Workshop, Sveriges Riksbank,
Sept. 9,2006

Model Uncertainty and Bayesian Model Averaging in Structural VAR Processes with Applications to:

(1) Stability of the 'Great Ratios in the USA' and
(2) Possible liquidity trap in the UK

Herman K. van Dijk
Erasmus University Rotterdam
The Netherlands

(joint work with Rodney W. Strachan)

http://www.econometric-institute.org/pubs
(EI report 2006-08)

Outline

1. **Motivation**: forecasting and policy advice using *un*conditional inference on SVAR models where the likelihood information dominates

2. **Bayesian Model Averaging** in SVAR's

3. Possible **irregular/non-smooth** shape of the posterior when information in the likelihood dominates and likelihood is nearly singular

4. Introduction of **Regularization/Smoothness** priors based on manifolds

5. Applications
(i) **Stability of US Great Ratios**
(ii) **Probability of UK liquidity trap**

1. Motivation

AIM 1: *Un*conditional Policy Advice with weak prior information.

CLIENT: "What will happen to prices if oil prices increase by x %?"

ANALYST: "We will have a permanent 2.5% increase in inflation …"

conditional upon the following:

Example: Macro Model of the UK Economy (Garrett, Lee, Pesaran & Shin 2003)

Nine variables in the vector $\boldsymbol{y}_t$:

$$\Gamma \boldsymbol{y}_t = \Pi_1 \boldsymbol{y}_{t-1} + \Pi_2 \boldsymbol{y}_{t-2} + \ldots + \delta d_t + \varepsilon_t$$

- Five cointegrating relations

- The overidentified cointegrating space is defined by:

Purchasing Power Parity	$= p_t - p_t^* - e_t$
Interest Rate Parity	$= r_t - r_t^*$
"Output gap"	$= y_t - y_t^*$
Fisher Inflation Parity	$= r_t$ - Δp_t
Money Market Equilibrium	$= h_t - y_t - dt$ - br_t

- At most a linear drift in $\boldsymbol{y_t}$ and a linear trend in one of the error correction terms

- Oil prices (p_t^o) are 'long-run forcing' $\Rightarrow$
  p_t^o weakly exogenous with respect to the above long run relatio

M_j is model *j* –
The restrictions are: Cointegration; deterministic processes; exogeneity; (over)identification. Lag length

There are 541 models assumed. Through a test-test-test strategy only one of these is used in the policy advice.

CLIENT: What is your confidence in the model used?

AIM 2: Unconditional predictive density forecasts.

The questions clients ask often require forecasts of entire densities rather than forecasts of the mean or variance.

Example is the liquidity trap: negative growth and negative inflation – removes interest rates as a policy tool.

CLIENT: “What is the probability of the liquidity trap occurring in the UK in the coming years?”

ANALYST:
Requires forecasts of the joint density of income growth (Δy_{T+h}) and inflation (Δp_{T+h}) at time *T+h* and the probability

$$P\left(\Delta y_{T+h} < 0, \Delta p_{T+h} < 0 \mid y\right) = \int_{-\infty}^{0}\int_{-\infty}^{0} \pi\left(\Delta y_{T+h}, \Delta p_{T+h} \mid y\right) d\Delta y_{T+h} \, d\Delta p_{T+h}$$

2. Implementation of Bayesian Model Averaging in SVAR's

In the best model (conditional) approach, we obtain inference on a feature of interest, say Δ, *given the model* from the conditional posterior $\pi(\Delta|M_j, y)$. Unconditional inference derives from

$$\pi(\Delta| y) = \sum_j \pi(\Delta|M_j, y)P(M_j| y)$$

$$\begin{aligned} E(\Delta| y) &= \int_{\Theta} \Delta\ \pi(\Delta| y)d\Delta \\ &= \sum_j E(\Delta|M_j, y)P(M_j| y) \end{aligned}$$

$P\left(M_j \mid y\right)$ is the posterior probability of model M_j.

- Δ – The object of interest (eg., a loss function, a response path to a shock, forecast, the probability of an event, …), is supplied by the client.
- M_j – The set of models, $M_j \in M \equiv \{ M_j : j = 1, 2, \ldots, m\}$ is decided by the analyst.

M – The Model Set: VAR with restrictions: Each bundle/combination of restrictions defines a model.

Intermezzo: On the shape of the likelihood/posterior when (near-)singularities occur in the likelihood.

Using the traditional approaches in the literature under diffuse priors,

- we either do not obtain well defined posteriors
- or our posteriors have no moments.
-

The source of the problem is how to implement COINTEGRATION

- Cointegration analysis is the search for stable relationships among unstable variables.
    - The relationships are defined by vectors of coefficients.

Let β be the matrix of cointegrating vectors.
Let ρ = sp(β) the cointegrating space, that is, the space spanned by the colums of β.

- We need to identify the vectors
  Tie β to a particular orientation in sp(β).
  – How we do this matters.

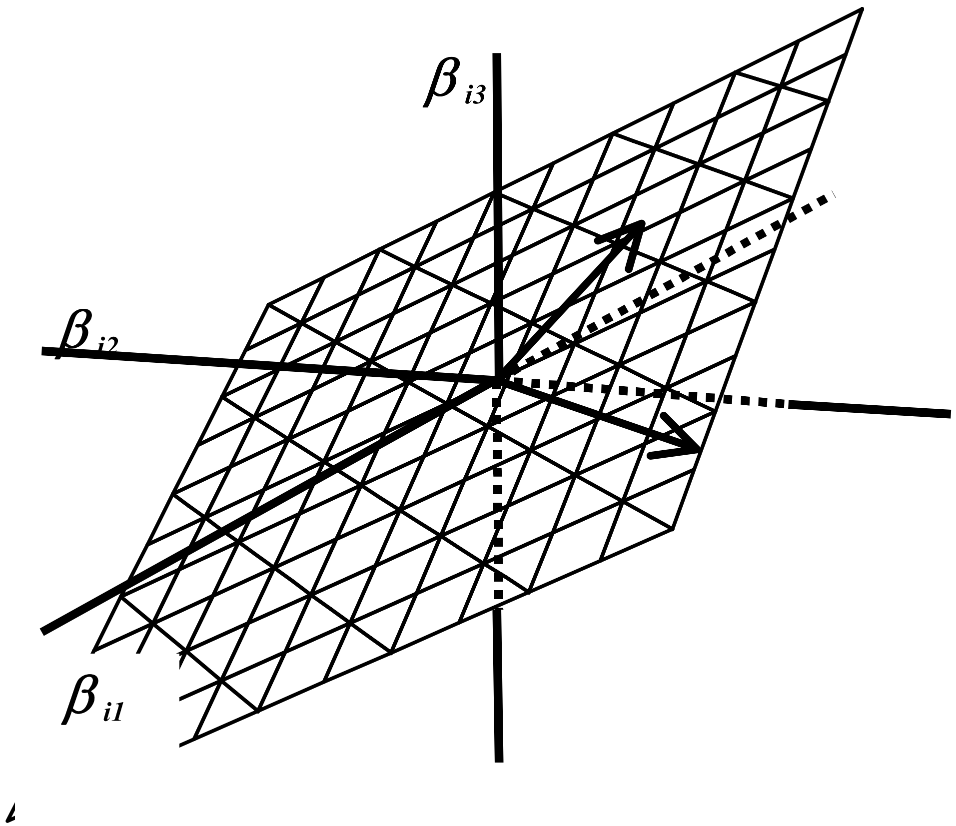
○ How this has been traditionally done
(linear identifying restrictions)
is the source of the problem.

The most common approach in the literature is to use linear restrictions. For example,

$$\beta = [I_r \quad B']'.$$

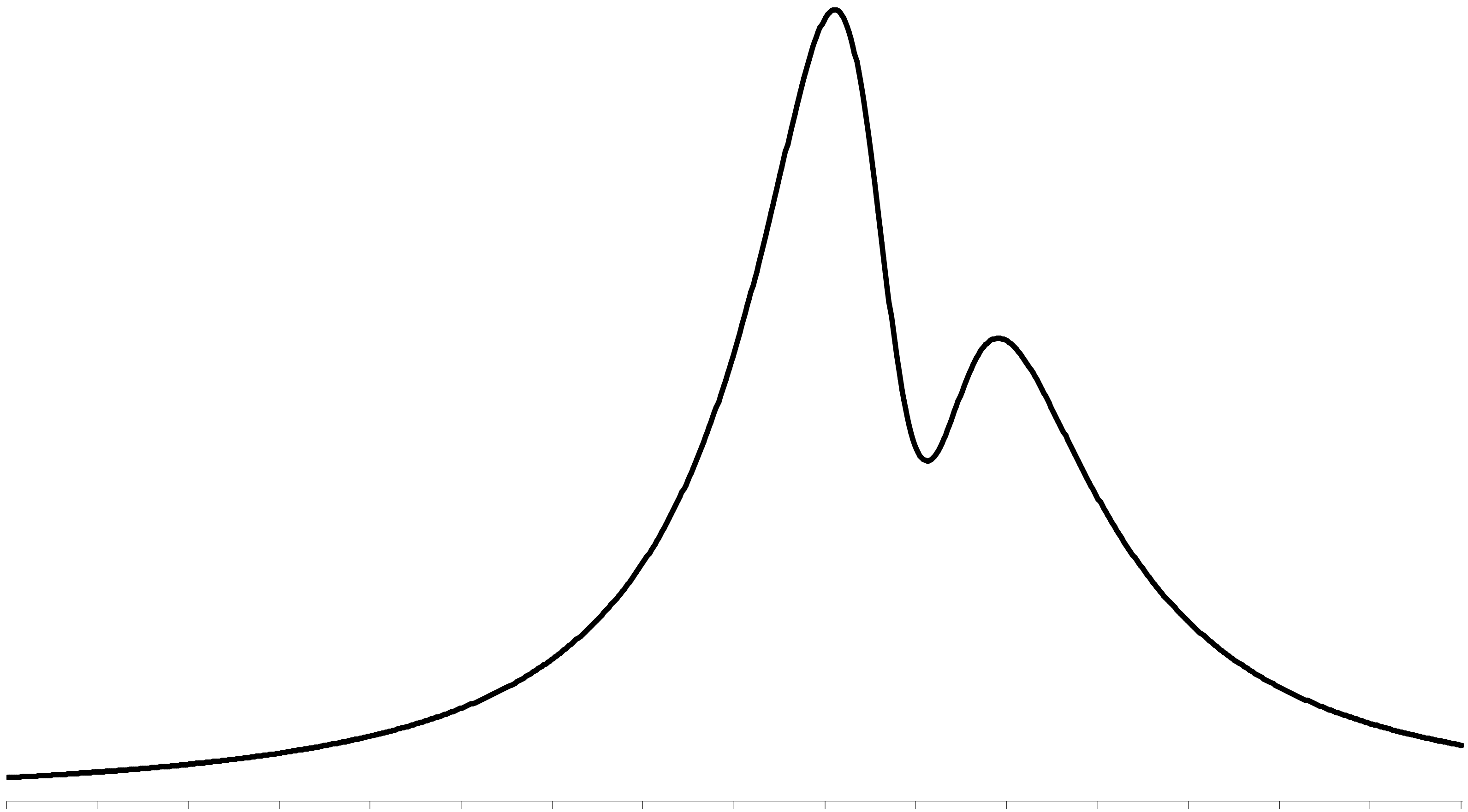
This ties β to the first r coordinate axes.

An advantage – if they are applied carefully, linear restrictions produce interpretable coefficients.



Example has two cointegrating relations in a three variable model.
Vector space is the plane and
Cointegration space is the collection of all planes through the origin.

Example of posterior density of cointegrating parameter: possibly bimodal and no moments (See already Dreze (1976))



PROBLEMS with the traditional approach of linear identifying restrictions:

1. Improper posterior, $\pi(\Delta|y)$ undefined (Kleibergen and van Dijk 1994, 1998)

2. No moments, $E(\Delta|y)$ undefined (KVD 1994 and Bauwens and Lubrano 1996, see graph and paper by Dreze (1976)

3. Exogeneity $\Rightarrow$ improper posterior $\pi(\Delta|y)$ undefined (Bauwens and Lubrano *unpublished notes*, Strachan and van Dijk 2004)

4. Invalid restrictions (JBES: Boswijk 1996, Luukkonen *et al.* 1999, Strachan 2003)

5. Such a restriction puts infinite prior weight in the direction of the space orthogonal to the normalisation (Strachan and Inder 2003, Strachan and van Dijk 2004).

PREVIOUS APPROACHES: to employing 'uninformative, regularization/smoothness' priors

Information Matrix (Jeffreys prior)

Kleibergen & van Dijk 1994 (ET)

Martin (2000)

Posterior probability not defined

Uniform prior on embedding model

Kleibergen & van Dijk 1998 (ET)

Kleibergen & Paap 2003 (JoE)

Paap and van Dijk 2003 (JBES)

Strachan 2003 (JBES)

Posterior probability not defined, except by using a penalty function.

Use SVD – techniques and ideas from Anderson 1951, Johansen (1988, 1991, 1992, 1995)

Uniform prior on cointegrating space

Villani 2000 – Cauchy prior with linear restrictions

Strachan & Inder 2004

Strachan & van Dijk 2003

– no restrictions on space

OUR APPROACH: A Prior on the cointegrating space

Following Villani (2000), we reconsider the parameter of interest

– the cointegrating space, ρ.

– NOT on the cointegrating vectors, β.

Specify the prior on the cointegrating space, ρ.

ρ is an element of the Grassman manifold,

$$\rho \in G_{r,n-r}.$$

We require a measure for $G_{r,n-r}$.

A semi-orthogonal matrix $V'V=I$ is an element of the Stiefel manifold, $V \in V_{r,n}$.

An r x r orthogonal matrix T is an element of the Orthogonal group, $T \in O(r)$.

If β is semi-orthogonal but has a particular (identifying) orientation in ρ

$$\beta T = V$$
$$(\beta' d\beta)\,(T' dT) = (V' dV)$$
$$dg_r^n d\nu_r^r = d\nu_r^n$$

(James, 1954).

$$\frac{\int_{V_{r,n}} d\nu_r^n}{\int_{O(r)} d\nu_r^r} = \int_{G_{r,n}} dg_r^n$$

Thus the uniform prior is:

$$\pi\left(\beta\right)dg_r^n$$
$$= \pi^{-(n-r)r/2} \prod_{j=1}^{r} \frac{\Gamma\left[\left(n-r-j\right)/2\right]}{\Gamma\left[j/2\right]} dg_r^n$$

The identification by the orientation of β is never implemented.

We work with β as $\beta \in V_{r,n}$.

BUT $dim(G_{r,n-r}) < dim(V_{r,n})$

Invariance of the posterior

$$\pi(V \mid y) d\nu_r^n = \pi(\beta T \mid y) dg_r^n d\nu_r^r$$
$$= \pi(\beta \mid y) dg_r^n d\nu_r^r$$

We integrate with respect to $V_{r,n}$ and adjust the resulting volume for $O(r)$.

$$\int_{V_{r,n}} \pi(V \mid y) d\nu_r^n \Big/ \int_{O(r)} d\nu_r^r$$
$$= \int_{G_{r,n}} \pi(\beta \mid y) dg_r^n$$

$\beta \in V_{r,n}$ simplifies the math and allows use of the Poincaré separation theorem to:

$$\pi(\beta \mid y) \propto \left|\beta' D_1 \beta\right|^{(T-n)/2} \left|\beta' D_0 \beta\right|^{(-T)/2}$$

- obtain modes of $\pi(\beta|y)$;
- establish that this posterior will be proper; and,
- establish existence of moments.

This is important since inference on ρ is derived from β.

ADVANTAGES of using this approach for model averaging (Orthonormalisation):

1. The posterior is now proper for all M_j of interest – including with exogeneity.

2. The estimates of the vectors are always valid.

3. All finite moments exist and posterior is unimodal (Poincaré separation theorem).

4. We can specify *coherent* (informative and uninformative) priors directly upon the cointegrating space.

5. We are now able to average across *all* models of interest to us.

6. A range of standard techniques are now available to obtain inference: (MC)MC; Laplace.

Let x_t be an $n \times 1$ vector, then a Vector Error Correction model (VECM) can be written as:

$$\begin{aligned} \Delta x_t &= \alpha\beta' x_{t-1} + d_t \mu + \Gamma(L)\, \Delta x_{t-1} + \varepsilon_t, \\ \varepsilon_t &\sim IIDN(0, \Sigma) \end{aligned}$$

where α and β are $n \times r$ matrices with $0 \leq r \leq n$ being the number of cointegrating relationships.

$\Gamma(L)$ is a matrix polynomial of degree p in the lag operator and d_t is the deterministic term

Models which differ in the number of cointegrating vectors (r), lag length (p) and the specification of deterministic terms (d).

Weak exogeneity of some of x_t with respect to (α, β) :

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ 0 \end{bmatrix}$$

Sampling methods: Koop, León-González, Strachan (2005)

Using the relation $\alpha\beta' = A\kappa\beta' = A\beta^{*'}$

$$\begin{aligned} \alpha &= A\kappa \quad \beta^* = \beta\kappa' \\ A'A &= I_r \quad \beta'\beta = I_r \end{aligned}$$

A Gibbs sampler over

$$\begin{aligned} \pi\left(\beta^*|A, y\right) &\sim Normal \\ \pi\left(\alpha|\beta, y\right) &\sim Normal \end{aligned}$$

$P\left(M_j|y\right)$: Estimated using the method of Gelfand and Dey (1994)

Candidate density is the **Matrix Angular Central Gaussian distribution** of Chikuse (1990).

I.e., draw β^* from a matrix Normal and take the orthogonal-triangular decomposition into

$$\beta^* = \beta\kappa \qquad \text{where} \qquad \beta'\beta = I_r.$$

4 Application I: The Great Ratios

(A story of sequential testing)

King, Plosser, Stock and Watson (1991) investigate stability of ratios of consumption to income and investment, i_t, to income .

Let c_t- log consumption, i_t- log investment, and y_t- log income.

Their theory of Balanced growth within a Real Business Cycle model implies

$$\begin{aligned} x_t &= (c_t, i_t, y_t) \sim I(1) \text{ but} \\ c_t - y_t &\sim I(0) \text{ and} \\ i_t - y_t &\sim I(0) \end{aligned}$$

Thus the theory implies $r = 2$ and $sp(\beta) \subseteq sp(H)$ where

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}$$

Theory would also suggest no trends in the ratios.

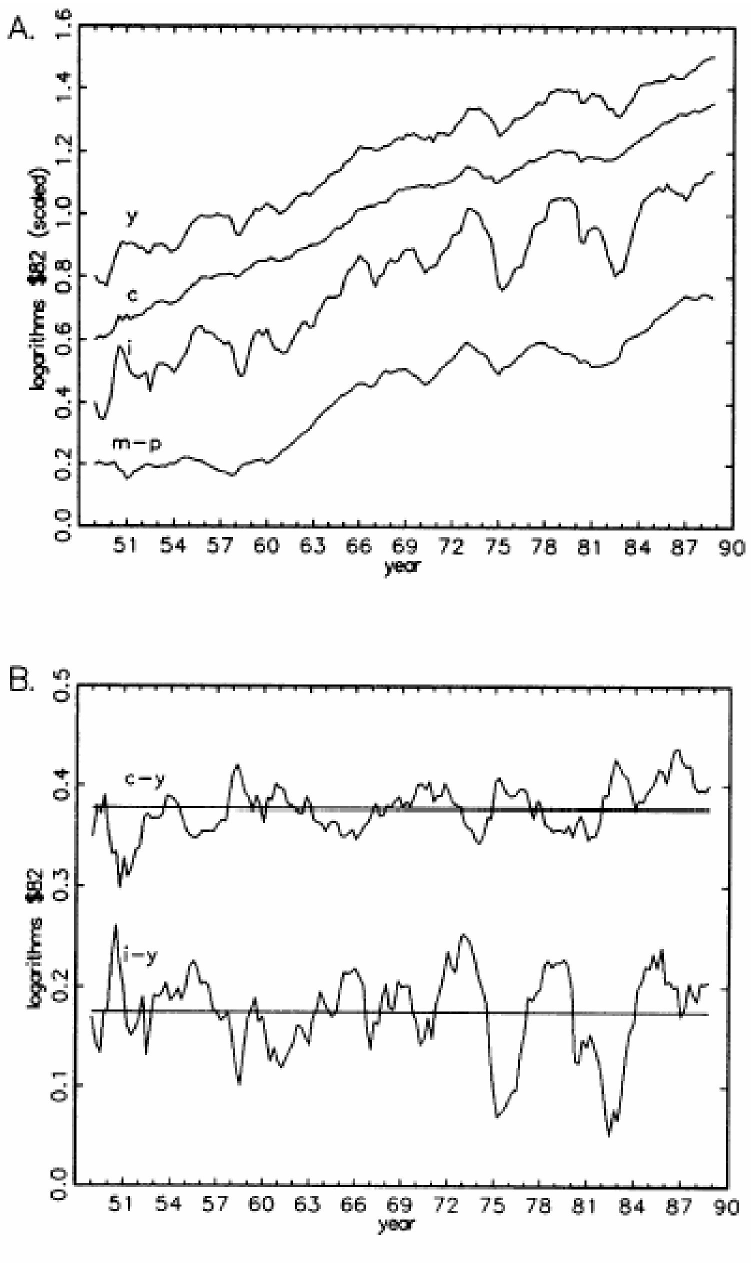


FIGURE 1. A) LOGARITHMS OF PRIVATE OUTPUT (y), CONSUMPTION (c), INVESTMENT (i), AND REAL MONEY BALANCES ($m - p$);
B) LOGARITHMS OF THE CONSUMPTION:OUTPUT ($c - y$) AND INVESTMENT:OUTPUT ($i - y$) RATIOS

Note: To facilitate graphing, constants were added to the logarithms of the variables. The horizontal lines in part B are the means of the (constant-adjusted) variables.

KPSW91 propose and find evidence that $r = 2$ and $sp(\beta) \subseteq sp(H)$ ...

if they assume 6 lags and allow for quadratic trends in x_t and linear trends in the ratios.

Using an extended data set (Paap & van Dijk, 2003) we find this result is sensitive to lag length, deterministic process, and order of testing.

Classical (Johansen) tests suggest $r = 0$.

The marginal probabilities of the alternative ranks do not support $r = 2$.

r	$P(r\|y)$
0	0.56
1	0.17
2	0.01
3	0.26

This agrees with the classical testing procedure and suggests we would NOT continue to conclude

$$(r = 2) \;\; AND \; sp(\beta) \subseteq sp(H) \,.$$

However, if we jointly estimate the support for the models

r	$P(r \cap sp(\beta) \subsetneq sp(H) \mid y)$	$P(r \cap sp(\beta) \subseteq sp(H) \mid y)$
0	0.23	0.00
1	0.07	0.03
2	0.00	0.57
3	0.10	0.00

Thus the implications of their model have 57% of the posterior probability mass.

APPLICATION 2: Macro Model for the UK
(Garratt, Lee, Pesaran, & Shin 2003)

(i) What impact will an oil price shock have upon UK prices?

1. Relative to the rest of the world
2. Long run inflation path

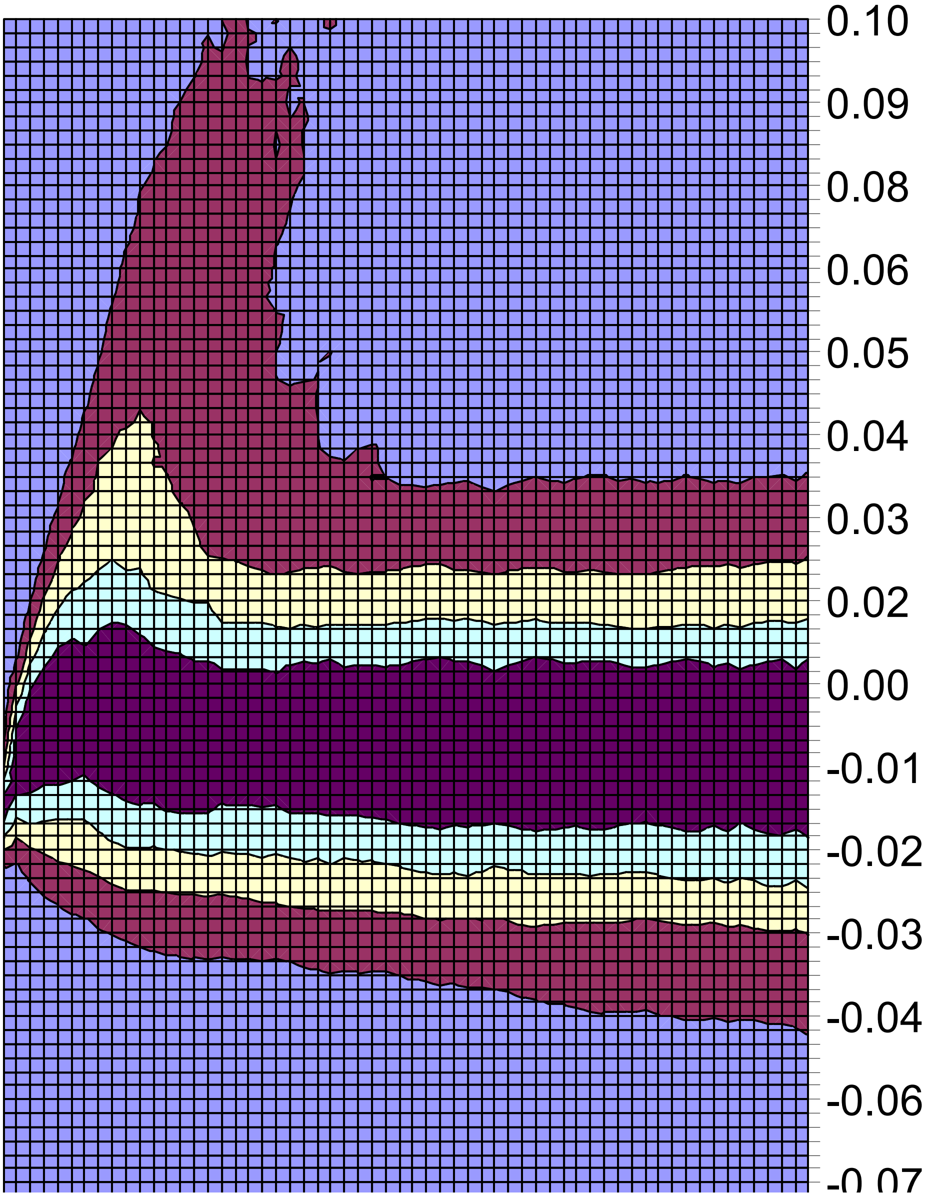
(ii) What impact will an oil price shock have upon the risk of the UK encountering the liquidity trap?

We consider:

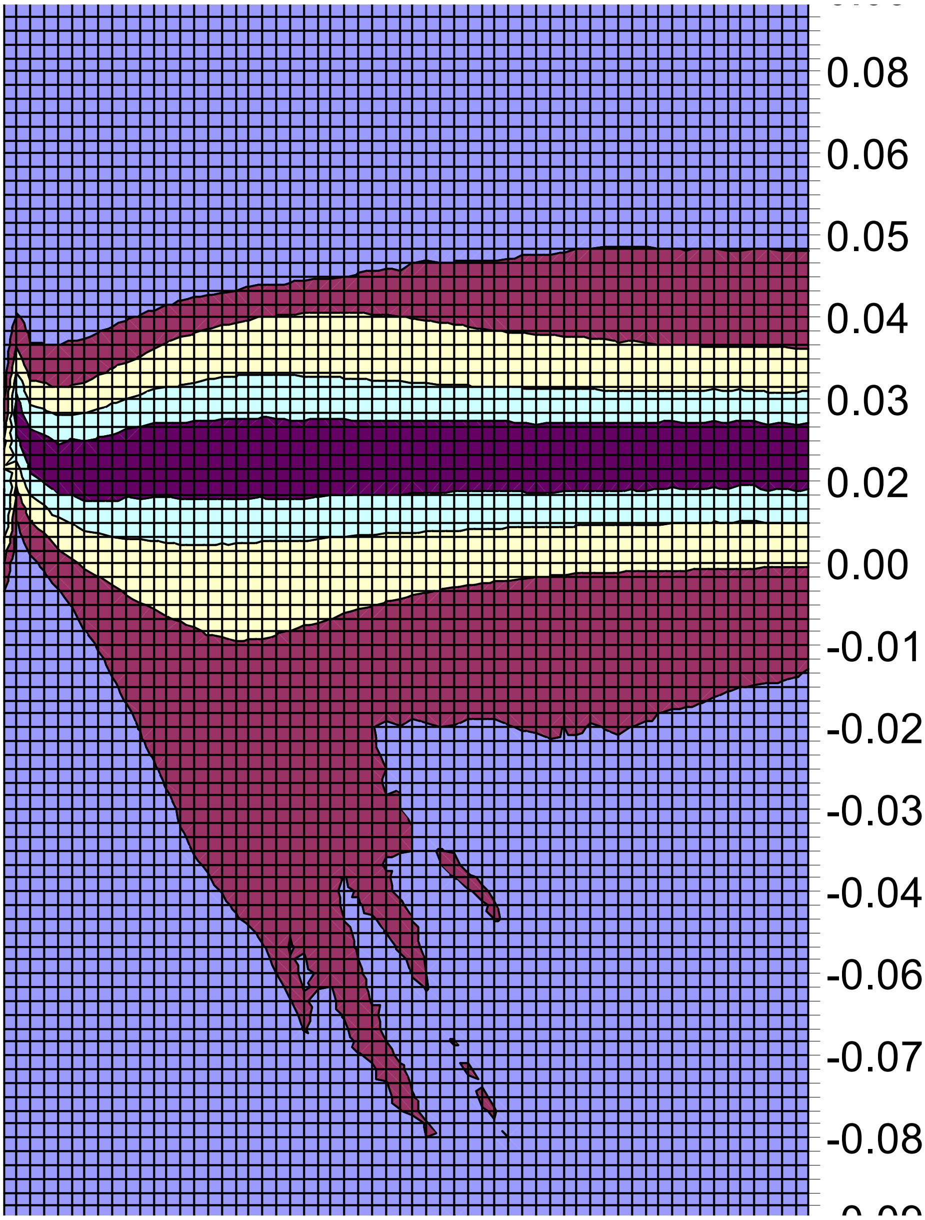
- Impulse Response Function
- Higher Posterior Density (HPD) Regions

NB: We allow oil prices to be weakly exogenous.

Strictly applied, the model set would have
36,864 models.
The restrictions we consider imply some
800 models.
Accounting for observational equivalence and 'impossible models' reduces this to
541 models.



Higher posterior density regions for the impulse response of relative UK prices (pt - pt*) to a shock in oil prices.



Higher posterior density regions
for the impulse response
of relative UK inflation (Δp_t)
to a shock in oil prices

sub (ii) What impact will an oil price shock have upon the risk of the UK encountering the liquidity trap?

- Forecast (joint) densities
- Forecasts probabilities of events

Deflation

$$P\left(\Delta p_{T+h} < 0 \mid y\right)$$

Negative Growth

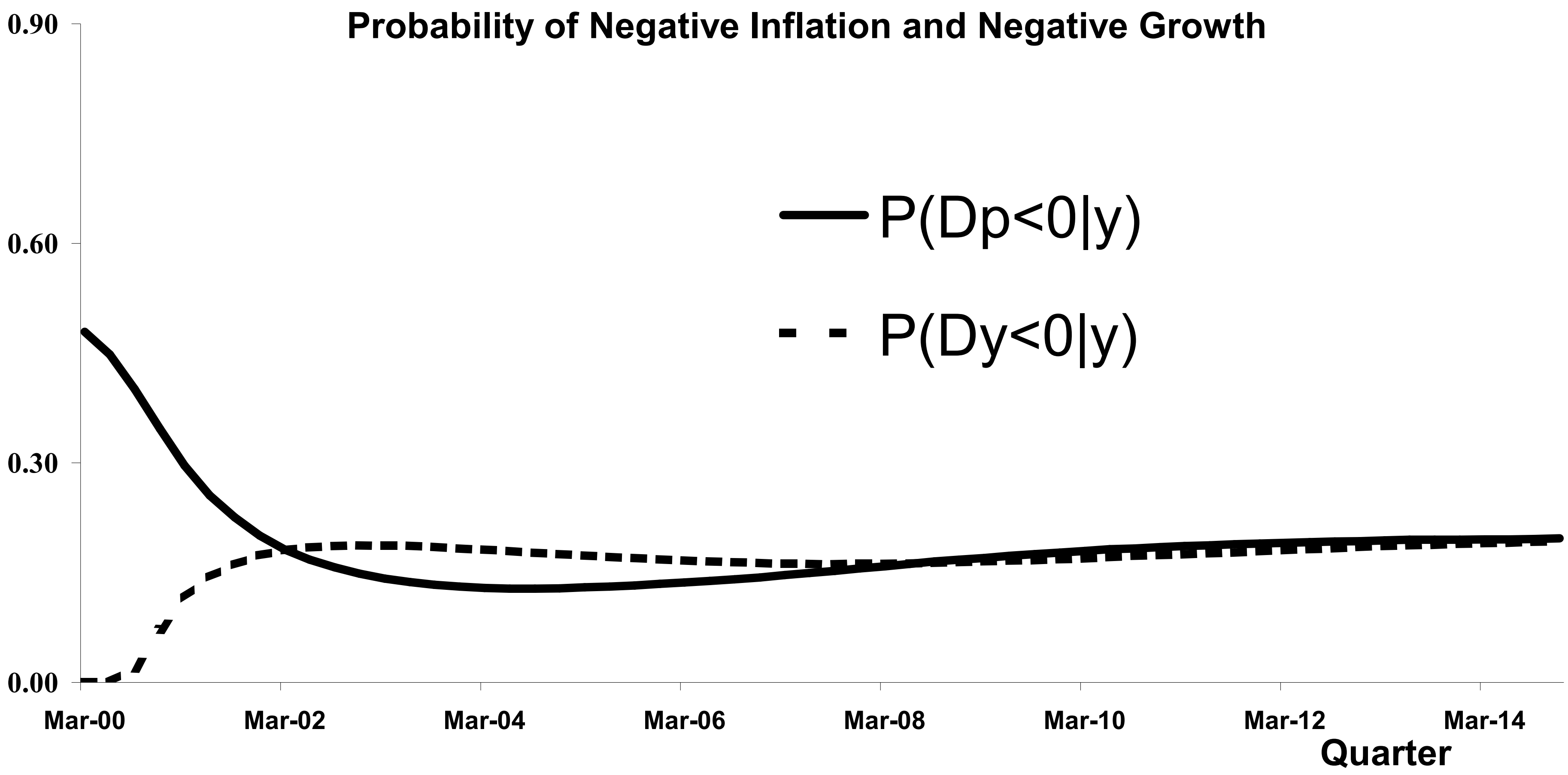
$$P\left(\Delta y_{T+h} < 0 \mid y\right)$$

Liquidity Trap

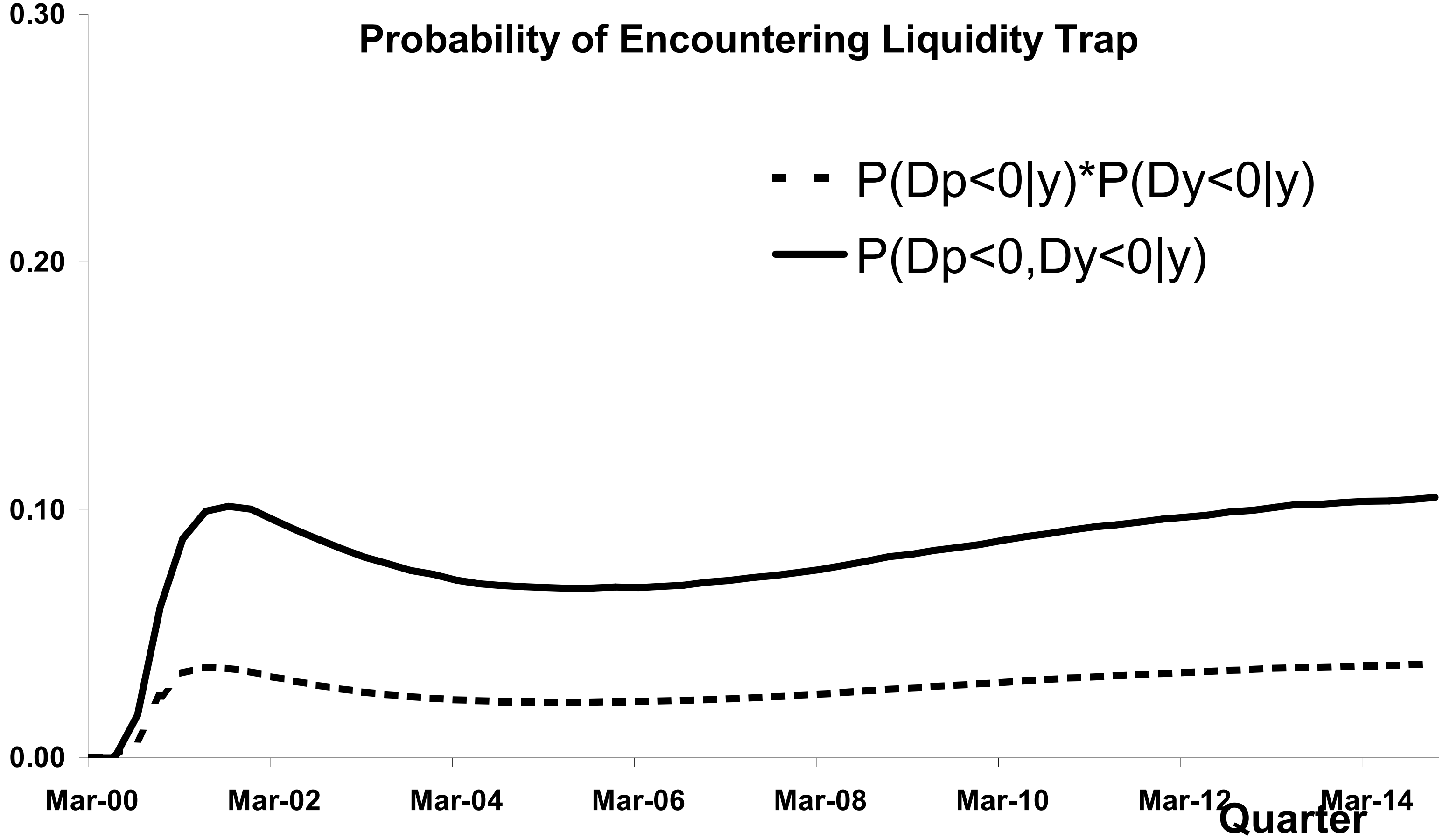
$$P\left(\Delta y_{T+h} < 0, \Delta p_{T+h} < 0 \mid y\right)$$

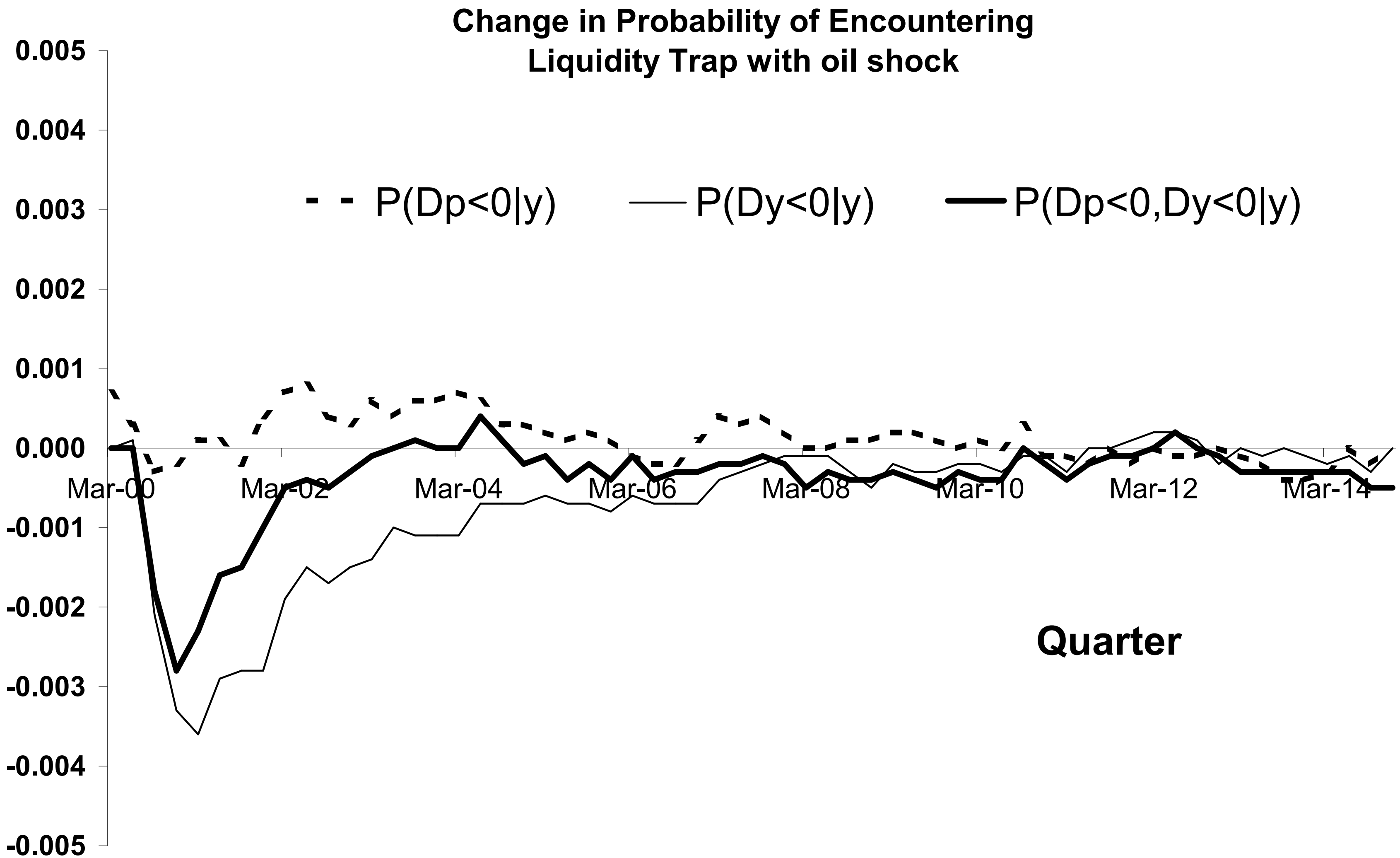
Liquidity Trap with oil price shock

$$P\left(\Delta y_{T+h} < 0, \Delta p_{T+h} < 0 \mid y, p_T^{oil} + \sigma_{oil}\right)$$



Probability of Negative Inflation and Negative Growth
0.90
0.60
0.30
0.00
P(Dp<0|y)
P(Dy<0|y)
Mar-00
Mar-02
Mar-04
Mar-06
Mar-08
Mar-10
Mar-12
Mar-14
Quarter


Probability of Encountering Liquidity Trap
P(Dp<0|y)*P(Dy<0|y)
P(Dp<0,Dy<0|y)
0.30
0.20
0.10
0.00
Mar-00
Mar-02
Mar-04
Mar-06
Mar-08
Mar-10
Mar-12
Mar-14
Quarter


Change in Probability of Encountering
Liquidity Trap with oil shock
0.005
0.004
0.003
0.002
0.001
0.000
-0.001
-0.002
-0.003
-0.004
-0.005
P(Dp<0|y)
P(Dy<0|y)
P(Dp<0,Dy<0|y)
Mar-00
Mar-02
Mar-04
Mar-06
Mar-08
Mar-10
Mar-12
Mar-14
Quarter

Summary and Topics for Further Research

Summary:

Our aim is unconditional inference, via $\pi\,(\Delta|y)$ and $E(\Delta|y)$, as we see this as of more use to our client.

We wish to average across a range of models and not be restricted in our priors.

Specifying a prior on the cointegrating space, rather than the cointegrating vectors.

We can employ a range of standard MCMC techniques.

Topics:

Time varying structure
Prior predictive information
Predictive odds