日本銀行金融研究所「金融研究」第10巻第1号（平成3年3月）

# Bayesian Vector AutoRegression

## ——その手法の整理と予測能力の検証

川　崎　能　典



## 1.　はじめに—目的・構成

本論文では、Doan, Litterman and Sims [1984]（以下 DLS）で提案されたモデルを、その本質的に重要な点だけに注目して、Bayesian Vector Autoregression（以下 BVAR）とはどのような procedure なのかを描出し、更に日本のマクロデータを用いてその予測能力を検討する。Vector Autoregression（ベクトル自己回帰）モデルについては改めてこの場で説明はしないが、[1) BVAR は予測用の VAR モデルである、と大ざっぱに考えてよい。手法のエッセンスは、2.の記述で十分言い尽くされていると思われるが、補論に紹介しておいたオリジナルの概要を見れば、DLS は必要以上に複雑に書かれていると感じられるであろう。

また、手法そのものは多分にアドホックであって、こうしたモデリングの根拠を彼らは empirical な予測能力に求めている。そもそも単純な（何の制約も置かずに方程式を推定する）VAR であっても、ある程度経済理論の立場からの批判は免れ得ない。VAR による表現を選んだ段階で、構造的な議論はもはや放棄されているからである。

経済学的解釈の付けにくい VAR モデルに更に（見方によってはアドホックな）制約を加える理由は一つ、モデルの推定に伴う母数過多（over-parametrization）の問題ゆえである。例えば DLS では米国のマクロ経済変数

本論文は著者が日本銀行金融研究所において委嘱されている研究活動の一環としてとりまとめたものである。本論文の作成にあたっては、矢島美寛（東京大学）、粕谷宗久（日本銀行金融研究所）、竹内惠行（福島大学）の各氏から有益なコメントを頂いた。また、本研究の早期段階においては、国友直人（東京大学）、北川源四郎（統計数理研究所）の両氏からご指導を頂いた。

1）本論文では、ベクトル自己回帰のことを VAR と、或いは何の制約も置かない単純なベクトル自己回帰という点を強調するときには OLS-VAR と記すことにする。

10系列（月次データ）を、ラグを6期までとって分析しているが、これを単純なVARで分析すると、誤差項の分散共散行列が仮に対角であると仮定しても、全部で600余りもパラメターを推定しなければならない。

このようにVARでは、一変量時系列と同じような感覚でモデリングを行うと、結果的に多くの母数を必要としてしまう。こうして推定されたモデルの予測能力がさほど高くないことは、極端な話、100個の観測値を100個のダミーで説明すればfitは完全であるがそのようなモデルが無意味で予測の用をなさないことを思い起こしてみれば直ちに理解できるであろう。

この点に対処すべく単純なVARを修正したものの一つにStructural VAR（SVAR）がある。SVARにおいては、分析に先立ち、経済理論の立場から変数の同時的関係に制約を課す（係数行列のある部分を0とおく）ことで推定すべき母数を減らしている。母数過多の解消という点では、BVARモデルもSVARと同じ方向にVARをmodifyしたものであると言えよう。[2] ではDLSはどのような制約を置いているのだろうか。

彼ら（DLS）のbeliefは「経済変数は程度の差こそあれ大概ドリフト付きの1階のランダムウォークプロセスだ」というものであり、モデルに対する制約もこの考えを反映したものとなっている。SVARとの比較で言うと、ある係数をdeterministicに0と置くのではなく、0の周りに正規分布しているという形で制約を加えるのである。途中の複雑な設定はともかく、結果から見れば、彼らの手法は10変数を10本のランダムウォークモデルにそれぞれ他のラグ付き変数の影響力も若干入れて予測するやりかたに他ならないのである。

本論文の構成を予め述べれば以下の通りである。

2.はBVAR予測方式の概説である。まず道具立てとして(1)でカルマンフィルター、(2)でTheil（タイル）の混合推定法（Mixed Estimation）を扱う。

カルマンフィルターとは時変パラメターモデルの一般的な定式化であり、一期分新たに観測値が得られるごとに未知母数（状態変数）の再推定を行い、一期先予測を繰り返すアルゴリズムである。

一方混合推定法とは、回帰モデルの推定に際して、回帰係数に確率的（nondeterministic）な線形制約を課す推定法である。この制約は「経済変数は程度の差こそあれ大概はドリフト付き一階のランダムウォークプロセスである」というbeliefを反映するものとなっている。

観測値に重きをおくか、先験的制約に重きをおくかは、すなわち回帰モデルの分散共分散行列と確率的線形制約の分散共分散行列の比率の問題であるが、BVARでは後者を未知母数として扱い、当該フィルタリング期間に「最良の予測」を得るようその未知母数を探索する。そのparametrizationを扱うのが(3)である。「最良の予測」といったが、予測精度の評価法については(4)で論じてある。

3.では、日本経済のマクロデータを用いてVARモデルを考え、とりわけ実質マネーサプライおよびその前期比伸び率に焦点をあ

2） ただ、SVARのmotivationが、まず誘導形を求めてからそれに対応する構造形モデルを求め、経済変数間の構造的解釈を試みることにあるのに対し、BVARのそれは予測能力の向上である。なお、SVARについては岩淵［1990］を参照。

て、どの程度の正確さで予測するか、四半期データで直近の5期間を予測させてみた。BVAR の予測能力については、theoretical にではなくあくまで empirical な議論しかできないのだが、少なくとも本論文で取り上げたケースでは、うまくいっていると結論してもよかろう。

2.、3.では手法や実証結果への論評は避けた。それらは4.にとりまとめてある。

## 2. BVAR Forecast Procedure

### (1) カルマンフィルター

BVAR では、経済現象をカルマンフィルターと呼ばれる可変パラメターモデルで定式化する。カルマンフィルターの手法そのものは特に目新しいものではないが、BVAR の基本的な tool の一つであるので、[3] どのような手法なのかをここで概観しておこう。基本的なカルマンフィルターモデル（一変量）は以下のように表される。

$$y_t = x_t\beta_t + u_t \tag{1}$$
$$\beta_t = \Phi\beta_{t-1} + v_t \tag{2}$$
$$(t = 1, \ldots, T)$$

ただし $y_t$ は被説明変数（スカラー）、$x_t$ は説明変数ベクトル $(1\times K)$、$\beta_t$ は係数ベクトル $(K\times 1)$ を表し、$u_t$ と $v_t(K\times 1)$ の各成分は平均0の正規分布に従いその分散は、

$$Var(u_t) = \sigma_u^2 \tag{3}$$
$$Var(v_t) = \sigma_v^2 R \tag{4}$$

であると仮定する。ここで $\Phi, R, \sigma_u^2, \sigma_v^2$ は既知、$\Phi, R$ は $K\times K$ 行列とする。すなわちパラメターが、連続的に加わるランダムショックにより確率的に変動すると仮定するのである。(1)式を観測方程式、(2)式を状態方程式という。

いま $t-1$期までの観測値が利用可能だとして、その $t-1$期までの情報を用いて推定した $\beta_{t-1|t-1}$ 並びにその分散共分散行列 $P_{t-1|t-1}$ が得られたら、$t$ 期における値を、

$$\beta_{t|t-1} = \Phi\beta_{t-1|t-1} \tag{5}$$
$$P_{t|t-1} = \Phi P_{t-1|t-1}\Phi' + \sigma_v^2 R \tag{6}$$

によって予測する。これらは(2)式より得られる自然な形である。一期先予測は(5)式を用いて、

$$\hat{y}_t = x_t\beta_{t|t-1}$$

として予測する。

次に $t$ 期の観測値が得られたら、$t$ 期での推定量を、$t-1$ 期までの観測値に基づくものと $t$ 期の観測値を利用する部分とに変形した次の更新方程式を用いて $\beta_{t|t}$、$P_{t|t}$ を推定する。

$$\beta_{t|t} = \beta_{t|t-1} + P_{t|t-1}x_t'(x_tP_{t|t-1}x_t' + \sigma_u^2)^{-1}(y_t - x_t\beta_{t|t-1}) \tag{7}$$
$$P_{t|t} = P_{t|t-1} + P_{t|t-1}x_t'(x_tP_{t|t-1}x_t' + \sigma_u^2)^{-1}x_tP_{t|t-1}. \tag{8}$$

この公式が便利なのは、観測値の増加に伴う行列演算の負荷を、既に得た推定量プラス新たな観測値に依存する部分という形で軽減していることにある。また、$\sigma_v^2 R=0$ かつ

3） カルマンフィルターに関する本節の説明は非常に概略的なものである。より詳しくは Harvey ［1989］、Brockwell and Davis ［1987］ 等を参照されたい。

$\Phi=I$（単位行列）の場合には、カルマンフィルターは逐次最小二乗法（recursive OLS）に帰着する（簡便さのため実用上こうする場合が多い）。

ところでカルマンフィルターのアルゴリズムでは、回帰係数ベクトルの初期値 $\beta_{t_0|t_0}$ 並びにその分散共分散行列 $P_{t_0|t_0}$ が必要である。従ってサンプル期間の一部を利用して、何らかの方法によりこれらの値を推定しなければならない。Bayesian VAR の特徴は、これらの推定に混合推定法（Mixed Estimation）を用いることである。では(2)でその混合推定法について一通りの解説を加えよう。

### (2) 混合推定法（Theil's Mixed Estimation）

カルマンフィルターとは、新たに観測値が得られるごとに(7)式、(8)式によって回帰係数の推定値ならびにその分散共分散行列を recursive に計算して一期先予測の繰り返しに利用するアルゴリズムである。この際問題となるのがアルゴリズムの初期値であるが、それはサンプルのある期間を通じて推定したものを与えるのが普通である。BVAR を特徴づける方法は、その初期値推定に用いられる混合推定法なのである。

まず下のような標準的回帰モデル、

$$y = X\beta + \epsilon \tag{9}$$

において、

$$E(\epsilon'\epsilon) = \sigma^2 V \tag{10}$$

としよう（観測値 $y$ と誤差項 $\epsilon$ は $N\times1$ ベクトル、$\beta$ は $K\times1$ 説明変数ベクトル、$X$ は $N\times K$ 説明変数行列であり、$V(N\times N)$ は非特異と仮定する)。[4] ここで観測値だけから $\beta$ を推定するなら、

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y \tag{11}$$

である（すなわち GLS estimator）。

しかし今ここで分析に先だって、未知パラメター $\beta$ の線形結合ないし未知パラメターそのものに関して、何らかの先験的情報が、non-deterministic に次のような制約式で表せるとしよう。

$$r = R\beta + v. \tag{12}$$

ただしここで、

$$E(v) = 0 \tag{13}$$

$$Var(v) = V_0 \tag{14}$$

である。（$R$ は既知の $K\times K$ 行列、$V_0$ も既知かつ非特異な $K\times K$ 行列と仮定する）。$r(K\times1$ ベクトル)の要素は、対応する $R\beta$ の各要素の予想値であり、$V_0$ の対角成分はその予想値に対する不確実性の尺度であると解釈できる。

混合推定とは、正規分布に従う観測値と事前分布が独立であるという仮定のもとで、(9)式と(12)式を組み合わせるのである。すなわち、

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix}\beta + \begin{bmatrix} \epsilon \\ v \end{bmatrix} \tag{15}$$

ただし、

4） 本節は混合推定をより一般的な枠組みで説明するため、誤差項の分散共分散行列を対角行列とは限らない形にしているが、実際の BVAR 分析では $V=I$（単位行列）である。

$$E\begin{pmatrix}\epsilon\\ v\end{pmatrix}=0 \tag{16}$$

$$Var\begin{pmatrix}\epsilon\\ v\end{pmatrix}=\begin{pmatrix}\sigma^2V & 0\\ 0 & V_o\end{pmatrix} \tag{17}$$

である。あとはGLS推定量の場合と全く同様の計算によって、

$$\hat{\beta}_M=\left(\frac{1}{\sigma^2}X'V^{-1}X+R'V_o^{-1}R\right)^{-1}\times\left(\frac{1}{\sigma^2}X'V^{-1}y+R'V_o^{-1}r\right) \tag{18}$$

を得る。ただし $\sigma^2$ が分析に先立って既知であるとは考えにくいので、$\hat{\beta}_M$ には $\sigma^2$ の何らかの推定値が使われることになる。例えば、観測値数が $N$、推定すべきパラメターが $K$ 個であるときに、何の制約も置かずに $\beta$ をOLSで推定した際の $\sigma^2$ の不偏推定量

$$\hat{\sigma}^2=\frac{1}{N-K}\,(y-X\hat{\beta})'V^{-1}(y-X\hat{\beta}) \tag{19}$$

を用いることが考えられるであろう。

### (3) 事前分布の与え方

混合推定とは、回帰モデルの推定にあたってその係数ベクトルに(12)式のようなnon-deterministicな制約を課す方法で、推定量は(18)式のように明示的に求めることができる。観測値に重きをおくか制約に重きをおくかは行列 $V_0$ の与え方次第であるが、BVARでは $V_0$ の要素を未知パラメターとして扱い、後で当該フィルタリング期間でのインサンプル予測が最良となるように $V_0$ の値を決めていく。本節で扱うのは、$V_0$ の各要素のparametrizationのやりかたである。

BVARはベイジアンと名はついているものの、分析の流れとしては通常のベイジアンとは若干方向が異なっている。ベイズ分析といったときに我々は、まず事前分布を想定する段階でその分布形はもちろん、平均・分散を特徴づけるパラメターも特定化し、その上でデータ観測後に得られる事後分布に関する損失関数の平均を最小化する値をもって未知パラメターの推定値とする、といった手順をイメージするであろう。ところがBVARでは、回帰係数ベクトルについての先験的制約は、平均1ないし0の正規分布というように「形」に関しては信念があるのだが、その「広がり」については信念がない。そこで事前分布の分散ないし標準偏差（すなわち制約の厳しさ）を三つの未知パラメターで表し、フィルタリング期間の予測誤差を極小化するという基準でその値を探索的に求めるのである。

具体的に標準偏差の定義式を述べる前に、それを構成する三つのパラメターが背負っている役割を一言ずつ説明しておこう。この種のパラメターをhyperparameter（超母数）と呼ぶことがあり、本論文でもしばしばこの用語を用いる。

- $\gamma$：先験的制約の厳しさ自体を全体的に制御
- $w$：Cross-lag項の影響をどの程度取り入れるかを制御
- $d$：2次以降のラグ付き変数の影響をどの程度取り入れるかを制御

今、$n$ 変数 $m$ 次のVARモデルを考えよう。このとき $S(i,j,k)$ を、第 $i$ 方程式における変数 $j$ の $k$ 期ラグ項（$x_{j,t-k}$）の係数が従う事前分布（正規）の標準偏差としよう。この $S(i,j,k)$ を、

$$S(i,j,k) \stackrel{\text{def}}{=} \frac{\{\gamma\cdot g(k;d)\cdot f(i,j)\}s_j}{s_i} \qquad (20)$$

と定義する。ただしここで、

$$f(i,j) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if} \quad i=j \\ w & \text{otherwise} \end{cases}$$

$$g(k;d) \stackrel{\text{def}}{=} k^{-d},\ g(0;d)=1$$
$$i,j=1,\ \ldots,\ n;\ k=1,\ \ldots,\ m$$

である。

(20)式右辺の各要素について順に説明を加えていこう。

$s_i$, $s_j$ は $x_{i,t}$, $x_{j,t}$ でそれぞれ一変量 AR$(m)$ を fit したときの残差系列の標準偏差の推定値である。これは cross-lag 項が説明変数に入った場合の変数間のスケールの違いを考慮したものである。

さて、hyperparameter 探索開始のためには、最初の $\gamma$、$w$、$d$ の値は分析者が指定しなくてはならない。最初に指定した $\gamma$、$w$、$d$ の良さについての評価（後述の Theil の $U$）が得られたら、それを改善するよう $\gamma$、$w$、$d$ の値を再探索していくのである。この出発点の選び方については3.(2)で触れるので、ここでは $\gamma$、$w$、$d$ の役割・特性についてもう少し説明を加えよう。

$\gamma$ は overall tightness と呼ばれているもので、$\gamma$ を小さくとれば事前分布は予め決めてあった事前分布の平均（つまりパラメターの予想値；0か1）の周りに厳しく縛られることになる。逆に $\gamma$ の値を非常に大きくとれば、制約自体は非常に緩いものになってしまうので、OLS 推定量に近付いていく。

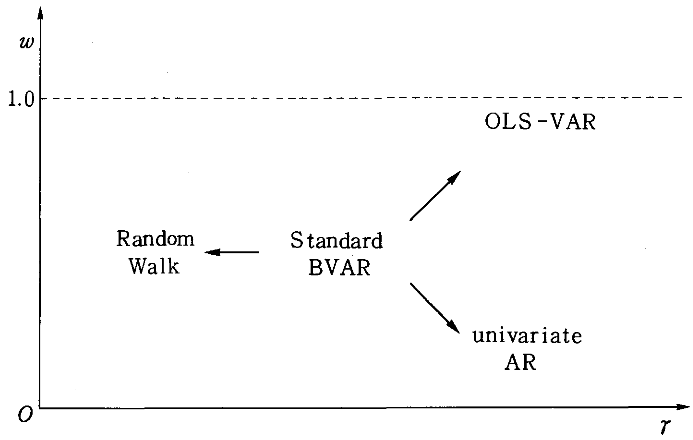
$w$ は $i\neq j$ の場合に指定するものだが、これは own-lag の $f(i,i)$ の値を1としたときに、cross-lag の $f(i,j)$ の値を own-lag のそれに対する相対比で与え、own-lag と cross-lag で分散の大きさに違いを持たせるための項である。なお、前節に述べた「経済変数は程度の差こそあれ大概ドリフト付き1階のランダムウォークプロセスだ」という DLS の考え方に沿えば、own-lag の1次項以外は、その係数が0の周りに厳しく縛られていくから、$w$ は1未満の値が妥当ということになろう。

$d$ はラグに伴う標準偏差の減衰を制御するパラメターである。現在から遠く離れた観測値については、その係数は0の周りに縛られていくわけだが、$d$ の値が0に近くなると $g(k;d)$ の減衰は緩やかに、$d$ の値が大きくなると急激に $g(k;d)$ は減少する（第2図を参照）。[5)]

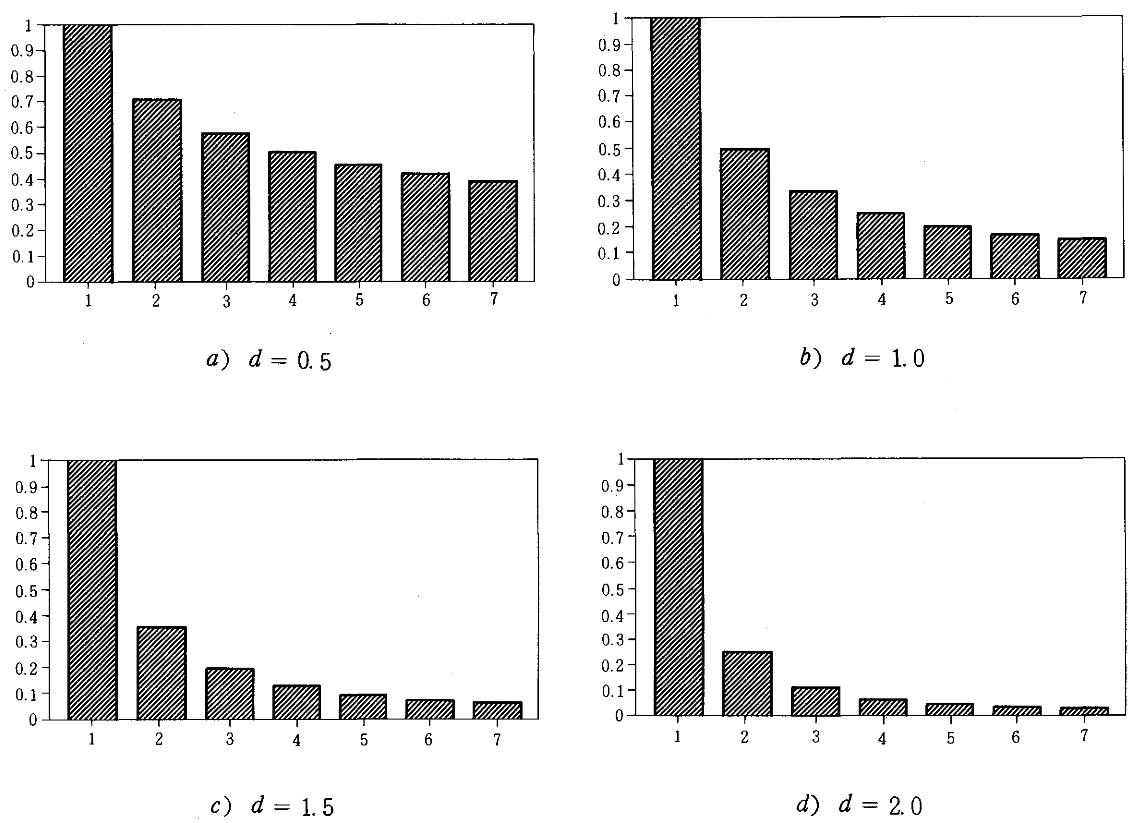
以上の説明を踏まえ、今 $d$ が正の適当な値（例えば1.0とか0.5）であるときに、$\gamma w$ の値の如何によって（とりわけ値が極端なとき）BVAR がどのようなモデルに近づくか整理したのが第1図である。

5) ラグについて注意しておくと、$d$ があるおかげで BVAR においては何期ラグをとるか予め深刻に考える必要がない。なぜなら、データの長さに問題がない限り多少長めにラグをとっておけば、$d$ に大きな値を指定することで、結果的にはラグを短くとることもできるからである。

第１図　BVARと他のモデルの位置関係




第２図　様々な$d$の値と$g(k\ ;\ d)$の動き



### (4) 簡単な例

混合推定法と事前分布の与え方について一通り説明が済んだところで、簡単な例として2変数 BVAR(2)モデルをとりあげてみよう:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} x_{t-2} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix}. \tag{21}$$

ただしここで $c_x, c_y$ は定数項、$\epsilon_t, \eta_t$ は独立に平均0の正規分布に従い、その分散共分散行列は、

$$Var \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix} \quad \forall\ t \tag{22}$$

と仮定する。ここで第1方程式 ($x$ − 方程式) は、

$$x_t = (x_{t-1},\ x_{t-2},\ y_{t-1},\ y_{t-2},\ 1) \begin{pmatrix} a_{11} \\ b_{11} \\ a_{12} \\ b_{12} \\ c_x \end{pmatrix} + \epsilon_t \tag{23}$$

と書ける。更に(23)式を $x_t = Z'_{t-1}\beta + \epsilon_t$ と記すことにすれば、今 $t = t_0$ までデータが利用可能だとして、

$$\begin{pmatrix} x_{t_0} \\ \vdots \\ x_2 \end{pmatrix} = \begin{pmatrix} Z'_{t_0-1} \\ \vdots \\ Z'_1 \end{pmatrix} \beta + \begin{pmatrix} \epsilon_{t_0} \\ \vdots \\ \epsilon_2 \end{pmatrix}, \tag{24}$$

つまり(24)式が、前節における $y = X\beta + \epsilon$ に相当する。この例では(10)式の $V$ は単位行列と仮定する。一方、パラメターに関する線形制約は、1次の own-lag の係数だけが平均1の正規分布に従い、他の変数は全て平均0の正規分布に従う、というものだったから、まとめて書けば以下のようになろう。

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} a_{11} \\ b_{11} \\ a_{12} \\ b_{12} \\ c_x \end{pmatrix} + v. \tag{25}$$

ただしここで $v$ については(13)式、(14)式が成立しており、この例に即して(14)式を具体的に記せば、

$$Var(v) = diag\left[\gamma^2, \left(\frac{\gamma}{2^d}\right)^2, \left(\frac{\gamma w s_i}{s_j}\right)^2, \left(\frac{\gamma w s_i}{2^d s_j}\right)^2, 10^{10}\right] \tag{26}$$

である。混合推定に現れる $r = R\beta + v$ の $R$ はここでは単位行列である。

$\gamma$、$w$、$d$ については(3)で説明した通りなのだが、ここでは定数項(の予想値)の分散について一言触れておかなければなるまい。$10^{10}$ という数字自体には意味はなく、十分大きな値をとることが重要なのである。定数項については、正規分布とはいっても殆ど非報知分布 (non-informative distribution) を想定していることになる。実際にはデータから標本平均を引きさってから $\hat{\beta}_M$ を推定し、それから定数項を推定すればこの項は不要である。

### (5) 予測精度の評価

BVAR で用いる tool の説明が一通り済んだところで、この節ではフィルタリング期間での予測精度の評価法に重点を移し、分析の流れに沿って procedure をながめることにしよう。

まず最初にやるべきは、暫定的にでも $\gamma$、$w$、$d$ を決定することである(その決定方法については3.(2)参照)。ともかく $\gamma$、$w$、$d$ の三つさえ具体的に決まれば、$S(i, j, k)$ が全

ての $i, j, k$ で確定して $V_0$ が数値的に得られ、混合推定によってフィルタリング期間直前までの VAR モデルが推定される。

こうして初期係数ベクトルが推定された時点で、$t_0+1$ 期の予測はこの係数を使ってなされる。そして新たに1期分観測値が得られるごとに更新された係数を用いて1期先予測を繰り返す (recursive OLS)。この際に用いられる予測誤差の指標は Root Mean Square Error (RMSE) を Cauchy-Schwarz の不等式を用いて unit-free にした指標で、Theil の $U$ と呼ばれている。

今サンプルサイズが $T$、フィルタリング期間は $t_0+1$ から $T$ までとしよう。このとき Theil の $U$ は次の式で定義される。

$$U=\frac{\sqrt{\frac{1}{T-t_0-1}\sum_{t=t_0+1}^{T}(\hat{Y}_t-Y_t)^2}}{\sqrt{\frac{1}{T-t_0-1}\sum_{t=t_0+1}^{T}\hat{Y}_t^2}+\sqrt{\frac{1}{T-t_0-1}\sum_{t=t_0+1}^{T}Y_t^2}} \qquad (27)$$

ここで $Y_t (t=1, ..., T)$ は観測値、$\hat{Y}_t$ は推定したモデルを使って事後的に予測した値 (シミュレーション値) である。分子は平均二乗誤差の平方根 (RMSE)、それを分母で割ることで $U$ の値が0と1の間に落ちるように規準化しているのである。$U=0$ なら fit は完全 (RMSE が0)、$U=1$ であれば fit は最悪 (この時は Cauchy-Schwarz の不等式で等式が成り立っている) ということになる。[6)]

こうして Theil の $U$ が得られたら、それで一回の探索は終了である。BVAR-procedure では、分析者の満足のいく $U$ の値を得るまで探索を繰り返すのである。どういう方向に探索を行うかを決める足掛りは必要であり、そのために $d$ を適当な値に固定した上、$\gamma$ と $w$ の組を4通りほど選びベンチマークとする。

ベンチマークの内容は具体的には3.(2)に示されるが、要するに第1図に登場した四つのモデルで Theil の $U$ を較べることで、どの方向に動かせばよいかの判断材料にするのである。

$\gamma$、$w$、$d$ ともパラメター集合の中を離散点でしか探索できないわけだが、どの程度細かいグリッドをとって探せばよいかについては、統一的な指標はなく、分析者によってまちまちである。ただ、あまり細かいグリッドで探索を行おうとすると、3次元パラメター空間内で各点を公平に扱いながら探索を行うことは計算量からいって事実上不可能である。

## 3. 予測能力検証

### (1) データセット

この章では日本経済のマクロデータを用いて行った予測について結果を報告する。まずデータセットから説明する。全て四半期デー

6) BVAR モデルの応用にあたっては、この Theil の $U$ の定義について若干の注意が必要である。Sims [1987]、Artis and Zhang [1990] 等をはじめ、BVAR 分析によく用いられているのが、回帰分析、時系列解析用演算パッケージ RATS Ver.3.00であるが、この RATS のプログラムではフィルタリング期間にはいると各方程式ごとに「BVAR で予測したときの RMSE と、$t_0$ 時点の実現値で毎期毎期予測したとき (naive, no-change forecast) の RMSE の比」が各期の 'Theil の $U$' として出力される。この定義は明らかに(27)式とは異なっており、この値を 'Theil の $U$' として用いるとしばしばおかしな (甚だしい場合は $U$ の値が1を超えるという、Cauchy-Schwarz の不等式からは有り得ない) ことが生じるが、これは RATS Ver.3.00 における Theil の $U$ の定義上の問題なのである。

| | $\gamma$ | $w$ | 特　　徴 |
|---|---|---|---|
| BR1 | 2.0 | 0.001 | 一変量 AR |
| BR2 | 0.1 | 0.001 | 一変量 BVAR ; near Random Walk |
| BR3 | 0.1 | 0.5 | 標準的 BVAR |
| BR4 | 2.0 | 1.0 | OLSVAR ; Unrestricted VAR |

タである。

RMSP：実質マネーサプライ。$M_2$+CD 平残、季調済を P で割った。

RGNP：実質 GNP。季調済。

P　：物価、GNP デフレータ。季調済名目 GNP を季調済実質 GNP で割って算出した。

R　：金利（5年もの金融債応募者利回り）

VOL：日経平均株価の変動率

データのサンプル期間は1967:1～1990:1、そのうち1967:1～1988:4までをrecursive OLS のための初期値の推定期間に使い、1989：1～1990：1の5期間を1期ずつ予測、5期平均で Theil の $U$ が極小となる $\gamma$、$w$、$d$ を探索する。

このデータセットを用いて2通りの予測を行った。BVAR による予測と OLS-VAR による予測である。また、今回の予測ターゲットは RMSP ならびにその前期比伸び率とした。なおここで注意すべきことは、RMSP の予測にとってベストな $\gamma$、$w$、$d$ が、害を及ぼさぬまでも他の変数の予測に何ら改善をもたらさないこともある、ということである。

ラグについては4期とした。最も BVAR においては hyperparameter $d$ の存在によって、最適なラグの大きさ自体が探索の対象となるので、ラグをどの程度とるかあまり神経質になる必要はない。もしパラメター探索の過程で $d$ があまりに小さくなる（例えば0.5を下回る）場合には、最初に決めたラグの大きさよりもう少し大きい値をとればよいのである。この点は BVAR の「実用上」のメリットと言ってもよいであろう。

また本論文では分析に、回帰分析・時系列解析用演算パッケージ RATS Ver.3.00を用いている。本論文に提示したやり方で実際に BVAR で予測を行ったり、結果を再生したいと思われる方は、RATS Manual の8.8（8章8節）～8.10、13.1～13.4、14.70、14.71、14.135等を重点的に参照されるとよい。

### (2) Hyperparameter 探索の手順

RATS Ver.3.00内に既に組まれてあったプログラム RUNTHEIL. PRC を若干改め、これを用いて4通りのパラメターセットで Benchmark Run（BR 1～4）を行い、探索の足掛りとした。4通りとも $d$ はデフォルト値の0、$\gamma$、$w$ については、順番に記せば上の表のようになる。

次に、Benchmark Run の結果を見つつ、適当に $w$ と $d$ を固定したままで Theil の $U$ を最小化する $\gamma$ を探す。$d$ は DLS に従い1に固定、$w$ は Benchmark Run で最小を達成したケースでの値、あるいは標準的 BVAR での値（0.5付近）に固定しておく。

上で最良の（と思われる）$\gamma$ が見つかったなら $\gamma$ をその値に固定、$d$=1.0 に固定したままで、Theil の $U$ を最小化する $w$ を探索す

る。最後に、$\gamma$、$w$ を上の値に固定した状態で $d$ の値をいろいろと変えてベストな $d$ を求めたら、ここで得られるモデルが最終予測用モデルである。また、当該フィルタリング期間に関してベストなインサンプル予測は既に行われている。

このようなパラメターの探索手順は、理論的にはアドホックというほかない。グリッド数の選び方・探索順序如何では、尤度が単峰型でないかぎりベストなモデルを選ぶ保証がないことは明らかである。ただ、DLS の提示した方法が、ある種の経済変数の予測問題については、理論的にはともかく経験的には実用に足ることを考えれば、このような数値的探索の有用性は認めざるを得ない。[7]

### (3) BVAR による予測

では2.(2)に記した順番で結果を見ていこう。次の表は Benchmark Run の結果である。これを見ると、89：1～90：1の5期間での予測モデルは、$\gamma$ が小さい方が良さそうだという見当はつく。

| | *Theil's U* |
|---|---|
| BR1 | 0.4160988 |
| BR2 | 0.3657719 |
| BR3 | 0.4919862 |
| BR4 | 0.4104005 |

まず $w$ と $d$ を「標準的な」位置（$w$=0.2、$d$=1.0 とした）に固定し、$\gamma$ を探索した結果が第1表である。標準的 BVAR では $w$=0.5 とするところだが、Benchmark Run の結果を見て若干小さめに0.2と設定した。$d$ については $d$=1.0 *or* 2.0 が経験的に言って良いあてはまりを持つという RATS Manual の示唆に従った。

$\gamma$=0.025 が一番良さそうだという結論を得た後に（第1表)、$w$ を探索した結果が第2表で、$w$=0.68 がもっともらしいとの結論に至った。$\gamma$、$w$ を上記のものに固定して decay parameter $d$ を探索した結果が第3表である。$d$=3.5 を過ぎると少数点以下4ケタでの改善となってしまうが、$d$ が大きすぎて後の解釈に不都合ということは一切ないので、ここでは探索した結果 Theil の $U$ の平均が一番小さかった $d$=6.5 を採用し、以上で最終予測用のモデルは決定された。

### (4) インサンプル予測結果——OLS-VAR との比較

まず、実質マネーサプライ(対数値)の89：1～90：1までの予測値を比較したのが第4表であり、第8表、第9表に状況が示されている。これに基づき実質マネーサプライの前期比伸び率の予測結果をまとめたのが第5表で

7） これまでのところ、こうした primitive な探索方法以外のパラメターサーチをした文献は一つしかない (Sims は最尤法で求めている)。参考までに他の文献での探索過程を紹介しておこう。DLS では、自らが設定した8個の未知パラメターのうち6個を先験的に固定し、比較的重要な2個のパラメターを4×4＝16の点で探している。中には、Artis and Zhang [1990] のように、探索を行わず標準的 BVAR に値をセットして、予測を行っただけのものもある。

第 1 表　γの探索

| γ | Theil's U | γ | Theil's U | γ | Theil's U |
|---|---|---|---|---|---|
| 0.01 | 0.3148126 | 0.04 | 0.2825749 | 0.3 | 0.4923796 |
| 0.015 | 0.2834675 | 0.05 | 0.3004101 | 0.325 | 0.4930551 |
| 0.02 | 0.2707122 | 0.09 | 0.3784021 | 0.35 | 0.4931194 |
| 0.022 | 0.2687319 | 0.1 | 0.3946121 | 0.375 | 0.4927565 |
| 0.024 | 0.2679050 | 0.15 | 0.4501679 | 0.4 | 0.4920932 |
| 0.025 | 0.2678567* | 0.175 | 0.4658322 | 0.425 | 0.4912274 |
| 0.026 | 0.2679960 | 0.2 | 0.4745037 | 0.45 | 0.4902149 |
| 0.028 | 0.2687835 | 0.225 | 0.4835343 | 0.475 | 0.4891099 |
| 0.03 | 0.2701386 | 0.25 | 0.4880636 | 0.5 | 0.4879446 |
| 0.035 | 0.2753972 | 0.275 | 0.4908357 | | |

第 2 表　*w* の探索

| *w* | Theil's U | *w* | Theil's U | *w* | Theil's U |
|---|---|---|---|---|---|
| 0.05 | 0.3993814 | 0.5 | 0.2678567 | 0.7 | 0.2649386 |
| 0.1 | 0.3721473 | 0.55 | 0.2663032 | 0.71 | 0.2649835 |
| 0.15 | 0.3425242 | 0.56 | 0.2660778 | 0.72 | 0.2650224 |
| 0.2 | 0.3181009 | 0.6 | 0.2654068 | 0.73 | 0.2650714 |
| 0.25 | 0.3003133 | 0.65 | 0.2649969 | 0.74 | 0.2651306 |
| 0.3 | 0.2880187 | 0.66 | 0.2649662 | 0.75 | 0.2651990 |
| 0.35 | 0.2796826 | 0.67 | 0.2649414 | 0.8 | 0.2656676 |
| 0.4 | 0.2740701 | 0.68 | 0.2649337* | 0.85 | 0.2663168 |
| 0.45 | 0.2703194 | 0.69 | 0.2649386 | 0.9 | 0.2671136 |

第 3 表　*d* の探索

| *d* | Theil's U | *d* | Theil's U | *d* | Theil's U |
|---|---|---|---|---|---|
| 0.5 | 0.3167890 | 2.0 | 0.2381781 | 4.5 | 0.2345449 |
| 0.75 | 0.2861045 | 2.5 | 0.2359649 | 5.0 | 0.2345109 |
| 1.25 | 0.2520206 | 3.0 | 0.2351218 | 5.5 | 0.2344945 |
| 1.5 | 0.2446630 | 3.5 | 0.2347714 | 6.0 | 0.2344865 |
| 1.0 | 0.2649337 | 4.0 | 0.2346164 | 6.5 | 0.2344865* |

第４表　RMSP（対数値）実測値と予測値

| *Period* | 実績値 | BVAR | OLS-VAR |
|---|---|---|---|
| 89 : 1 | 15.1365 | 15.1344 | 15.1314 |
| 89 : 2 | 15.1448 | 15.1448 | 15.1353 |
| 89 : 3 | 15.1701 | 15.1691 | 15.1483 |
| 89 : 4 | 15.1898 | 15.1925 | 15.1737 |
| 90 : 1 | 15.2303 | 15.2240 | 15.2081 |

第５表　RMSP伸び率（%）、実測値と予測値

| *Period* | 実績値 | BVAR | OLS-VAR |
|---|---|---|---|
| 89:1 | 2.04 | 1.84 | 1.54 |
| 89:2 | 0.84 | 0.83 | -0.11 |
| 89:3 | 2.52 | 2.45 | 0.34 |
| 89:4 | 1.97 | 2.27 | 0.37 |
| 90:1 | 4.05 | 3.47 | 1.85 |

ある。

グラフについて説明しておくと、まず第６表、第７表は初期値推定期間（68：1～88：4）での BVAR と OLS-VAR のフィットの様子を示したグラフである。こう見ると BVAR のほうが一見してあてはまりがよくない。しかし予測期間も含めた直近３年分の動きを見ると（第８表、第９表)、BVAR の方がこの場合明らかに予測能力が高い。通常の VAR は初期値推定期間全体では説明力が高いのに対し、１期先予測は比較的大きく underestimate している。そもそも BVAR はフィルタリング期間で局所的には最良な予測を行う初期値を与えるのだから、普通の VAR と較べてこの程度の差は生じてもおかしくはない。この場合、BVAR の予測能力が優れていたというよりも、RMSP という変数に関して「ドリフトつき１階のランダムウォーク」という制約が(唯一とは言わないが)適当なものだった、と結論づけたらよいであろう。

第６表、第７表と第８表、第９表を対比しつつ、もう少し検討を進めよう。第６表、第７表を見ると、石油ショック規模の大きな外生的ショックがもたらす変動に対しては、ドリフト付き1階のランダムウォークという単純なモデルでは完全に fit に「遅れ」が生じてしまい、このようなケースでの予測には役立たないことがわかる。

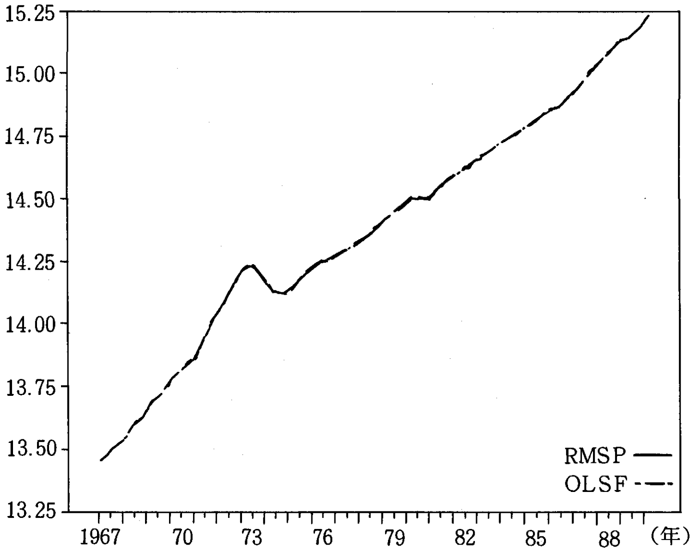
一方、第８表、第９表は89：1～90：1までのフィルタリング期間とその直前の数期間のグラフである。89：1～89：2にかけての伸び率の落込みは新短期プライムレート導入が、89：2～89：3にかけての伸長は新金融商品の導入が大きく関与していると思われるが、BVAR はこの程度の外生的要因による変動はモデルの時変構造の中でとらえていることになる。

なお、OLS-VAR との対比において注意すべき点を指摘しておこう。

BVAR においては、初期値推定期間は長く、フィルタリング期間は最後の数期間に、というのが原則であろう。これは理論的にそう言えるということではなく、検証結果からある程度類推してのことである。BVAR はフィルタリング期間直近の動きを数少ないパラメターで捉えてくれるところに優位性があるのであって、仮に長期に亘ってフィルタリングを行えばアウトサンプル予測の際に初期値の優位性は失われてしまうかもしれない。

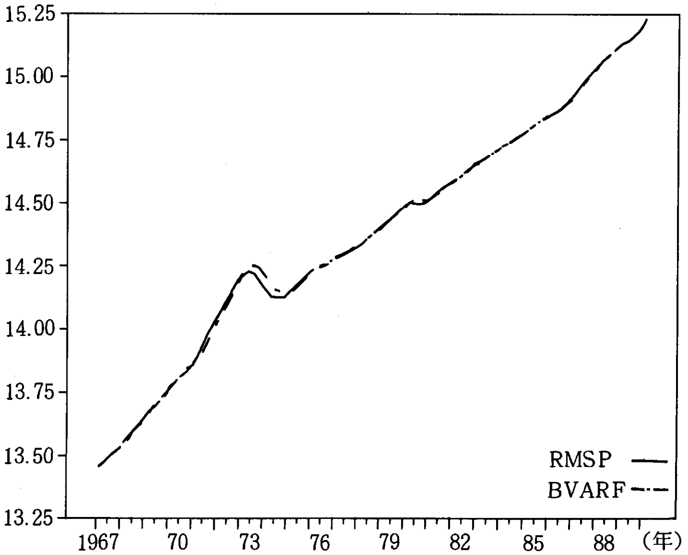
OLS-VAR の場合は事情が逆である。OLS で推定された初期値は、推計に利用したデータの動きを平均的に記述しており長期にわたる変動を観察するには都合がよいが、フィルタリングを始める時点での初期値がパラメターの真の値から大きく離れている場合（例えば過去の outlier の影響によって）には、

第6表 OLSFIT



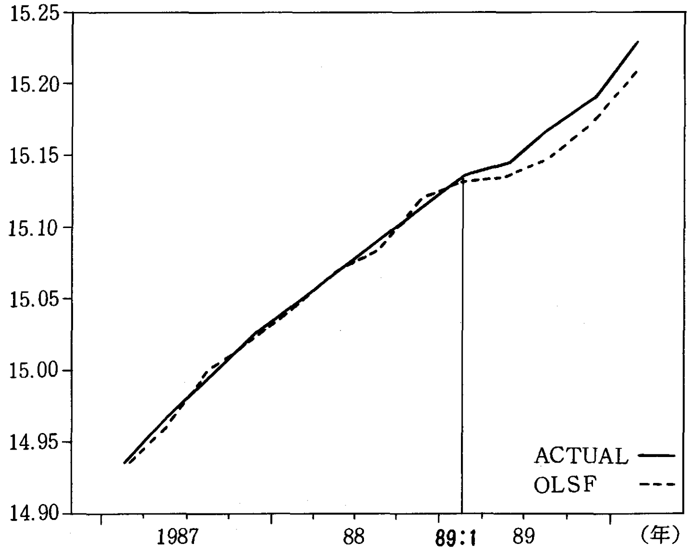

(注) 1. 68：1〜88：4での実質マネーサプライの実測地と内挿値。
2. グラフ縦軸は実質マネーサプライ（$M_2$＋CD）の対数値。
3. 実線が実測値、破線が内挿値。

第7表 B-VARFIT



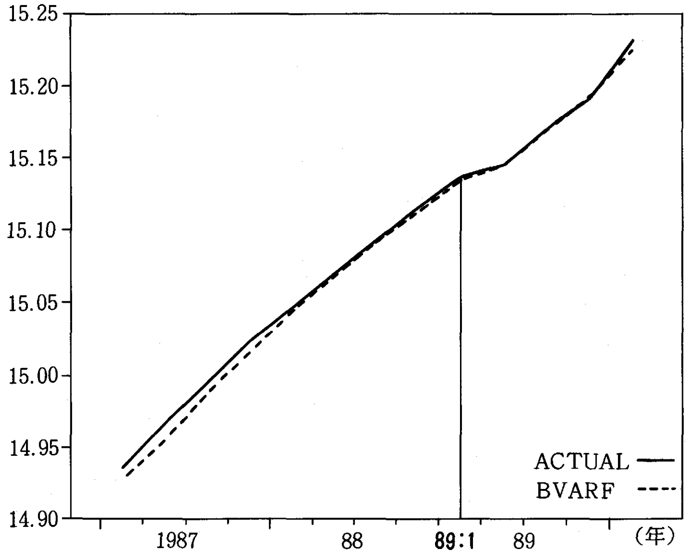

(注) 1. 68：1〜88：4での実質マネーサプライの実測地と内挿値。
2. グラフ縦軸は実質マネーサプライ（$M_2$＋CD）の対数値。
3. 実線が実測値、破線が内挿値。

第8表　OLSFIT




(注)1.　88：4までフィットさせ、89：1～90：1は1期先予測の繰り返し。
2.　グラフ縦軸は実質マネーサプライ（$M_2$+CD）の対数値。
3.　実線が実測値、点線が内挿値。

第9表　BVARFIT




(注)1.　88：4までフィットさせ、89：1～90：1は1期先予測の繰り返し。
2.　グラフ縦軸は実質マネーサプライ（$M_2$+CD）の対数値。
3.　実線が実測値、点線が内挿値。

数期にわたって1期先予測の精度が悪くなることがある。

## 4. まとめと問題点――結びに代えて

2.、3.では客観的な紹介を心がけたこともあり、本章では今回浮き彫りにされたBVARの特徴・問題点・限界等を論じてみよう。

まずは手法の位置づけについて、ネーミングという角度から再考してみよう。呼び名にこだわってみれば、BVARをベイジアンたらしめているのはhyperparameterの使用であると思われる。この点を明らかにしておくために、ここで簡単にいくつかのベイジアンの立場を確認しておこう。

ベイジアンの手法は、事前分布をどう解釈するかによって以下の四つの立場に分類される。第1は、事前分布は統計家の主観的な確率判断であるとするもので、このような立場を主観的ベイズ（subjective Bayes）という。第2は、繰り返し実験が可能な状況のもとで、事前分布を未知パラメターの実現の仕方の頻度分布と考える立場であり、これを経験的ベイズ（empirical Bayes）という。第3は、統計的決定理論の枠組みのなかで事前分布を便宜的な重み関数と捉えるやりかたで、あえて呼ぶとすれば決定理論的ベイズ（decision theoretic Bayes）ということになろう。[8)]

第4の立場は、未知パラメターを含むパラメトリックモデルに対し、それらの未知パラメターが満たすべき制約を表す確率分布を考え、その制約条件の強さを新たな未知パラメター（hyperparameter）として持つ事前分布を考える、というものである。一般的な名称かどうかは不明だが、このような立場を仮にここでは操作的ベイズ（operational Bayes）と呼んでおこう。[9)]

BVARの創始者達の認識がどうであるかは明らかでないが、ベイジアンというネーミングは、カルマンフィルターや混合推定という手法から来るというよりはむしろ、ここで分類した第4の意味でのベイジアンなのだと解釈するのが妥当であろう。[10)]

最後に、手法としてのBVARについて、多少批判的な立場から若干の評価を加えて本論文の結びとする。

① BVARの枠組みは多変量だが、ある変数についてドリフトつき1階のランダムウォークという制約が適当であればあるほど予測パフォーマンスは良くなり、推計式自体は一変量に帰着していく。もし全ての変数についてこの制約が適当であるとしたら、一つのBVARシステムに一組のhyperparameter setで全変数を精度良く予測できる理想的状況となる。

---

8） 統計的決定理論については、竹村［1989］ならびにその文献リストが有用。

9） このような立場については、赤池［1989］を参照されたい。

10） ベイジアンというネーミングについては、2.(2)の(12)式を事前分布とみて、混合推定法がベイジアンたる由縁かと判断される向きも多いであろう。しかし混合推定については、ベイズ的アプローチというよりは、（確率的）線形制約つき回帰モデルと理解すればそれで十分である。というのも、推定量は事後分布を使わずにより簡単な操作で明示的に導くことができるからである。さすればBayesian VARとは、アルゴリズムの初期値推定にOLSやGLSでなく混合推定を用いる、カルマンフィルターの一つのバリエーションであると解釈することもできる。

しかしこのときシステムは、殆ど一変量予測の「寄せ集め」に帰着し、多変量という最初の設定は意味を失ってしまうように思われる。つまりBVARは、その出発点において二つの相容れない事柄（多変量というフレームワークとドリフトつき1階のランダムウォークという制約）を抱えこんでいるのである。

この点をもって矛盾であるとする批判に強く抗するだけのものを、BVARはその内部に持ち合わせていない。しかし、仮にある変数の特定期間において、ドリフト付き1階のランダムウォークという制約を与えることが予測の見地から適当であるとしても、このことは当該変数が真にランダムウォークプロセスに従っていることを意味しない。従って、そのような意味での観測者の「留保」を尊重するのであれば、1次のown-lagの係数とドリフト項以外を全てdeterministicに0とおかずに、それらが予測に寄与する可能性を残しておいても直ちに不整合という訳ではあるまい。

② 2.(2)では混合推定法を、ある種の経済変数に対する事前の信念を確率的制約として表すためのツールとして捉えたが、別の角度から混合推定を眺めることができる。

ラグ4期のOLS-VARで、推定されたRMSPのラグ付き変数の係数和はほぼ1であった。このことから、もし今RMSPが真にドリフト付き1階のランダムウォークプロセスに従っているとしたら、OLS-VARでは多重共線が生じているとみてよい。そしてBVARは、hyperparameter探索の過程で自動的に多重共線性を回避していると考えられる。

ところで、多重共線性のベイジアン的解決として知られる手法にリッジ回帰（ridge regression）があるが、そのフレームワークを見ると混合推定法はリッジ回帰と形式的に同じ操作をしていることがわかる。具体的には2.(2)、(12)式で、$r=0$、$R=I$、$V_0$は単位行列の定数倍とすれば、mixed estimator (18)はridge estimator

$$\hat{\beta}_R = (X'X + kI)^{-1}X'y \tag{28}$$

に帰着するからである。

リッジ回帰では、適当に$k$を探索することで多重共線を避け推定量の安定性を確保するが、混合推定においては、$k$の選択が(18)式の$V_0$の選択すなわちhyperparameter $\gamma$、$w$、$d$の選択に相当しているのである。

推定量の安定性といったが、(18)式、(28)式のような推定量の族は、不偏性を放棄する代わりにOLS推定量に較べて小さい分散を得るというトレードオフが特徴である。リッジ回帰あるいは混合推定においては、$k$ないし$V_0$がある条件を満たす限り推定量の平均二乗誤差（MSE）をOLSのそれより小さくすることができる。[11)]

もっとも、モデルの予測能力を向上させることだけを念頭におくのであれば多重共

11) その条件については、例えばJugde, et al. [1985], p. 60, p. 914を参照。問題は、それらの条件が回帰モデルの真の母数に依存するため、代わりに何らかの推定量を使わざるを得ないことである。この真の母数に関する推定の正しさが保証されない限り条件そのものもstochasticであり、この場合のMSEの改善は解析的には保証されない。

線性は困難な問題ではなく、主成分回帰やリッジ回帰といったバイアスを持つ推定量の族によって回避できることは周知である。この点を踏まえて論じればBVARのオリジナリティーとは、統計学的に従来から認知されていた手法を組み合わせた一つの予測用 procedure に過ぎないことになる。そうだとすれば、評価すべきは手法としてのオリジナリティーの有無ではなく、procedure としての実用性であると考えるべきであろう。

③ 予測モデルとしては、ドリフト付き1階のランダムウォークは外生的変動に対して脆弱である、という意見もあるかもしれない。ただ3.の終わりにも述べたように、石油ショック規模の大きな変化ならともかく、大局的には単調増加傾向にあるマネーサプライの動きに若干の上下があっても、BVAR はそれに適応するだけの能力はあると言っていい。

その他当然批判があろう点として、hyperparameter の探索に非常に手間を要するにも拘らず探索の結果得られたモデルについてはベストなものであるという理論的保証がないこと、また混合推定の結果得られる方程式については何の経済学的解釈も導き出せないのは勿論、係数の推定値の有意性も議論できないこと、等があろう。

しかしこの手法の目的そのものが、解析的にというよりは経験的に精度のよい予測式を探すための実用的な procedure の提示という点にあるとすれば、このような意味での没理論性について深刻に批判しても始まるまい。手法としての実用性について納得できる範囲内で分析者を満足させる予想を提供しようというのが、この分析そのものの目的なのであるから。

## 補論. Doan-Litterman-Sims Procedure

BVAR は、北米のミネソタ大学周辺の数人の研究者を中心に考案・応用されている手法であって、広範に認知されているとは言い難い。また、どういった procedure をもって標準的な BVAR procedure とするかという点についても合意は存在しないようである。

数多くはない BVAR についての文献の中で、必ずと言っていいほど refer されるのが Doan, Litterman and Sims [1984] である。そこで本論文でも補論として DLS に記された手法の概要を紹介しておきたい。

DLS と本論文2.との差異は2点あり、一つは制約を課す時に用いる hyperparameter の数、もう一つは予測誤差の評価法である。本論文2.では hyperparameter の個数は3であったが、DLS では以下に述べる $\pi_1$～$\pi_8$ の8個である。ちなみに Sims [1990] は12個の hyperparameter を用意し、最尤法でそれらの値を決めているが、BVAR の本質は2.に登場した三つの hyperparameter に集約されていると思われる。また、本論文2.に述べた hyperparameter との関連で言えば、$\gamma$ が $\pi_5$ に、$w$ が $\pi_2$ に対応し、$d$ は1に固定されている。

### ① DLS の parametrization

では、事前分布（正規）の形状等を定める $\pi_1$～$\pi_8$ の八つのパラメターを、全く DLS 通りに書き出しておこう（第A-1表参照）。

では以下で更に具体的に、$\pi_1$～$\pi_8$ がどのような形で用いられるのかみていこう。

$n$ 変数、$m$ 次の VAR モデルを考えよう。

第A-1表　パラメターとその役割

| パラメター | その制御対象・役割 |
|---|---|
| $\pi_1$ | Relative tightness on own lags |
| $\pi_2$ | Relative tightness on lags of other variables |
| $\pi_3$ | Relative tightness on constant term |
| $\pi_4$ | Differential tightness among other variables |
| $\pi_5$ | Overall tightness |
| $\pi_6$ | Loneness on sums of coefficients |
| $\pi_7$ | Tightness on time variation |
| $\pi_8$ | Rate of coefficient decay toward proir mean |

ここで、$Z_{i,t-1}$ を $x_{it}$ の説明変数ベクトル（$(mn+1)\times 1$ ベクトル;定数項も説明変数に含めているため要素には1が入っている）、$\theta_{it}$ を第 $i$ 方程式の定数項も含めた係数を並べた行ベクトルとすると、第 $i$ 方程式は、

$$x_{it} = Z'_{i,t-1}\,\theta_{it} + \epsilon_{it} \qquad \text{(A-1)}$$

と書ける。以後、この $i$ 番目の方程式について prior の入れ方等を考察していく。

フィルタリング期間に入れば $\theta_{it}$ が時変であることは先に断わっておいた。では $\theta_{it}$ がどのように変化するのか、そのモデルを記述しよう。

$\theta_{i0}$ を時点0における初期ベクトルとする。今 $\theta_{i0}$ が、平均 $\bar{\theta}_{i0}$、分散共分散行列 $\Sigma_{i0}$ の多変量正規分布に従っているとする。すなわち、

$$\theta_{i0} \sim N_{mn+1}\,(\bar{\theta}_i,\ \Sigma_{i0})\,. \qquad \text{(A-2)}$$

ここで $\Sigma_{i\,0}$ はパラメター $\pi_1 \sim \pi_8$ の関数である。さて、$\theta_{it}$ は次のようなプロセスに従っていると仮定する。

$$\theta_{it} = \pi_8\theta_{i,t-1} + (1-\pi_8)\,\bar{\theta}_i + \mu_{it}, \qquad \text{(A-3)}$$

ただし、$\mu_{it} \sim N(o,\ \pi_7\ \Sigma_{i0})$, $\pi_7 > 0$, $\mu_{it}$ と $\epsilon_{it}$ は $\forall i, t$ で独立とする。$\pi_8$ は $\theta_{it}$ が prior の平均 $\bar{\theta}_i$ に収束する減衰率である。仮に $\pi_8=1$ であれば、第 $i$ 方程式の係数ベクトルがランダムウォークプロセスに従っていることになる。一方 $\pi_7$ は、パラメターベクトル $\theta_{it}$ に許される time variation の程度を示している。

$\pi_1 \sim \pi_5$ は $\Sigma_{i0}$ の中身、つまり個々の係数についての tightness condition を与えるパラメターである。$a^i_{i,k}$ $(k=1, 2, \ldots, m)$ を第 $i$ 方程式における own lag 変数の係数、$a^i_{j,k}$ $(k=1, 2, \ldots, m)$ を第 $i$ 方程式における cross -lag 変数の係数としよう。一つの方程式内では前者が $m$ 個、後者が $m(n-1)$ 個、あわせて $mn$ 個の係数が存在する。$a^i_{i,k}$, $a^i_{j,k}$ の prior の分散、ならびに定数項 $c^i$ の分散を以下のように parameterize する。

$$Var(a^i_{i,k}) = \frac{\pi_5 \cdot \pi_1}{k \cdot \exp(\pi_4 w^i_i)} \qquad \text{(A-4)}$$

$$Var(a^i_{j,k}) = \frac{\pi_5 \cdot \pi_2 \cdot \sigma_i^2}{k \cdot \exp(\pi_4 w^i_j) \cdot \sigma_j^2} \qquad \text{(A-5)}$$

$$(k = 1,\ 2,\ \ldots,\ m)$$

$$Var(c^i) = \pi_5 \cdot \pi_3 \cdot \sigma_i^2 \qquad \text{(A-6)}$$

未定義のままいきなり使用してしまったfactor から説明しよう。$\sigma_i^2$ ($i$=1, ..., $m$) は(A-1) で定義したVAR システムの誤差項 $\epsilon_t$ の分散である。当然読者はここでおかしいと思われたであろう。なぜなら通常の回帰分析であれば係数の推定の際に残差分析も推定される筈なのに、ここでは推定する前に既にしかも prior のパラメターの中にあらわれているからである。しかし、各変数ごとにデータの単位が異なることを考慮すれば、(A-7)式で scale factor として $\sigma_i^2$, $\sigma_j^2$ を入れざるを得ない。

予め $\sigma_i^2$ ($i$=1, ..., $n$) を決めるためにDLS が行っている方式は以下の如くである。今考えているモデルが VAR($m$) であったとしたら、各変数で1変量 AR($m$) をfit し、パラメターを最少二乗法で推定した際の残差分散を、あたかも多変量で推定した $\sigma_i^2$ ($i$=1, ..., $n$) であるかのように用いるのである。

次に $w_j^i$ であるが、これは DLS では相対比重 (relative weights) と呼ばれているもので、第 $i$ 方程式において、変数 $j$ のラグ付き変数が0であるかどうかについて、分析者の事前の知識を反映するためのものである。

(A-4) 式或は (A-5) 式の形からわかるように、$w_j^i$ を大きく指定すればするほど、その係数は prior mean に近いと思っていることになる。DLS では、たいていの変数について、

$$w_i^i = 0,\ w_j^i = 1\ (j \neq i)$$

として分析している。例外的なケースについて言及しておくと利子率、為替レートの2変数については、

$$w_i^i = 1,\ w_j^i = 2\ (j \neq i)$$

としている。このウエイトは、これらの変数がよりランダムウォーク的だというDLS の belief を反映している訳である。更に株価については、

$$w_i^i = 1,\ w_j^i = 5\ (j \neq i)$$

としている。これは株価がランダムウォークだという強い belief を反映している。しかしながら、2とか5という値については全く恣意的に定められている。

(A-4) 式と (A-5) 式を見れば、ラグ $k$ の値が大きくなるにつれて、prior の分散はどんどん小さくなっていくことも一目瞭然である。現在から遠くはなれた観測値については、その係数は0のまわりに縛られていくのである。

また、先に表にした各パラメターの役割も $\pi_1$〜$\pi_5$ については (A-4) 式〜(A-6) 式を見れば、うなづけるところであろう。

$\pi_7$, $\pi_8$ によって係数ベクトルの時系列としての大まかな構造（状態方程式）が $\pi_1$〜$\pi_5$ によって個々の係数の分散構造（状態方程式の分散共分散行列）が与えられたところで、DLS は更に次のような制約をモデルに課す : own-lag の係数の総和は1に近い筈であり、cross-lag 変数に関しては係数の総和は0に近い筈である。
第 $i$ 方程式における変数 $j$ (cross-lag 変数) の係数ベクトルを $\vartheta$ で表そう。このとき、

$$\psi = \frac{\pi_6 \cdot \sigma_j^2}{\sigma_j^2} \times \underbrace{(1,\ 1,\ \ldots,\ 1)'}_{\text{m 個}}$$

とおき、ダミー観測値 $u=\psi\vartheta'$ を導入し、新たに混合推定の制約式を加えるのである

($Var(u)=1$ と仮定)。

2.(4)の例を再び取り上げれば、次の線形制約式が新たに課されることになる。

$$1=\left(1,1,\frac{\pi_6\sigma_x^2}{\sigma_y^2},\frac{\pi_6\sigma_x^2}{\sigma_y^2},0\right)\begin{pmatrix}a_{11}\\b_{11}\\a_{12}\\b_{12}\\c_x\end{pmatrix}+u.$$

結局この制約は、再び各変数にランダムウォーク仮説を課しているに等しい（1階ないし2階のランダムウォークモデルで、説明変数の係数和が1となっていることを想起されたい)。

### ② モデルの選択

先にも述べたが、ここでのモデルの良さの基準は、特定のフィルタリング期間において予測誤差が小さいかどうかである。この節では、予測誤差を小さくすることがモデルの尤度を大きくするということと、実際にカルマンフィルターによる計算時に用いる分散共分散行列の形等を明らかにしておこう。

$t$ 期までデータが観測されたとして、第 $i$ 方程式の係数ベクトル $\theta_{it}$ が次のように多変量正規分布に従っていると仮定する。

$$\theta_{it}\mid t\sim N\ (\hat{\theta}_{it},\ \Sigma_{it}) \tag{A-7}$$

ここで '$t$' は、$t$ 時点の観測値も含むこれまでの全ての観測値。(9)式ならびに (A-5)式から、

$$\theta_{i,t+1}|t\sim N(\pi_8\hat{\theta}_{it}+(1-\pi_8)\bar{\theta}_i,\ \pi_8^2\Sigma_{it}+\pi_7\Sigma_{i0}). \tag{A-8}$$

ここで $\theta_{it}$ が得られたら、

$$\theta_{i,t+1}|t,\theta_{i,t}\sim N(\pi_8\theta_{i,t}+(1-\pi_8)\bar{\theta}_i,\ \pi_7\Sigma_{i0}). \tag{A-9}$$

また (A-1)式から当然、

$$x_{i,t+1}|t,\theta_{it}\sim N(Z_{it}'\theta_{t+1},\ \sigma_i^2). \tag{A-10}$$

この三つの式から、カルマンフィルターによる推定・予測方法は明らかである。$t$ 期まで観測値が得られた時の $x_{i,t+1}$ の予測量の分布は、

$$x_{i,t+1}|t\sim N\Big(Z_{it}'(\pi_8\hat{\theta}_{it}+(1-\pi_8)\bar{\theta}_i),\ \sigma_i^2+Z_{it}(\pi_8^2\Sigma_{it}+\pi_7\Sigma_{i0})Z_{it}'\Big) \tag{A-11}$$

である。

さて、このときの分散、つまり1期先予測の分散を $s_{it}^2$ と書こう。

$$s_{it}^2=Z_{it}(\pi_8^2\Sigma_{it}+\pi_7\Sigma_{i0})\Sigma_{it}'+\sigma_i^2 \tag{A-12}$$

$x_{i,t+1}$ が観測されたときの予測誤差を、

$$\hat{\epsilon}_{i,t+1}=x_{i,t+1}-Z_{it}'(\pi_8\hat{\theta}_{it}+(1-\pi_8)\bar{\theta}_i) \tag{A-13}$$

としよう。(A-13)から、$x_{i,t+1}$ の（条件つき）対数密度関数は、

$$\begin{aligned}&\log f(x_{i,t+1}\mid x_{it},\ \dots,\ x_{i0})\\&=-\frac{1}{2}\log 2\pi s_{it}^2-\frac{\hat{\epsilon}_{i,t+1}^2}{2s_{it}^2}\\&\propto-\frac{1}{2}\left(\log_{it}^2+\frac{\hat{\epsilon}_{i,t+1}^2}{s_{it}^2}\right)\end{aligned}$$

従って $T$ 個観測値が得られたとすれば、その尤度は上の式を $t$ に関して和をとったものに比例している。[12)]

12) hyperparameter $\pi$ 条件付きの下、標本点での密度関数の高さ。

ここでもし$\Sigma_{i0}$が$h$だけスカラー倍されると、$\Sigma_{it}(\forall t)$も同じく$h$だけスカラー倍される。これは、$\Sigma_{it}$が$\Sigma_{i0}$の定数倍で表せることからすぐにわかる。今更に、$\sigma_i^2$も同じく$h$だけスカラー倍されるとしよう。このとき$x_{i,t+1}$予測値ならびに予測誤差は不変である。

しかし尤度そのものは不変ではない。定数項の部分を除けば、

$$L(x_0,\ \ldots,\ x_{T-1})$$
$$=-\frac{1}{2}\left(\sum_{t=0}^{T-1}\log s_{it}^2+T\log h+\sum_{t=0}^{T-1}\frac{\hat{\epsilon}_{i,t+1}^2}{hs_{it}^2}\right) \tag{A-14}$$

となる。$h$は局外母数であるので、$h$に関して尤度を最大化してやるためにこれを$h$で微分して0おくと、

$$\hat{h}=\frac{1}{T}\sum_{t=0}^{T-1}\frac{\hat{\epsilon}_{i,t+1}^2}{s_{it}^2} \tag{A-15}$$

を得る。これを（A-14）に代入して、

$$L=-\frac{1}{2}\left[\sum_{t=0}^{T-1}\log s_{it}^2+T\log\left(\frac{1}{T}\sum_{t=0}^{T-1}\frac{\hat{\epsilon}_{i,t+1}^2}{s_{it}^2}\right)+T\right] \tag{A-16}$$

ここで$\bar{s}_i^2$を、$s_{i0}^2, \ldots, s_{i,T-1}^2$の幾何平均とすると、

$$L=-\frac{1}{2}\log(s_{i0}\ldots s_{i,T-1})$$
$$-\frac{T}{2}-\frac{T}{2}\log\left(\frac{1}{T}\sum_{t=0}^{T-1}\frac{\hat{\epsilon}_{i,t+1}^2}{s_{it}^2}\right)$$
$$=-\frac{1}{2}\log(\bar{s}_i^2)^T-\frac{T}{2}\log\left(\frac{1}{T}\sum_{t=0}^{T-1}\frac{\hat{\epsilon}_{i,t+1}^2}{s_{it}^2}\right)-\frac{T}{2}$$
$$=-\frac{T}{2}\log\left(\frac{1}{T}\sum_{t=0}^{T-1}\delta_t\,\hat{\epsilon}_{i,t+1}^2\right)-\frac{T}{2}$$

ただしここで、

$$\delta_t=\frac{\bar{s}_i^2}{s_{it}^2}$$

である。以上のことから1期先予測の二重誤差を荷重平均が小さければ小さい程、尤度は大きいということになる。

しかしながらDLSは、モデルの適合度の尺度にこの尤度を直接に利用してはいない（つまりこの尤度関数から、尤度を最大にする$\pi$を求める等の操作はしていない）。むしろここまでの議論から、予測誤差を極小にする基準を設けてそれに沿って予測パフォーマンスの良いパラメターを選べば、それは尤度の大きさ（最大とは限らないかもしれないが）を保証することにもなる、というロジックのようである。

### ③ 適合度の基準

DLSの適合度の基準は以下の通りである。$t$期において変数ベクトル$X$の$k$期先の値を予測したものを$\hat{X}_{t+k|t}$、実現値を$X_{t+k}$、予測誤差を$\hat{\epsilon}_{t+k|t}$と書く。即ち、

$$\hat{\epsilon}_{t+k|t}=\hat{X}_{t+k|t}-X_{t+k}$$

この予測誤差ベクトルをサンプルの各時点で計算し、その cross-product $\hat{\epsilon}\hat{\epsilon}'$を作る（これはサイズが方程式数×方程式数の行列である）。更にこの行列を$t$について和をとる。

$$E_k=\sum_{t=1}^{T-k}\hat{\epsilon}_{t+k|t}\hat{\epsilon}'_{t+k|t}$$

こうしてできた行列$E_k$の行列式の対数値がDLSの用いたfitの尺度である。この値が小さい方がモデルとしてすぐれていると判断される。

ただ、この評価方法自体ポピュラーなも

のではないので、$\log(det\ E_k)$ の値を示されて、それが改善しているからといって、それがどの程度の改善なのか直感的には把握し難い。

また、当り前のことだが $\log(det\ E_k)$ がいわば多変量の尺度であるのに対し、2.(5)に登場したTheilの$U$は各方程式ごとにこの値が計算されることに注意されたい。

以　上

[東京大学大学院経済学研究科]

**【参考文献】**


赤池弘次、「事前分布の選択とその応用」、鈴木雪夫・国友直人（編）、『ベイズ統計学とその応用』、第3章、1989年

岩淵純一、「金融変数が実態変数に与える影響について── Structural VAR モデルによる再検証」、『金融研究』、第9巻第3号、日本銀行金融研究所、1990年10月

竹村彰通、「ベイズ決定方式と許容性」、鈴木雪夫・国友直人（編）、『ベイズ統計学とその応用』、第4章、1989年

日本銀行調査統計局、「可変パラメターモデルとカルマンフィルターについて」、検討資料、1983年

Artis, M.J., and W. Zhang, "BVAR Forecasts of the World Economy," CEPR, No.380, 1990.

Brockwell, P.J., and R.A. Davis, *Time Series; Theory and Methods*, Springer Verlag, 1987.

Doan, T., R. Litterman, and C. Sims, "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, Vol.3, 1984, pp.1-144 (with comments).

Harvey, A., *The Econometric Analysis of Time Series*, 2nd ed., Philip Allan, 1990.

————, *Forecasting, Structural Time Series Models and Kalman Filter*, Cambridge University Press, 1989.

Judge, G.G., W.E. Griffith, R.C. Hill, H. Lütkepohl, and T. C. Lee, *The Theory and Practice of Econometrics*, 2nd ed., John Wiley, 1985.

Kahn, G.A., "The Changing Interest Sensitivity of the U.S. Economy," *Economic Review*, Federal Reserve Bank of Kansas City, Novemver 1989.

Litterman, R., "Forecasting with Bayesian Vectorautoregression-Five Years Experience," *Journal of Business and Economic Statistics*, Vol.4, No.1, 1986, pp.25-38.

Runkle, D.E., "Vector Autoregressions and Reality," *Journal of Business and Economic Statistics*, Vol.5, No.4, 1987, pp.437-454, (with discussions).

Sims, C., Comment on Runkle's Paper (above), 1987.

————, "Macroeconomics and Reality," *Econometrica*, Vol.48, No.1, 1980, pp.1-48.

————, "A Nine Variable Probabilistic Macroeconomic Forecasting Model," Invited Paper at Far Easter Meeting of Econometric Society, 1990.

Theil, H., *Principles of Econometrics*, John Wiley, 1971.

Todd, R.M., "Improving Economic Forecasting with Bayesian Vectorautoregressions," *Quarter Review*, Vol.4, Federal Reserve Bank of Minneapolis, Fall 1984, pp.18-29.