{A}^h_{\text{Lasso}}|$ represents the cardinality of the active set, and $||\cdot||_1$ is the ℓ_1-norm. The regularization parameter $\hat{\lambda}_{hT}$ is chosen via 10-fold cross-validation (CV). As discussed in the paper we adapt the standard CV procedures to take account of the persistence and changing variance of time series data. In the standard procedure the CV subsets (folds) are typically chosen randomly, but since these are time series where order matters, we retain the order of the time series data in construction of subsets using all the data, not dropping observations between subsets. In addition, the standard 10-fold procedure chooses the $\hat{\lambda}_{hT}$ that minimises the pooled MSE over the ten subsets. But when variances differ substantially over subsets, pooling is not appropriate, instead we follow Chudik, Kapetanios, and Pesaran (2018, CKP), and use the average of the $\hat{\lambda}_{hT}$ chosen in each subset. Full details are provided in the online simulation appendix to CKP (2018).

We considered a number of Lasso models where we did not apply selection to the AR2 or ARX variables which were included in the $\boldsymbol{z}_t$. Then Lasso was applied to the active set having removed the effect of $\boldsymbol{z}_t$. This gives models AR2-LASSO; ARX-LASSO. We now consider GOCMT-based Models. Like OCMT, GOCMT allows for the multiple testing nature of the procedure (K separate tests - with K large) by increasing the level of significance with K. But it also includes the principal component of the standardised values of the active set, $\hat{\boldsymbol{\varkappa}}_t$, to allow for the correlations between the variables. In the first stage, K separate OLS regressions are computed entering the variables from the active set one at a time, together with any $\boldsymbol{z}_t$ and $\hat{\varkappa}_t$:

$$y_{t+h} = c_h + \mathbf{a}_h' \boldsymbol{z}_t + \mathbf{b}_h' \hat{\varkappa}_t + \phi_{jh} x_{jt} + e_{j,h,t+h},\ t = 1, 2, ..., T, \text{ for } j = 1, 2, ..., K,$$

Denote the t-ratio of ϕ_{jh} by $t_{\hat{\phi}_{j,(1)}}$. Then variable j is selected if

$$\widehat{\mathcal{J}}_{j,(1)} = I\left[\left|t_{\hat{\phi}_{j,(1)}}\right| > c_p(K)\right],\ \text{for } j = 1, 2, ..., K,$$

where the critical value is given by

$$c_p(K, \delta) = \Phi^{-1}\left(1 - \frac{p}{2K^{\delta}}\right).$$

p is the nominal size set to 5%, $\Phi^{-1}(\cdot)$ is the inverse of a standard normal distribution function and δ is a fixed constant set close to 1. We experiment with two values for δ, $\delta = 1$ and $\delta = 1.5$.

The principal component $\hat{\varkappa}_t$ is used just to control for correlations across the variables in the active set at the selection stage, it is not included in the final forecast regression. The AR2-OCMT model includes the AR2 in the $\boldsymbol{z}_t$ and applies GOCMT to the rest of the active set, similarly for ARX-OCMT.

Thus we have a set of 9 models: (1) AR2, (2) ARX, (3) LASSO, (4) AR2-LASSO, (5) ARX-LASSO, (6) AR2-OCMT ($\delta = 1$), (7) ARX-OCMT ($\delta = 1$), (8) AR2-OCMT ($\delta = 1.5$), (9) ARX-OCMT ($\delta = 1.5$).

S-3 Variable Selection

This section sets out the variables chosen by each selection method for each horizon by the quarter in which the sample ended. The abbreviations for the variables are given in Table 1, in the paper.

S-3.1 Variables selected by OCMT

S-3.1.1 For h=1 quarter ahead models

This sub-section sets out the variables chosen by OCMT for $h = 1$.

AR2-OCMT selects DWUK for samples ending in 2021q2, 2021q4 and 2022q1 and no other additional variables. ARX-OCMT ($\delta = 1$), AR2-OCMT ($\delta = 1.5$), and ARX-OCMT ($\delta = 1.5$) do not select any additional variables.

S-3.1.2 For h=2 quarter ahead models

This sub-section sets out the variables chosen by OCMT for $h = 2$.

Table S.1: Selected Variables by AR2-OCMT ($\delta = 1$) 2q Ahead

#	Ending Year					
	2019q3	2019q4	2020q1	2020q2	2020q3	2020q4
1	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
2	DEPUK4	DEPUK4	DEPUK4	DEPUK4	DEPUK4	DEPUK4
3	·	·	·	·	·	DWUK4
	2021q1	2021q2	2021q3	2021q4	2022q1	2022q2
1	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
2	DEPUK4	DEPUK4	DEPUK4	DEPUK4	DEPUK4	DEPUK4
3	DWUK4	DWUK4	DWUK4	DWUK4	DWUK4	DWUK4
	2022q3	2022q4	2023q1			
1	DMUK4	DMUK4	DMUK4			
2	DWUK4	DWUK4	DWUK4			
	$\delta = 1$					

ARX-OCMT $\delta = 1$ selects just DMUK4 for every sample.

Table S.2: Selected Variables by AR2-OCMT ($\delta = 1.5$) 2q Ahead

#	Ending Year					
	2019q3	2019q4	2020q1	2020q2	2020q3	2020q4
1	·	·	·	·	·	DMUK4
	2021q1	2021q2	2021q3	2021q4	2022q1	2022q2
1	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
2	·	DWUK4	DWUK4	DWUK4	DWUK4	DWUK4
	2022q3	2022q4	2023q1			
1	DWUK4	·	·			
2	·	·	·			
	$\delta = 1.5$					

ARX-OCMT $\delta = 1.5$ selects DMUK4 for 2020q4-2022q2.

S-3.1.3 For h=4 quarter ahead models

This sub-section sets out the variables chosen by OCMT selection method for $h = 4$.

Table S.3: Selected Variables by AR2-OCMT ($\delta = 1$) 4q Ahead

#	Ending Year					
	2019q1	2019q2	2019q3	2019q4	2020q1	2020q2
1	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
2	DEPUK4	DEPUK4	DEPUK4	DEPUK4	DEPUK4	DEPUK4
3	DPSUK4	DPSUK4	DPSUK4	DPSUK4	DPSUK4	DPSUK4
	2020q3	2020q4	2021q1	2021q2	2021q3	2021q4
1	DMUK4	DMUK4	DMUK4	DDYUK4	DEMUK4	LRUK4
2	DEPUK4	DEPUK4	DEPUK4	DGAPUK8	LRUK4	DMUK4
3	DPSUK4	DPSUK4	DPSUK4	DGAPUK12	DMUK4	DPMAT4
4	·	·	·	DEMUK4	DPMAT4	DPSUK4
5	·	·	·	RUK4	DPSUK4	LRUS4
6	·	·	·	LRUK4	LRUS4	DWUK4
7	·	·	·	DMUK4	DWUK4	·
8	·	·	·	DPMAT4	·	·
9	·	·	·	DPSUK4	·	·
10	·	·	·	RUS4	·	·
11	·	·	·	LRUS4	·	·
12	·	·	·	DWUK4	·	·
	2022q1	2022q2	2022q3	2022q4	2023q1	
1	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	
2	DPMAT4	DPMAT4	DPMAT4	DPMAT4	DPMAT4	
3	DPSUK4	DPSUK4	DPSUK4	DPSUK4	DWUK4	
4	LRUS4	DWUK4	DWUK4	DWUK4	·	
5	DWUK4	·	·	·	·	
	$\delta = 1$					

Table S.4: Selected Variables by ARX-OCMT ($\delta = 1$) 4q Ahead

#	Ending Year					
	2019q1	2019q2	2019q3	2019q4	2020q1	2020q2
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
	2020q3	2020q4	2021q1	2021q2	2021q3	2021q4
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
	2022q1	2022q2	2022q3	2022q4	2023q1	
1	DEMUK4	DMUK4	DMUK4	DMUK4	DMUK4	
2	DMUK4	·	·	·	·	
3	DPMAT4	·	·	·	·	
	$\delta = 1$					

Table S.5: Selected Variables by AR2-OCMT ($\delta = 1.5$) 4q Ahead

#	Ending Year					
	2019q1	2019q2	2019q3	2019q4	2020q1	2020q2
1	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
	2020q3	2020q4	2021q1	2021q2	2021q3	2021q4
1	DMUK4	DMUK4	DMUK4	DGAPUK8	DMUK4	DMUK4
2	·	·	·	DGAPUK12	DPSUK4	DPMAT4
3	·	·	·	DEMUK4	LRUS4	DPSUK4
4	·	·	·	LRUK4	·	·
5	·	·	·	DMUK4	·	·
6	·	·	·	DPMAT4	·	·
7	·	·	·	DPSUK4	·	·
8	·	·	·	LRUS4	·	·
	2022q1	2022q2	2022q3	2022q4	2023q1	
1	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	
2	DPMAT4	DPMAT4	DPMAT4	DWUK4	DWUK4	
3	DPSUK4	DWUK4	DWUK4	·	·	
4	DWUK4	·	·	·	·	
	$\delta = 1.5$					

Table S.6: Selected Variables by ARX-OCMT ($\delta = 1.5$) 4q AheadA

#	Ending Year					
	2019q1	2019q2	2019q3	2019q4	2020q1	2020q2
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
	2020q3	2020q4	2021q1	2021q2	2021q3	2021q4
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
	2022q1	2022q2	2022q3	2022q4	2023q1	
1	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	
	$\delta = 1.5$					

S-3.2 Variables selected by Lasso

S-3.2.1 For h=1 quarter ahead models

This sub-section sets out the variables chosen by Lasso for $h = 1$.

Table S.7: Selected Variables by Lasso 1q Ahead

	Ending Year						
#	2019q4	2020q1	2020q2	2020q3	2020q4	2021q1	2021q2
1	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4
2	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4
3	DLRUK4	DLRUK4	DRUK4	DLRUK4	DLRUK4	DEMUK4	DLRUK4
4	DMUK4	DMUK4	DLRUK4	DMUK4	DMUK4	DRUK4	DMUK4
5	DPMUK4	DPMUK4	DMUK4	DPMUK4	DPMUK4	DLRUK4	DPMUK4
6	DPSUK4	DPSUK4	DDPOIL4	DPSUK4	DPSUK4	DMUK4	DPSUK4
7	DDPSUK4	DDPSUK4	DPMUK4	DDPSUK4	DWUK4	DDEQUK4	DDPSUK4
8	DWUK4	DWUK4	DPSUK4	DWUK4	·	DPMUK4	DWUK4
9	·	·	DDPSUK4	·	·	DPSUK4	·
10	·	·	DDYSUK4	·	·	DDPSUK4	·
11	·	·	DGAPSUK12	·	·	DWUK4	·
12	·	·	DWUK4	·	·	·	·
λ	0.1094	0.1093	0.0736	0.1098	0.1101	0.0741	0.1099

#	2021q3	2021q4	2022q1	2022q2	2022q3	2022q4	2023q1
1	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4
2	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4
3	DLRUK4	DLRUK4	DLRUK4	DLRUK4	DLRUK4	DLRUK4	DLRUK4
4	DMUK4	DMUK4	DMUK4	DPMUK4	DPMUK4	DPMUK4	DPMUK4
5	DPMUK4	DPMUK4	DPMUK4	DPSUK4	DPSUK4	DPSUK4	DPSUK4
6	DPSUK4	DPSUK4	DPSUK4	DDPSUK4	DDPSUK4	DDPSUK4	DDPSUK4
7	DDPSUK4	DDPSUK4	DDPSUK4	DWUK4	DWUK4	DWUK4	DWUK4
8	DWUK4	DWUK4	DWUK4	·	·	·	·
λ	0.1096	0.1091	0.1091	0.1104	0.1128	0.1159	0.1184

Note: $\hat{\lambda}_{hT}$ is the estimate of the Lasso penalty parameter computed by 10-fold cross-validation.

Table S.8: Selected Variables by AR2-Lasso 1q Ahead

	Ending Year						
#	2019q4	2020q1	2020q2	2020q3	2020q4	2021q1	2021q2
1	DEMUK4	DMUK4	DEMUK4	DEMUK4	DMUK4	DPSUK4	DMUK4
2	DMUK4	DPSUK4	DLRUK4	DDWUK4	DPSUK4	DDPSUK4	DDEQUK4
3	DPSUK4	DDPSUK4	DMUK4	DLRUK4	DDPSUK4	DWUK4	DPSUK4
4	DDPSUK4	DGAPSUK8	DPMUK4	DMUK4	DWUK4	·	DDPSUK4
5	DGAPSUK8	DGAPSUK12	DPSUK4	DDEQUK4	·	·	DWUK4
6	DGAPSUK12	DWUK4	DDPSUK4	DPMUK4	·	·	·
7	DWUK4	·	DGAPSUK8	DPSUK4	·	·	·
8	·	·	DGAPSUK12	DDPSUK4	·	·	·
9	·	·	DWUK4	DWUK4	·	·	·
$\hat{\lambda}_{hT}$	0.0883	0.0913	0.0823	0.0767	0.0899	0.0955	0.0919

#	2021q3	2021q4	2022q1	2022q2	2022q3	2022q4	2023q1
1	DMUK4	DLRUK4	DLRUK4	DLRUK4	DLRUK4	DLRUK4	DLRUK4
2	DDEQUK4	DMUK4	DMUK4	DDEQUK4	DMUK4	DDEQUK4	DDEQUK4
3	DPSUK4	DDEQUK4	DDEQUK4	DPMUK4	DDEQUK4	DPMUK4	DPMUK4
4	DDPSUK4	DPMUK4	DPMUK4	DPSUK4	DPMUK4	DPSUK4	DPSUK4
5	DWUK4	DPSUK4	DPSUK4	DDPSUK4	DPSUK4	DDPSUK4	DDPSUK4
6	·	DDPSUK4	DDPSUK4	DWUK4	DDPSUK4	DWUK4	DWUK4
7	·	DWUK4	DWUK4	·	DWUK4	·	·
$\hat{\lambda}_{hT}$	0.0952	0.0927	0.0878	0.0874	0.0787	0.0904	0.0840

Note: $\hat{\lambda}_{hT}$ is the estimate of the Lasso penalty parameter computed by 10-fold cross-validation.

Table S.9: Selected Variables by ARX-Lasso 1q Ahead

	Ending Year						
#	2019q4	2020q1	2020q2	2020q3	2020q4	2021q1	2021q2
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DVUK4	DVUK4	DVUK4	DVUK4	DDWUK4	DVUK4	DVUK4
3	DDWUK4	DDWUK4	DDWUK4	DDWUK4	DLRUK4	DDWUK4	DDWUK4
4	DLRUK4	DLRUK4	DLRUK4	DLRUK4	DMUK4	DLRUK4	DLRUK4
5	DMUK4	DMUK4	DMUK4	DMUK4	DDEQUK4	DMUK4	DMUK4
6	DYCHINA4	DYCHINA4	DDEQUK4	DDEQUK4	·	DDEQUK4	DDEQUK4
7	DGAPSUK8	DGAPSUK8	DGAPSUK8	·	·	DPMUK4	DPUS4
8	DGAPSUK12	DGAPSUK12	DGAPSUK12	·	·	DPUS4	DWUK4
9	·	·	·	·	·	DWUK4	·
$\hat{\lambda}_{hT}$	0.0659	0.0678	0.0665	0.0702	0.0796	0.0562	0.0639

#	2021q3	2021q4	2022q1	2022q2	2022q3	2022q4	2023q1
1	DEMUK4	DEMUK4	DDYUK4	DDYUK4	DDYUK4	DEMUK4	DEMUK4
2	DLRUK4	DLRUK4	DEMUK4	DEMUK4	DEMUK4	DDVUK4	DDVUK4
3	DMUK4	DMUK4	DDVUK4	DDVUK4	DDVUK4	DDWUK4	DLRUK4
4	DDEQUK4	DDEQUK4	DDUUK	DDUUK	DDUUK	DLRUK4	DMUK4
5	·	DWUK4	DLRUK4	DLRUK4	DLRUK4	DMUK4	DDEQUK4
6	·	·	DMUK4	DMUK4	DMUK4	DDEQUK4	DPMUK4
7	·	·	DDEQUK4	DDEQUK4	DDEQUK4	DDPMAT4	DYCHINA4
8	·	·	DDPMAT4	DDPMAT4	DDPMAT4	DPMUK4	DWUK4
9	·	·	DPMUK4	DPMUK4	DPMUK4	DYCHINA4	·
10	·	·	DYCHINA4	DYCHINA4	DYCHINA4	DDPUS4	·
11	·	·	DPUS4	DDPUS4	DDPUS4	DWUK4	·
12	·	·	DWUK4	DWUK4	DWUK4	·	·
$\hat{\lambda}_{hT}$	0.0752	0.0714	0.0422	0.0457	0.0456	0.0550	0.0624

Note: $\hat{\lambda}_{hT}$ is the estimate of the Lasso penalty parameter computed by 10-fold cross-validation.

S-3.2.2 For h=2 quarter ahead models

This sub-section sets out the variables chosen by Lasso for $h = 2$.

Table S.10: Selected Variables by Lasso 2q Ahead

	Ending Year							
#	2019q3	2019q4	2020q1	2020q2	2020q3	2020q4	2021q1	2021q2
1	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4
2	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4
3	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
4	DPSUK4	DPSUK4	DPSUK4	DPSUK4	DPSUK4	DPSUK4	DPSUK4	DPMUK4
5	DWUK4	DWUK4	DWUK4	DWUK4	DWUK4	DWUK4	DWUK4	DPSUK4
6	·	·	·	·	·	·	·	DWUK4
$\hat{\lambda}_{hT}$	0.3113	0.3114	0.2494	0.2813	0.3133	0.2521	0.2529	0.2521

#	2021q3	2021q4	2022q1	2022q2	2022q3	2022q4	2023q1
1	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4
2	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4	DDPUK4
3	DLRUK4	DMUK4	DMUK4	DMUK4	DLRUK4	DLRUK4	DLRUK4
4	DMUK4	DPMUK4	DPMUK4	DPMUK4	DMUK4	DMUK4	DMUK4
5	DPMUK4	DPSUK4	DPSUK4	DPSUK4	DPMUK4	DPMUK4	DPMUK4
6	DPSUK4	DWUK4	DWUK4	DWUK4	DPSUK4	DPSUK4	DPSUK4
7	DWUK4	·	·	·	DDPSUK4	DDPSUK4	DDPSUK4
8	·	·	·	·	DWUK4	DPUS4	DPUS4
9	·	·	·	·	·	DWUK4	DWUK4
$\hat{\lambda}_{hT}$	0.2196	0.2490	0.2475	0.3112	0.2221	0.2297	0.2366

Note: $\hat{\lambda}_{hT}$ is the estimate of the Lasso penalty parameter computed by 10-fold cross-validation.

Table S.11: Selected Variables by AR2-LASSO 2q Ahead

	Ending Year					
#	2019q3	2019q4	2020q1	2020q2	2020q3	2020q4
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DVUK4	DVUK4	DVUK4	DVUK4	DVUK4	DVUK4
3	DDWUK4	DDWUK4	DDWUK4	DDWUK4	DDWUK4	DDUUK
4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DDWUK4
5	DPMAT4	DPMUK4	DPMAT4	DPMAT4	DPMAT4	DLRUK4
6	DPMUK4	DDPMUK4	DPMUK4	DPMUK4	DPMUK4	DMUK4
7	DDPMUK4	DPSUK4	DDPMUK4	DDPMUK4	DDPMUK4	DPMAT4
8	DEPUK4	DDPSUK4	DEPUK4	DEPUK4	DEPUK4	DDPMETAL4
9	DPSUK4	DDYCHINA4	DPSUK4	DPSUK4	DPSUK4	DDPMUK4
10	DDPSUK4	DGAPSUK8	DDPSUK4	DDPSUK4	DDPSUK4	DEPUK4
11	DYCHINA4	DWUK4	DDYCHINA4	DDYCHINA4	DDYCHINA4	DPSUK4
12	DDYCHINA4	·	DGAPSUK8	DGAPSUK8	DGAPSUK8	DDPSUK4
13	DGAPSUK8	·	DWUK4	DWUK4	DWUK4	DGAPSUK8
14	DWUK4	·	·	·	·	DWUK4
$\hat{\lambda}_{hT}$	0.0966	0.1298	0.1020	0.1084	0.0955	0.0941

#	2021q1	2021q2	2021q3	2021q4	2022q1	2022q2
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DVUK4	DVUK4	DVUK4	DDWUK4	DDUUK4	DDUUK4
3	DDUUK	DDUUK	DDUUK	DLRUK4	DLRUK4	DLRUK4
4	DDWUK4	DDWUK4	DDWUK4	DMUK4	DMUK4	DMUK4
5	DLRUK4	DLRUK4	DLRUK4	DPMAT4	DPMAT4	DPMAT4
6	DMUK4	DMUK4	DMUK4	DPMUK4	DPMUK4	DPMUK4
7	DPMAT4	DPMAT4	DPMAT4	DDPMUK4	DDPMUK4	DDPMUK4
8	DDPMETAL4	DDPMETAL4	DDPMUK4	DPSUK4	DPSUK4	DPSUK4
9	DDPMUK4	DDPMUK4	DEPUK4	DDPSUK4	DDPSUK4	DDPSUK4
10	DEPUK4	DEPUK4	DPSUK4	DWUK4	DWUK4	DWUK4
11	DPSUK4	DPSUK4	DDPSUK4	·	·	·
12	DDPSUK4	DDPSUK4	DWUK4	·	·	·
13	DWUK4	DWUK4	·	·	·	·
$\hat{\lambda}_{hT}$	0.1005	0.0955	0.1043	0.1164	0.1127	0.1237

#	2022q3	2022q4	2023q1
1	DEMUK4	DEMUK4	DDUUK4
2	DDUUK4	DDUUK4	DMUK4
3	DLRUK4	DLRUK4	DPMUK4
4	DMUK4	DMUK4	DDPMUK4
5	DPMAT4	DPMAT4	DPSUK4
6	DPMUK4	DPMUK4	DDPSUK4
7	DDPMUK4	DDPMUK4	DWUK4
8	DPSUK4	DPSUK4	·
9	DDPSUK4	DDPSUK4	·
10	DWUK4	DWUK4	·
$\hat{\lambda}_{hT}$	0.1232	0.1252	0.1406

Note: $\hat{\lambda}_{hT}$ is the estimate of the Lasso penalty parameter computed by 10-fold cross-validation.

Table S.12: Selected Variables by ARX-Lasso 2q Ahead

#	Ending Year					
	2019q3	2019q4	2020q1	2020q2	2020q3	2020q4
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DVUK4	DVUK4	DVUK4	DVUK4	DVUK4	DVUK4
3	DDWUK4	DDWUK4	DDWUK4	DDWUK4	DDWUK4	DDUUK4
4	DLRUK4	DLRUK4	DLRUK4	DLRUK4	DLRUK4	DDWUK4
5	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DLRUK4
6	DPMAT4	DPMAT4	DPMAT4	DPMAT4	DDPMUK4	DMUK4
7	DDPMAT4	DDPMUK4	DDPMUK4	DDPMUK4	DEPUK4	DPMAT4
8	DPMUK4	DEPUK4	DEPUK4	DEPUK4	DYCHINA4	DDPMETAL4
9	DDPMUK4	DYCHINA4	DYCHINA4	DYCHINA4	DGAPSUK8	DDPMUK4
10	DEPUK4	DDYCHINA4	DDYCHINA4	DDYCHINA4	DDPUS4	DEPUK4
11	DYCHINA4	DGAPSUK8	DGAPSUK8	DGAPSUK8	·	DYCHINA4
12	DDYCHINA4	DDPUS4	DDPUS4	DDPUS4	·	DGAPSUK8
13	DGAPSUK8	·	·	·	·	DPUS4
14	DDPUS4	·	·	·	·	DDPUS4
15	DWUK4	·	·	·	·	DWUK4
$\hat{\lambda}_{hT}$	0.0757	0.0925	0.0920	0.0920	0.1036	0.0713

#	2021q1	2021q2	2021q3	2021q4	2022q1	2022q2
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DVUK4	DVUK4	DVUK4	DVUK4	DDUUK	DDUUK
3	DDUUK	DDUUK	DDUUK	DDUUK	DDWUK4	DLRUK4
4	DDWUK4	DDWUK4	DDWUK4	DDWUK4	DLRUK4	DMUK4
5	DLRUK4	DLRUK4	DLRUK4	DLRUK4	DMUK4	DPMUK4
6	DMUK4	DMUK4	DMUK4	DMUK4	DPMAT4	DDPMUK4
7	DPMAT4	DPMAT4	DPMAT4	DPMAT4	DPMUK4	DYCHINA4
8	DDPMETAL4	DDPMETAL4	DDPMETAL4	DDPMETAL4	DDPMUK4	DWUK4
9	DDPMUK4	DDPMUK4	DDPMUK4	DDPMUK4	DEPUK4	·
10	DEPUK4	DEPUK4	DEPUK4	DEPUK4	DYCHINA4	·
11	DYCHINA4	DYCHINA4	DYCHINA4	DYCHINA4	DDPUS4	·
12	DPUS4	DPUS4	DPUS4	DPUS4	DWUK4	·
13	DDPUS4	DDPUS4	DDPUS4	DWUK4	·	·
14	DWUK4	DWUK4	DWUK4	·	·	·
$\hat{\lambda}_{hT}$	0.0699	0.0666	0.0739	0.0921	0.0964	0.1092

#	2022q3	2022q4	2023q1
1	DEMUK4	DEMUK4	DEMUK4
2	DDUUK	DDUUK	DDUUK
3	DDWUK4	DDWUK4	DDWUK4
4	RUK4	RUK4	DLRUK4
5	DLRUK4	DLRUK4	DMUK4
6	DMUK4	DMUK4	DPMUK4
7	DPMUK4	DDPMAT4	DDPMUK4
8	DDPMUK4	DPMUK4	DYCHINA4
9	DYCHINA4	DDPMUK4	DWUK4
10	DDPUS4	DYCHINA4	·
11	DWUK4	DDPUS4	·
12	·	DWUK4	·
$\hat{\lambda}_{hT}$	0.1046	0.0974	0.1191

Note: $\hat{\lambda}_{hT}$ is the estimate of the Lasso penalty parameter computed by 10-fold cross-validation.

S-3.2.3 For h=4 quarter ahead models

This sub-section sets out the variables chosen by Lasso for $h = 4$.

Table S.13: Selected Variables by Lasso 4q Ahead

#	Ending Year					
	2019q1	2019q2	2019q3	2019q4	2020q1	2020q2
1	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4
2	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
3	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
4	DPMAT4	DPMAT4	DPMAT4	DPMAT4	DPMAT4	DPMAT4
5	DPSUK4	DPSUK4	DPSUK4	DPSUK4	DPSUK4	DPSUK4
6	DWUK4	DWUK4	DWUK4	DWUK4	DWUK4	DWUK4
$\hat{\lambda}_{hT}$	0.3345	0.3346	0.3823	0.3824	0.3829	0.3610
	2020q3	2020q4	2021q1	2021q2	2021q3	2021q4
1	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4
2	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
3	DPMAT4	DPMAT4	DPMAT4	DPMAT4	DPSUK4	DPMAT4
4	DPSUK4	DPSUK4	DPSUK4	DPSUK4	·	DPSUK4
5	DWUK4	·	DWUK4	·	·	DWUK4
$\hat{\lambda}_{hT}$	0.3874	0.3897	0.4401	0.4394	0.4613	0.4296
	2022q1	2022q2	2022q3	2022q4	2023q1	
1	DPUK4	DPUK4	DPUK4	DPUK4	DPUK4	
2	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	
3	DPMAT4	DPMAT4	DPMAT4	DPMAT4	DPMAT4	
4	DPSUK4	DPSUK4	DPSUK4	DPSUK4	DPSUK4	
5	DWUK4	DWUK4	DWUK4	DPUS4	DPUS4	
6	·	·	·	DWUK4	DWUK4	
$\hat{\lambda}_{hT}$	0.4506	0.4632	0.4669	0.4163	0.4240	

Note: $\hat{\lambda}_{hT}$ is the estimate of the Lasso penalty parameter computed by 10-fold cross-validation.

Table S.14: Selected Variables by AR2-Lasso 4q Ahead

#	Ending Year					
	2019q1	2019q2	2019q3	2019q4	2020q1	2020q2
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
3	DPMAT4	DPMAT4	DPMAT4	DPMAT4	DPMAT4	DPMAT4
4	DDPMUK4	DDPMUK4	DDPMUK4	DDPMUK4	DDPMUK4	DDPMUK4
5	DPSUK4	DEPUK4	DEPUK4	DEPUK4	DEPUK4	DEPUK4
6	DDPSUK4	DPSUK4	DPSUK4	DPSUK4	DPSUK4	DPSUK4
7	DGAPSUK8	DDPSUK4	DDPSUK4	DDPSUK4	DDPSUK4	DDPSUK4
8	·	DGAPSUK8	DGAPSUK8	DDYCHINA4	DDYCHINA4	DDYCHINA4
9	·	·	·	DGAPSUK8	DGAPSUK8	DGAPSUK8
$\hat{\lambda}_{hT}$	0.2357	0.2233	0.2237	0.1999	0.1998	0.2014
	2020q3	2020q4	2021q1	2021q2	2021q3	2021q4
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4	DMUK4
3	DPMAT4	DPMAT4	DPMAT4	DPMAT4	DPMAT4	DPMAT4
4	DDPMUK4	DDPMUK4	DDPMUK4	DDPMUK4	DDPMUK4	DDPMUK4
5	DEPUK4	DEPUK4	DEPUK4	DPSUK4	DPSUK4	DPSUK4
6	DPSUK4	DPSUK4	DPSUK4	DDPSUK4	DDPSUK4	DDPSUK4
7	DDPSUK4	DDPSUK4	DDPSUK4	·	·	·
8	DDYCHINA4	DDYCHINA4	DGAPSUK8	·	·	·
9	DGAPSUK8	DGAPSUK8	·	·	·	·
$\hat{\lambda}_{hT}$	0.1901	0.1902	0.1780	0.2158	0.2170	0.2106
	2022q1	2022q2	2022q3	2022q4	2023q1	
1	DEMUK4	DMUK4	DMUK4	DDUUK	DDUUK	
2	DMUK4	DPMAT4	DPMAT4	DMUK4	DMUK4	
3	DPMAT4	DDPMUK4	DDPMUK4	DPMAT4	DPMAT4	
4	DDPMUK4	DPSUK4	DPSUK4	DDPMUK4	DDPMUK4	
5	DPSUK4	DDPSUK4	DDPSUK4	DPSUK4	DPSUK4	
6	DDPSUK4	DDYSUK4	DWUK4	DDPSUK4	DDPSUK4	
7	·	DWUK4	·	DYUS4	DWUK4	
8	·	·	·	DWUK4	·	
$\hat{\lambda}_{hT}$	0.2735	0.2513	0.2736	0.2493	0.2522	

Note: $\hat{\lambda}_{hT}$ is the estimate of the Lasso penalty parameter computed by 10-fold cross-validation.

Table S.15: Selected Variables by ARX-Lasso 4q Ahead

#	Ending Year					
	2019q1	2019q2	2019q3	2019q4	2020q1	2020q2
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DMUK4	DMUK4	DVUK4	DMUK4	DMUK4	DVUK4
3	DPMAT4	DPMAT4	DMUK4	DPMAT4	DPMAT4	DMUK4
4	DGAPSUK8	DDPMUK4	DPMAT4	DDPMUK4	DDPMUK4	DPMAT4
5	·	DEPUK4	DDPMUK4	DEPUK4	DEPUK4	DDPMUK4
6	·	DGAPSUK8	DEPUK4	DGAPSUK8	DGAPSUK8	DEPUK4
7	·	·	DGAPSUK8	·	·	DDYCHINA4
8	·	·	·	·	·	DGAPSUK8
9	·	·	·	·	·	DPUS4
10	·	·	·	·	·	DDPUS4
$\hat{\lambda}_{hT}$	0.2349	0.2238	0.2020	0.2139	0.2127	0.1579
	2020q3	2020q4	2021q1	2021q2	2021q3	2021q4
1	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4	DEMUK4
2	DVUK4	DVUK4	DVUK4	DMUK4	DVUK4	DVUK4
3	DMUK4	DMUK4	DMUK4	DPMAT4	DMUK4	DMUK4
4	DPMAT4	DPMAT4	DPMAT4	DDPMUK4	DPMAT4	DPMAT4
5	DDPMUK4	DDPMUK4	DDPMUK4	·	DDPMUK4	DDPMUK4
6	DEPUK4	DEPUK4	DEPUK4	·	DEPUK4	·
7	DDYCHINA4	DDYCHINA4	DGAPSUK8	·	DPUS4	·
8	DGAPSUK8	DGAPSUK8	·	·	·	·
9	DPUS4	DPUS4	·	·	·	·
10	DDPUS4	DDPUS4	·	·	·	·
$\hat{\lambda}_{hT}$	0.1471	0.1583	0.1923	0.2166	0.1847	0.1930
	2022q1	2022q2	2022q3	2022q4	2023q1	
1	DEMUK4	DEMUK4	DMUK4	DMUK4	DMUK4	
2	DVUK4	DMUK4	DPMAT4	DPMAT4	DPMAT4	
3	DMUK4	DPMAT4	·	·	·	
4	DPMAT4	DDPMUK4	·	·	·	
5	DDPMUK4	DDYSUK4	·	·	·	
$\hat{\lambda}_{hT}$	0.2017	0.2330	0.2932	0.2719	0.3251	

Note: $\hat{\lambda}_{hT}$ is the estimate of the Lasso penalty parameter computed by 10-fold cross-validation.

S-4 Realizations and Forecasts

S-4.1 h=1 quarter ahead forecasts

Table S.16: One Quarter Ahead Forecast Results

Models	Actual	AR2	ARX	Lasso	Lasso-AR2
2020q1	1.6808	1.4930	1.5005	1.4411	1.1366
2020q2	0.7819	1.9051	1.8708	1.7274	1.2898
2020q3	0.7678	0.7901	0.3358	−1.6184	−1.7781
2020q4	0.7147	1.0091	0.9193	1.0883	0.9987
2021q1	0.9211	0.9482	0.8014	1.2458	1.0556
2021q2	2.0533	1.1989	1.3852	1.7410	1.4759
2021q3	2.6215	2.4556	2.7557	3.1343	3.0982
2021q4	4.3095	2.8118	2.9337	3.1241	3.0738
2022q1	5.3960	4.6293	4.8813	4.8406	5.0132
2022q2	7.6259	5.4458	5.7835	5.8914	6.0747
2022q3	8.3877	7.8385	8.1711	8.0610	8.3580
2022q4	8.9856	8.0723	8.2705	8.6325	8.7409
2023q1	8.6111	8.5777	8.3988	8.7458	9.0455
2023q2		7.9164	7.2945	7.6445	7.8217
RMSFE		0.9141	0.7884	0.9696	0.9617
	Lasso-ARX	OCMT-AR2 $\delta = 1$	OCMT-ARX $\delta = 1$	OCMT-AR2 $\delta = 1.5$	OCMT-ARX $\delta = 1.5$
2020q1	1.3148	1.4930	1.5005	1.4930	1.5005
2020q2	1.8793	1.9051	1.8708	1.9051	1.8708
2020q3	−0.9434	0.7901	0.3358	0.7901	0.3358
2020q4	1.4325	1.0091	0.9193	1.0091	0.9193
2021q1	1.1040	0.9482	0.8014	0.9482	0.8014
2021q2	1.7652	1.1989	1.3852	1.1989	1.3852
2021q3	2.6691	3.0457	2.7557	2.4556	2.7557
2021q4	3.0872	2.8118	2.9337	2.8118	2.9337
2022q1	4.9254	4.5468	4.8813	4.6293	4.8813
2022q2	5.9782	5.4612	5.7835	5.4458	5.7835
2022q3	8.6872	7.8385	8.1711	7.8385	8.1711
2022q4	9.1017	8.0723	8.2705	8.0723	8.2705
2023q1	9.3315	8.5777	8.3988	8.5777	8.3988
2023q2	7.6670	7.9164	7.2945	7.9164	7.2945
RMSFE	0.8750	0.9233	0.7884	0.9141	0.7884

Notes: The "Actual" column gives the realized inflation rate for quarter t, the other columns the forecasts for each model. RMSFE is the Root Mean Square Forecast Error. Lasso models are estimated using the code of Chudik, Kapetanios, and Pesaran (2018).

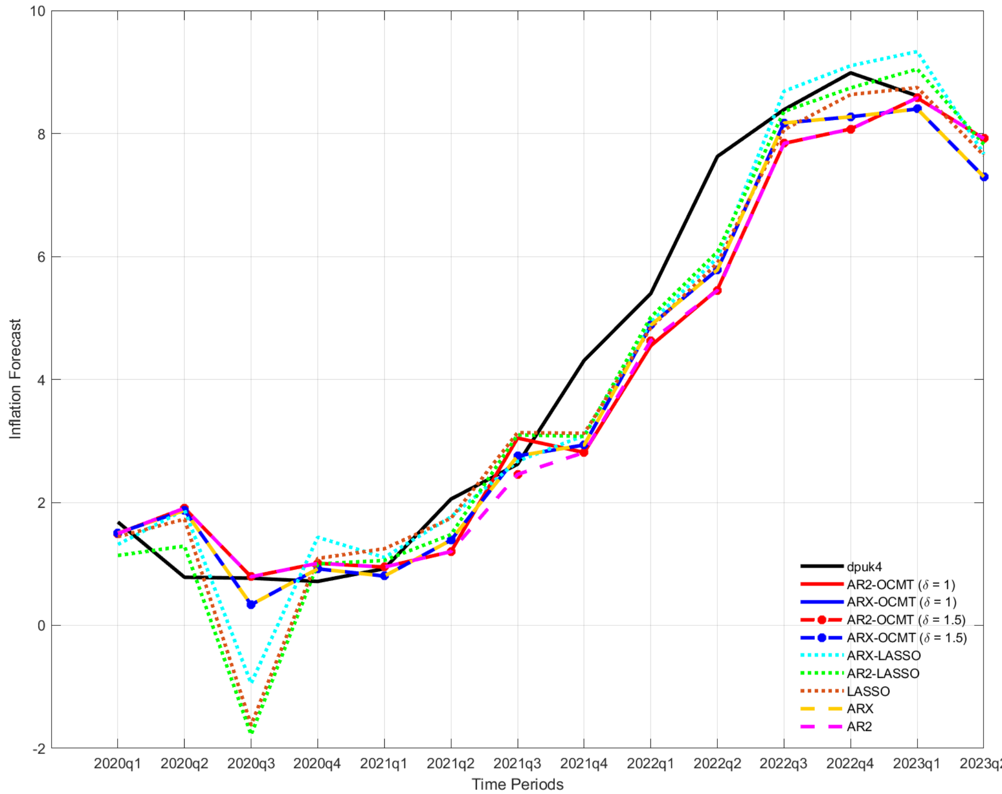

Figure S.1: One Quarter Ahead Forecast Plots

S-4.2 h=2 quarters ahead forecasts

Table S.17: Two Quarters Ahead Forecast Results

Models	Actual	AR2	ARX	Lasso	Lasso-AR2
2020q1	1.6808	1.9854	1.6773	1.7494	1.8994
2020q2	0.7819	1.6115	1.5979	1.4380	0.8428
2020q3	0.7678	2.0705	1.9977	1.8217	2.0103
2020q4	0.7147	0.8941	0.1859	0.5350	–4.7147
2021q1	0.9211	1.2182	1.0423	1.3422	7.7188
2021q2	2.0533	1.1590	0.9049	1.5683	1.6895
2021q3	2.6215	1.4199	1.6725	1.8149	1.9367
2021q4	4.3095	2.6839	3.0931	3.6585	2.0096
2022q1	5.3960	2.9218	3.1026	3.3691	2.2883
2022q2	7.6259	4.7324	5.1079	4.7079	4.9191
2022q3	8.3877	5.3660	5.8973	5.6290	5.9572
2022q4	8.9856	7.7515	8.2645	7.4422	7.9237
2023q1	8.6111	7.7117	8.0551	8.1659	8.4389
2023q2		8.1422	7.8532	7.9158	7.4686
2023q3		7.3340	6.2890	6.1918	5.7513
RMSFE		1.6039	1.3813	1.4109	2.8719
	Lasso-ARX	OCMT-AR2 $\delta = 1$	OCMT-ARX $\delta = 1$	OCMT-AR2 $\delta = 1.5$	OCMT-ARX $\delta = 1.5$
2020q1	1.8915	1.7638	1.4778	1.9854	1.6773
2020q2	1.0312	1.4532	1.5288	1.6115	1.5979
2020q3	2.2963	1.9038	1.8882	2.0705	1.9977
2020q4	–3.5975	1.1311	0.3795	0.8941	0.1859
2021q1	8.4157	1.6797	1.5628	1.2182	1.0423
2021q2	2.1084	1.5683	1.3765	1.6344	1.3765
2021q3	1.9160	1.8149	2.1371	1.8554	2.1371
2021q4	1.9221	3.3262	3.3351	3.5873	3.3351
2022q1	2.2122	3.0279	3.0481	3.1239	3.0481
2022q2	4.0165	4.4689	4.9381	4.4676	4.9381
2022q3	6.3907	5.2246	5.6954	5.2253	5.6954
2022q4	8.4832	7.0244	7.8871	6.9260	7.8871
2023q1	9.1288	6.8038	7.5799	6.9298	8.0551
2023q2	8.0515	7.2864	7.4267	8.1422	7.8532
2023q3	5.8963	6.3556	5.6673	7.3340	6.2890
RMSFE	2.9231	1.6880	1.4217	1.6662	1.4037

Notes: The "Actual" column gives the realized inflation rate at period t, the other columns the forecasts for each model. RMSFE is the Root Mean Square Forecast Error. Lasso models are estimated using the code of Chudik, Kapetanios, and Pesaran (2018).

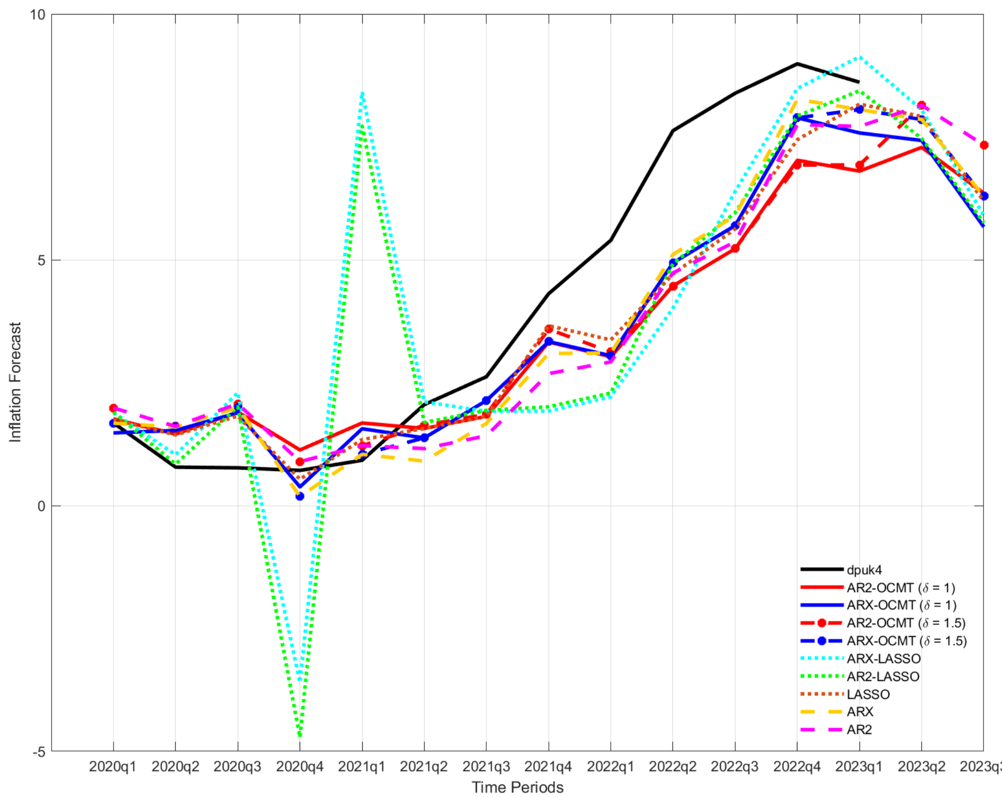

Figure S.2: Two Quarters Ahead Forecast Plots

S-4.3 h=4 quarters ahead forecasts

Table S.18: Four Quarters Ahead Forecast Results

Models	Actual	AR2	ARX	Lasso	Lasso-AR2
2020q1	1.6808	2.3062	1.6093	1.6462	1.8149
2020q2	0.7819	2.3417	2.2183	1.8204	1.9359
2020q3	0.7678	2.2620	1.7780	1.5496	1.4110
2020q4	0.7147	2.0491	2.0380	1.5778	0.8652
2021q1	0.9211	2.1876	2.0822	1.5795	1.4660
2021q2	2.0533	1.6754	0.5406	0.5873	–7.2149
2021q3	2.6215	1.6277	1.3694	2.0021	8.3843
2021q4	4.3095	1.5822	1.1737	2.3696	1.1065
2022q1	5.3960	1.6856	2.1203	2.7836	1.4914
2022q2	7.6259	2.3353	2.9872	3.3813	2.8290
2022q3	8.3877	2.6974	2.9742	2.5152	2.8806
2022q4	8.9856	3.6744	4.2203	3.4427	3.5490
2023q1	8.6111	4.3493	5.1543	4.2743	4.6459
2023q2		5.7823	6.6329	4.7789	5.1198
2023q3		6.2075	6.7753	4.7462	5.4047
2023q4		6.6574	6.2714	4.8914	4.0569
2024q1		6.3149	4.6141	3.7130	2.6772
RMSFE		3.2524	2.9883	3.0131	4.3440
	Lasso-ARX	OCMT-AR2	OCMT-ARX	OCMT-AR2	OCMT-ARX
		$\delta = 1$	$\delta = 1$	$\delta = 1.5$	$\delta = 1.5$
2020q1	1.7270	1.6437	1.3401	1.7873	1.3401
2020q2	1.9359	1.6977	1.9417	1.8175	1.9417
2020q3	1.6059	1.5965	1.4484	1.7514	1.4484
2020q4	0.9909	1.6540	1.9467	1.7858	1.9467
2021q1	–0.1663	1.7455	1.7464	1.8576	1.7464
2021q2	–4.6085	1.7584	0.4562	2.0774	0.4562
2021q3	8.6921	2.2141	1.3667	2.5984	1.3667
2021q4	1.5211	2.0870	0.7506	2.4724	0.7506
2022q1	1.8425	2.2670	1.5652	2.5388	1.5652
2022q2	2.8290	8.3146	2.4832	3.6200	2.4832
2022q3	0.6644	3.5159	2.7159	2.9876	2.7159
2022q4	2.1661	4.2029	3.7504	3.1511	3.7504
2023q1	3.1894	4.8976	4.6815	4.0785	4.7542
2023q2	4.9977	4.5467	5.7679	4.4758	5.7679
2023q3	5.4954	4.6832	5.6688	4.6143	5.6688
2023q4	4.9713	4.7016	5.2734	4.9634	5.2734
2024q1	2.9576	3.6366	3.1912	4.2441	3.1912
RMSFE	4.5021	2.4643	3.2470	2.9795	3.2402

Notes: The "Actual" column gives the realized inflation rate at period t, the other columns the forecasts for each model. RMSFE is the Root Mean Square Forecast Error. Lasso models are estimated using the code of Chudik, Kapetanios, and Pesaran (2018).

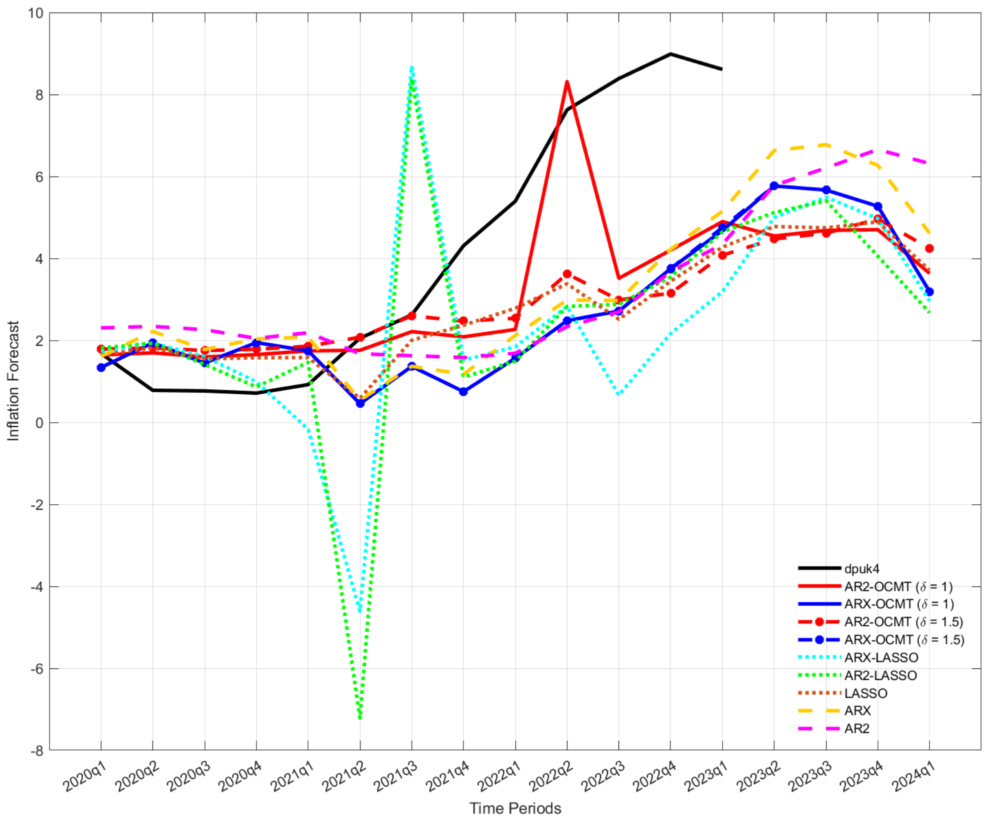

Figure S.3: Four Quarters Ahead Forecast Plots

References

Chudik, Kapetanios, and Pesaran (2018) "A One Covariate at a Time, Multiple Testing Approach to Variable Selection in High-Dimensional Linear Regression Models," *Econometrica*, Vol. 86, No. 4, 1479-1512.