

# Department of Economics

## A Review of Forecasting Techniques for Large Data Sets

Jana Eklund and George Kapetanios

Working Paper No. 625

March 2008

ISSN 1473-0278



# A Review of Forecasting Techniques for Large Data Sets \*

Jana Eklund<sup>†</sup>  
Bank of England

George Kapetanios<sup>‡</sup>  
Queen Mary, University of London

March 14, 2008

## Abstract

This paper provides a review which focuses on forecasting using statistical/econometric methods designed for dealing with large data sets.

**Keywords:** Macroeconomic forecasting, factor models, forecast combination, principal components,

**JEL classification:** C22, C53, E37, E47.

## 1 Introduction

In recent years there has been increasing interest in forecasting methods that utilise large data sets. There is an awareness that there is a huge quantity of information available in the economic arena which might be useful for forecasting, but standard econometric techniques are not well suited to extract this in a useful form. This is not an issue of mere academic interest. Lars Svensson described what central bankers do in practice in Svensson (2005). ‘Large amounts of data about the state of the economy and the rest of the world ... are collected, processed, and analyzed before each major decision.’ In an effort to assist in this task, econometricians began assembling large macroeconomic data sets and devising ways of forecasting with them.

---

\*The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England.

<sup>†</sup>jana.eklund@bankofengland.co.uk

<sup>‡</sup>g.kapetanios@qmul.ac.uk

In the last few years a large number of methods which are either new or new to econometrics has been proposed to deal with forecasting using large data sets. This review aims to provide a brief discussion of the available methods. Given the recent and evolving nature of this literature this review is bound to be incomplete. The need for new methods in the face of the availability of large data sets arises out of the fact that, given time series observations for a large data set, which at time  $t$  is denoted by the  $N$ -dimensional vector  $\mathbf{x}_t$ , it is either inefficient or downright impossible to incorporate  $\mathbf{x}_t$  in a single forecasting model and estimate it using standard econometric techniques.

We assume that primary interest focuses on forecasting a single variable  $y_t$  which may or may not be included in  $\mathbf{x}_t$ . Broadly speaking, the available methodologies for forecasting with large data sets fall into four groups: The first group consists of estimation strategies that allow estimation of a single equation model that utilises the whole of  $\mathbf{x}_t$ . This is perhaps the most diverse group ranging from factor-based methods to Bayesian regression. The methods of the second group involve inherently two steps: In the first step some form of variable selection is undertaken. The variables that are chosen are then most likely to be used in a standard forecasting model. Of course, if the resulting data set is too large, it may still be analysed using methods designed for large data sets. These first two groups of methods inevitably overlap. However, we feel that the step of variable selection is, and involves methods that are, sufficiently distinct to merit separate mention and treatment. The third group of methods involves the use of subsets of  $\mathbf{x}_t$  in distinct forecasting models and the production of multiple forecasts for  $y_t$ , which are then averaged to produce a final forecast. The distinctive feature of this group is the explicit use of model and forecast averaging. Finally, the fourth and perhaps most innovative group of methods departs from the convention of forecasting a single variable. For this group the aim is to forecast the whole of  $\mathbf{x}_t$  (which is now assumed to contain  $y_t$ ). Thus, use of multivariate models is inevitable. As is clear, specially designed estimation methods need to be employed, as the size of the data set,  $\mathbf{x}_t$ , does not allow use of standard econometric techniques.

As the above makes clear, our review will focus on statistical/econometric methods for dealing with large data sets. This ignores the large literature on traditional macro-forecasting

models which usually involve many more variables than traditional econometric models. These more traditional models, typically use economic theory, essentially, to restrict the size of the data set. Therefore, they offer an alternative means of analysing large data sets to the second group of methods discussed above. In practise, policymakers consider both traditional large scale macroeconomic models and statistical models of large data sets in producing forecasts. Typically, statistical models perform better at shorter horizons whereas macroeconomic models perform better at longer horizons and during periods of structural change.

The review is organised as follows: The next four sections deal with each of the above four groups of forecasting methods. Finally, Section 6 concludes.

## 2 Forecasting Using the Whole Data set

This group of methods is, in some sense, the most heterogeneous as it includes a wide variety of approaches in terms of specifying a forecasting equation. The framework here is provided by the following equation

$$y_t = \boldsymbol{\alpha}'\mathbf{x}_t + \epsilon_t \tag{1}$$

The aim is to determine  $\boldsymbol{\alpha}$  without restricting any elements of it to be equal to zero. We view such restrictions to be of sufficiently different nature to be dealt with as part of variable selection methods in the next section. Of course, there are inevitable overlaps which we highlight where appropriate. In the following subsections we present alternative approaches for determining  $\boldsymbol{\alpha}$ .

### 2.1 Ordinary Least Squares Estimation

OLS is the standard estimation method for determining  $\boldsymbol{\alpha}$ . There is little need for further comment on this estimation method apart from observing that as  $N$  increases the estimator's behaviour becomes increasingly erratic and of course estimation of  $\boldsymbol{\alpha}$  is impossible when  $N$  exceeds the number of observations,  $T$ . We note however that estimation of  $\boldsymbol{\alpha}'\mathbf{x}_t$  is possible via the use of generalised inverses even in this case.

## 2.2 Bayesian Regression Estimation

Bayesian regression is a standard tool for providing inference for  $\boldsymbol{\alpha}$  and there exist a large variety of approaches for implementing Bayesian regression. We will provide a brief exposition of this method. A starting point is the specification of a prior distribution for  $\boldsymbol{\alpha}$ . Once this is in place standard Bayesian analysis proceeds by incorporating the likelihood from the observed data to obtain a posterior distribution for  $\boldsymbol{\alpha}$  which can then be used for a variety of inferential purposes, including, of course, forecasting. A popular and simple implementation of Bayesian regression results in a shrinkage estimator for  $\boldsymbol{\alpha}$  given by  $(\mathbf{X}'\mathbf{X} + v\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}$  where  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_T)'$ ,  $\mathbf{y} = (y_1, \dots, y_T)'$  and  $v$  is a shrinkage scalar parameter. As the name suggests this shrinkage estimator shrinks the OLS estimator, given by  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  towards zero, thus enabling a reduction in the variance of the resulting estimator. This is a major feature of Bayesian regression that makes it useful in forecasting when large data sets are available. This particular implementation of Bayesian regression implies that elements of  $\boldsymbol{\alpha}$  are small but different from zero ensuring that all variables in  $\mathbf{x}_t$  are used for forecasting. Other implementations use prior distributions for  $\boldsymbol{\alpha}$  that imply that only a few of the variables in  $\mathbf{x}_t$  are actually used in forecasting, thereby closely relating Bayesian regression to variable selection methods. A recent application of Bayesian regression for forecasting is by De Mol, Giannone, and Reichlin (2007).

## 2.3 Factor Methods

Factor methods have been at the forefront of developments in forecasting with large data sets and in fact started this literature with the influential work of Stock and Watson (2002). The defining characteristic of most factor methods is that relatively few summaries of the large data sets are used in forecasting equations which thereby become standard forecasting equations as they only involve a few variables. Implicitly,  $\boldsymbol{\alpha}'\tilde{\mathbf{x}}_t = \tilde{\boldsymbol{\alpha}}'\boldsymbol{\Lambda}'\tilde{\mathbf{x}}_t$  where  $\tilde{\mathbf{x}}_t$  may be equal to  $\mathbf{x}_t$  or may involve other variables such as, e.g., lags and leads of  $\mathbf{x}_t$ .  $\boldsymbol{\Lambda}'\tilde{\mathbf{x}}_t$  is referred to as the factors. The main difference between different factor methods relate to how  $\boldsymbol{\Lambda}$  is estimated. The following subsections give brief overviews of three factor methods that are available in the literature.

### 2.3.1 Principal Components

The use of principal components for the estimation of factor models is, by far, the most popular factor extraction method. It has been popularised by Stock and Watson (2002), in the context of large data sets, although the idea had been well established in the traditional multivariate statistical literature. The method of principal components is simple. The  $N \times r$  matrix of linear combinations,  $\mathbf{\Lambda}$ , relating to  $r$  factors, is estimated by the eigenvectors corresponding to the  $r$  largest eigenvalues of the second moment matrix  $\mathbf{X}'\mathbf{X}$ . Data are usually normalised to have zero mean and unit variance prior to the application of principal components.

### 2.3.2 Dynamic Principal Components

Principal component estimation of the factor structure is essentially a static exercise as no lags or leads of  $\mathbf{x}_t$  are considered. Dynamic principal components which, as a method of factor extraction, has been suggested in a series of papers by Forni, Hallin, Lippi and Reichlin (see, e.g., Forni, Hallin, Lippi, and Reichlin (2000, 2004) among others) is designed to address this issue. Dynamic principal components are extracted in similar fashion to static principal components but, instead of the second moment matrix, the spectral density matrices of the data at various frequencies are used. These are then used to construct estimates of the common component of the data set which is a function of the unobserved factors. This method uses leads of the data and as a result its application to forecasting has been problematic for obvious reasons. Recent work by the developers of the method has addressed this issue (see, e.g., Forni, Hallin, Lippi, and Reichlin (2005)).

### 2.3.3 Subspace Methods

The third method of factor extraction assumes a parametric state space model for the data set,  $\mathbf{x}_t$ . This follows earlier work by Stock and Watson (1989) who used state space models to extract factors via the Kalman filter and maximum likelihood estimation, for small data sets. Conventional wisdom suggested that such methods would be too computationally intensive for large data sets. Borrowing work from the engineering literature which again focused on small data sets, Kapetanios and Marcellino (2003) suggest using subspace algorithms to estimate factors from a state space model. This essentially uses OLS estimation

to obtain estimates of the matrix coefficient in a multivariate regression of leads of  $\mathbf{x}_t$  on lags of  $\mathbf{x}_t$ . Then, a reduced rank approximation to this estimated coefficient matrix provides estimates for the factors. In a Monte Carlo study Kapetanios and Marcellino (2003) found that subspace estimation compared favourably to static and dynamic principal components.

### 3 Forecasting Using a Subset of the Data set

This section considers a group of methods that involve two steps. In the first step, some form of variable selection reduces the dimensionality of the original large data set to a degree that enables use of a standard forecasting model, which, for our purposes and simplicity, will be assumed to be a standard linear forecasting regression. As the second step is standard, we focus on the first step of variable selection. Note, however, that the second step may be in fact a method designed for large data sets such as those methods discussed in the rest of the sections of this review. Usually, the selection of any subset of a set of variables involves a consideration and evaluation of all possible subsets with a view to selecting the most appropriate subset. Such evaluation is usually carried out using some information criterion such as the Bayesian information criterion. In the context of large data sets this is not possible. To appreciate this, it is worth simply noting that for a set of  $N$  variables there exist  $2^N$  possible subsets. For values of  $N$ , exceeding, say 30 or 40, evaluation of all subsets is not possible with current computing technology. Therefore, in this section we mainly focus on methods that search the space of variable subsets efficiently.

#### 3.1 General-to-Specific Variable Selection

One of the most popular variable selection approaches for regression analysis is the general-to-specific variable selection approach popularised in a number of papers by David Hendry and his co-authors. A useful self contained account of this approach may be found in Hoover and Perez (1999). The main algorithm of that paper provides a tractable formalisation of the general-to-specific methodology advocated by Hendry and his co-authors and discussed in some detail in a number of paper such as, e.g., Hendry (1995, 1997); Krolzig and Hendry (2001) (see also Brüggemann, Krolzig, and Lütkepohl (2003) for an application of this methodology to model reduction in VAR processes). The salient features of the algorithm may be summarised as follows: A general regression specification is considered and

tested for misspecification using a battery of specification tests such as tests for residual autocorrelation and ARCH and tests for structural breaks. Then, a sequential testing procedure is used to remove insignificant regressors from this specification making sure that resulting specifications are acceptable using misspecification tests. Although, this approach is not able to handle very large data sets, since a regression involving all variable needs to be estimated, recent work by Hendry has relaxed this limitation.

### **3.2 Simulated Annealing**

In this and the next subsection, we present two efficient algorithms that search the space of all possible regression specifications and choose the one that optimises a objective function such as, e.g., an information criterion.

The first algorithm is known as simulated annealing and its properties as a variable selection devise have been analysed in Kapetanios (2007) who proposed this approach. Simulated annealing is a generic term used to refer to a family of powerful optimisation algorithms. In essence, it is a method that uses the objective function to create a nonhomogeneous Markov chain that asymptotically converges to the maximum or minimum of the objective function. It is especially well suited for functions defined in discrete spaces like information criteria. The concept is originally based on the manner in which liquids freeze or metals recrystallize in the process of annealing. In an annealing process a melt, initially at high temperature and disordered, is slowly cooled so that the system at any time is approximately in thermodynamic equilibrium. As cooling proceeds, the system becomes more ordered and approaches a ‘frozen’ ground state. The analogy to an optimisation problem is as follows: The current state of the thermodynamic system is analogous to the current solution to the optimisation problem, the energy equation for the thermodynamic system is analogous to the objective function, and the ground state is analogous to the global optimum. Kapetanios (2007) who provides a detailed algorithm and Monte Carlo evidence suggests that this is one of the most promising methods of those reviewed in this section.



### 3.3 Genetic Algorithms

Another group of powerful optimisation methods are genetic algorithms. Genetic algorithms have been used widely for optimising discontinuous and multimodal functions. Genetic algorithms iterate towards a solution through a process reminiscent of the Darwinian process of natural selection. Candidate solutions to the optimisation problem, which appear promising are allowed to thrive whereas less promising solutions are less likely to be considered further. As far as economic applications of the algorithm are concerned, we note the work of Dorsey and Mayer (1995), Marimon, McGratten, and Sargent (1990) and Ostermark (1999). The work of Kapetanios (2007) who proposed this algorithm as a variable selection device suggests that this is also a promising avenue for variable selection.

It is also worth noting that other objective functions may be used in place of information criteria which rely on in-sample fit. For example, out-of-sample forecasting ability in the form of, e.g., root-mean square forecast error, can be used as an objective function. Both genetic algorithms and simulated annealing can be used to optimise such objective functions as discussed in the empirical work of Kapetanios (2007).

### 3.4 Boosting

Boosting refers to a relatively new set of algorithms originating in the machine learning literature. Boosting is, in some sense, the opposite of the general-to-specific methodology for variable selection, that we discussed earlier. A generic boosting algorithm for linear regression may be characterised as follows: The researcher, faced with a set of variables  $\mathbf{x}_t$  to be used for fitting a linear regression for  $y_t$ , starts with a very simple regression model (possibly one containing only a constant). The first step is to estimate univariate regression models containing each of the variables in  $\mathbf{x}_t$  on its own. The variable that minimises a suitably chosen loss function (such as, e.g., the sum of squared residuals) is chosen for inclusion in the final model. Then, the residuals from this univariate regression are treated as a new variable to be explained by the rest of the variables in  $\mathbf{x}_t$ . Once again univariate regressions containing each one of the remaining  $\mathbf{x}_t$  variables are used and the best-fitting variable in  $\mathbf{x}_t$  is chosen next. The process is repeated for either a prespecified number of steps or until some criterion, such as, e.g., an information criterion suggest the process should

stop. This set of algorithms has been recently shown to have good theoretical properties (see, e.g., Buhlmann (2006)).

### **3.5 LASSO and Least Angle Regression**

LASSO and least angle regression (LAR) are algorithms for fitting linear regression models that are closely related to boosting. Starting with LASSO we note that it is a shrinkage estimator and in this sense it is also related to Bayesian regression discussed in Section 2. LASSO minimises the sum of squared residuals as OLS does, but under the restriction that some norm of the estimated vector of regression coefficients is smaller than some prespecified threshold, implying a varying degree of shrinkage. LAR on the other hand works similarly to boosting, but at each step of the algorithm described for boosting in the previous subsection, the chosen variable is not fully included in the regression model but the coefficient associated with this variable is increased as much as is needed for the variable not to be the most correlated (or the one minimising the loss function). Then, the variable is included in the regression model with the required coefficient and the search for a new variable to include starts afresh. LAR is computationally easier than LASSO.

### **3.6 Variable preselection for factor analysis**

In a recent paper Boivin and Ng (2006) note that forecasting using factor analysis may be problematic if the idiosyncratic parts of the series in the data set (the part that is not explained by the factors) are highly correlated with each other. They suggest preselecting a subset of the variables in the large data set so that the resulting subset does not have highly correlated idiosyncratic parts. The subset of series selected is still analysed using factor analysis as it is typically still too large for use in traditional econometric forecasting models. In their empirical analysis, they find that their suggested preselection methods provide improvement in forecasting performance compared to standard factor analysis.

## 4 Forecast Averaging

### 4.1 Bayesian model averaging (BMA)

BMA can be viewed as a Bayesian approach to combination forecasting. Model averaging summarizes the different possible relationships between the predicted variable and the predictor variables. With appropriately chosen weights, this should lead to more efficient extraction of information. Model averaging also has the advantage of providing robustness against misspecification, and model uncertainty can easily be accounted for, if the model averaging is conducted in a Bayesian setting, i.e. the weights are the posterior probabilities of the models. In BMA, the posterior probability associated with the model being correct serves as the weight assigned to each model in the forecast combination. Bayesian model averaging can be used to combine forecasts from the set of models that can be constructed using various combinations of the predictors. The averaging over many different competing models incorporates model as well as parameter uncertainty into conclusions about parameters and predictions. A good reference on recent BMA work is Fernandez, Ley, and Steel (2001). Recently, Eklund and Karlsson (2007) use an out-of-sample measure when combining forecasts and show that the forecast combination with weights based on models' predictive performance outperforms forecast combination with in-sample weights.

### 4.2 Frequentist Model Averaging

As an alternative to Bayesian model averaging, there is a sizable literature, competently summarised by Burnham and Anderson (1998), on a frequentist information theoretic approach in an analogous vein. In this context, information theory suggest ways of constructing model confidence sets. Given the existence of a set of models, relative model likelihood can be defined. Model weights within this framework have been suggested by Akaike in a series of papers (see Akaike (1978, 1979)) and expounded further by Burnham and Anderson (1998). In practical terms such weights are easy to construct using standard information criteria such as Akaike's information criterion. Kapetanios, Labhard, and Price (2007) have considered this way of model averaging as an alternative to Bayesian model averaging for forecasting.

Similarly to the work of Eklund and Karlsson (2007), Kapetanios, Labhard, and Price (2006) use an out-of-sample measure of fit in standard information criteria when constructing weights for forecast combination in an information theoretic approach. They find that the proposed method performs well and, in some respects, outperforms other averaging methods considered.

## 5 Forecasting the Whole Data set

This section draws heavily on the work of Carriero, Kapetanios, and Marcellino (2007) and provides a number of approaches for forecasting large data sets.

### 5.1 Reduced Rank Regression (RR)

The starting point for these models are standard  $VAR(p)$  models. When applied to a large data set, VAR models result in a large number of insignificant coefficients. Therefore, in order to obtain a more parsimonious model, one might impose rank reduction, i.e to assume that  $rk(\mathbf{B}') = r < N$  where  $\mathbf{B}$  is the VAR coefficient matrix. This is equivalent to the parametric specification

$$\mathbf{Y}_t = \boldsymbol{\alpha} \left[ \sum_{i=1}^p \boldsymbol{\beta}'_i \mathbf{Y}_{t-i} \right] = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_t + \mathbf{e}_t, \quad (2)$$

where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_p)'$  are respectively a  $N \times r$  and a  $M \times r$  matrices. The model (2) was studied by Velu, Reinsel, and Wichern (1986). In (2), it is assumed that the true rank of the matrices  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are identical and equal to  $r$  which is thus referred to as the rank of the system (2). Given the assumed system rank  $r$ , Velu, Reinsel, and Wichern (1986) suggested an estimation method for the parameters  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  that may be shown to be quasi-maximum likelihood (see also Reinsel and Velu (1998)).

All the above work assumes implicitly that these models are applied to a relatively small number of variables. Recently, Carriero, Kapetanios, and Marcellino (2007) have suggested that RR models may be useful for forecasting large data sets as a whole. They find that the parsimony imposed by RR models is useful for forecasting.

## 5.2 Bayesian VAR (BVAR)

Factor models and Reduced Rank Regressions are both based on the idea of reducing dimensionality by imposing a structure which summarizes the information contained in a large set of predictors by focussing on some relevant linear combinations of them. An alternative route to obtain a more parsimonious model might be to impose exclusion restrictions on the predictors. However, excluding some variables from a regression is likely to be relatively ad hoc, unless a coherent statistical framework is adopted to do so. Bayesian VAR models and Multivariate Boosting provide a solution to this problem.

Bayesian VAR models work similarly to Bayesian regression but relate to VAR models and as a result are applied to all equations of a VAR model simultaneously. They allow the imposition of restrictions on the data, but also a degree of data dependent coefficient determination. The exclusion restrictions are imposed as priors, so if some a-priori excluded variable turns out to be relevant in the data, the posterior estimate would contain such information. This provides a way of solving the curse of dimensionality problem without resorting to ad-hoc exclusion of some variables.

Bayesian VAR models have been used on relatively small sets of variables but Banburra, Giannone, and Reichlin (2007) and Carriero, Kapetanios, and Marcellino (2007) have applied them to the problem of forecasting large data sets with encouraging results.

## 5.3 Multivariate Boosting (MB)

Multivariate boosting is an extension of boosting as discussed in subsection 3.4. It relates to a multivariate regression which for simplicity we assume to be a VAR model. Starting by setting the VAR coefficient  $\mathbf{B}$  matrix equal to zero, multivariate boosting sets recursively individual coefficients of  $\mathbf{B}$  to non zero values depending on how well these coefficient values explain the whole vector of dependent variables. The main difference between single equation boosting and multivariate boosting relates to the fact that whereas in univariate boosting attention is focused on the fit of a single equation regression model, multivariate boosting looks at measure of multivariate fit. Otherwise the two methods are basically identical. Lutz and Buhlmann (2006) discusses multivariate boosting and provides theoretical results suggesting that this method is applicable to very large data sets. Carriero, Kapetanios, and Marcellino (2007) apply this method to macroeconomic forecasting with mixed results.

## 6 Conclusion

This paper aims to provide a brief and relatively non-technical overview of the state-of-the-art of forecasting with large data sets. We classify existing methods into four groups depending on whether data sets are used wholly or partly, whether a single model or multiple models are used and whether a small subset or the whole data set is being forecast. We mainly provide brief descriptions of the methods and short recommendations where appropriate, without going into detailed discussions of their merits or demerits. Further references provided may help in this respect.

## References

- AKAIKE, H. (1978): “A Bayesian Analysis of the minimum AIC Procedure,” *Annals of the Institute of Statistical Mathematics*, 30.
- (1979): “A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting,” *Biometrika*, 66.
- BANBURRA, M., D. GIANNONE, AND L. REICHLIN (2007): “Bayesian VARS with Large Panels,” *CEPR Discussion Paper No. 6326*.
- BOIVIN, J., AND S. NG (2006): “Are More Data Always Better for Factor Analysis?,” *Journal of Econometrics*, 132, 169–194.
- BRÜGGEMANN, R., H. M. KROLZIG, AND H. LÜTKEPOHL (2003): “Comparison of Model Reduction Methods for VAR Processes,” *Technical Report 2003-W13, Nuffield College, University of Oxford*.
- BUHLMANN, P. (2006): “Boosting for high-dimensional linear models,” *Annals of Statistics*, 34, 559–583.
- BURNHAM, K. P., AND D. R. ANDERSON (1998): *Model selection and inference*. Berlin: Springer Verlag.

- CARRIERO, A., G. KAPETANIOS, AND M. MARCELLINO (2007): “Forecasting Large Datasets with Reduced Rank Multivariate Models,” *Queen Mary, University of London Working Paper No. 617*.
- DE MOL, C., D. GIANNONE, AND L. REICHLIN (2007): “Forecasting using a large number of predictors: is Bayesian regression a valid alternative to principal components?,” *Mimeo, European Central Bank*.
- DORSEY, R. E., AND W. J. MAYER (1995): “Genetic Algorithms for Estimation Problems with Multiple Optima, Nondifferentiability and Other Irregular Features,” *Journal of Business and Economic Statistics*, 13(1), 53–66.
- EKLUND, J., AND S. KARLSSON (2007): “Forecast Combination and Model Averaging Using Predictive Measures,” *Econometric Reviews*, 26(2-4), 329–363.
- FERNANDEZ, C., E. LEY, AND M. F. J. STEEL (2001): “Benchmark Priors for Bayesian Model Averaging,” *Journal of Econometrics*, 100, 381–427.
- FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): “The Generalised Factor Model: Identification and Estimation,” *Review of Economics and Statistics*, 82, 540–554.
- (2004): “The Generalised Factor Model: Consistency and Rates,” *Journal of Econometrics*, 119, 231–255.
- FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2005): “The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting,” *Journal of the American Statistical Association*, 100(471), 830–840.
- HENDRY, D. F. (1995): *Dynamic Econometrics*. Oxford University Press.
- (1997): “On Congruent Econometric Relations: A comment,” *Carnegie-Rochester Conference Series on Public Policy*, 47, 163–190.
- HOOVER, K. D., AND S. J. PEREZ (1999): “Data Mining Reconsidered: Encompassing and the General-To-Specific Approach to Specification Search,” *Econometrics Journal*, 2, 167–191.

- KAPETANIOS, G. (2007): “Variable Selection in Regression Models using Non-Standard Optimisation of Information Criteria,” *Computational Statistics and Data Analysis*, Forthcoming.
- KAPETANIOS, G., V. LABHARD, AND S. PRICE (2006): “Forecasting Using Predictive Likelihood Model Averaging,” *Economics Letters*, 91(3), 373 – 379.
- KAPETANIOS, G., V. LABHARD, AND S. PRICE (2007): “Forecasting using Bayesian and Information Theoretic Model Averaging: An Application to UK Inflation,” *Journal of Business and Economic Statistics*, Forthcoming.
- KAPETANIOS, G., AND M. MARCELLINO (2003): “A Comparison of Estimation Methods for Dynamic Factor Models of Large Dimensions,” *Queen Mary, University of London Working Paper No. 489*.
- KROLZIG, H. M., AND D. F. HENDRY (2001): “Computer Automation of General-to-Specific Model Selection Procedures,” *Journal of Economic Dynamics and Control*, 25(6–7), 831–866.
- LUTZ, R. W., AND P. BUHLMANN (2006): “Boosting for High-Multivariate Responses in High-Dimensional Linear Regression,” *Statistica Sinica*, 16, 471–494.
- MARIMON, R., E. MCGRATTEN, AND T. J. SARGENT (1990): “Money as a Medium of Exchange in an Economy with Artificially Intelligent Agents,” *Journal of Economic Dynamics and Control*, 14, 329–373.
- OSTERMARK, R. (1999): “Solving Irregular Econometric and Mathematical Optimization Problems with a Genetic Hybrid Algorithm,” *Computational Economics*, 13(2), 103–115.
- REINSEL, G. C., AND R. P. VELU (1998): *Multivariate Reduced Rank Regression*. Lecture Notes in Statistics No. 136. Springer: New York.
- STOCK, J. H., AND M. W. WATSON (1989): “New Indices of Coincident and Leading Indicators,” in *NBER Macroeconomics Annual 1989*, ed. by O. J. Blanchard, and S. Fischer. Cambridge, M.I.T. Press.



- (2002): “Macroeconomic Forecasting Using Diffusion Indices,” *Journal of Business and Economic Statistics*, 20, 147–162.
- SVENSSON, L. E. O. (2005): “Monetary Policy with Judgment: Forecast Targeting,” *International Journal of Central Banking*, 1(1), 1–54.
- VELU, R. P., G. C. REINSEL, AND D. W. WICHERN (1986): “Reduced Rank Models for Multiple Time Series,” *Biometrika*, 73, 105–118.

**This working paper has been produced by  
the Department of Economics at  
Queen Mary, University of London**

**Copyright © 2008 Jana Eklund and George Kapetanios  
All rights reserved**

**Department of Economics  
Queen Mary, University of London  
Mile End Road  
London E1 4NS  
Tel: +44 (0)20 7882 5096  
Fax: +44 (0)20 8983 3580  
Web: [www.econ.qmul.ac.uk/papers/wp.htm](http://www.econ.qmul.ac.uk/papers/wp.htm)**