Sparse Seemingly Unrelated Regression Modelling: Applications in Econometrics and Finance

By Hao Wang

Department of Statistical Science, Duke University,
Durham, North Carolina 27708-0251, U.S.A.

hao@stat.duke.edu

Summary

We present a sparse seemingly unrelated regression (SSUR) model to generate substantively relevant structures in the high-dimensional distributions of seemingly unrelated model (SUR) parameters. This SSUR framework includes prior specifications, posterior computations using Markov chain Monte Carlo methods, evaluations of model uncertainty, and model structure searches. Extensions of the SSUR model to dynamic models embed general structure constraints and model uncertainty in dynamic models. A simulated example illustrates the model and highlights questions regarding model uncertainty, searching, and comparison. The model is then applied to three real-world examples in macroeconomics and finance according to which its identified structures have practical significance.

Some key words: Gaussian graphical model; Hyper-inverse Wishart distribution; Marginal likelihood; Markov chain Monte Carlo; Performance evaluation; Variable selection

1 Introduction

This article develops a sparse seemingly unrelated regression (SSUR) model with Gaussian errors; that is, a set of regressions in which both regression coefficients and error precision matrix have many zeros. Zeros in regression coefficients arise when each response possibly only depends on a subset of different predictors; zeros in a precision matrix arise when the error terms satisfy a set of conditional independence restrictions consistent with an underlying graphical model (Whittaker, 1990; Lauritzen, 1996). We study and propose a fully Bayesian analysis of the SSUR model, and provide effective methods for marginal likelihood computation using a specified subset of variables and a specified graphical model to structure the covariance matrix. This enables the simultaneous selections of variables and the covariance matrix as well as comparison of posterior inferences with respect to subsets and conditional independence structures.

Seemingly unrelated regression models (SUR) are frequently used in econometric, financial and sociological modelling (Zellner, 1962, 1971; Box & Tiao, 1973; Srivastava & Giles, 1987). Recent advances in Markov chain Monte Carlo methods enable Bayesian analyses of more complex variations of the SUR model (Chib & Greenberg, 1995; Smith & Kohn, 2000; Griffiths, 2001). We build on prior work in non-sparse Bayesian SUR model analysis and develop MCMC methods for model fitting and computation for the sparse SUR
model. We note that the synthesis of sparse regression models and sparse covariance matrix models has been considered by Cripps et al. (2003) and George et al. (2008), though with a different practical focus than the SUR model. These analysis indirectly model parsimonious covariance matrices on their re-parameterised forms; they address model uncertainty by computing posterior model probabilities without attempting to calculate the marginal likelihood by introducing a model indicator into a list of unknown parameters. To use these methods, one must specify all competing models, and carefully choose some tuning parameters to ensure that the chain mixes well in model space. We directly model the sparse inverse covariance matrix through use of conjugate priors; this leads to an efficient posterior sampling (Carvalho et al., 2007) with marginal likelihood calculated using Monte Carlo methods. Our examples show that these marginal likelihood approximations are adequate and useful in assessing alternative models.

In the context of dynamic SUR models, the graphical modelling of the covariance matrix of multivariate data appears in Carvalho & West (2007a,b) and Wang et al. (2009). Our extension of the dynamic SSUR models generalises this earlier work on the dynamic matrix-variate graphical model; we provide a fully Bayesian inference and model comparison related to both regression coefficient linear equality constraints and error intra-dependencies in the cross-sectional structure of the time series.

2 Model and priors

2.1 The basic model

To introduce the SSUR model, we begin with the usual SUR model. Consider $p$ univariate dependent variables $y_{i,t}$ following individual regressions:

$$
y_{i,t} = X_{i,t}'\beta_i + e_{i,t}, \quad t = 1, 2, \cdots, T
$$

where $X_{i,t}$ is the $n_i$-vector of observations on $n_i$ explanatory variables with possibly a constant term for individual $i$ at time $t$, $\beta_i = (\beta_{i1}, \cdots, \beta_{in_i})$ is a $n_i$-vector of unknown coefficients, and $e_{i,t}$ is a random error. Combine the model as follows: (a) $y_t = (y_{1,t}, \cdots, y_{p,t})'$, the $p \times 1$ observation vector; (b) $X_t = \text{diag}(X_{1,t}, X_{2,t}, \cdots, X_{p,t})$, the $n \times p$ matrix of observations on explanatory variables at time $t$ with $n = \sum_{i=1}^{p} n_i$; (c) $\beta_t = (\beta_{1}', \cdots, \beta_{p}')'$, the $n$-vector of coefficients; and (d) $e_t = (e_{1}, \cdots, e_{p})'$, the $p$-vector of errors distributed as $N(0, V)$. Then the model is

$$
y_t = X_t'\beta_t + e_t, \quad t = 1, 2, \cdots, T.
$$

The SUR model assumes the errors are contemporaneous correlated but not autocorrelated. In other words, let $e = (e_1', \cdots, e_T')'$, the SUR model assumes $\text{cov}(e) = I_T \otimes V$.

2.2 Variable selection in SUR

Sparse regressions involve a set of higher-dimensional regressions with which one models the relationship between $y_i$ and $X_i$ for $i = 1, \cdots, p$, and there is uncertainty about
which subset of $X_i$ to use. Brown et al. (1998) extend the variable selection of multiple regression models to multivariate regression models. However, their model assumes all responses have the same predictors, and thus must generate a subset of predictors appropriate for all responses. The SUR model is broader and employs multivariate regression as a special case.

We now proceed to consider Bayesian approaches to SUR model variable selection. The variable selection problem arises when there is an unknown subset of $X_i$ with regression coefficients so small in predicting $y_i$ that it becomes preferable to ignore them. We let $\gamma = (\gamma_{i1}, \cdots, \gamma_{ip})$ index each of these $2^p$ possible subset choices, where $\gamma_{ij} = 0$ or 1 according to whether $\beta_{ij}$ is small or large, respectively.

The forms of prior distributions for parameters $\beta$ given $\gamma$ is given as $\beta = N(m_0, H_\gamma)$. $m_0$ and $H_\gamma$ must be specified after structuring. One convenient choice of $m_0$ is a zero vector. The class of $H_\gamma$ may take the form $H_\gamma = D_\gamma R_\gamma D_\gamma$, following the univariate regression form of George & McCulloch (1993). Here $D_\gamma$ is a $n \times n$ diagonal matrix and $R_\gamma$ is a correlation matrix. The element of $D_\gamma$ corresponding to $\beta_{ij}$ is $\tau_{ij0}$ when $\gamma_{ij} = 0$ and $\tau_{ij1}$ when $\gamma_{ij} = 1$. Particular considerations about $\tau_{ij0}, \tau_{ij1}$ and $R_\gamma$ are discussed by George & McCulloch (1993, 1997). One convenient choice for $R$ is $I$, under which the elements of $\beta$ are a priori independent. Another choice is a block diagonal matrix in which each block corresponds to the covariance matrix of the $n_i$-vector of $\beta_i$.

2.3 Structured covariance matrix

The role of covariance matrix $V$ in inferring $\beta$ is one of the most important features of SUR models. The non-diagonality of the error covariance matrix usually entails that individual regression estimates using univariate linear model are sub-optimal; joint estimations of SUR that exploits the correlation between errors across equations may improve the level of model inference. Motivation for our work relates to the increased dimension and complexity of the error covariance matrix. In this context, the covariance matrix must be understood in terms of structure and parsimony.

Substantial progress has been made on Bayesian covariance modelling by imposing structures. Structures are typically obtained by restricting the elements of a reparameterisation for $V$ (Daniels & Poursalhadi, 2002; Smith & Kohn, 2002; Chen & Dunson, 2003). Although a conditional conjugate posterior can usually be obtained on reparameterised space, the main issue in these approaches is that the induced prior on $V$ depends on the ordering of elements. We directly and parsimoniously model $V$ by considering its restrictions induced by graphical model structuring.

The terminology for graphical models, which can be found in Lauritzen (1996); Jones et al. (2005), and Carvalho & West (2007a), is defined as follows. Consider the error vector $e_t$ where $e_t \sim N(0, V)$ with precision $\Lambda = V^{-1}$, where the $ij$-th element is $\lambda_{ij}$. Then an undirected graph $G$ on nodes $\{1, \cdots, p\}$ has edges between pairs of indices $(i, j)$ for which $\lambda_{ij} \neq 0$; $\Lambda$ has off-diagonal zeros corresponding to conditional independencies among elements in $e_t$. We focus here on decomposable graphs $G$. The theory of graphical models can be applied here to define conditional factorizations of
SUR model density over graphs. For notational clarity, we suppress subscript \( t \). Now for any graph \( G \), we have

\[
p(y \mid X, \beta, V, G) = \prod_{P \in \mathcal{P}} p(y_P \mid X_P, \beta_P, V_P) \left/ \prod_{S \in \mathcal{S}} p(y_S \mid X_S, \beta_P, V_S) \right. \tag{3}
\]

where \( \mathcal{P} \) is the set of complete prime components, or cliques, of \( G \) and \( \mathcal{S} \) is the set of separators. For each subgraph \( g \in \{\mathcal{P}, \mathcal{S}\} \), \( y_g \) is \( |g| \)-vector defined as \( y_g = \{y_i : i \in g\}' \), \( X_g \) is the corresponding design matrix defined as \( X_g = \text{diag}\{X_i : i \in g\} \), \( \beta_g = \{\beta_i' : i \in g\}' \), and \( V_g \) the corresponding sub-matrix of \( V \). Each term in Eq. (3) is multivariate normal, \( y_g \sim \mathcal{N}(X_g'\beta_g, V_g) \) with \( \Lambda_g = V_g^{-1} \) having no off-diagonal zeros.

Hyper-inverse Wishart priors are conjugate for covariance matrices in multivariate normal graphical models (Dawid & Lauritzen, 1993). On decomposable graphs, the implied priors on sub-covariance matrices on all components and separators are inverse Wishart. Using the hyper-inverse Wishart notation from Giudici & Green (1999) and Jones et al. (2005), we assume the prior for \( V \) is \( \text{HIW}(d, D) \) with the following density function

\[
p(V) = \prod_{P \in \mathcal{P}} p(V_P \mid b, D_P) \left/ \prod_{S \in \mathcal{S}} p(V_S \mid d, D_S) \right. \]

where each component is an inverse Wishart density.

### 3 Posterior and Marginal Likelihood Computation

#### 3.1 Gibbs sampling on a given graph and variable index

Assume a SSUR model of Section 2.1 and priors of Section 2.2 and 2.3, and write \( Y = (y_1, \cdots, y_T) \) for the full set of data. It is easy to see that, on any specified graph \( G \) and index vector \( \gamma \), the posterior \( p(\beta, V \mid Y) \) has conditionals:

\[
(V \mid \beta, Y) \sim \text{HIW}_{G}\{b + T, D + \sum_{t=1}^{T} (y_t - X'_t\beta)(y_t - X'_t\beta)'\}, \tag{4}
\]

\[
(\beta \mid V, Y) \sim \mathcal{N}\{C(\sum_{t=1}^{T} X_tV^{-1}y_t + H^{-1}m_0), C\} \tag{5}
\]

where \( C = \{\sum_{t=1}^{T} X_tV^{-1}X'_t + H^{-1}\}^{-1} \) These form the basis of an efficient Gibbs sampler to generate from the full posterior \( p(\beta, V \mid Y) \). The Gibbs iterates involve sampling from the hyper-inverse Wishart and multivariate normal distribution. Simulation of the former is based on Carvalho et al. (2007).
3.2 Marginal likelihood approximation

Exploration of uncertainty about regression and graphical structures involves consideration of the marginal likelihood function over structures; namely

\[
p(Y) \equiv p(Y|\gamma, G) = \int p(Y|\beta, V)p(\beta)p(V)d\beta \, dV
\]

over \((\gamma, G)\); the priors in the integrand depend on the index vectors and graphs although we drop that in the notation for clarity. The integral cannot be evaluated but we can generate useful approximations via use of Candidate’s formula (Besag, 1989; Chib, 1995). Write \(\Theta = \{\beta, V\}\) for all parameters, and suppose that we can evaluate \(p(\theta | Y)\) for some subset of parameters \(\theta \in \Theta\); Candidate’s formula gives the marginal likelihood via the identity \(p(Y) = p(Y | \theta)/p(\theta | Y)\). Applying this requires that we estimate components of the numerator or denominator. Choosing \(\theta\) to maximally exploit analytic integration is key, and different choices that integrate over different subsets of parameters will lead to different, parallel approximations of \(p(Y)\) that can be compared. We use (A): \(\theta = V\), and (B): \(\theta = \beta\), giving two approximations based on marginalisation over desirably disjoint parameter subsets. Other choices might be considered though with less analytic tractability.

The marginal likelihood is theoretically given by each of

(A) \(p(Y) = p(Y,V)/p(V | Y)\) at any chosen value of \(\theta = V\), and

(B) \(p(Y) = p(Y,\beta)/p(\beta | Y)\) at any value of \(\theta = \beta\).

We estimate the components of these equations that have no closed form, then plug-in chosen values \(\beta, V\) such as approximate posterior means, to provide two estimates of \(p(Y)\). For (A), the numerator terms, \(p(Y,V) = p(Y | V)p(V)\), are each easily computed at any \(V\). The denominator term may be approximated by

\[
p(V | Y) = \int p(V | Y, \beta)p(\beta | Y)d\beta \approx \frac{1}{M} \sum_{j=1}^{M} p(V | Y, \beta^{(j)})
\]

where the sum is over posterior draws \(\beta^{(j)}\); this is easy to compute as it is a sum of the product of hyper-inverse Wishart densities. For (B), the numerator can be analytically evaluated as \(p(Y | \beta)\). The density function in the denominator is approximated as

\[
p(\beta | Y) = \int p(\beta | V,Y)p(V | Y)dV \approx \frac{1}{M} \sum_{j=1}^{M} P(\beta | Y, V^{(j)})
\]

where the sum over posterior posterior draws \(V^{(j)}\) can be easily performed, with terms given by normal density evaluations.
### 3.3 Model space priors for variable selection and graphs

For the model space priors, we use beta-binomial priors for both variable space and graphical model space. The prior probability for a subset of variables is given by

\[
p(\gamma | w_\gamma) = \prod_{i=1}^{p} w_{\gamma,i}^{k_{\gamma,i}} (1 - w_{\gamma,i})^{(n_i - k_{\gamma,i})}
\]

and each variable inclusion probability for regression \(i\) has a beta prior \(w_{\gamma,i} \sim \text{Be}(a, b)\). This structure yields

\[
p(\gamma) = \prod_{i=1}^{p} \text{B}(a + k_{\gamma,i}, b + n_i - k_{\gamma,i}) / \text{B}(a, b)
\]

where \(\text{B}(a, b)\) is the beta function. The prior on graphical model space is

\[
p(G | w_G) = w_G^{k_G} (1 - w_G)^{(m - k_G)}
\]

and \(w_G \sim \text{Be}(c, d)\), for a graph \(G\) having \(k_G\) edges out of \(m = \frac{2p(p-1)}{2}\) possible ones. The default uniform priors on \(w_{\gamma,i}\)'s and \(w_G\) imply a marginal priors

\[
p(\gamma) = \prod_{i=1}^{p} \frac{k_{\gamma,i}!(n_i - k_{\gamma,i})!}{(n_i + 1)n_i!} = \prod_{i=1}^{p} \frac{1}{n_i + 1} \left(\frac{n_i}{k_{\gamma,i}}\right)^{-1}
\]

and

\[
p(G) = \frac{k_G!(m - k_G)!}{(m + 1)m!} = \frac{1}{m + 1} \left(\frac{m}{k_G}\right)^{-1}
\]

Our choice of the model space prior is based on the consideration that the fully Bayesian priors have automatic adjustment for multiple testing as the numbers of possible variables and edges grow (Scott & Carvalho, 2008; Scott & Berger, 2008).

### 4 Stochastic model search

#### 4.1 Direct Metropolis-Hastings-within-Gibbs algorithms

We now extend Markov chain Monte Carlo for variable selection (George & McCulloch, 1993; Geweke, 1996; George & McCulloch, 1997; Madigan & York, 1995; Raftery et al., 1997; Brown et al., 1998) and multivariate graphical models (Giudici & Green, 1999; Jones et al., 2005) to learning on \((\gamma, G)\) in the above SSUR analysis. This relies on the computation of the unnormalised posterior over graphs, \(p(\gamma, G | Y) \propto p(Y | \gamma, G)p(\gamma, G)\), for any specified model \((\gamma, G)\). Jones et al. (2005) discuss performance of various stochastic search methods in single multivariate graphical models; for modest dimensions, they recommend simple local-move Metropolis-Hastings. Here, given a current pair \((\gamma, G)\), we can apply local moves in \(G\) space based on the conditional posterior \(p(G | Y, \gamma)\), and vice-versa. A candidate \(G'\) is sampled from a proposal distribution \(q(G'; G)\) and accepted with probability

\[
\alpha = \min\{1, p(G' | Y, \gamma)q(G; G')/p(G | Y, \gamma)q(G'; G)\};
\]

our examples use the simple random add/delete edge move proposal of (Jones et al., 2005). We then couple this with a similar step using \(p(\gamma | Y, G)\) at each iteration. This requires a Markov chain analysis on each variable and graph pair visited in order to evaluate marginal likelihood, so implying a substantial computational burden.

#### 4.2 Indirect Metropolis-Hastings-within-Gibbs sampling algorithms

We can also simulate \(G, V, \beta\) and \(\gamma\) without eliminating the values of parameters \(\beta\) and \(V\) using the following Metropolis-Hastings-within-Gibbs sampler:
(a) \((G \mid Y, \beta, \gamma) \propto H(b, D, G) / H\{b + T, D + \sum_{t=1}^{T} (y_t - X_t'\beta)(y_t - X_t'\beta)'\}, G\) is sampled through local move Metropolis-Hastings algorithm.

(b) \((V \mid Y, \beta, G, \gamma)\) is the same as in Eq. (4)

(c) \((\beta \mid Y, V, \gamma, G)\) is the same as in Eq. (5)

(d) \((\gamma_i \mid Y, \gamma_{-i}\beta, V, G) \sim Bern\{u_{i1}/(u_{i0}+u_{i1})\}\), where \(u_{i0} = p(\beta \mid \gamma_{-i}, \gamma_i = 1)p(\gamma_{-i}, \gamma_i = 1)\), \(u_{i0} = p(\beta \mid \gamma_{-i}, \gamma_i = 0)p(\gamma_{-i}, \gamma_i = 0)\), and \(\gamma_{-1} = (\gamma_1, \cdots, \gamma_{i-1}, \gamma_{i+1}, \cdots, \gamma_n)\).

This indirect model search algorithm does not require Markov chain analysis for marginal likelihood approximations at each step, and hence it is much faster than the direct search. However, it is hard to assess whether such stochastic search can avoid becoming stuck in a posterior mode in which case the use the empirical frequency to represent model evidence is less informative as compared with marginal likelihoods. Nevertheless, an initial run of this faster indirect sampling method could provide some useful \(\gamma\) and \(G\) values in order to start the direct model search based on the marginal likelihood.

5 Example: A simulated random sample

A sample of size \(T = 60\) was drawn from a \(p = 6\) SUR model,

\[
\begin{align*}
y_1 &= 1.3x_1 - 0.5x_3 + e_1 \\
y_2 &= 0.9x_1 - 0.3x_2 + 0.5x_3 + e_2 \\
y_3 &= x_1 + 0.5x_2 + 0.7x_3 + e_3 \\
y_4 &= 0.8x_4 - 0.6x_5 + e_4 \\
y_5 &= x_4 + 0.7x_5 + e_5 \\
y_6 &= 1.1x_4 - 0.6x_5 + e_6
\end{align*}
\]

where \(x_i\)'s are draws from i.i.d. \(N(0, 1)\), and the error covariance matrix is the autocovariance matrix of a stationary \(AR(1)\) process with AR parameter 0.6 and innovation variance 1. To perform variable selection, we added six noisy variables to each regression equation, and so \(n = 51\).

First consider an analysis on the true subset of variables and the graph under priors with \(m_0 = 0, \tau_0 = 0.01, \tau_1 = 10, b = 3\) and \(D = 0.0001I_6\). Convergence is rapid and apparently fast-mixing in this simulation as well as in other simulated examples. Parallel checking for assessing the dual approximation of marginal likelihood, in Fig. 1, shows an implementation check and illustrates the concordance of the two, parallel marginal likelihood estimates; these are very close and differ negligibly on the log probability scale. When compared with each other, method (A) generates more stable estimates across differing Monte Carlo sample sizes than method (B). This is probably because the posterior standard error of \(V\) is greater than that of \(\beta\).
Figure 1: Log-marginal likelihood values on the true model in the simulation example of Section 5. The two estimates (solid line: method (A); dashed line: method (B)) of Section 3.2, were successively re-evaluated at differing simulation sample sizes. The plot confirms the concordance even at low samples sizes, and suggests accuracy in terms of differences on the log-likelihood scale.

Consider model uncertainty with model space priors in Section 3.3. I first ran an indirect stochastic search sampler with 10000 full iterations starting with the full model and $\beta = 0$. The median probability model of $\gamma$ is the true subset of variables excluding $x_2$ and $x_3$ in the regression of $y_2$, and the median probability model of $G$ is the true underlying band diagonal graph. Repeat explorations suggest stability in the marginal likelihood estimation when smaller Monte Carlo sample sizes are used, and we use 2,000 draws within each step of the model search. The direct add-delete Metropolis-within-Gibbs was run for 5,000 iterations starting with the median probability model found in the initial indirect search. The most probable model visited, $(\hat{\gamma}, \hat{G})$, is the true subset of variables and true underlying graph; these are local modes and also have the largest marginal likelihood. This model was first visited after 203 direct Markov chain steps.

We also estimated the variable and edge inclusion probability using the top 30 models identified. The variables and edges in the modes $(\hat{\gamma}, \hat{G})$ generally have higher posterior inclusion probability than those not included; the lowest probability included variable and edge have probability 0.56 and 0.62 respectively, while the highest probability excluded variable and edge have probability 0.07 and 0.02 respectively. Thus, models discovered by highest posterior probability and by aggregating high probability models are not dramatically different. Further, the approximate posterior mean of the proportion of variables and edges, a measure of sparsity, are about 26%, 29% for $\gamma, G$ respectively.
Example: Relations among stock returns, interest rates, real activity, and inflation

In this example, we use a special form of the SUR model, the vector autoregressive model (VAR), to investigate the relations and dynamic interactions among stock returns, interest rates, real activity, and inflation in the postwar United States. The data are monthly real stock returns (SRE), real interest rates (IRE), industrial production growth (IPG) and inflation rates (INF). Real returns (SRE and IRE) are computed as nominal returns less the expected inflation rate. For comparison with the results in Lee (1992), the sample period for this study is from January 1947 to December 1987; the data appear in Fig. 2.

Figure 2: Monthly data on real stock returns (SRE), real interest rates (IRE), industrial production growth rates (IPG) and inflation rates (INF). The data set consists of 492 monthly rates for each of these four time series, over the period of 41 years: 01/1947 to 12/1987.

We choose a lag length of 6 months for the VAR model. For variable selections, we use the default semiautomatic priors $\tau_{ij0} = 1/10\hat{\sigma}_{ij}$ and $\tau_{ij1} = 10\hat{\sigma}_{ij}$, where $\hat{\sigma}_{ij}$ is the standard error associated with the unconstrained generalised least squares estimate of $\beta_{ij}$. For covariance selection, we use the flat prior $b = 3$ and $D = 0.0001I_4$. The initial indirect stochastic search was run for 10000 steps, followed by a 10000 step run of direct stochastic search using marginal likelihood approximation based on 2000 Monte Carlo draws within each step. The marginal likelihood allows us to compute the exact relative probabilities using $p(\gamma, G \mid Y)$. The relative probabilities of the 200 most probable models are displayed in order in Fig. 3. This relative probability distribution...
is rather peaked, suggesting that a small subset of models are far more promising than others.

The following most probable models were reported with fitted VAR; the posterior standard errors appear in parentheses:

\[
\begin{align*}
    y_{1,t} &= 0.248(0.044)y_{1,t-1} + e_1 \\
    y_{2,t} &= 0.890(0.020)y_{2,t-1} - 0.076(0.011)y_{4,t-1} + e_2 \\
    y_{3,t} &= 0.002(0.001) + 0.360(0.041)y_{3,t-1} + 0.047(0.014)y_{1,t-2} + 0.058(0.014)y_{3,t-3} + e_3 \\
    y_{4,t} &= 0.001(0.0002) + 0.430(0.045)y_{4,t-1} + 0.184(0.045)y_{4,t-2} + 0.157(0.040)y_{4,t-5} + e_4.
\end{align*}
\]

The estimated error adjacency matrix and covariance matrix are as follows:

\[
\begin{pmatrix}
    1 & 1 & 0 & 0 \\
    1 & 1 & 0 & 1 \\
    0 & 0 & 1 & 0 \\
    0 & 1 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
    0.00113 & -4.86 \times 10^{-6} & 0 & -1.78 \times 10^{-6} \\
    -4.86 \times 10^{-6} & 7.75 \times 10^{-7} & 0 & 2.84 \times 10^{-7} \\
    -1.78 \times 10^{-6} & 2.84 \times 10^{-7} & 0.000105 & 0 \\
    8.35 \times 10^{-6} & 0 & 0 & 8.35 \times 10^{-6}
\end{pmatrix}.
\]

Real stock returns strongly signal positive first lag autocorrelation. Real interest rates appear to be highly positively autocorrelated, and are lead by inflation rates with a negative sign. Industrial product growth is positively autocorrelated, and is lead by real stock returns. Finally, inflation is also positively autocorrelated. The estimated error graph suggests that real stock returns are conditionally independent of inflation rates given real interest rates.
Figure 4: Impulse response of real stock returns (SRE), real interest rates (IRE), real industrial production growth (IPG) and inflation rate (INF) to shocks in each variable.
We further address the relations and interactions among these four variables by examining the percentage of 24-month orthogonal forecast error variance explained by innovations in each variable as shown in Table 1 as well as the impulse response functions displayed in Fig. 4.

1. Real stock returns and real activity: Real stock returns appear to explain a substantial fraction (8.54%) of the variance in real activity, which responds positively to shocks in stock returns. Figure 4 shows that the response of industrial growth to shocks in real stock returns is significantly positive, peaks after five month and becomes negligible after ten months. This observation confirms the view that the stock market signals changes in real activity, and this correlation between stock returns and real activity is positive (Fama, 1981; Geske & Roll, 1983; Lee, 1992).

2. Real stock returns and inflation: Real stock returns fail to Granger-cause inflation rates, since all of the coefficients on the lagged values of stock returns are zeros in the equation for inflation rates. Furthermore, the fourth column of Table 1 suggests that innovation in real stock returns explains none of the forecast error variance of inflation. This finding is generally compatible with the view that the negative observed relations between stock returns and inflation rates might be a proxy for other possible macroeconomic relations (Fama, 1981; Geske & Roll, 1983). In addition, Figure 4 suggests there is no signal of a consistent negative response of inflation to shocks in stock returns.

3. Real interest rates and inflation: Contrary to the findings presented in Lee (1992), there is no indication that real interest rates Granger-cause inflation in this data set. Moreover, as is shown by column 4 in Table 1, innovations in real interest rates do not explain any of the forecast variance of inflation. However, inflation appears to explain a substantial fraction (31.8%) of forecast error variance of real interest rates.

4. Inflation and real activity: Table 1 indicates that inflation only has negligible explanatory power (0.00003%) for real activity in the presence of real stock returns. Figure 4 shows that this weak relation between inflation and real activity is negative.

7 Mutual fund performance

7.1 Alpha and the SUR model

A mutual fund’s historical performance can be summarised by estimating its alpha. This term is defined as the intercept in a regression of the fund’s excess return on the excess return of one or more passive benchmarks. It is usually estimated by applying an ordinary least square analysis to the regression

\[ y_{0,t} = \alpha_0 + x_{0,t}' \beta_0 + e_{0,t}, \quad t = 1, 2, \ldots, T \]
where $y_{0,t}$ is the fund’s return at time $t$, $x_t$ is a $k \times 1$ vector of benchmark returns at time $t$, and $\alpha_0$ is the fund’s alpha. The choice of benchmarks is often guided by a pricing model, such as the capital asset pricing model (CAPM) (Sharpe, 1964; Lintner, 1965; Mossin, 1966) and the Fama-French three factor model (Fama & French, 1993). The recent work of Pástor & Stambaugh (2002) has explored the role of nonbenchmark passive assets in estimating a fund’s alpha using a seemingly unrelated regression model. Suppose there are $p$ nonbenchmark passive returns $y_{i,t}$ besides the $k$ benchmark returns $x_{i,t}$. Then the SUR model used to estimate the mutual fund’s $\alpha_0$ is written as

$$
\begin{align*}
    y_{0,t} &= \alpha_0 + x_t' \beta_0 + e_{0,t} \\
    y_{i,t} &= \alpha_i + x_t' \beta_i + e_{i,t}, i = 1, \cdots, p
\end{align*}
$$

where $e_t = (e_{0,t}, e_{1,t}, \cdots, e_{p,t})$ is correlated contemporaneously and not autocorrelated. The basic idea is that a more precise estimate of $\alpha_0$ is provided through a more precise estimate of $\alpha_i$ when $e_{0,t}$ is correlated with $e_{i,t}$’s for all $i = 1, \cdots, p$. Note that many mutual funds have relatively short histories as compared with passive assets. Given the more accurate estimate of $\alpha_i (i = 1, \cdots, p)$ computed from a longer sample period, the $\alpha_0$ estimated from a SUR model is more precise than the $\alpha_0$ estimated solely based on a single regression model.

### 7.2 Alpha and the SSUR model

Some interesting questions arise in evaluating mutual fund performance using SUR models. First, as is observed by Pástor & Stambaugh (2002), the assumption of pricing power of benchmark assets on nonbenchmark assets is critical in estimating a fund’s $\alpha$ in a SUR model. In particular, if in each case the benchmark assets are assumed to have no pricing ability on the nonbenchmark assets, i.e. $\alpha_i \neq 0 (i = 1, \cdots, p)$, then the estimate of a $\alpha_i$ from a longer sample period is more precise than the estimate of $\alpha_i$’s from the same period of a mutual fund’s available history. Given the correlation between $e_{i,t}$ and $e_{0,t}$, the same can be said of the estimate of $\alpha_0$ based on the SUR model relative to the estimate of $\alpha_0$ from a single regression. Otherwise, if benchmark assets price other nonbenchmark assets, i.e. $\alpha_i = 0 (i = 1, \cdots, p)$, then the better performance of an estimate of $\alpha_0$ based on the SUR model as compared to that based on a single regression is attributed to additional information about sampling error provided by the seemingly unrelated regressions of nonbenchmark assets. Pástor & Stambaugh (2002) address the assumption of pricing power by separately applying SUR models to such situations. However, as is shown below, within the SSUR framework, the uncertainty about the pricing power of benchmark assets on nonbenchmark assets can be incorporated naturally. The second interesting question concerns the strictness of the SUR model assumption, that is, a fund’s return is assumed to be contemporaneously correlated with all nonbenchmark returns given the benchmark returns. For certain types of managed funds, perhaps only the errors from a subset of nonbenchmark assets are relevant in explaining a fund’s returns. Including too many correlated nonbenchmark
assets to estimate a fund’s alpha will mean a potentially high misspecification risk. Hence the possibility that a SUR model can account for the subset of nonbenchmark assets correlated with a fund is very compelling.

We also note that a fund’s history is very likely to be shorter than the passive assets. In order to extend the basic SSUR model to allow one equation to have fewer observations than the others, we re-parameterise the models in Eq. (7). Suppose returns on passive assets including benchmark or nonbenchmark assets are constructed for the period from 1 to \( T \) and a mutual fund only has a history from \( t_0 \) to \( T \) where \( t_0 \geq 1 \). Notice that \( e_{0,t} = \sum_{i=1}^{p} \epsilon_{i,t}\theta_i + \epsilon_{0,t} = \sum_{i=1}^{p}(y_{i,t} - \alpha_i - x'_i\beta_i)\theta_i + \bar{e}_{0,t} \) if the errors are correlated contemporaneously. Eq. (7) can then be rewritten as

\[
y_{0,t} = \tilde{\alpha}_0 + x'_t\tilde{\beta}_0 + \sum_{i=1}^{p} y_{i,t}\theta_i + \bar{e}_{0,t} \quad t = t_1, \ldots, T
\]

\[
y_{i,t} = \alpha_i + x'_i\beta_i + \bar{e}_{i,t} \quad i = 1, \ldots, p, t = 1, \ldots, T
\]

where \( \tilde{\alpha}_0 = \alpha_0 - \sum_{i=1}^{p} \alpha_i \theta_i \), \( \tilde{\beta}_0 = \beta_0 - \sum_{i=1}^{p} \beta_i \theta_i \) and \( \bar{e}_{0,t} \sim N(0, \bar{\sigma}^2) \) is uncorrelated with the error vector \( (\epsilon_{1,t}, \ldots, \epsilon_{p,t}) \), which is distributed as \( N(0, \Sigma) \). We further assume throughout this section that the benchmark assets \( x_t \) are included in every possible model. For Eq. (8), the models for different subsets of nonbenchmark assets may be represented by a vector of binary variables, \( \gamma = (\gamma_{00}, \gamma_{01}, \ldots, \gamma_{0p})' \), where \( \gamma_{ij} \) is an indicator of the inclusion of intercept \( \tilde{\alpha}_0 \), when \( j = 0 \), or nonbenchmark asset \( y_j \), when \( j \geq 1 \). For the \( p \) equations in Eq. (9), we index each of the possible benchmark assets’ pricing abilities by \( \gamma = (\gamma_1, \ldots, \gamma_p)' \) where \( \gamma_i = 0 \) or 1 according to whether \( \alpha_i \) is small or large, respectively. We use \( G \) to denote the graph underlying \( \Sigma \), which is the error covariance matrix of nonbenchmark passive assets.

Two interesting questions can now be addressed by incorporating model uncertainty regarding the choice of the triple \((\gamma, \tilde{G}, \bar{\sigma})\), which is denoted by \( M_k \) for one particular specification. Define \( \Theta_0 = (\tilde{\alpha}_0, \tilde{\beta}_0, \theta_1, \ldots, \theta_p, \bar{\sigma}^2) \), \( \Theta_1 = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_p, \Sigma) \), \( Y_0 = (y_{0,t_1}, \ldots, y_{0,T}) \), and \( Y_1 = (y_{1,1}, \ldots, y_{1,T}, \ldots, y_{p,1}, \ldots, y_{p,T}) \). The likelihood function for \((\Theta_0, \Theta_1)\) can be factorised as

\[
l(\Theta_0, \Theta_1) = p(Y_0, Y_1 | \Theta_0, \Theta_1) = p(Y_0 | Y_1, \Theta_0)p(Y_1, \Theta_1).
\]

For each model \( M_k \), we assume the prior \( p(\Theta_0, \Theta_1 | M_k) = p(\Theta_0 | \gamma_0)p(\Theta_1 | \gamma, G) \) where \( p(\Theta_0 | \gamma_0) \) can be the fully conjugate variable selection prior, and \( p(\Theta_1 | \gamma, G) \) is the prior discussed in Sections 2.2 and 2.3. Coupling the likelihood and separable priors yields the full marginal likelihood of the data under model \( M_k \):

\[
p(Y | M_k) = \int p(Y | \Theta_0, \Theta_1)p(\Theta_0 | M_k)p(\Theta_1 | M_k)d\Theta_0d\Theta_1
\]

\[
= \int p(Y_0 | Y_1, \Theta_0)p(\Theta_0 | \gamma_0)d\Theta_0 \int p(Y_1 | \Theta_1)p(\Theta_1 | \gamma, G)d\Theta_1.
\]
The posterior distribution for model $M_k$ is

$$p(M_k \mid Y) = \frac{p(M_k)p(Y \mid M_k)}{\sum_k p(M_k)p(Y \mid M_k)}.$$  

If the fund’s alpha, $\alpha_0$, is of interest, we first transform back to $\alpha_0$ by using

$$\alpha_0 = \bar{\alpha}_0 \gamma_{00} + \sum_{i=1}^p \alpha_i \gamma_i \theta_i \gamma_{0i}. $$

Under model $M_k$, we then represent the posterior distribution for $\alpha_0$ using a mixture distribution over all models,

$$p(\alpha_0 \mid Y) = \sum_k p(\alpha_0 \mid Y, M_k)p(M_k \mid Y)$$

where $p(\alpha_0 \mid Y, M_k)$ is the posterior distribution of $\alpha_0$ under model $M_k$.

### 7.3 Vanguard managed funds

To evaluate the efficacy of the model, it is applied to a collection of 15 actively managed Vanguard mutual funds, using all available monthly returns through December 2008, available from the Center for Research in Security Prices (CRSP) mutual fund database. The names of the fund, the associated NASDAQ tickers and relevant inception dates are available in Table 2.

Our set of benchmark and nonbenchmark assets consists of nine portfolios constructed passively. Monthly returns on these passive assets are available from January 1927 through December 2008. The sample period for any given mutual fund is a much shorter subset of this overall period. We specify the benchmark series as the excessive market returns (MKT), and so the alpha is exclusively defined with respect to just MKT. The first two of the nonbenchmark passive portfolios are the Fama-French factors, namely, SMB and HML, which are the payoffs on long-short spreads constructed by sorting stocks according to the market capitalisation and the book-to-market ratio. The third nonbenchmark series, which is denoted by MOM, is the momentum factor. The remaining five nonbenchmark assets, denoted by IP1, IP2, IP3, IP4 and IP5, are the value-weighted returns for five industrial portfolios. All data and detailed descriptions of these nine series are publicly available at the data library of Professor Kenneth R. French.

For priors on $\Theta_0$, I assume $\theta_i \sim N(0, \hat{\sigma}^2 \hat{\sigma}^2_i/100)$ if $\gamma_{0i} = 0$, and $\theta_i \sim N(0, 100\hat{\sigma}^2 \hat{\sigma}^2_i)$ if $\gamma_{0i} = 1$, where $\hat{\sigma}_i$ is the standard error of unconstrained OLS estimator $\hat{\theta}_i$, coupled with inverse gamma prior on $\hat{\sigma}^2$, $\hat{\sigma}^2 \sim IG(3/2, 9/2)$. For priors on $\Theta_1$, I choose $\alpha_i \mid \gamma_i = 0 \sim N(0, 0.03^2)$, and $\alpha_i \mid \gamma_i = 1 \sim N(0, 1)$ for monthly $\alpha_i$’s and $i \geq 1$ in Eq. (9). This choice of hyperparameters is in line with the view that a yearly return of 0.36% in

---

excess of the compensation for the risk borne may possibly be ignored; moreover, these excess yearly returns would be within 12%, The prior on the error covariance matrix was specified to provide weak prior knowledge, with $b = 3$ and $D = 0.0001I_8$. Finally, I assume a uniform prior for $\gamma_0$, and a model space prior as in Section 3.3 for $(\gamma, G)$. In each of the 15 funds, the model space of $\gamma_0$ has size $2^{10} = 1024$, which is small enough to be enumerated in a row. The model space of $(\gamma, G)$ is of size $2^8 \times 2^{28}$. To explore this model space, the add-delete Metropolis-Hastings-within-Gibbs sampler was run for 20,000 steps based on the marginal likelihood approximation from the 2,000 Monte Carlo sample.

For models indexed by $(\gamma, G)$, the most probable model is that $\gamma_i = 1$ for all $i = 1, \ldots, p$, with the residual graph pictured in Fig. 5. This is also the median probability model. This modal model seems to suggest that the eight nonbenchmark assets are not all perfectly priced by the benchmark asset. The residual graph indicates a great deal of conditional independencies among error terms.

For models indexed by $\gamma_0$, Table 3 shows the inclusion probabilities for eight nonbenchmark assets for each of the 15 aggressive Vanguard funds. As can be seen, the errors between each one of the Vanguard managed funds and eight nonbenchmark assets are contemporaneously correlated in different ways. The number of nonbenchmark asset regression equations that is related to a fund’s regression equations varies from 0 (for US growth) and 8 (for Equity-Income). We note that a fund’s contemporaneously dependencies on nonbenchmark assets seem to reflect a fund’s portfolio composition. For example, the Capital Opportunity Fund seeks companies with long-term growth and has a 44.6% holding on the information technology sector as of May, 2008. The error of this fund is related to the error of nonbenchmark assets representing market capitalisation (SML) and high technology (IP3).

Table 4 reports the estimates of monthly $\alpha_0$’s within each fund based on the OLS, SUR and SSUR models for a five-year period, a ten-year period and the period since a fund’s inception. The SSUR estimates are nontrivially different from their OLS and SUR counterparts. In particular, the $\alpha_0$’s tend towards zeros under the SSUR model. This is not surprising since the SSUR model assumes a positive probability for $\alpha_0 = 0$. One important issue in fund performance evaluation is whether the managed fund adds value beyond the standard passive benchmarks. We address this issue by computing the standard error of the three estimators of a fund’s alpha. In Table 5 we examine the three standard errors. These standard errors reflects the precision of inferences about $\alpha_0$. Two results are worth noting. First, the SUR standard errors are generally smaller than their OLS counterparts. This observation is compatible with that in Pástor & Stambaugh (2002). Second, with few exceptions, the SSUR model seems to reduce the standard error even more than the SUR model. Recall that the standard error of the SSUR estimates takes into account of structure uncertainty. The reduced standard errors seem to suggest that there is a great deal of sparsity within the SUR models and that identifying this sparsity can help provide more precise estimates of $\alpha_0$’s. Examining the results in Table 4 and 5 together, we find only a few funds have estimated $\alpha_0$ that are two standard errors away from 0. This suggest that most of
the 15 mutual funds do not generate excessive returns beyond the passive benchmark assets.

Table 6 reports the estimates of $\beta_0$’s within each fund based on the OLS, SUR and SSUR models for a five-year period, a ten-year period and the period since a fund’s inception. First, we note that quite different $\beta_0$’s are generated when we use OLS versus the two SUR-type models. Second, the difference in $\beta_0$’s between the SUR and SSUR models is substantially less than that for the OLS model and SUR or SSUR models. This means that nonbenchmark assets play an important role in estimating the $\beta_0$’s, and that imposing structures seems to affect $\beta_0$’s less than adding nonbenchmark assets. The manner in which nonbenchmark assets provide information is illustrated most dramatically in the cases of higher-beta and lower-beta funds. For example, the Capital Opportunity Fund and Growth Equity funds have $\beta_0$’s of 1.20 based on OLS model, while these figures decrease to 1 according to the two SUR models. The Dividend Growth and Equity-Income funds have $\beta_0$’s of about 0.53 and 0.67 if estimated using OLS, while these figures are around 0.75 and 0.86 if estimated using the two SUR models. In Table 7 we examine the standard deviations of the three estimators. As evident in the table, these standard deviations are very close to each other.

Figure 5: Highest log posterior graph of errors of nonbenchmark assets from the analysis

8 Extensions to linear equality restrictions and dynamic SUR models

In this section, we consider two important extensions of the SSUR model given in Section 2. First, many economic applications of SUR models involve linear restrictions
on the coefficients. For example, the same coefficients may appear in more than one equation, and so one may want to hypothesise that all equations have the same coefficient vector (Min & Zellner, 1993). In general, the main problem involves assessing the evidence in favour of a reduced model of the kind \( A\beta = b \), where \( A \) is a \( r \times n \) matrix and \( b \) is a \( r \)-vector. Second, we assume that the regression parameters are time varying. In particular, the parameter vector \( \beta \) at time \( t \) is denoted by \( \beta_t \), and so the model is re-specified as

\[
y_t = X'_t\beta_t + e_t, \quad \nu_t \sim N(0, V) \tag{11}
\]

\[
\beta_t = \beta_{t-1} + w_t, \quad w_t \sim N(0, W_t).
\]

with the initial prior \( \beta_0 \sim N(m_0, C_0) \). Carvalho & West (2007a,b); Wang et al. (2009) used graphical model structuring for a covariance matrix in a class of dynamic SUR models that has been widely used to study financial time series (Quintana & West, 1987; Quintana, 1992; West & Harrison, 1997; Quintana et al., 2003). Their dynamic graphical model which leads to conjugate analysis requires that each univariate series \( y_{r,t} \) must have the same predictors. In addition, \( W_t \) must be separable by a Kronecker product. Here we consider the general dynamic SUR model in Eq. (11) with two additional restrictions (a): \( A\beta_t = b \) for all \( t \), and (b): \( V \) is constrained by one decomposable graph \( G \).

Under the existence of the linear equality restriction (a), we reorder the elements in \( \beta_t \) so that the restrictions can be written as

\[
A\beta_t = (A_1, A_2) \left( \begin{array}{c} \beta_{1,t} \\ \beta_{2,t} \end{array} \right) = b
\]

implying \( \beta_{1,t} = A_{1}^{-1}(b - A_2\beta_{2,t}) \). Here \( A_1 \) is \( r \times r \) and nonsingular; \( A_2 \) is \( r \times (n - r) \); and \( \beta_{1,t} \) and \( \beta_{2,t} \) are \( r \) and \( n - r \) sub-vectors of \( \beta_t \) respectively. Correspondingly, the observation equation in a dynamic SUR model can be written as

\[
y_t = X'_t\beta_t + e_t = X'_{1,t}\beta_{1,t} + X'_{2,t}\beta_{2,t} + e_t, \quad \nu_t \sim N(0, V)
\]

or

\[
\bar{y}_t = \tilde{X}'_{2,t}\beta_{2,t} + e_t \quad \nu_t \sim N(0, V) \tag{12}
\]

where \( z_t = y_t - X'_{1,t}A_{1}^{-1}b \) and \( \tilde{X}'_{2,t} = X'_{2,t} - X'_{1,t}A_{1}^{-1}A_2 \). Thus, we can work on Eq. (12) as a SUR model without linear constraints.

Suppose \( W_t \) is specified \emph{a priori}. The inputs for the Gibbs sampler are as follows. Given \( V \), sampling the joint distribution of \( (\beta_0|T, Y, V) \) is conducted using the forward filtering backward sampling algorithm that is detailed in West & Harrison (1997). The simulation of \( V \) for a specified graph is based on its full conditional distribution \( (V | \beta_{0:T}, Y) \sim \text{HIW}_G\{b + T, D + \sum_{t=1}^T(y_t - X'_t\beta_t)(y_t - X'_t\beta_t)'\} \). Furthermore, if Bayes factors to discriminate between models are of interests, these draws of \( (\beta_{0:T}, V) \) allow us to approximate the marginal likelihood in a similar manner to that in Section 3.2.
8.1 Example: Annual output growth rate data

An example concerns the choice of the pooled and the unpooled models for predicting annual output growth rates for industrialised countries. The data are taken from the IMF International Financial Statistics database for five countries, namely, Australia, Canada, Japan, the UK and the USA. We fit the dynamic SUR model in this illustration following the model considered by Min & Zellner (1993) and Chib & Greenberg (1995).

In our model framework, we set $y_{i,t}$ as the annual output growth rate for the $i$th country in the $t$th year. We let $X_{i,t} = (1, y_{i,t-1}, y_{i,t-2}, y_{i,t-3}, SR_{i,t-1}, SR_{i,t-2}, GM_{i,t-1}, MSR_{t-1})'$ where $SR_{i,t}$ is the rate of growth of real stock prices, $GM_{i,t}$ is the rate of growth of real money, and $MSR_t$ is the median of $SR_{i,t}$ in year $t$. Therefore, for each country $i$ at year $t$, $\beta_{i,t}$ is a vector of eight regression coefficients. The pooled model corresponds to the choice of $\beta_t = \beta_{1,t} = \cdots = \beta_{5,t}$, while the unpooled model corresponds to the choice of $\beta_t = (\beta'_{1,t}, \cdots, \beta'_{5,t})'$.

It is worth noting that the Gibbs sampler for the dynamic SUR applies only when the sequence of state evolution variance matrices $W_t$ are specified. This is different from the dynamic matrix-variate linear models in which $W_t$ is depending on $V$ through a discount factor. In the general dynamic SUR model, if $W_t$ depends on $V$ through a discount factor, then such dependencies prevent the conditional distribution $(V | Y, \beta_0: T)$ from maintaining a tractable form. To identify a reasonable sequence of $W_t$, the following strategy is used. First fit a static SUR model and estimate $V$ using the Gibbs sampler from Section 3.1. Then specify $W_t$ as $W_t = (1 - \delta)/\delta C_{t-1}$, where $C_{t-1}$ is the sequentially updated covariance matrix of $(\beta_{i,t-1} | D_{t-1})$ using an off-line estimated value for $V$ from the static SUR model. In the analysis below, $\delta = 0.98$, and the four models represent different combinations of linear constraints and graphs $G$, where the linear constraint means either a pooled or an unpooled model and $G$ is either a full graph or an empty graph. For each case, the Gibbs sampler was run 20 times each of which generated 50,000 draws from the posterior distribution after discarding the first 5,000 draws. The iterations began with the specification of values of $m_0 = 0$, $C_0 = 10I_{40}$, $b = 3$ and $D = 0.0001I_5$. The results are summarised in Table 8, where for each of four models, the log of the marginal likelihoods from two approximations are reported along with their numerical standard errors. Based on this table, it appears that the marginal likelihood is precisely estimated in all the fitted models. As expected, the more complex model has a larger numerical standard error associated with the estimation; for example, the numerical standard error of the unpooled model with a full graph is 147 times greater than that of the pooled model with an empty graph. These marginal likelihoods support the conclusion of Min & Zellner (1993); Chib & Greenberg (1995), who argue that a pooled model is better than an unpooled model. Moreover, these marginal likelihoods indicate that the error covariance matrix may be diagonal.
The SUR models are common in econometric studies. It is recognized that the conventional unconstrained SUR models may be over-parametrised. To remedy this problem, we have introduced a Bayesian analysis of the sparse seemingly unrelated regression (SSUR) model. The main innovations include inferences via Markov chain Monte Carlo simulations for specific constraints of regression coefficients and errors, evaluations of the marginal likelihoods of restrictions using coupled Candidate’s formula approximations, and the extension of sparse modelling to dynamic SUR models.

Regarding the use of the conjugate hyper-inverse Wishart prior for the covariance matrix, Rajaratnam et al. (2008) provide theoretical support for the method used to estimate higher-dimensional covariance and precision matrices in Gaussian graphical models; in our models, this prior induces tractable and computationally accessible posteriors, leads to an adequate mixing of Markov chain simulations, and produces different approximations to marginal likelihoods of restrictions using the Candidate’s formula.

The three real-world examples illustrate the important practical potential of the structured model. The first real-world example investigates the causal relations and dynamic interactions among stock returns, interest rates, real activity, and inflation. Compared to previous studies, the SSUR model analysis identifies the important signals of dynamic relations among the variables without imposing a priori restrictions. The second real-world example applies SSUR modelling to estimate the mutual fund performance measures using information from returns on seemingly unrelated nonbenchmark assets. The SSUR model is shown to be able to address several important practical considerations; for most funds in my examples, the SSUR model provides more precise estimates of alpha than both the standard OLS estimates and the original SUR estimates in Pastor & Stambaugh (2002). The third example highlights the possibility of marginal likelihood estimation in a dynamic setting with general linear equality constraints.

A number of methodological issues remain. First, our examples are in modest dimensional problems where local move Metropolis-Hastings methods for the variable selection and graphical model components of the analysis can be expected to be effective. To scale to higher dimensions, alternatives computational strategies such as shotgun stochastic search over graphs (Dobra et al., 2004; Jones et al., 2005; Hans et al., 2007) become relevant. There is also potential for computationally faster approximations using expectation-maximisation style and variational methods (Jordan et al., 1999). Another issue involves the consideration of the structures in more complicated models with dynamic error covariance matrices such as those models examined in West & Harrison (1997) and Carvalho & West (2007a,b).

The web page http://stat.duke.edu/~hw27 provides freely available MATLAB code that implements the method described here.
REFERENCES


<table>
<thead>
<tr>
<th>Name</th>
<th>Ticker</th>
<th>5 Year</th>
<th>10 Year</th>
<th>Inception Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap Opp</td>
<td>VHCOX</td>
<td>-2.34</td>
<td>5.71</td>
<td>08/1995</td>
</tr>
<tr>
<td>Dividend Growth</td>
<td>VDIGX</td>
<td>-1.09</td>
<td>-3.07</td>
<td>05/1992</td>
</tr>
<tr>
<td>Equity-Income</td>
<td>VEIPX</td>
<td>-2.35</td>
<td>-1.30</td>
<td>03/1988</td>
</tr>
<tr>
<td>Explorer</td>
<td>VEXPX</td>
<td>-6.16</td>
<td>0.20</td>
<td>12/1967</td>
</tr>
<tr>
<td>Growth &amp; Income</td>
<td>VQNPX</td>
<td>-6.09</td>
<td>-4.41</td>
<td>12/1986</td>
</tr>
<tr>
<td>Growth Equity</td>
<td>VGEQX</td>
<td>-8.13</td>
<td>-7.71</td>
<td>03/1992</td>
</tr>
<tr>
<td>Mid Cap Growth</td>
<td>VMGRX</td>
<td>-4.08</td>
<td>0.89</td>
<td>12/1997</td>
</tr>
<tr>
<td>Morgan Growth</td>
<td>VMRGX</td>
<td>-5.75</td>
<td>-4.23</td>
<td>12/1968</td>
</tr>
<tr>
<td>PRIMECAP</td>
<td>VPMCX</td>
<td>-1.38</td>
<td>0.52</td>
<td>11/1984</td>
</tr>
<tr>
<td>Selected Value</td>
<td>VASVX</td>
<td>-2.62</td>
<td>1.83</td>
<td>02/1996</td>
</tr>
<tr>
<td>Strategic Equity</td>
<td>VSEQX</td>
<td>-6.72</td>
<td>-10.95</td>
<td>08/1995</td>
</tr>
<tr>
<td>US Growth</td>
<td>VWUSX</td>
<td>-6.72</td>
<td>-10.95</td>
<td>01/1959</td>
</tr>
<tr>
<td>US Value</td>
<td>VUVLX</td>
<td>-5.28</td>
<td>-1.07</td>
<td>-1.07</td>
</tr>
<tr>
<td>Windsor</td>
<td>VWNDX</td>
<td>-7.19</td>
<td>-1.57</td>
<td>4.70</td>
</tr>
<tr>
<td>Windsor II</td>
<td>VWNFX</td>
<td>-3.68</td>
<td>-2.14</td>
<td>5.22</td>
</tr>
<tr>
<td>Market</td>
<td>-</td>
<td>-4.43</td>
<td>-3.51</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2: Summary statistics of 15 Vanguard funds
### Table 3: Exact (to 2 decimal places) inclusion probabilities for 8 nonbenchmark assets for each of 15 aggressive Vanguard funds

<table>
<thead>
<tr>
<th>Name</th>
<th>SML</th>
<th>HML</th>
<th>MOM</th>
<th>IP1</th>
<th>IP2</th>
<th>IP3</th>
<th>IP4</th>
<th>IP5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap Opp</td>
<td>1.00</td>
<td>0.04</td>
<td>0.03</td>
<td>0.05</td>
<td>0.26</td>
<td>0.94</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>Dividend Growth</td>
<td>0.10</td>
<td>1.00</td>
<td>0.04</td>
<td>0.06</td>
<td>0.74</td>
<td>0.09</td>
<td>0.34</td>
<td>0.05</td>
</tr>
<tr>
<td>Equity-Income</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.97</td>
<td>1.00</td>
<td>0.92</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Explorer</td>
<td>1.00</td>
<td>1.00</td>
<td>0.25</td>
<td>0.06</td>
<td>0.36</td>
<td>0.15</td>
<td>0.18</td>
<td>0.21</td>
</tr>
<tr>
<td>Growth &amp; Income</td>
<td>1.00</td>
<td>0.29</td>
<td>0.12</td>
<td>1.00</td>
<td>0.18</td>
<td>0.61</td>
<td>0.35</td>
<td>0.63</td>
</tr>
<tr>
<td>Growth Equity</td>
<td>0.07</td>
<td>1.00</td>
<td>1.00</td>
<td>0.82</td>
<td>0.06</td>
<td>1.00</td>
<td>0.07</td>
<td>0.13</td>
</tr>
<tr>
<td>Mid Cap Growth</td>
<td>1.00</td>
<td>0.60</td>
<td>1.00</td>
<td>0.35</td>
<td>0.04</td>
<td>0.14</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Morgan Growth</td>
<td>1.00</td>
<td>0.60</td>
<td>1.00</td>
<td>0.49</td>
<td>1.00</td>
<td>0.87</td>
<td>1.00</td>
<td>0.22</td>
</tr>
<tr>
<td>PRIMECAP</td>
<td>1.00</td>
<td>0.15</td>
<td>0.04</td>
<td>0.60</td>
<td>0.58</td>
<td>1.00</td>
<td>0.31</td>
<td>0.32</td>
</tr>
<tr>
<td>Selected Value</td>
<td>0.99</td>
<td>0.94</td>
<td>1.00</td>
<td>0.93</td>
<td>1.00</td>
<td>0.06</td>
<td>0.22</td>
<td>0.48</td>
</tr>
<tr>
<td>Strategic Equity</td>
<td>1.00</td>
<td>0.85</td>
<td>0.06</td>
<td>0.28</td>
<td>1.00</td>
<td>0.08</td>
<td>0.08</td>
<td>0.98</td>
</tr>
<tr>
<td>US Growth</td>
<td>0.02</td>
<td>0.14</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td>US Value</td>
<td>0.07</td>
<td>1.00</td>
<td>0.17</td>
<td>0.45</td>
<td>0.08</td>
<td>0.05</td>
<td>0.21</td>
<td>0.15</td>
</tr>
<tr>
<td>Windsor</td>
<td>0.19</td>
<td>1.00</td>
<td>0.89</td>
<td>1.00</td>
<td>0.99</td>
<td>0.07</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Windsor II</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.98</td>
<td>1.00</td>
<td>0.10</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

### Table 4: Estimated monthly α’s from each of the three models: the least square estimates from the OLS, and the posterior mean estimates from the SUR and SSUR models. An asterisk symbol (*) flags an estimated $α_0$ that is two standard errors away from 0

<table>
<thead>
<tr>
<th>Name</th>
<th>OLS</th>
<th>SUR</th>
<th>SSUR</th>
<th>OLS</th>
<th>SUR</th>
<th>SSUR</th>
<th>OLS</th>
<th>SUR</th>
<th>SSUR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap Opp</td>
<td>0.27</td>
<td>0.22</td>
<td>0.07</td>
<td>*0.88</td>
<td>*0.81</td>
<td>*0.56</td>
<td>0.34</td>
<td>0.33</td>
<td>0.18</td>
</tr>
<tr>
<td>Dividend Growth</td>
<td>0.16</td>
<td>0.31</td>
<td>0.08</td>
<td>-0.08</td>
<td>-0.20</td>
<td>-0.02</td>
<td>0.05</td>
<td>-0.11</td>
<td>0.13</td>
</tr>
<tr>
<td>Equity-Income</td>
<td>0.06</td>
<td>0.31</td>
<td>0.20</td>
<td>0.07</td>
<td>0.09</td>
<td>0.15</td>
<td>0.14</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>Explorer</td>
<td>-0.03</td>
<td>-0.03</td>
<td>0.06</td>
<td>0.43</td>
<td>0.08</td>
<td>0.16</td>
<td>-0.05</td>
<td>-0.16</td>
<td>0.03</td>
</tr>
<tr>
<td>Growth &amp; Income</td>
<td>-0.16</td>
<td>-0.11</td>
<td>-0.04</td>
<td>-0.10</td>
<td>-0.01</td>
<td>0.05</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Growth Equity</td>
<td>-0.25</td>
<td>-0.09</td>
<td>-0.04</td>
<td>-0.20</td>
<td>-0.12</td>
<td>0.00</td>
<td>-0.20</td>
<td>-0.13</td>
<td>-0.02</td>
</tr>
<tr>
<td>Mid Cap Growth</td>
<td>0.11</td>
<td>0.14</td>
<td>0.02</td>
<td>0.61</td>
<td>0.38</td>
<td>0.23</td>
<td>0.55</td>
<td>0.43</td>
<td>0.25</td>
</tr>
<tr>
<td>Morgan Growth</td>
<td>-0.07</td>
<td>-0.09</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.05</td>
<td>0.04</td>
<td>-0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>PRIMECAP</td>
<td>0.24</td>
<td>0.22</td>
<td>0.06</td>
<td>*0.36</td>
<td>*0.34</td>
<td>0.15</td>
<td>0.23</td>
<td>0.18</td>
<td>*0.21</td>
</tr>
<tr>
<td>Selected Value</td>
<td>0.14</td>
<td>-0.01</td>
<td>-0.11</td>
<td>0.39</td>
<td>0.17</td>
<td>0.17</td>
<td>0.09</td>
<td>-0.11</td>
<td>0.15</td>
</tr>
<tr>
<td>Strategic Equity</td>
<td>-0.01</td>
<td>-0.23</td>
<td>0.00</td>
<td>0.34</td>
<td>0.10</td>
<td>*0.23</td>
<td>0.14</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>US Growth</td>
<td>-0.20</td>
<td>0.03</td>
<td>0.02</td>
<td>*- 0.53</td>
<td>-0.32</td>
<td>-0.17</td>
<td>0.31</td>
<td>0.29</td>
<td>0.01</td>
</tr>
<tr>
<td>US Value</td>
<td>-0.10</td>
<td>-0.11</td>
<td>0.00</td>
<td>0.31</td>
<td>0.20</td>
<td>*0.17</td>
<td>0.31</td>
<td>0.20</td>
<td>*0.17</td>
</tr>
<tr>
<td>Windsor</td>
<td>-0.23</td>
<td>-0.25</td>
<td>-0.19</td>
<td>0.15</td>
<td>0.06</td>
<td>0.14</td>
<td>0.14</td>
<td>-0.12</td>
<td>-0.11</td>
</tr>
<tr>
<td>Windsor II</td>
<td>0.02</td>
<td>0.07</td>
<td>0.00</td>
<td>0.04</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.13</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
</tbody>
</table>
Table 5: Standard errors of each of the three estimates of monthly $\alpha$’s.

<table>
<thead>
<tr>
<th>Name</th>
<th>OLS</th>
<th>SUR</th>
<th>SSUR</th>
<th>OLS</th>
<th>SUR</th>
<th>SSUR</th>
<th>OLS</th>
<th>SUR</th>
<th>SSUR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap Opp</td>
<td>1.73</td>
<td>1.03</td>
<td>1.09</td>
<td>1.23</td>
<td>1.04</td>
<td>1.02</td>
<td>1.20</td>
<td>1.05</td>
<td>1.07</td>
</tr>
<tr>
<td>Dividend Growth</td>
<td>0.72</td>
<td>0.80</td>
<td>0.75</td>
<td>0.56</td>
<td>0.83</td>
<td>0.80</td>
<td>0.53</td>
<td>0.74</td>
<td>0.75</td>
</tr>
<tr>
<td>Equity-Income</td>
<td>0.72</td>
<td>0.83</td>
<td>0.80</td>
<td>0.60</td>
<td>0.85</td>
<td>0.86</td>
<td>0.67</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>Explorer</td>
<td>1.23</td>
<td>1.20</td>
<td>1.19</td>
<td>1.20</td>
<td>1.24</td>
<td>1.23</td>
<td>1.11</td>
<td>1.02</td>
<td>1.04</td>
</tr>
<tr>
<td>Growth &amp; Income</td>
<td>0.96</td>
<td>1.00</td>
<td>0.99</td>
<td>0.93</td>
<td>0.93</td>
<td>0.95</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>Growth Equity</td>
<td>1.11</td>
<td>1.00</td>
<td>1.02</td>
<td>1.29</td>
<td>1.00</td>
<td>1.00</td>
<td>1.25</td>
<td>0.98</td>
<td>0.96</td>
</tr>
<tr>
<td>Mid Cap Growth</td>
<td>1.15</td>
<td>1.05</td>
<td>1.07</td>
<td>1.37</td>
<td>1.20</td>
<td>1.25</td>
<td>1.36</td>
<td>1.16</td>
<td>1.23</td>
</tr>
<tr>
<td>Morgan Growth</td>
<td>1.07</td>
<td>0.99</td>
<td>1.02</td>
<td>1.16</td>
<td>0.98</td>
<td>0.99</td>
<td>1.06</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>PRIMECAP</td>
<td>0.96</td>
<td>0.93</td>
<td>0.95</td>
<td>1.04</td>
<td>0.94</td>
<td>0.93</td>
<td>1.07</td>
<td>0.97</td>
<td>1.00</td>
</tr>
<tr>
<td>Selected Value</td>
<td>0.96</td>
<td>1.00</td>
<td>0.98</td>
<td>0.71</td>
<td>1.02</td>
<td>1.01</td>
<td>0.77</td>
<td>1.10</td>
<td>1.12</td>
</tr>
<tr>
<td>Strategic Equity</td>
<td>1.23</td>
<td>1.21</td>
<td>1.24</td>
<td>0.95</td>
<td>1.17</td>
<td>1.17</td>
<td>0.98</td>
<td>1.18</td>
<td>1.19</td>
</tr>
<tr>
<td>US Growth</td>
<td>0.98</td>
<td>0.90</td>
<td>0.90</td>
<td>1.19</td>
<td>0.97</td>
<td>0.98</td>
<td>1.09</td>
<td>0.96</td>
<td>1.08</td>
</tr>
<tr>
<td>US Value</td>
<td>0.93</td>
<td>0.97</td>
<td>0.95</td>
<td>0.82</td>
<td>1.03</td>
<td>1.01</td>
<td>0.82</td>
<td>1.03</td>
<td>1.01</td>
</tr>
<tr>
<td>Windsor</td>
<td>0.99</td>
<td>1.00</td>
<td>1.01</td>
<td>0.87</td>
<td>1.07</td>
<td>1.08</td>
<td>0.93</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td>Windsor II</td>
<td>0.89</td>
<td>0.98</td>
<td>0.96</td>
<td>0.69</td>
<td>1.00</td>
<td>0.99</td>
<td>0.81</td>
<td>1.02</td>
<td>1.02</td>
</tr>
</tbody>
</table>

Table 6: Estimated $\beta$’s from each of the three models: the least square estimates from the OLS, and the posterior mean estimates from the SUR and SSUR models.
Table 7: Standard errors of each of the three estimates of $\beta$'s.

<table>
<thead>
<tr>
<th>Name</th>
<th>5 year OLS</th>
<th>SUR</th>
<th>SSUR</th>
<th>10 year OLS</th>
<th>SUR</th>
<th>SSUR</th>
<th>Since Inception OLS</th>
<th>SUR</th>
<th>SSUR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap Opp</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>Dividend Growth</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>Equity-Income</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Explorer</td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
<td>0.06</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Growth &amp; Income</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Growth Equity</td>
<td>0.06</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Mid Cap Growth</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.09</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Morgan Growth</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>PRIMECAP</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Selected Value</td>
<td>0.05</td>
<td>0.06</td>
<td>0.05</td>
<td>0.07</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Strategic Equity</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>US Growth</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>US Value</td>
<td>0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Windsor</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Windsor II</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table 8: Summary of results for output growth rate data. Each of the two approximations were run for 20 times. The mean and the numeric standard error of log marginal likelihoods from these 20 runs are reported for each of the approximations.