New Developments in VAR Modeling
Prepared for RATS User's Group Meeting
Trinity College, Dublin
April 22, 2004

By Tom Doan, Estima

Sims and Zha, "Error Bands for Impulse Responses," *Econometrica*, September 1999

Bernanke and Mihov, "Measuring Monetary Policy", *QJE*, August 1998

Faust, "The Robustness of Identified VAR Conclusions About Money", *Carnegie-Rochester Conference Series on Public Policy*, December 1998

Uhlig, "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," Tilburg University discussion paper, 2000.

Sims and Zha

Sims and Zha describe a procedure for estimating error bands for overidentified structural VAR's. The MC integration procedures widely used for computing error bands for impulse response functions aren't justified (in small samples) when applied to overidentified structural VAR's; that is, you can't draw a covariance matrix from an unrestricted distribution and then restrict it to a subspace.

Basic Idea: (Dates back to a footnote in Kloek and Van Dijk 1978 *Econometrica* 1978) Take the asymptotic Normal distribution obtained from computing the posterior mode (ML estimates if using a flat prior), and use that as the importance density. This relatively simple procedure was (apparently) unsuccessful in this case. SZ instead used a complex Metropolis-Gibbs sampling scheme. I'll argue that (a) the problems stem from the behavior of this particular model and (b) even then, the simpler importance sampling procedure could have been made to work.

Importance sampling: choose a density g with convenient MC properties and compute the expectation of h using

$$\begin{aligned} E_f(h(x)) &= \int h(x)f(x)dx \\ &= \int h(x)\big(f(x)/g(x)\big)g(x)dx \\ &= E_g\left(hf/g\right) \end{aligned}$$

In practice, we only work with the kernels f^* and g^* for the two densities, so the expected value is estimated by $\hat{h} = \sum h(x_i)w(x_i)\big/\sum w(x_i)$ where $w(x_i) = f^*(x_i)/g^*(x_i)$

If *f* has fatter tails than *g*, then the relative weights will be high on a handful of draws, and the estimate will converge very slowly. In high dimension problems (SZ have 19), it's possible for the tails to be fat in only a few directions, so a scheme which seems to work fine with 10,000 draws might finally run into difficulties if you go to 100,000.

The SZ model for the contemporaneous part is

$$\begin{bmatrix} a_{11} & a_{12} & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} & 0 & 0 \\ a_{31} & 0 & a_{33} & 0 & 0 & a_{36} \\ a_{41} & 0 & a_{43} & a_{44} & 0 & a_{46} \\ a_{51} & 0 & a_{53} & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \end{bmatrix} \begin{bmatrix} u_R \\ u_M \\ u_Y \\ u_P \\ u_U \\ u_I \end{bmatrix} = \begin{bmatrix} v_{MS} \\ v_{MD} \\ v_Y \\ v_P \\ v_U \\ v_I \end{bmatrix}$$

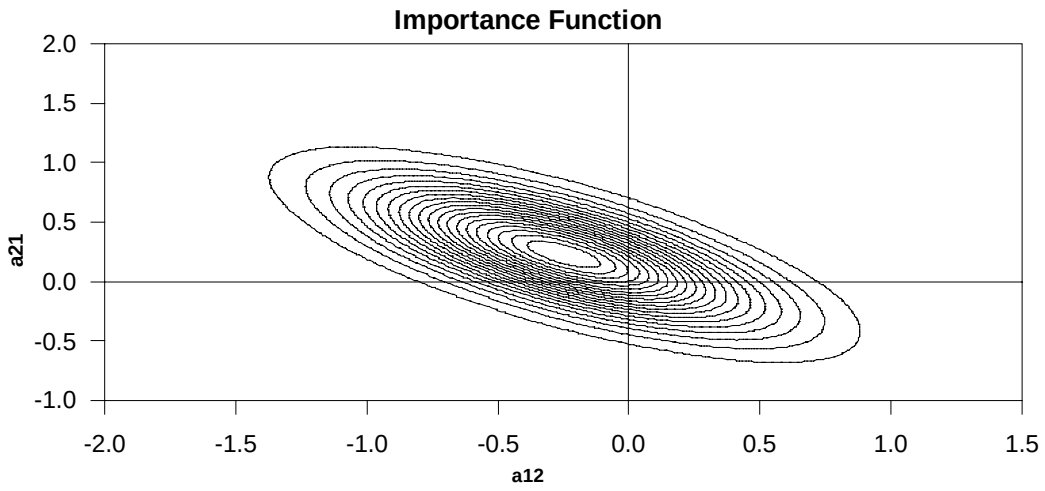
where the *v* processes are orthonormal. The first two equations are "identified" (economically, not statistically) as money supply and money demand, respectively. Statistical identification comes from requirement that v's be uncorrelated.

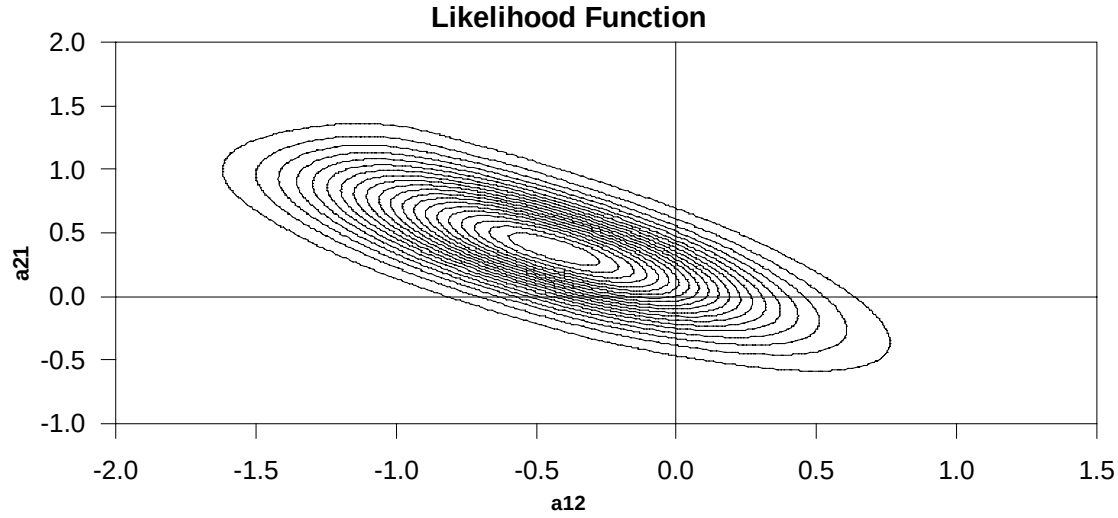
$Au = v \Rightarrow A\Sigma A' = \mathrm{I} \Rightarrow A'A = \Sigma^{-1}$. The model is overidentified, it won't be the case that this will hold exactly, but if we write out $A'A$ we get

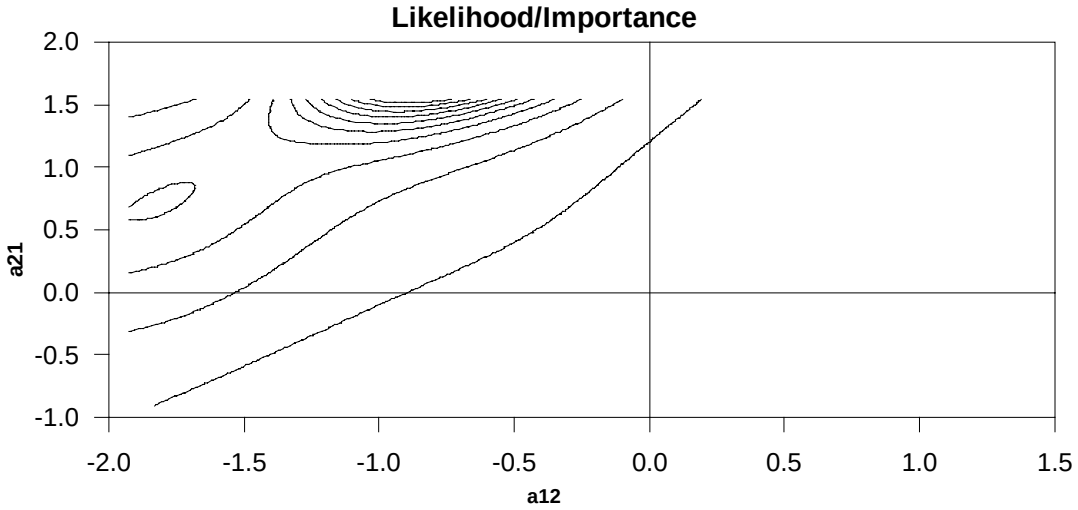
$$\begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} & a_{51} & 0 \\ a_{12} & a_{22} & 0 & 0 & 0 & 0 \\ 0 & a_{23} & a_{33} & a_{43} & a_{53} & 0 \\ 0 & a_{24} & 0 & a_{44} & a_{54} & 0 \\ 0 & 0 & 0 & 0 & a_{55} & 0 \\ 0 & 0 & a_{36} & a_{46} & a_{56} & a_{66} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} & 0 & 0 \\ a_{31} & 0 & a_{33} & 0 & 0 & a_{36} \\ a_{41} & 0 & a_{43} & a_{44} & 0 & a_{46} \\ a_{51} & 0 & a_{53} & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \end{bmatrix}$$

Multiply out the second row: the fifth and sixth columns produce 0. This leaves 19 conditions to solve for 19 free parameters, and the overidentifying test will just be testing whether those two elements of the inverse are, in fact, zero. The restrictions eliminate two conditions which might help pin down the difference between the first two equations.

Sims and Zha prefer the form of model given above where the model is normalized by making the orthogonal shocks have unit variances. The more common normalization is to peg one of the *a* coefficients in each row to one and allow the variances of the orthogonal shocks to be free. If we adopt that form, we get the following contours for the marginal joint density of the importance and the likelihood (smoothed on a 100x100 grid) of the two key coefficients:







This is the ratio of the likelihood to the importance function, that is, the weights. The weights are much higher in the one area, and this is likely to be a bit conservative because of the smoothing. It's also clear from the likelihood that the "economic identification" of the two equations is tenuous at best. A non-trivial part of the likelihood is outside the northwest quadrant.

SZ never describe the procedure for doing importance sampling on the model with standard normalization. With *n*=number of variables, *p*=number of regressors per VAR equation, and *T*=number of data points, the marginal density for Σ (using the Jeffrey's prior) $|\Sigma|^{-(n+1)/2}$) is

(1) $$|\Sigma|^{-(T-p)/2} \exp\left(-\frac{T}{2} trace\, \Sigma^{-1}\hat{\Sigma}\right)$$

The model states that

(2) $$\Sigma = A^{-1}\Lambda A^{-1\prime}$$

where Λ is diagonal. SZ point out that this doesn't not necessarily give an integrable density, that is, we can't necessarily use a flat prior on A and Λ. The prior which suggests itself is proportional to

(3) $\left|\Lambda\right|^{-\delta}$

Rewrite (1) as

(4) $\left|AA'\right|^{(T-p)/2} \Lambda^{-(T-p)/2-\delta} \exp\left(-\frac{T}{2} trace\, \Lambda^{-1}(A\hat{\Sigma}A')\right)$

By inspection

(5) $\Lambda_i^{-1} \mid A \sim \Gamma\left(\frac{T}{2}(A\hat{\Sigma}A')^{-1}_{ii}, \frac{T-p}{2}+\delta+1\right)$

If we integrate over Λ, we'll get an integrating constant proportional to

(6) $\prod_i \left(\frac{T}{2}\left(A\hat{\Sigma}A'\right)_{ii}\right)^{-\left(\frac{(T-p)}{2}+\delta+1\right)}$

It can be shown that a sufficient condition for this to be integrable over A is that $\delta > \frac{n-3}{2}$, somewhat less stringent than the $\delta = \frac{n+1}{2}$ given in SZ. (There are cases where it's integrable with a flat prior on Λ as well.)

Because $\Lambda \mid A$ has a convenient distribution, the importance sampling can be confined just to A. Given a value of δ, locate the posterior mode for A. Take the AN density. Now there are two adjustments which can help the behavior of the importance sampling procedure:

(a) Scale up the covariance matrix (slightly)

(b) Use a multivariate t rather than a Normal

Both of these work to fatten the tails. If the likelihood is not too ill-behaved, this will likely produce a respectable yield of the MC process. This is the setup code:

```
compute delta=3.5
cvmodel(parmset=simszha,dfc=ncoef,pdf=delta,$
  dmatrix=marginalized,method=bfgs) vmat afrml
dec rect saxx
```

axbase is the maximizing vector of coefficients.

```
compute [vector] axbase=%parmspeek(simszha)
```

saxx is a factor of the (estimated) inverse Hessian at the final estimates. This gets scaled slightly to fatten up the tails a bit.

```
compute saxx=1.2*%decomp(%xx)
```

scladjust is used to prevent overflows when computing the weight function

```
compute scladjust=%funcval
```

nu is the degrees of freedom for the multivariate Student used in drawing A's

```
compute nu=30.0
```

And this is the working code inside the loop:

Do a draw for the coefficients from a multivariate t density and "poke" it back into the parmset so the AFRML can get the new values.

```
compute grandom =nu/%rangamma(nu)
compute au      =%ran(sqrt(grandom))
compute %parmspoke(simszha,axbase+saxx*au)
```

Compute (log kernels of) the true marginal posterior density

```
compute a       =afrml(1)
compute dhat    =a*vmat*tr(a)
compute ddiag   =%xdiag(dhat)
compute pdensity=.5*(%nobs-ncoef)*log(%det(a*tr(a)))-$
 (.5*(%nobs-ncoef)+delta+1)*%sum(%log(ddiag))
```

and the importance function.

```
compute idensity=-((nu+nfree)/2.0)*log(nu+%dot(au,au))
```

Compute the weight value by exp'ing the difference between the two densities, with scale adjustment terms to prevent overflow.

```
compute weight  =exp(pdensity-scladjust-idensity-$
 ((nu+nfree)/2.0)*log(nu))
```

Conditioned on A, make a draw for the D matrix

```
ewise d(i)      =(%nobs/2.0)*$
    ddiag(i)/%rangamma(.5*(%nobs-ncoef)+delta+1)
```

Combine D and A to generate the draw for a factor of sigma.

```
compute swish   =inv(a)*%diag(%sqrt(d))
```

The following are the "effective sample sizes" ($\left(\sum w_i\right)^2 \Big/ \left(\sum w_i^{\,2}\right)$) for 10000 draws of the importance sampler for different settings for the degrees of freedom of the t and scale factors for the (matrix square root of the) asymptotic variance of the estimators. (Draws were done separately for each setting). Yields in the 20% range are fairly good for a model that isn't as well-behaved as we might like. For a smaller and more sharply estimated model, we would expect results much better than this.

	Normal	20	10	5	1
1.0	13	446	2881	1455	695
1.1	1067	660	1255	2071	1217
1.2	505	1633	2204	2201	633

Bernanke-Mihov

Bernanke and Mihov attempt to isolate the monetary policy shock in a six variable model. This includes six variables: three macro variables representing the overall economy (GDP or a proxy, deflator or a proxy and commodity price index) and three representing the monetary sector: Federal Funds rate, total reserves and non-borrowed reserves. The first step in the SVAR is to "sweep" out the contemporaneous effects of the macroeconomic sector, thus assuming that none of the shocks originating in the monetary sector have a contemporaneous effect on the macro variables. This can be done most easily using the "%SWEEP" function or one of its relatives:

```
compute sweepvcv=%sweeptop(%sigma,NNONPOLICY)
compute [symm] moneyvcv=%xsubmat(sweepvcv,$
  NNONPOLICY+1,NVAR,NNONPOLICY+1,NVAR)
```

This isolates the 3x3 matrix of desired residual monetary sector shocks. The authors then propose a general class of SVAR models on the monetary variables.

$$\begin{bmatrix} 1 & 0 & \alpha \\ 1 & -1 & -\beta \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} u_{TR} \\ u_{NBR} \\ u_{FF} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \varphi_d & \varphi_b & 1 \end{bmatrix} \begin{bmatrix} v^d \\ v^b \\ v^s \end{bmatrix}$$

As written, this is underidentified by one parameter. The structure includes as restrictions upon this a number of possible "reaction" functions for v^s. All but one of these result in a model which is overidentified by one.

This is done using the "A-B" form in CVMODEL.

```
dec frml[rect] ab phi
nonlin alpha beta phid phib
frml ab = ||1.0,0.0,alpha|1.0,-1.0,-beta|0.0,1.0,0.0||
frml phi = ||1.0,0.0,0.0|0.0,1.0,0.0|phid,phib,1.0||
*
disp 'JI (Just Identified)'
```

```
nonlin alpha=0 beta phid phib
compute phid=1.0,phib=-1.0,beta=0.0
cvmodel(pmethod=simplex,piters=10) moneyvcv ab phi

disp 'FFR Model'
nonlin alpha beta phid=1 phib=-1
cvmodel(pmethod=simplex,piters=10) moneyvcv ab phi
*
disp 'NBR Model'
nonlin alpha beta phid=0 phib=0
compute alpha=.03,beta=.014
cvmodel(pmethod=simplex,piters=100,factor=f) moneyvcv ab
phi
*
disp 'NBR/TR Model (Orthogonalized NBR)'
nonlin alpha=0 beta phid phib=0
cvmodel(pmethod=simplex,piters=10) moneyvcv ab phi
*
disp 'BR Model'
nonlin alpha beta phid=1 phib=alpha/beta
cvmodel(pmethod=simplex,piters=10) moneyvcv ab phi
```

The authors also estimate a Markov switching model for the just identified SVAR model. A minor quibble: the initial step of running the lagged VAR and working off the residuals isn't really justified here, because the efficiency of equation by equation OLS depends upon a covariance matrix fixed over the sample, which we won't have here. However, given that there are around 500 coefficients in the VAR, it's a safe bet that it won't be feasible to allow them to vary between the states.

A somewhat more significant quibble is that they allow the three structural parameters to vary between states, but not the variances. It's highly unlikely that there could be an interesting shift in the structure without the variances of the orthogonal shocks changing as well. The restriction of fixed variances will be overwhelmingly rejected by the data.

The covariance matrix of the policy sector no longer forms sufficient statistics for the residuals; we need the residuals transformed to sweep out the contemporaneous effects of the other three. The rectangular block of sweepvcv in the rows for the policy variables and columns for the nonpolicy has the -1 x the coefficients of the required transformations. Since we're supposed to subtract off anyway, you can just apply those values directly:

```
dec vect[series] upolicy(NPOLICY) unon(NNONPOLICY)
do i=1,NNONPOLICY
   set unon(i) = u0(i)
end do i
do i=1,NPOLICY
   compute [vector] sweepco=%xsubmat(sweepvcv,$
        NNONPOLICY+i,NNONPOLICY+i,1,NNONPOLICY)
   set upolicy(i) date1(IPERIOD) date2(IPERIOD) =
        u0(NNONPOLICY+i)+%dot(sweepco,%xt(unon,t))
end do i
```

Since the variances won't concentrate out, we switch to an alternative form which takes the *u*'s to an identity matrix. The standard deviation scale factors will go into the "B" part of the Au=Bv expression.

Note that the Markov switching procedures all use an $(n-1)xn$ representation for the free parameters in the transition matrix, hence the dimensions of p.

```
dec rect p(1,2)
nonlin p beta phib_1 phid_1 d1_1 d2_1 d3_1 $
   phib_2 phid_2 d1_2 d2_2 d3_2
source c:\rats\600exam\markov.src
dec frml[rect] phi_1 phi_2
frml ab = ||1.0,0.0,0.0|1.0,-1.0,-beta|0.0,1.0,0.0||
frml phi_1 = $
||d1_1,0.0,0.0|0.0,d2_1,0.0|d1_1*phid_1,d2_1*phib_1,d3_1||
frml phi_2 = $
||d1_2,0.0,0.0|0.0,d2_2,0.0|d1_2*phid_2,d2_2*phib_2,d3_2||
*
```

The sigma matrices and determinants depend only upon the coefficients, and not data, so we compute the (inverse of the) sigma matrix during the initialization, along with the ergodic probabilities of the two states.

```
frml init = (pstar=%mcergodic(p)),$
            (siginv_1=%innerxx(inv(phi_1(1))*ab(1))),$
            (siginv_2=%innerxx(inv(phi_2(1))*ab(1))),$
            (det_1=log(abs(%det(siginv_1)))),$
            (det_2=log(abs(%det(siginv_2))))
*
function StateF time
type vector StateF
type integer time
*
local vector u
compute u=%xt(upolicy,time)
compute StateF=$
||1.0/sqrt(2*%pi)*exp(.5*det_1-.5*%qform(siginv_1,u)),$
  1.0/sqrt(2*%pi)*exp(.5*det_2-.5*%qform(siginv_2,u))||
end

compute p=||.5,.5||
compute phid_1=0.0,phib_1= 0.0,$
  d1_1=sqrt(moneyvcv(1,1)),d2_1=sqrt(moneyvcv(2,2)),$
  d3_1=sqrt(moneyvcv(3,3))
compute phid_2=1.0,phib_2=-1.0,$
  d1_2=d1_1,d2_2=d2_1,d3_2=d3_1
*
set phist = 0.0
frml logl = $
  f=StateF(t),pstar=%msupdate(f,%mcstate(p,pstar),fpt),$
  phist(t)=pstar(1),log(fpt)
maximize(start=init,trace,pmethod=simplex,piters=50,$
 method=bhhh,iters=400) logl date1(iperiod) date2(iperiod)
graph
# phist date1(iperiod) date2(iperiod)
```

Faust and Uhlig

Faust and Uhlig both eschew the development of a full structural VAR and instead see what inference can be drawn by looking at individual shocks. The root behind this is the following, part of which is implicit (though unstated) in both author's results:

Proposition. If Σ is an nxn positive definite symmetric matrix, and **x** is a non-zero $nx1$ vector, then

a) there exists a factor of $\Sigma = \mathbf{FF}'$ where the first column of **F** is a scale multiple of **x**

b) there exists a factor of $\Sigma = \mathbf{P}^{-1}\mathbf{P}^{-1\prime}$ where the first row of **P** is a scale multiple of $\mathbf{x}'$

Proof. For (a) factor $\Sigma = \mathbf{SS}'$ (any factor will do). Generate an orthonormal matrix **U** with a scale multiple of $\mathbf{S}^{-1}\mathbf{x}$ as the first column. Then **SU** gives the desired factor. For (b), generate an orthonormal matrix **V** with a scale of $\mathbf{S'x}$ as the first column. **SV** is the desired factor.

Forcing a column fixes the initial responses. Forcing a row in the inverse fixes as one of the orthogonalized shocks a particular linear combination of the non-orthogonalized innovations. For instance, in a two variable system, x={1,1} applied to (a) would mean that the innovation being set would hit both variables equally in the first period. Applied to (b) means that the innovation is the sum of the non-orthogonal innovations in the two variables. These factorizations can be computed using the ForcedFactor procedure.

A well-known example of this is the Blanchard-Quah factorization. While BQ name the two shocks as "demand" and "supply", they're actually demand and whatever shock is required to complete an orthogonal factorization. If one were to apply the above

procedure (to the 2x2 case), if **C(1)** is the long-run response matrix, the "demand" shock is defined to be a solution to $\mathbf{C}(1)\mathbf{x} = \begin{bmatrix} 0 \\ * \end{bmatrix}$. In their 3x3 model, King, Plosser, Stock and Watson define a shock which loads equally onto all three variables long-run. You can solve $\mathbf{C}(1)\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ for x to determine that.

Faust's paper is based upon the following: given any chosen (orthonormal) factorization of the covariance matrix, the fraction of variance of a variable at any horizon explained by a linear combination α of the orthonormal shocks can be written as the quadratic form $\alpha' V_h \alpha$ where V_h is obtained from the IRF's for the original factorization. Since we still need this to define a unit variance shock, we need $\alpha'\alpha = 1$. The approach is to solve the problem

(7) $\max_{\alpha} \alpha' V_h \alpha$ subject to $\alpha'\alpha = 1$ and $C_R \alpha \geq 0$

where the final restrictions are on the impulse responses (sign restrictions, signs of changes between periods). In other applications, simply looking at the solution to this without those sign restrictions might be interesting; in Faust's case, since he is trying to look at the more minor effect of (potential) monetary policy shocks, the restrictions are critical.

The v_h matrix can be obtained in RATS using the following (this is getting the 108 step horizon variance decomp for GDP):

Compute impulse response function for any decomposition of sigma

```
impulse(responses=irf,model=bmmodel,$
  decomp=%decomp(%sigma),noprint) * 108
```

Pull out responses of GDP

```
dec vect[series] gdpresp(nvar)
do i=1,nvar
   set gdpresp(i) = irf(1,i)
end do i
```

The CMOM of this will give the quadratic form matrix which will give the variance from the weights on the orthogonal components. The forecast error variance is the sum of the diagonal elements. Scale the matrix by that, so the matrix sums to one on the diagonal.

```
cmom
# gdpresp
compute vh=%cmom/%sum(%xdiag(%cmom))
```

The solution to (7) without the further constraints is just the eigenvector (normalized to unit length) for the maximal eigenvalue. Since that gives a linear combination of already orthonormalized shocks, to get the original impulse vector, we need to premultiply that by that first factor:

```
eigen vh eigval eigvect
compute x=%decomp(%sigma)*%xcol(eigvect,1)
```

The following is an example of imposing a restriction on the impulse responses: here that at the 4th step, the response of interest rates (variable 6) is non-negative and the response of prices (variable 2) is non-positive. This is done with FIND with restrictions. Faust solves a set of restricted eigenvalue problems assuming (in turn) each subset of constraints is binding.

```
compute [rect] step4 = %xt(irf,4)
compute rcond = %xrow(step4,6)
compute pcond = %xrow(step4,2)
dec vect alpha(6)
nonlin alpha %normsqr(alpha)==1 %dot(rcond,alpha)>=0.0 $
 %dot(pcond,alpha)<=0.0
compute alpha=%xcol(eigvect,1)
find(method=bfgs,trace) max %qform(vh,alpha)
```

```
end find
compute x=%decomp(%sigma)*alpha
impulse(shock=x,model=bmmodel) * 108
```