Lecture 1: Vector Autoregressions: An Introduction

James M. Nason

Centre for Applied Macroeconomic Analysis
and
Virginia Center for Economic Policy
February 11, 2024

Sims: Macroeconomics and Reality

Estimating Unrestricted VARs

Fundamentals and VARs

Impulse Response Functions

Forecast Error Variance Decompositions

Macroeconometrics Prior to 1980

- Macroeconomists estimated "structural" models in the tradition of Jan Tinbergen (http://en.wikipedia.org/wiki/Jan_Tinbergen), Trygve Haavelmo (http://en.wikipedia.org/wiki/Trygve_Haavelmo), Art Goldberger (http://en.wikipedia.org/wiki/Arthur_Goldberger) and Lawrence Klein (http://en.wikipedia.org/wiki/Lawrence_Klein) since the 1920s.
  1. Tinbergen: The concept of a macro model mixing behavioral equations and resource and adding up constraints.
  2. Haavelmo: Simultaneity/simultaneous systems and model evaluation.
  3. Goldberger and Klein: The first Keynesian econometric model useful for policy analysis.
- The goal was to use Keynesian macroeconometrics models to conduct business cycles analysis and policy evaluation.

Keynesian Macroeconometric Practice

- Estimated Keynesian macro model required aggregate data and armies of graduate students to calculate regression estimates using mechanical desk calculators and slide rules.
  1. National Income and Product Accounts data were not released by government statistical agencies until the early 1950s.
  2. Mainframe computers become available for academic research only in the mid 1950s.

- Haavelmo recognized estimates of early macro models were plagued by simultaneity problems $\Longrightarrow$ 2SLS, 3SLS, etc.

- These estimators imply identifying restrictions that are not obvious.

- For example, Keynesian models were built from separate consumption, investment, wage, price, government, financial, and monetary blocks in the tradition of Goldberger and Klein.

Sims' Critique of Keynesian Econometric Practice

- Sims (1980, p. 1): "... the identification claimed for existing large-scale models is incredible." $\Longrightarrow$ Exclusion restrictions are ad hoc.
  1. This is especially problematic in the Keynesian macro model procedure of estimating a block of equations at a time.
  2. Theory often offers little advice to guide the specification of model dynamics $\Longrightarrow$ lengths of leads and lags.
  3. Keynesian macro models violate the rational expectations hypothesis.

- Identification is about the likelihood of a model conditional on its parameter vector.
  1. A model is identified if there is not another parameterization (*i.e.*, model) yielding a likelihood that is a scalar of the first.
  2. The likelihood principle: The likelihood contains all the evidence about model parameters that can be extracted from the sample data; see Berger and Wolpert (1988, The Likelihood Principle, Beachwood, OH: Institute of Mathematical Statistics), `http://www.jstor.org/discover/4355509?sid=21105069390931&uid=2134&uid=2&uid=70&uid=4`

Another Intellectual Tradition in Macroeconometrics

- There is another tradition macro draws on to study business cycles and evaluate monetary and fiscal policies.
- Eugen Slutzky shows sequences of unforecastable shocks can produce business cycles in an article published in Russian in 1927 and republished in English in *Econometrica* in 1937.
  1. His insight is that summing white noise shocks or draws from a Gaussian process generates time series with periodicity resembling business cycles.
  2. See `http://en.wikipedia.org/wiki/Eugen_Slutsky`.
- Ragnar Frish is responsible for the study of dynamics, innovation analysis, and impulse response functions (IRFs).
  1. These are tools that can be used to conduct business cycle analysis and policy evaluation.
  2. See `http://en.wikipedia.org/wiki/Ragnar_Frisch`.
- Slutzky and Frisch's contributions are mostly forgotten by Keynesian macro.

Vector Autoregressions

- Sims (1980) describes an alternative class of empirical macro models, which builds on the tradition of Slutzky (1937) and Frisch (1933).
  1. The alternative class of models is vector autoregressions (VARs).
  2. Sims argues VARs avoid the arbitrary identification schemes applied to Keynesian macro models.

- A VAR describes the dynamics of a vector (*i.e.*, multivariate) time series, y_t.
  1. For example, $y_t = [RGDP_t \;\; \pi_t \;\; UR_t \;\; M1_t \;\; R_{Short,t}]'$.
  2. The auto- and cross-covariance functions define the dynamics of y_t.

- These lecture notes sample the VAR literature that follows in the wake of Sims (1980).

Vector Autoregressions

- Sims describes an alternative class of empirical macro models.
- The alternative class of models is vector autoregressions (VARs).
- Sims argues VARs avoid the arbitrary identification schemes applied to Keynesian macro models.
- A VAR describes the dynamics of a vector (*i.e.*, multivariate) time series, y_t.
  1. For example, $y_t = [RGDP_t \;\; \pi_t \;\; UR_t \;\; M1_t \;\; R_{Short,t}]'$.
  2. The auto- and cross-covariance functions define the dynamics of y_t.
- These lecture notes sample the VAR literature that follows in the wake of Sims (1980).

An Unrestricted VAR

- Consider a $n \times 1$ vector process at date t, y_t that is assumed to be a pth-order Gaussian vector autoregression (VAR), p a finite integer,

$$y_t = \mathbf{c} + \mathbf{B}(\mathbf{L}) y_{t-1} + \varepsilon_t,$$

  where $\mathbf{c}$ is a $n \times 1$ vector of intercepts, $\mathbf{B}(\mathbf{L}) = \sum_{j=1}^{p} \mathbf{B}_j \mathbf{L}^{j-1}$, $\mathbf{B}_j$ is a $n \times n$ matrix of slope coefficients, $\varepsilon_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{\Omega})$, and $\mathbf{\Omega}$ is a $n \times n$ (positive definite) covariance matrix of the Gaussian error process ε_t.

- Data from $t = -p+1, -p+2, \ldots, -1, 0, 1, 2, \ldots, T$ exists. Estimation of the intercepts, $\mathbf{c}$, matrices of slope parameters, $\mathbf{B}_j$, $j = 1, \ldots, p$, and error covariance matrix, $\mathbf{\Omega}$, requires the first p observations as conditioning information for the sample that runs from $t = 1, \ldots, T$.

- Since y_t is Gaussian, maximum likelihood estimation (MLE) dominates other estimators. Form the conditional likelihood

$$f_{y_T, \ldots, y_1 | y_0, \ldots, y_{-p+1}} \left(y_T, \ldots, y_1 \middle| y_0, \ldots, y_{-p+1}; \boldsymbol{\beta} \right),$$

  and maximize it with respect to $\boldsymbol{\beta}$, where $\boldsymbol{\beta} = \left(\mathbf{c}, \mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_p, \mathbf{\Omega} \right)$.

The Likelihood of an Unrestricted VAR

- Assuming normality of a VAR(p)'s errors yields the conditional density of y_t

$$y_t \left| y_{t-1}, \; y_{t-2}, \; \ldots, \; y_{-p+1} \right. \sim \mathcal{N}\left(\mathbf{c} + \mathbf{B}(\mathbf{L})y_{t-1}, \mathbf{\Omega}\right).$$

- Define $\mathbf{X}_t \equiv \begin{bmatrix} 1 & y_{t-1} & y_{t-2} & \ldots & y_{t-p} \end{bmatrix}' \Longrightarrow \mathbf{X}_t$ is a $(np+1) \times 1$ column vector) and $\mathbf{\Theta}' \equiv \begin{bmatrix} \mathbf{c} & \mathbf{B}_1 & \mathbf{B}_2 & \ldots, & \mathbf{B}_p \end{bmatrix}$ is a $n \times (np+1)$ matrix.

  Conditional density $\Longrightarrow y_t \left| y_{t-1}, \; y_{t-2}, \; \ldots, \; y_{-p+1} \right. \sim \mathcal{N}\left(\mathbf{\Theta}'\mathbf{X}_t, \mathbf{\Omega}\right).$

- This operation converts the VAR into a 'static' regression model. Regress y_t on $\mathbf{X}_t \Longrightarrow$ the joint (natural) log density of y_t is

$$\ln\left[f_{y_t \left| y_{t-1}, \ldots, y_{-p+1}\right.}\left(y_t \left| y_{t-1}, \; \ldots, \; y_{-p+1}; \boldsymbol{\beta}\right.\right)\right]$$

$$= \; -0.5\left[n\ln[2\pi] \; - \; \ln\left[\left|\mathbf{\Omega}^{-1}\right|\right] + \left(y_t \; - \; \mathbf{\Theta}'\mathbf{X}_t\right)'\mathbf{\Omega}^{-1}\left(y_t \; - \; \mathbf{\Theta}'\mathbf{X}_t\right)\right].$$

- Other than assuming Gaussian errors, ε_t, and lag length p no restrictions are imposed on the parameters of the VAR $\Longrightarrow$ the VAR is unrestricted.

An Estimator of the Intercept and Slope Coefficients

- Construct the sample log likelihood of the VAR:

$$\mathfrak{L}\left(\mathbf{\Theta}, \mathbf{\Omega} \middle| y_1, \ldots, y_T\right) \equiv \sum_{t=1}^{T} \ln\left[f_{y_t \middle| y_{t-1}, \ldots, y_{-p+1}}\left(y_t \middle| y_{t-1}, \ldots, y_{-p+1}; \boldsymbol{\beta}\right)\right]$$

- Differentiate with respect to $\mathbf{\Theta}$ to compute the (conditional) MLE of $\mathbf{\Theta}$.

$$\widehat{\mathbf{\Theta}}' = \left[\sum_{t=1}^{T} y_t \mathbf{X}_t'\right]\left[\sum_{t=1}^{T} \mathbf{X}_t \mathbf{X}_t'\right]^{-1}.$$

- Row-by-row, the MLE $\mathbf{\Theta}$ is the ordinary least squares (OLS) estimator. Line-by-line the ℓth row of $\mathbf{\Theta}$ is

$$\widehat{\mathbf{\Theta}}_\ell' = \left[\sum_{t=1}^{T} y_{\ell,t} \mathbf{X}_t'\right]\left[\sum_{t=1}^{T} \mathbf{X}_t \mathbf{X}_t'\right]^{-1}.$$

- Consistent estimates of the parameters, the elements of the $\mathbf{B}_j$s, can be computed by OLS equation-by-equation for the unrestricted VAR(p).

An Estimator of the Covariance Matrix

- Remaining parameters of the unrestricted VAR(p) to estimate are in $\mathbf{\Omega}$.
- Return to the conditional log likelihood of this VAR, substitute $\hat{\mathbf{\Theta}}$ into the joint density of y_t and sum across all T observations to produce

$$\mathcal{L}\left(\mathbf{\Theta}, \mathbf{\Omega} \middle| y_1, \ldots, y_T\right) = -\frac{1}{2}\left[nT\ln[2\pi] - T\ln\left[\left|\mathbf{\Omega}^{-1}\right|\right] + \sum_{t=1}^{T}\left(\hat{\varepsilon}_t'\mathbf{\Omega}^{-1}\hat{\varepsilon}_t\right)\right],$$

  where the residuals, estimates of the errors ε_t, $\hat{\varepsilon}_t = y_t - \hat{\mathbf{\Theta}}'\mathbf{X}_t$.

- Differentiate this expression with respect to $\mathbf{\Omega}$ to compute the MLE of $\mathbf{\Omega}$

$$\hat{\mathbf{\Omega}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t'.$$

- MLE of an unrestricted VAR's covariance matrix, $\hat{\mathbf{\Omega}}$, is the covariance matrix of the OLS residuals.

The Likelihood of an Unrestricted VAR at $\widehat{\Theta}$ and $\widehat{\Omega}$

- The log likelihood evaluated at $\widehat{\boldsymbol{\Theta}}$ and $\widehat{\boldsymbol{\Omega}}$ is

$$\mathcal{L}\left(\boldsymbol{\Theta}, \boldsymbol{\Omega} \middle| y_1, \ldots, y_T\right) = -\frac{1}{2}\left[nT\ln[2\pi] - T\ln\left[\left|\widehat{\boldsymbol{\Omega}}^{-1}\right|\right] + \sum_{t=1}^{T}\left(\widehat{\varepsilon}_t'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\varepsilon}_t\right)\right].$$

- Next, apply the trace$(\cdot)$ operator, which sums all the diagonal elements of a square matrix, to $\sum_{t=1}^{T}\widehat{\varepsilon}_t'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\varepsilon}_t = \text{trace}(\sum_{t=1}^{T}\widehat{\varepsilon}_t'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\varepsilon}_t)$ $= \text{trace}(\widehat{\boldsymbol{\Omega}}^{-1}\sum_{t=1}^{T}\widehat{\varepsilon}_t'\widehat{\varepsilon}_t) = \text{trace}(\widehat{\boldsymbol{\Omega}}^{-1}T\widehat{\boldsymbol{\Omega}}) = \text{trace}(T\mathbf{I}_n) = nT.$

- The result is

$$\mathcal{L}\left(\boldsymbol{\Theta}, \boldsymbol{\Omega} \middle| y_1, \ldots, y_T\right) = -\frac{T}{2}\left(n(1 + \ln[2\pi]) - \ln\left[\left|\widehat{\boldsymbol{\Omega}}^{-1}\right|\right]\right).$$

Notes on Estimating Unrestricted VARs

- A problem is these MLEs are conditional on the lag length of the VAR $\Longrightarrow$ p is a parameter of the VAR.
- If the unrestricted VAR approximates a VARMA(q_1, q_2), increase p until the serial correlation in ε_t is eliminated.
- As p increases, the number of VAR parameters rises at rate n^2 $\Longrightarrow$ given T, degrees-of-freedom (df) falls providing less power for hypotheses tests.
- There exists a trade-off between increasing the lag length of the VAR to whiten residuals and the loss of power (*i.e.*, information) to evaluate a VAR.

Information Criteria for Choosing p

- The Akaike information criterion (AIC) captures this tradeoff

$$\mathbf{Min}_p \, \text{AIC}(p) \;=\; \ln\left(\left|\hat{\mathbf{\Omega}}\right|\right) \;+\; \frac{2pn^2}{T}.$$

- A similar rule for choosing p is the Bayesian Information Criterion (BIC)

$$\mathbf{Min}_p \, \text{BIC}(p) \;=\; \ln\left(\left|\hat{\mathbf{\Omega}}\right|\right) \;+\; pn^2\frac{\ln(T)}{T}.$$

- The Hannah-Quinn information criterion (HQC) selects p by

$$\mathbf{Min}_p \, \text{HQC}(p) \;=\; \ln\left(\left|\hat{\mathbf{\Omega}}\right|\right) \;+\; 2pn^2\frac{\ln\ln(T)}{T}.$$

- Minimize the AIC, BIC, and HQC by selecting p from $p = 1, \ldots, K$, where K is a large integer.
  1. As p increases, $\ln\left(\left|\hat{\mathbf{\Omega}}\right|\right)$ falls or is unchanged, but the penalty terms rise.
  2. AIC assumes the VAR(p) is true, but the BIC and HQC do not $\Longrightarrow$ BIC and HQC produce consistent estimates of p.
  3. $\Longrightarrow$ AIC selects too large a p while BIC and HQC often yield more conservative p.

A Likelihood Ratio Test to Choose p

- Another way to choose p is to compute a likelihood ratio (LR) test of the null hypothesis of a VAR(p) against the alternative of a VAR$(p+i)$.
- The null hypothesis is the $i \times n^2$ elements of $\mathbf{B}_{p+i}$ equal zero $\Longrightarrow$ a joint test of restrictions imposed on the VAR(p) compared with the VAR$(p+i)$.
- The LR test statistic is $2\left(\hat{\mathcal{L}}_A - \hat{\mathcal{L}}_0\right) = T\left(\ln\left[\left|\hat{\mathbf{\Omega}}_0\right|\right] - \ln\left[\left|\hat{\mathbf{\Omega}}_A\right|\right]\right)$, where A and 0 denote the alternative and null hypothesis, respectively.
- To compute the LR test, estimate a VAR(p) and a VAR$(p+i)$.
  1. Construct the covariance matrix of the residuals of the VAR(p) and VAR$(p+i)$.
  2. Under the null of the VAR(p), call its estimated covariance matrix $\hat{\mathbf{\Omega}}_0$.
  3. The VAR$(p+i)$ is the alternative with estimated covariance matrix $\hat{\mathbf{\Omega}}_A$.
- The LR test is asymptotically distributed $\chi^2(n \times i)$, where the df $= n \times i$, the number of restrictions $\Longrightarrow$ the difference in the number of coefficients of the VAR$(p+i)$ and VAR(p).

Sims' Likelihood Ratio Test to Choose p

- Sims (1980) proposes a degrees-of-freedom correction for the LR test

$$2\left(\hat{\mathcal{L}}_A - \hat{\mathcal{L}}_0\right) = (T - k)\left(\ln\left[\left|\hat{\boldsymbol{\Omega}}_0\right|\right] - \ln\left[\left|\hat{\boldsymbol{\Omega}}_A\right|\right]\right),$$

  to correct for small sample bias, where $k = 1 + ni$ represents the number of coefficients excluded from the VAR_0. Subtracting k from T, reduces the size of the LR statistic, which makes the test more conservative w/r/t the choice of $p \Longrightarrow$ the LR test is less likely to reject the null hypothesis for the sample sizes typically encountered in macro.

- The last issue for the choice of p concerns how to perform the sequence of LR tests of $p = 1, 2, \ldots$. At this moment, the accepted procedure is to move from the general null hypothesis to more specific null hypotheses. This moves p from a large integer K, say for quarterly macro data of 12 or 16, to smaller values. The first occurrence of a LR test statistic at or below the appropriate significance level determines p.

The Autocovariance Function of a VAR

- The claim of no restrictions other than ε_t is Gaussian is applied to the VAR(p), $y_t = \mathbf{c} + \mathbf{B}(\mathbf{L})\, y_{t-1} + \varepsilon_t$, is not quite correct.
- Several restrictions have to be imposed on the $\mathbf{B}_j$s to guarantee consistency and efficiency $\Longrightarrow \sqrt{T}\left(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\right) \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Omega} \otimes \left[\sum_{t=1}^{T} \mathbf{X}_t \mathbf{X}_t'\right]^{-1}\right)$.
- These restrictions ensure that y_t is stationary, which give
  1. the unconditional mean: $\mathbf{E} y_t = \mu_y \Longrightarrow$ define $x_t \equiv y_t - \mu_y$, and
  2. the unconditional jth autocovariance matrix: $\mathbf{E}\left\{x_t x_{t-j}'\right\} = \boldsymbol{\Gamma}_j$,
  3. $\Longrightarrow \mu_y$ and $\boldsymbol{\Gamma}_j$ are finite and independent of time.
  4. But $\boldsymbol{\Gamma}_j \neq \boldsymbol{\Gamma}_{-j} \Longrightarrow \boldsymbol{\Gamma}_j = \mathbf{E}\left\{x_t x_{t-j}'\right\} \neq \mathbf{E}\left\{x_t x_{t+j}'\right\} = \boldsymbol{\Gamma}_{-j}$.
  5. Rather than lag the jth autocovariance, lead it j periods $\Longrightarrow$ $\mathbf{E}\left\{x_{t+j} x_t'\right\} = \boldsymbol{\Gamma}_j$ and take transposes to find $\mathbf{E}\left\{x_t x_{t+j}'\right\} = \boldsymbol{\Gamma}_j' = \boldsymbol{\Gamma}_{-j}$

The Companion Form of a VAR

- Computing the unconditional mean and especially the covariance generating process of y_t when its data generating process (DGP) is the Gaussian VAR(p), $y_t = \mathbf{c} + \mathbf{B}(\mathbf{L})y_{t-1} + \varepsilon_t$, appears difficult.
- The unconditional mean is $\mu_y = \left[\mathbf{I}_n - \mathbf{B}(\mathbf{1})\right]^{-1}\mathbf{c}$, where $\mathbf{B}(\mathbf{1}) = \sum_{j=1}^{p}\mathbf{B}_j$. However, this calculation assumes that $\mathbf{I}_n - \mathbf{B}(\mathbf{1})$ is not singular.
- Assume μ_y exists $\Longrightarrow$ the "demeaned" VAR(p) is $x_t = \mathbf{B}(\mathbf{L})x_{t-1} + \varepsilon_t$.
- The demeaned VAR(p) can be represented as a VAR(1), $z_t = \mathbf{F}z_{t-1} + v_t$, where $z_t = \begin{bmatrix} x_t' & x_{t-1}' & \dots & x_{t-p+1}' \end{bmatrix}'$, $v_t = \begin{bmatrix} \varepsilon_t' & \mathbf{0}_{1\times p} & \dots & \mathbf{0}_{1\times p} \end{bmatrix}'$,

$$\mathbf{F} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \dots & \mathbf{B}_{p-1} & \mathbf{B}_p \\ \mathbf{I}_n & \mathbf{0}_{n\times n} & \dots & \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \mathbf{I}_n & \dots & \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} & \dots & \mathbf{I}_n & \mathbf{0}_{n\times n} \end{bmatrix}, \text{ and } \mathbf{E}\left\{v_t v_t'\right\} = \mathcal{Q}.$$

- The VAR(1), $z_t = \mathbf{F}z_{t-1} + v_t$, is the companion form of the VAR(p) and $\mathbf{F}$ is the companion matrix.

The VMA(∞) of a VAR(1)

- The companion form of a VAR has some useful features.
  1. The vector MA(∞) is $\mathcal{Z}_t = [\mathbf{I}_n - \mathbf{F}\mathbf{L}]^{-1}\mathcal{V}_t = \sum_{j=0}^{\infty}\mathbf{F}^j\mathcal{V}_{t-j}$,
  2. which assumes that $\mathbf{F}$ is not singular.
  3. Non-singularity of $\mathbf{F}$ rests on it having no roots or eigenvalues > 1 in absolute value $\Longrightarrow$ search for $|\lambda|$s < 1 that set $\left|\mathbf{F} - \lambda\mathbf{I}_{n^2}\right| = 0$.
  4. This is equivalent to finding the determinant of the VAR(p) $\left|\lambda\mathbf{I}_n - \lambda^{p-1}\mathbf{B}_1 - \lambda^{p-2}\mathbf{B}_2 - \ldots - \lambda\mathbf{B}_{p-1} - \mathbf{B}_p\right| = 0$.

- A VAR(p) has autocovariance matrices independent of time if its eigenvalues are outside the unit circle.
  1. The impact on $\mathcal{Y}_t$ of a unit increase in ε_t decays to zero in a finite span of time.
  2. The VAR(p) is covariance stationary when its $\mathbf{B}_j$s are restricted to have λs $\in (-1, 1)$.

Computing the Autocovariances of a VAR(1)

- Pass the covariance operator through the VAR(1)

$$\begin{aligned}
\mathbf{E}\left\{z_t z_t'\right\} &= \mathbf{E}\left\{\left(\mathbf{F}z_{t-1} + v_t\right)\left(\mathbf{F}z_{t-1} + v_t\right)'\right\} \\
&= \mathbf{F}\,\mathbf{E}\left\{z_{t-1} z_{t-1}'\right\}\mathbf{F}' + \mathbf{E}\left\{v_t v_t'\right\} \\
\Sigma_z &= \mathbf{F}\,\Sigma_z \mathbf{F}' + \mathcal{Q}.
\end{aligned}$$

- The covariance matrix of z_t, Σ_z, is nonlinear function of $\mathbf{F}$ and $\mathcal{Q}$ (because of the quadratic term $\mathbf{F}\,\Sigma_z\mathbf{F}'$).
  1. The $\text{vec}(\cdot)$ operator linearizes $\Sigma_z = \mathbf{F}\,\Sigma_z\mathbf{F}' + \mathcal{Q}$, which stacks columns of the matrix to form a vector.
  2. $\Longrightarrow \text{vec}(\Sigma_z) = \text{vec}(\mathbf{F}\,\Sigma_z\mathbf{F}') + \text{vec}(\mathcal{Q})$.
  3. Since $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec}(\mathbf{B})$, $[\mathbf{I}_{n^2} - \mathbf{F} \otimes \mathbf{F}]\text{vec}(\Sigma_z) = \text{vec}(\mathcal{Q})$.
     $\Longrightarrow$ Lag zero autocovariances are $\text{vec}(\Sigma_z) = [\mathbf{I}_{n^2} - \mathbf{F} \otimes \mathbf{F}]^{-1}\text{vec}(\mathcal{Q})$.
  4. Lag j autocovariances are $\mathbf{E}\left\{z_t z_{t-j}'\right\} = \mathbf{E}\left\{\mathbf{F}z_{t-1} z_{t-j}'\right\} + \mathbf{E}\left\{v_t z_{t-j}'\right\}$
     $\Longrightarrow \Sigma_{z,j} = \mathbf{F}\,\Sigma_{z,j-1} = \mathbf{F}^j\Sigma_z$.

Computing the Autocovariances of the VMA(∞) of the VAR(1)

- Pass the covariance operator through VMA(∞), $\mathcal{Z}_t = \sum_{j=0}^{\infty} \mathbf{F}^j \mathcal{V}_{t-j}$

$$\begin{aligned}
\Sigma_{\mathcal{Z}} &= \sum_{j=0}^{\infty} \mathbf{E}\left\{\mathbf{F}^j \mathcal{V}_{t-j} \mathcal{V}_{t-j}' \mathbf{F}^{j\prime}\right\} \\
&= \sum_{j=0}^{\infty} \mathbf{F}^j \mathcal{Q} \mathbf{F}^{j\prime} \\
\text{vec}(\Sigma_{\mathcal{Z}}) &= \sum_{j=0}^{\infty} \left(\mathbf{F}^j \bigotimes \mathbf{F}^j\right) \text{vec}(\mathcal{Q}) \\
&= \left[\mathbf{I}_{n^2} - \mathbf{F} \bigotimes \mathbf{F}\right]^{-1} \text{vec}(\mathcal{Q}).
\end{aligned}$$

Computing the Autocovariances of a VMA(∞) in General

- Given the VAR(p), $\mathbf{y}_t = \mathbf{c} + \mathbf{B}(\mathbf{L})\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$, is stationary, the implied VMA(∞) is $\mathbf{y}_t = \boldsymbol{\mu}_y + \sum_{\ell=0}^{\infty} \mathbf{C}_\ell \boldsymbol{\varepsilon}_{t-\ell}$, where $\mathbf{C}(\mathbf{L}) = \left[\mathbf{I}_n - \mathbf{B}(\mathbf{L})\right]^{-1}$ and $\mathbf{C}_0 \equiv \mathbf{I}_n$.
- The autocovariances are

$$\begin{aligned}
\mathbf{E}\left\{\left(\mathbf{y}_t - \boldsymbol{\mu}_y\right)\left(\mathbf{y}_t - \boldsymbol{\mu}_y\right)'\right\} &= \sum_{\ell=0}^{\infty} \mathbf{E}\left\{\mathbf{C}_\ell \boldsymbol{\varepsilon}_{t-\ell}\boldsymbol{\varepsilon}_{t-\ell}'\mathbf{C}_\ell'\right\} \\
\boldsymbol{\Gamma}_0 &= \sum_{\ell=0}^{\infty} \mathbf{C}_\ell \boldsymbol{\Omega} \mathbf{C}_\ell'. \\
\mathbf{E}\left\{\left(\mathbf{y}_t - \boldsymbol{\mu}_y\right)\left(\mathbf{y}_{t-s} - \boldsymbol{\mu}_y\right)'\right\} &= \sum_{\ell=0}^{\infty} \mathbf{E}\left\{\mathbf{C}_\ell \boldsymbol{\varepsilon}_{t-\ell}\boldsymbol{\varepsilon}_{t-\ell-s}'\mathbf{C}_\ell'\right\} \\
\boldsymbol{\Gamma}_s &= \sum_{\ell=0}^{\infty} \mathbf{C}_{s+\ell} \boldsymbol{\Omega} \mathbf{C}_\ell'.
\end{aligned}$$

Restrictions on the VMA(∞) to Guarantee Stationarity

- Restrictions are needed on the $\mathbf{C}_j$s of the VMA(∞) to guarantee $\mathbf{\Gamma}_s$ is independent of time $s = 0, 1, \ldots, k, \ldots$.
  1. The $\mathbf{C}_j$s are nonlinear functions of the $\mathbf{B}_j$s, which implies invertibility.
  2. Next, y_t and ε_t have bounded fourth moments to ensure μ_y and the diagonals of $\mathbf{\Gamma}_0$ are ergodic (*i.e.*, the sample average converges to the average of sample averages as $T \longrightarrow \infty$).

- The $\mathbf{C}_\ell$s are absolutely summable, $\sum_{\ell=0}^{\infty} \left| C_{\ell,i,j} \right| < \infty$, $i, j = 1, 2, \ldots, n$, or element by element the sum of the absolute values of the $\mathbf{C}_\ell$s are finite.
  1. Absolute summability places strong restrictions on the $\mathbf{C}_\ell$s in the speed of convergence to finite values element by element.
  2. A weaker restriction is squared summability, $\sum_{\ell=0}^{\infty} C_{\ell,i,j}^2 < \infty$.
  3. In either case, element by element the $\mathbf{C}_\ell$s have smaller and smaller increments as $\ell \longrightarrow \infty$.

VMA(∞)s and Fundamentalness

- The implied VMA(∞), $y_t = \mu_y + \sum_{\ell=0}^{\infty} \mathbf{C}_\ell \varepsilon_{t-\ell}$, has an important feature.
- Consider constructing a forecast of y_{t+1}, $\mathbf{E}_t y_{t+1}$.
  1. The forecast innovation is $y_{t+1} - \mathbf{E}_t y_{t+1} = \varepsilon_{t+1}$.
  2. Similarly, the h-step ahead innovation is $y_{t+h} - \mathbf{E}_t y_{t+h} = \varepsilon_{t+h}$.
  3. The innovation, or news, about y_{t+h} between dates t and $t+h$ is ε_{t+h} given the VMA(∞) is the true DGP of y_t.
  4. This explains the assumption that $\mathbf{C}_0$ is the identity matrix.
  5. Since the only news about y_{t+h} between dates t and $t+h$ is ε_{t+h}, it is **fundamental** for y_{t+h}.
- Knowledge of fundamental errors is necessary to produce y_t.

VMA(∞)s and Fundamentalness: A Counterexample

- Start with $y_t = \mu y + \sum_{\ell=0}^{\infty} \mathbf{C}_\ell \varepsilon_{t-\ell}$.
- Multiply ε_t by $\mathbf{G}$ to produce $\nu_t \equiv \mathbf{G}\varepsilon_t$, where $\mathbf{G}$ is nonsingular.
- Use this definition in $y_t = \mu y + \sum_{\ell=0}^{\infty} \mathbf{C}_\ell \mathbf{G}^{-1}\mathbf{G}\varepsilon_{t-\ell} = \mu y + \sum_{\ell=0}^{\infty} \mathbf{K}_\ell \nu_{t-\ell}$, where $\mathbf{K}_\ell \equiv \mathbf{C}_\ell \mathbf{G}^{-1} \implies$ a VMA(∞) is not unique.
- Suppose that $\mathbf{\Omega}_\nu = \mathbf{E}\left\{\mathbf{G}\varepsilon_t \varepsilon_t' \mathbf{G}'\right\} = \mathbf{G}\mathbf{\Omega}\mathbf{G}'$ is diagonal.
- Forecasts of y_t produced by $\sum_{\ell=0}^{\infty} \mathbf{K}_\ell \nu_{t-\ell}$ do not yield fundamental errors.
- Although $\mathbf{\Omega}_\nu$ is diagonal,
  1. $\mathbf{G}$ is not necessarily a triangular matrix (*i.e.*, endows the elements of ε_t with a recursive ordering).
  2. Instead, ν_t consists of linear combinations of the elements of ε_t.
  3. These linear combinations of the elements of ε_t $(= [\varepsilon_{1,t} \ \ldots \ \varepsilon_{n,t}]')$ are not the errors fundamental for y_t.
  4. $\implies \left|\mathbf{K}_0 \mathbf{\Omega}_\nu \mathbf{K}_0'\right| \not> \left|\mathbf{K}_\ell \mathbf{\Omega}_\nu \mathbf{K}_\ell'\right|$, but $\left|\mathbf{C}_0 \mathbf{\Omega} \mathbf{C}_0'\right| > \left|\mathbf{C}_\ell \mathbf{\Omega} \mathbf{C}_\ell'\right|$, $\ell \geq 1$.

VMA(∞)s and Fundamentalness: The Wold Decomposition Theorem

- Fundamentalness suggests a VMA(∞) is a mapping that recovers ε_t given the data, y_t, or having knowledge of ε_t the data can be produced.
- **Wold Decomposition Theorem** states minimum necessary and sufficient conditions for a VMA(∞) to be a fundamental representation of y_t and ε_t.
- **Wold Decomposition Theorem**: Any mean zero, covariance stationary process, $\{y_t\}_{t=-\infty}^{\infty}$, can be represented as $y_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j} + \kappa_t$, $C_0 = \mathbf{I}_n$, conditional on the restrictions
  1. ε_t is mean zero, $\mathbf{E}\varepsilon_t = 0$, linearly unpredictability given past history, $\mathbf{E}\{\varepsilon_{t+h} \mid \varepsilon_t, \varepsilon_{t-1}, \ldots, y_t, y_{t-1}, \ldots\} = 0$, $h \geq 1$, is serially uncorrelated with its own history, $\mathbf{E}_t\{\varepsilon_t \varepsilon_{t-j}'\} = 0$, and history of y_t, $\mathbf{E}_t\{\varepsilon_t y_{t-j}'\} = 0$, $j \geq 1$, and homoskedastic, $\mathbf{E}\{\varepsilon_t \varepsilon_t'\} = \mathbf{\Omega}$,
  2. the roots of $\mathbf{C}(\mathbf{L})$ all lie on or outside the unit circle $\Longrightarrow \mathbf{C}(\mathbf{L})^{-1}$ exists,
  3. sequence of $\mathbf{C}_j$s are square summable, $\sum_{\ell=0}^{\infty} C_{j,i,\ell}^2 < \infty$, $i, \ell = 1, 2, \ldots, n$,
  4. and κ_t is linearly deterministic (*i.e.*, any class of known, fixed functions can enter the stochastic process y_t).

The Wold Decomposition Theorem: Dos and Don'ts

- The WDT is a framework that *(a)* connects stochastic difference equations to a general class of stationary stochastic processes (*i.e.*, VARMA models) and *(b)* provides restrictions on the building block of the stochastic process, the white noise process ε_t.

- The WDT does not need $\varepsilon_t \sim IID$ and/or normal (*i.e.*, Gaussian).
  1. Linear unpredictability of ε_t given past history places strong restrictions on the class of models relevant for the WDT.
  2. $\Longrightarrow$ Only linear regressions, $y_t = \mathbf{c} + \sum_{j=1}^{p} \mathbf{B}_j y_{t-j} + \varepsilon_t$, $p < \infty$, matter for the WDT and for constructing ε_t.
  3. But, in general, the linear regression could have time-varying parameters $\Longrightarrow$ $\mu_{y,t}$, $\mathbf{B}_{j,t}$, and $\mathbf{\Omega}_t$.

- Proofs of the WDT do not depend on assumptions about y_t.
  1. There is nothing about linearity or stationarity $\Longrightarrow$ $\mathbf{E}\{\cdot\}$ does not require y_t to be a linear or stationary stochastic process a priori.
  2. If y_t is not stationary, alter the sample data to render it so.

Moving Beyond the Wold Decomposition Theorem: Granger-causality

- The WDT is a device for recovering the fundamental errors of y_t using a VMA(∞), or the converse.
- Relationships between the elements of y_t are about something else. An example is Granger-causality (GC).
- GC is about *the ability of the history of one time series to predict or forecast the future path of another time series.*
- There is no economic structural interpretation to be given to a finding of GC or the lack thereof.
- Although evidence of GC does not provide information about fundamentalness, GC can be informative about the dynamic responses of left hand side variables to fundamental errors.

Statistical and Economic Predictability

- The concept underlying GC is about prediction.
  1. GC exploits the current response of a scalar time series $y_{1,t}$,
  2. to movements in the history of a second scalar time series, $y_{2,t}$
  3. $\Longrightarrow$ GC is about bivariate relationships.

- **Definition**: $y_{2,t}$ Granger-causes $y_{1,t}$ if $y_{2,t}$ contributes to forecasts of $y_{1,t}$, given the past history of $y_{1,t}$, $y_{1,t-1}$, $\ldots$, $y_{1,t-j}$.

- Once the ability of a variable's own past to forecast itself is removed, use GC to ask, "Is there any variation left in the future path of $y_{1,t}$ for the history of $y_{2,t}$, $y_{2,t-1}$, $\ldots$, $y_{2,t-j}$ to forecast?"

- Only two variables are being considered here. When $y_{2,t}$ is a vector, the concept of GC is more difficult to analyze.

An Example of Granger-Causality

- Suppose $y_{1,t}$ and $y_{2,t}$ are scalars $\Longrightarrow$

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} B_{11}(\mathbf{L}) & B_{12}(\mathbf{L}) \\ B_{21}(\mathbf{L}) & B_{22}(\mathbf{L}) \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}.$$

- Granger-causality implies $y_{2,t}$ does not Granger-cause $y_{1,t}$ if $B_{12}(\mathbf{L}) = 0$

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} B_{11}(\mathbf{L}) & 0 \\ B_{21}(\mathbf{L}) & B_{22}(\mathbf{L}) \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}.$$

- $y_{1,t}$ has a univariate AR(p) representation because $B_{12}(\mathbf{L}) = 0$, which gives the VAR matrix a lower triangular representation

$$\begin{bmatrix} 1 - B_{11}(\mathbf{L})\mathbf{L} & 0 \\ -B_{21}(\mathbf{L})\mathbf{L} & 1 - B_{22}(\mathbf{L})\mathbf{L} \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}.$$

- Since $[y_{1,t} \;\; y_{2,t}]'$ is covariance stationary and has a VAR(p) representation, the vector MA(∞) process is

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \widetilde{B}(\mathbf{L})^{-1} \begin{bmatrix} 1 - B_{22}(\mathbf{L})\mathbf{L} & B_{12}(\mathbf{L})\mathbf{L} \\ B_{21}(\mathbf{L})\mathbf{L} & 1 - B_{11}(\mathbf{L})\mathbf{L} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix},$$

  where $\widetilde{B}(\mathbf{L}) = [1 - B_{11}(\mathbf{L})\mathbf{L}][1 - B_{11}(\mathbf{L})\mathbf{L}] - B_{12}(\mathbf{L})B_{21}(\mathbf{L})\mathbf{L}^2$.

- A lack of GC from $y_{2,t}$ to $y_{1,t}$ holds, given the Wold MA matrix lag polynomial is lower triangular $\Longrightarrow$ $y_{1,t}$ possesses a univariate Wold representation, $y_{1,t} = \widetilde{B}(\mathbf{L})[1 - B_{22}(\mathbf{L})\mathbf{L}]\varepsilon_{1,t}$.

- A projection of $y_{1,t}$ on itself and $y_{2,t}$ is equivalent to a projection of $y_{1,t}$ only on itself $\Longrightarrow$ $y_{1,t}$ is only a function of $\varepsilon_{1,t}$.

- The null hypothesis of GC is a statement that some subset of VAR coefficients equals zero. **That is, the null hypothesis of a test of Granger-causality is that Granger-causality does not exist.**

Comments about Granger-Causality

- Put most simply, *a test for Granger-causality reveals information about the ability of one time series to forecast another time series conditional on the variables that enter the VAR*(p).
- When this information is combined with economic theory, it might be possible to make statements about the direction of Granger-causality.
- GC tests are often not robust to small changes in the VAR.
- There are examples in which the implications of tests for Granger-causality are reversed either by introducing new variables to the VAR(p) or simply by changing the lag length of the VAR(p) to $p + i$ for some finite i.
- When additional variables are added to the VAR, say, a single variable $y_{3,t}$, then, the response of $y_{1,t}$ to the history of $y_{2,t}$ through the history of $y_{3,t}$ must be considered (the only way to avoid this problem in VARs with $n > 2$ is to restrict $p = 1$; see Dufour and Renault (1998, "Short-run and long-run causality in time series: Theory," *Econometrica* 66, 1099–1125) and Dufour, Pelletier, and Renault (2006, "Short-run and long-run causality in time series: Inference," *Journal of Econometrics* 132, 337–362).

The Shape of the Response of y_t to a Shock to $\varepsilon_{i,t}$

- Perhaps the most common use of VARs is to ask about the response of a left hand side (dependent) variable to some (unobserved) shock.
- The tool that answers this question is the impulse response function (IRF).
- An IRF is the dynamic response or multiplier of a dependent variable to a one unit change in the ith innovation, $\varepsilon_{i,t}$, of the VAR.
- It is easy to compute IRFs.
- When economists attempt to give economic meaning to an IRF, there are often difficulties.
- Suggests the need to employ economic theory to provide structural interpretations to an IRF and its associated shock.

Computing Impulse Response Functions, I

- A VAR(p) possesses a VMA(∞) representation $\mathbf{y}_t = \sum_{i=0}^{\infty} \mathbf{C}_i \boldsymbol{\varepsilon}_{t-i}$, $\mathbf{C}_0 = \mathbf{I}_n$, given the invertibility of $\mathbf{I}_n - \mathbf{B}(\mathbf{L})$ and ignore the intercept $\boldsymbol{\mu}_{\mathbf{y}}$.
- By implication, the h-step ahead conditional expectation is

$$\mathbf{E}_t \mathbf{y}_{t+h} = \sum_{j=0}^{\infty} \mathbf{C}_{j+h} \boldsymbol{\varepsilon}_{t-j},$$

- $\mathbf{E}_t \mathbf{y}_{t+h}$ provides information about the expected response of any element of $\mathbf{y}_{t+h}$ to any element of $\boldsymbol{\varepsilon}_{t-j} \implies$ the information is embedded in the $\mathbf{C}_j$s.
- The idea behind the IRF is the response of an element of $\mathbf{y}_t$ to one of the fundamental shocks, an element of $\boldsymbol{\varepsilon}_t$.
- The Wold representation is a difficult way to compute IRFs.
- A simple method for generating IRFs is to write the VAR(p) as a VAR(1).

Computing Impulse Response Functions, II

- The VAR(p)'s companion form, $\mathcal{Z}_t = \mathbf{F}\mathcal{Z}_{t-1} + \mathcal{V}_t$, yields the conditional expectation or h-step ahead forecast

$$\mathbf{E}_t \mathcal{Z}_{t+h} = \mathbf{F}^h \mathcal{Z}_t, \quad h = 1, 2, \ldots.$$

- The VAR(1) is also associated with the VMA(∞), $\mathcal{Z}_t = \sum_{j=0}^{\infty} \mathbf{F}^j \mathcal{V}_{t-j}$.

- The IRF of $y_{i,t}$ to $\varepsilon_{\ell,t+j}$ is

$$\mathbf{IRF}_{i,\ell}(h) \equiv \frac{\partial y_{i,t+h}}{\partial \varepsilon_{\ell,t}} = \left[\mathbf{F}^h\right]_{i,\ell}, \quad h = 1, 2, \ldots, H,$$

where $[\cdot]_{i,\ell}$ denotes the (i, ℓ) element of the matrix $\mathbf{F}^h$, $i, \ell = 1, 2, \ldots, n$.

Notes about IRFs

- The Wold representation answers the question the IRF asks.
  1. $\Longrightarrow$ Information about the response of $y_{i,t}$, $i = 1, \ldots, n$,
  2. to a one unit change in the ℓth fundamental shock, $\varepsilon_{\ell,t}$.
- The $\mathbf{IRF}_{i,\ell}(h)$ captures this information at the h-step ahead horizon $\Longrightarrow$ need the sequence of matrix powers of $\mathbf{F}$, $\mathbf{F}^j$, and $\mathbf{\Omega}$.
- IRFs are forecasting statements tracing the dynamic shape of the response of $y_{i,t+h}$ to a one unit change in $\varepsilon_{\ell,t}$, $h = 0, 1, 2, \ldots, H \Longrightarrow$ the $\mathbf{F}^j$s.
- But at $j = 0$, $\mathbf{F}^j = \mathbf{I}_n \Longrightarrow \mathbf{\Omega}$ is only other information available to compute the impact response at lag zero.
- Nothing is revealed about how fundamental shocks drive the variability of fluctuations (*i.e.*, forecasting the magnitude) in y_{t+h}.

Identifying IRFs

- The discussion about IRFs ignores the problem of separating the impact on $y_{i,t}$ of a one unit change in $\varepsilon_{\ell,t}$ from the response of $\varepsilon_{m,t}$ to the change in $\varepsilon_{\ell,t}$, $m \neq \ell$.
- No method exists to isolate or identify these one unit movements, say, of $\varepsilon_{\ell,t}$ independently of $\varepsilon_{m,t}$ grounded only in the data, y_t.
- The properties of y_t alone will not help to construct a set of shocks that identify the economic responses of the data to a one unit movement in $\varepsilon_{\ell,t}$.
- The point is that to give a meaningful economic interpretation to IRFs the shocks have to be identified $\Longrightarrow$ compute the dynamic impact of $\varepsilon_{\ell,t}$ on $y_{i,t}$, $i = 1, \ldots,$ n, conditional on an estimated VAR(p).
- Separating or isolating the effect of $\varepsilon_{\ell,t}$ from $\varepsilon_{m,t}$ implies these errors are orthogonal, but $\mathbf{\Omega}$ is unrestricted other than it is positive definite.

The Cholesky Decomposition Orthogonalizes VAR Innovations

- One way to identify VAR shocks is
  1. to orthogonalize the innovations of a VAR(p).
  2. $\Longrightarrow$ use the covariance matrix of the VAR innovations, $\mathbf{\Omega}$.

- Since $\mathbf{\Omega}$ is a (symmetric) positive definite matrix, the Cholesky decomposition produces a triangular representation.
  1. Let $\mathbf{D}$ be the Cholesky decomposition of $\mathbf{\Omega}$, where $\mathbf{\Omega}^{0.5} = \mathbf{D}$ $\Longrightarrow \mathbf{\Omega} = \mathbf{D}\mathbf{D}'$.
  2. Hence, $\mathbf{D}$ is a lower triangular matrix $\Longrightarrow$ zeros above the diagonal.
  3. However, the Cholesky decomposition is only one of several decompositions of positive definite matrices.

The Cholesky Decomposition

- Let a 3×3 symmetric matrix $\mathbf{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ be positive definite.
- Its Cholesky decomposition is $\mathbf{A}^{0.5} = \begin{bmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{bmatrix}$, where $\ell_{11} = \sqrt{a_{11}}$,

  $\ell_{21} = \frac{a_{21}}{\ell_{11}}$, $\ell_{31} = \frac{a_{31}}{\ell_{11}}$, $\ell_{22} = \sqrt{a_{22} - \ell_{21}^2}$, $\ell_{32} = \frac{a_{32} - \ell_{31}\ell_{21}}{\ell_{22}}$, and

  $\ell_{33} = \sqrt{a_{33} - \ell_{31}^2 - \ell_{32}^2}$, which is a set of recursive equations.
- The general formulas for the Cholesky decomposition of a $n \times n$ symmetric positive definite matrix are $\ell_{ii} = \sqrt{a_{ii} - \sum_{j=1}^{i-1} \ell_{ij}^2}$ and $\ell_{ij} = \frac{a_{ij} - \sum_{\ell=1}^{j-1} \ell_{i\ell}\ell_{j\ell}}{\ell_{jj}}$, where $i > \ell$, the sums are zeros when $i = 1$, and $i, j = 1, 2, \ldots, n$.

Constructing Structural Shocks for a VAR

- Orthogonalized shocks give meaningful economic content to IRFs.
- The structural shocks $\eta_t \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n) \implies$ orthogonalized shocks.
  1. Structural shocks are uncorrelated and have unit variance.
  2. The mapping from the reduced-form errors, ε_t, to η_t is $\varepsilon_t = \mathbf{D}\eta_t$.
  3. Reduced-form errors are linear combinations of the structural shocks.
- Calculate a linear combination of η_t using $\eta_t = \mathbf{D}^{-1}\varepsilon_t \implies y_t = \sum_{j=0}^{\infty} \mathbf{G}_j \eta_{t-j}$, where the IRFs are $\mathbf{G}_j = \mathbf{C}_j\mathbf{D}$, $j = 0, 1, \ldots, \infty$, and $\mathbf{C}_0 = \mathbf{I}_n$ yielding $\mathbf{G}_0 = \mathbf{D}$.
- An implication of the Cholesky decomposition is

$$\mathbf{D}\mathrm{E}\left\{\eta_t \eta_t'\right\}\mathbf{D}' = \mathrm{E}\left\{\varepsilon_t \varepsilon_t'\right\} \implies \mathbf{D}\boldsymbol{\Omega}_\eta\mathbf{D}' = \boldsymbol{\Omega} \implies \mathrm{E}\left\{\eta_t \eta_t'\right\} = \mathbf{I}_n.$$

Cholesky or Recursive Ordering

- The problem is the Cholesky decomposition is not unique $\Longrightarrow$ there exist a (countably) infinite ways to rotate $\mathbf{\Omega}$ to generate $\mathbf{D}$.
- Sims (1980) works with a Wold representation of a VAR(p) that imposes a lower triangular restriction on $\mathbf{G}_0 = \mathbf{D} \Longrightarrow$ a recursive identification.
- The bivariate MA(∞) used to study Granger-causality is an example $\Longrightarrow B_{12}(\mathbf{L}) = 0$.
- This example shows Sims proposes to order the elements of ε_t recursively.
  1. Only $\varepsilon_{1,t}$ is fundamental for $y_{1,t}$,
  2. a linear combination of $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ is fundamental for $y_{2,t}$,
  3. a linear combination of $\varepsilon_{1,t}$, $\varepsilon_{2,t}$, and $\varepsilon_{3,t}$ is fundamental for $y_{3,t}$,
  4. ... to a linear combination of $\varepsilon_{1,t}, \ldots, \varepsilon_{n-1,t}$, is fundamental for $y_{n,t}$.
  5. A Cholesky decomposition of $\mathbf{\Omega} \Longrightarrow$ a recursive identification same as just-identified IV, where $\varepsilon_{1,t}, \ldots, \varepsilon_{n-1,t}$ are the instruments.
- Giving an economic interpretation to this identification scheme, or any scheme, is central to the study and implementation of structural VARs.

Sims' Recursive Ordering and Identification

- The Cholesky decomposition generates a lower triangular representation, which is Sims' orthogonalization.

- Sims' orthogonalization or recursive ordering yields a just-identified system.
  1. Remember that $\mathbf{\Omega}$ is a covariance matrix $\Longrightarrow$ it is symmetric.
  2. The dimensions of $\mathbf{\Omega}$ give a $n \times n$ matrix $\Longrightarrow$ n^2 elements.
  3. A Cholesky decomposition of $\mathbf{\Omega}$ is also a $n \times n$ matrix, but it places zeros above (or below) the diagonal.
  4. $\Longrightarrow$ There are $0.5n(n+1)$ elements left unrestricted by a Cholesky decomposition $\Longrightarrow$ count the unrestricted diagonal elements.
  5. A recursive identification restricts $0.5n(n-1)$ $[= n^2 - 0.5n(n+1)]$ elements of $\mathbf{D}$ to $= 0$ $\Longrightarrow$ $\mathbf{D}$ is a lower (or upper) triangular matrix.

- A recursive identification produces a just-identified structural VAR grounded in hypotheses that cannot be tested directly.

Notes about Identified IRFs

- An economic interpretation of a SVAR identification scheme is an econometric measurement of movements in $y_{i,t}$ generated by a small changes in $\eta_{\ell,t}$.
- However, there are free parameters when computing identified IRFs.
- The height of IRFs at impact has to be restricted a priori.
  1. Under a recursive identification, the impact matrix, $\mathbf{G}_0$ $(= \mathbf{D})$, is a nonlinear function of $\mathbf{\Omega}$.
  2. Choice of $\mathbf{G}_{0,i,\ell}$, $\ell = i$, on which to normalize the remaining diagonal elements of $\mathbf{G}_0$ is not arbitrary $\Longrightarrow$ the choice can affect (the shape of) the likelihood of the SVAR.
  3. A "rule" for $\mathbf{G}_{0,i,\ell}$ that leaves the likelihood unchanged sets $\mathbf{G}_{0,\ell,\ell} > 0$, $\ell = 1, \ldots, n \Longrightarrow$ "estimate" $\hat{\mathbf{G}}_{0,\ell,\ell} = \left|\mathbf{G}_{0,\ell,\ell}\right|$.
- Normalization of the impact matrix of an identified set of IRFs is more difficult under non-recursive identification schemes; see Waggoner and Zha (2003, "Likelihood preserving normalization in multiple equation models," *Journal of Econometrics* 114, 329–347).

Confidence Bands for IRFs

- Easy to compute standard errors of $\mathbf{B}_j$s and $\mathbf{\Omega} \Longrightarrow$ OLS or SUR estimators.
- Difficult to obtain good numerical estimates of IRF covariance matrices.
  1. The problem is not a lack of an asymptotic distribution for the IRFs; see Mittnik and Zadrozny (1993, "Asymptotic distributions of impulse responses, step responses, and variance decompositions of estimated linear dynamic models," *Econometrica* 61, 857–870).
  2. They show $\mathbf{IRF}(h)$ is asymptotically normal, given IRFs are nonlinear functions of the $\mathbf{B}_j$s and $\mathbf{\Omega}$.
  3. The variance of $\mathbf{IRF}(h)$ is asymptotically distributed $\chi^2(1)$.
- Numerical computation of IRF covariance matrices rely either on an asymptotic approximation or simulation methods.

Asymptotic Approximation of Confidence Bands for IRFs

- The δ-method is used to construct an approximation of the asymptotic distribution of the IRF covariance matrix.
- Using the definition $\mathbf{IRF}(h) \equiv \mathbf{IRF}(h, \beta)$, the δ-method employs

$$\sqrt{T}\left(\widehat{\mathbf{IRF}}_{i,\ell}(h) - \mathbf{IRF}_{i,\ell}(h)\right) \sim \mathcal{N}\left(0, \frac{\partial \mathbf{IRF}_{i,\ell}(h, \boldsymbol{\beta})}{\partial \mathrm{vec}(\boldsymbol{\beta})} \boldsymbol{\Omega}_{\boldsymbol{\beta}} \frac{\partial \mathbf{IRF}_{i,\ell}(h, \boldsymbol{\beta})}{\partial \mathrm{vec}(\boldsymbol{\beta})}'\right).$$

- Approximate standard errors by the ratio of the total derivative of the date $t+h$ forecast error to the total derivative of $\sum_{j=0}^{H} \mathbf{F}^j$.
- These standard errors of IRFs have
  1. poor small sample properties for sample sizes, T, typical in macro,
  2. empirical distributions far from χ^2 in large samples, given persistence in y_t (*i.e.*, near unit roots),
  3. and inherent properties of the estimated standard errors of the $\mathbf{B}_j$s and $\boldsymbol{\Omega}$, which are often large relative to the point estimates.

Confidence Bands for IRFs using Bootstrap Simulation

- The bootstrap is a nonparametric method for generating an empirical distribution of the IRFs from $h = 0, 1, \ldots, H$.
  1. Nonparametric $\Longrightarrow$ does not invoke parametric assumptions about $\mathbf{B}_j$ and/or ε_t to generate synthetic samples of the IRFs.
  2. Empirical distributions $\Longrightarrow$ small sample distributions built from estimated VAR.

- Bootstrapped IRF confidence bands resample $\widehat{\varepsilon}_t$ with replacement to create K synthetic samples of length T of the VAR's errors.
  1. Random shuffling of $\{\widehat{\varepsilon}_t\}_{t=0}^{T}$ produces $\left\{\left\{\widetilde{\varepsilon}_t^k\right\}_{t=1}^{T}\right\}_{k=1}^{K}$.
  2. Generate $\left\{\left\{\widetilde{y}_t^k\right\}_{t=1}^{T}\right\}_{k=1}^{K}$ using $\left\{\left\{\widetilde{\varepsilon}_t^k\right\}_{t=1}^{T}\right\}_{k=1}^{K}$ and the OLS estimates of the VAR(p) $\Longrightarrow \widehat{\mathbf{B}}_j$, $j = 1, 2, \ldots, p$ $\Longrightarrow$ do not resample the $\widehat{\mathbf{B}}_j$s.
  3. Estimate the VAR(p) on the K synthetic samples, $\left\{\left\{\widetilde{y}_t^k\right\}_{t=1}^{T}\right\}_{k=1}^{K}$, construct $\left\{\widetilde{G}_0^k\right\}_{k=1}^{K}$ as described above to produce the bootstrapped empirical distribution $\left\{\left\{\widetilde{\mathbf{IRF}}^k(h)\right\}_{h=0}^{H}\right\}_{k=1}^{K}$.

Notes on Bootstrapped Confidence Bands for IRFs

- There are problems when computing bootstrapped IRF confidence bands.
  1. Bootstrapped distributions of $\mathbf{IRF}(h)$ are often not symmetric.
  2. Non-normal distributions suggest reporting percentiles instead of standard deviations, especially for the sample sizes found in macro.
  3. Bootstrapped IRF confidence bands require $\hat{\varepsilon}_t \sim \mathcal{WN}$, otherwise biased confidence bands.
  4. Resampling is sensitive to the units in which y_t is measured.
  5. Change method for rendering y_t stationary $\Longrightarrow$ alter shape of bootstrapped IRF confidence bands, which affects inference.
  6. OLS estimates of the $\mathbf{B}_j$s are biased downward $\Longrightarrow$ this bias is transmitted into the bootstrapped IRFs.
  7. Set K large enough to produce small simulation error; see Andrews and Buchinsky (2000, "A three-step method for choosing the number of bootstrap repetitions," *Econometrica* 68, 23–51).
  8. See Pesavento and Rossi (JAE, 2006 and JEDC, 2007) for proposed solutions in small samples.

Building Blocks for Bayesian Confidence Bands for IRFs, I

- Remember a VAR(p) has a static regression representation, $\mathbf{y}_t = \mathbf{\Theta}'\mathbf{X}_t + \boldsymbol{\varepsilon}_t$, where $\mathbf{\Theta}$ is $n \times (np+1)$, $\mathbf{X}_t$ is $(np+1) \times 1$, and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{\Omega})$.
- Denote $\mathbb{Y}_{T\times n} = [\mathbf{y}_1\ \mathbf{y}_2\ \ldots\ \mathbf{y}_T]'$, $\mathbb{X}_{T\times np} = [\mathbf{X}_1\ \mathbf{X}_2\ \ldots\ \mathbf{X}_T]'$, $\mathbb{B}_{(np+1)\times n} = [\mathbf{B}_1'\ \mathbf{B}_2'\ \ldots\ \mathbf{B}_p']'$, and $\mathbb{\Xi}_{T\times n} = [\boldsymbol{\varepsilon}_1\ \ \boldsymbol{\varepsilon}_2\ \ldots\ \boldsymbol{\varepsilon}_T]' \Longrightarrow$ a static simultaneous equations system $\mathbb{Y} = \mathbb{X}\mathbb{B} + \mathbb{\Xi}$.
- The regression for $\mathbf{y}_{j,t}$ is the jth row of this system $\Longrightarrow \mathbb{Y}_j = \mathbb{X}\mathbb{B}_j + \mathbb{\Xi}_j$.
- Columns of the j (= 1, 2, ..., n) regressions are stacked in ascending order to obtain $\mathbf{y} = \left(\mathbf{I}_n \otimes \mathbb{X}\right)\mathbf{b} + \boldsymbol{\xi} \equiv \mathbf{X}\mathbf{b} + \boldsymbol{\xi}$, where $\mathbf{y}$ and $\boldsymbol{\xi}$ are $nT \times 1$ column vectors, $\mathbf{b} = \text{vec}(\mathbb{B}')$, and $\mathbf{X}$ is a $nT \times n(np+1)$ matrix $\Longrightarrow$ organize regressions in blocks of rows.
- Given several assumptions, show the likelihood of the stacked system of regressions of the VAR(p) crosses the conditional (on the $\widehat{\mathbf{B}}_j$s) normality of $\mathbf{b}$ with the Wishart distribution of $\mathbf{\Omega}_{\boldsymbol{\xi}}^{-1}$, where $\mathbf{E}\left\{\boldsymbol{\xi}\boldsymbol{\xi}'\right\} = \mathbf{\Omega}_{\boldsymbol{\xi}} = \mathbf{\Omega} \otimes \mathbf{I}_T$.
- Bayesian Monte Carlo simulation generates the distribution of IRFs given the data, $\mathbf{y}$, the VAR(p), and prior information on $\mathbf{b}$ and $\mathbf{\Omega}_{\boldsymbol{\xi}}$.

Building Blocks for Bayesian Confidence Bands for IRFs, II

- The likelihood of the VAR(p), $\mathbf{y} = (\mathbf{I}_n \otimes \mathbb{X})\mathbf{b} + \boldsymbol{\xi}$, $\boldsymbol{\xi} \sim \mathcal{N}\left(\mathbf{0}_{nT\times 1}, \boldsymbol{\Omega}_{\boldsymbol{\xi}}\right)$, is

$$\begin{aligned}\mathcal{H}\left(\mathbf{b}, \boldsymbol{\Omega}_{\boldsymbol{\xi}} \middle| \mathbf{y}\right) \quad &\propto \quad \left|\boldsymbol{\Omega}_{\boldsymbol{\xi}} \otimes \mathbf{I}_T\right|^{-0.5} \\ &\times \exp\left\{-\frac{1}{2}\left[\mathbf{y} - \left(\mathbf{I}_n \otimes \mathbb{X}\right)\mathbf{b}\right]'\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-1} \otimes \mathbf{I}_T\right)\left[\mathbf{y} - \left(\mathbf{I}_n \otimes \mathbb{X}\right)\mathbf{b}\right]\right\}.\end{aligned}$$

- The goal is to factor $\mathcal{H}\left(\mathbf{b}, \boldsymbol{\Omega}_{\boldsymbol{\xi}} \middle| \mathbf{y}\right)$ $\left(= \exp\left\{\mathcal{L}\left(\mathbf{b}, \boldsymbol{\Omega}_{\boldsymbol{\xi}} \middle| \mathbf{y}\right)\right\}\right)$ into recognizable prior distributions from which posterior (synthetic) distributions can be drawn by Monte Carlo simulations.

- The term inside $\exp\{\cdot\}$ can be factored as

$$\begin{aligned}&\left[\mathbf{y} - \left(\mathbf{I}_n \otimes \mathbb{X}\right)\mathbf{b}\right]'\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-1} \otimes \mathbf{I}_T\right)\left[\mathbf{y} - \left(\mathbf{I}_n \otimes \mathbb{X}\right)\mathbf{b}\right] \\ &= \quad \left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbf{I}_T\right)\left[\mathbf{y} - \left(\mathbf{I}_n \otimes \mathbb{X}\right)\mathbf{b}\right]'\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbf{I}_T\right)\left[\mathbf{y} - \left(\mathbf{I}_n \otimes \mathbb{X}\right)\mathbf{b}\right] \\ &= \quad \left[\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbf{I}_T\right)\mathbf{y} - \left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbb{X}\right)\mathbf{b}\right]'\left[\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbf{I}_T\right)\mathbf{y} - \left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbb{X}\right)\mathbf{b}\right].\end{aligned}$$

Building Blocks for Bayesian Confidence Bands for IRFs, III

- The OLS estimator of $\hat{\mathbf{b}} = \left[\boldsymbol{\Omega}_{\xi}^{-1} \bigotimes \mathbf{X}'\mathbf{X}\right]^{-1}\left[\boldsymbol{\Omega}_{\xi}^{-1} \bigotimes \mathbf{X}\right]'\mathbf{y}$.
- Add and subtract $\hat{\mathbf{b}}$ from $\mathbf{b}$ in $\left[\left(\boldsymbol{\Omega}_{\xi}^{-0.5} \bigotimes \mathbf{I}_T\right)\mathbf{y} - \left(\boldsymbol{\Omega}_{\xi}^{-0.5} \bigotimes \mathbf{X}\right)\mathbf{b}\right]$

$$\left(\boldsymbol{\Omega}_{\xi}^{-0.5} \bigotimes \mathbf{I}_T\right)\mathbf{y} - \left(\boldsymbol{\Omega}_{\xi}^{-0.5} \bigotimes \mathbf{X}\right)\left(\mathbf{b} - \hat{\mathbf{b}} + \hat{\mathbf{b}}\right)$$

$$= \left(\boldsymbol{\Omega}_{\xi}^{-0.5} \bigotimes \mathbf{I}_T\right)\mathbf{y} - \left(\boldsymbol{\Omega}_{\xi}^{-0.5} \bigotimes \mathbf{X}\right)\hat{\mathbf{b}} + \left(\boldsymbol{\Omega}_{\xi}^{-0.5} \bigotimes \mathbf{X}\right)\left(\mathbf{b} - \hat{\mathbf{b}}\right).$$

- Use this result to reconstruct the likelihood of the VAR(p),

$$\begin{aligned}\mathcal{N}\left(\mathbf{b}, \boldsymbol{\Omega}_{\xi} \middle| \mathbf{y}\right) \propto \left|\boldsymbol{\Omega}_{\xi} \bigotimes \mathbf{I}_T\right|^{-0.5} \exp\Bigg\{ & -\frac{1}{2}\left[\left(\boldsymbol{\Omega}_{\xi}^{-0.5} \bigotimes \mathbf{I}_T\right)\mathbf{y} - \left(\boldsymbol{\Omega}_{\xi}^{-0.5} \bigotimes \mathbf{X}\right)\hat{\mathbf{b}}\right]' \\ & \times \left[\left(\boldsymbol{\Omega}_{\xi}^{-0.5} \bigotimes \mathbf{I}_T\right)\mathbf{y} - \left(\boldsymbol{\Omega}_{\xi}^{-0.5} \bigotimes \mathbf{X}\right)\hat{\mathbf{b}}\right] \\ & + \left(\mathbf{b} - \hat{\mathbf{b}}\right)'\left(\boldsymbol{\Omega}_{\xi}^{-1} \bigotimes \mathbf{X}'\mathbf{X}\right)\left(\mathbf{b} - \hat{\mathbf{b}}\right)\Bigg\}.\end{aligned}$$

Building Blocks for Bayesian Confidence Bands for IRFs, IV

- The reconstructed likelihood

$$\begin{aligned}
\mathscr{H}\left(\mathbf{b}, \boldsymbol{\Omega}_{\boldsymbol{\xi}} \middle| \mathbf{y}\right) \propto & \left|\boldsymbol{\Omega}_{\boldsymbol{\xi}} \otimes \mathbf{I}_T\right|^{-0.5} \exp\left\{-\tfrac{1}{2}\left[\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbf{I}_T\right)\mathbf{y} - \left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbb{X}\right)\hat{\mathbf{b}}\right]'\right. \\
& \times\left[\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbf{I}_T\right)\mathbf{y} - \left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbb{X}\right)\hat{\mathbf{b}}\right] \\
& \left.-\tfrac{1}{2}\left(\mathbf{b}-\hat{\mathbf{b}}\right)'\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-1} \otimes \mathbb{X}'\mathbb{X}\right)\left(\mathbf{b}-\hat{\mathbf{b}}\right)\right\} \\
= & \left|\boldsymbol{\Omega}_{\boldsymbol{\xi}}\right|^{-0.5m} \exp\left\{-\tfrac{1}{2}\left(\mathbf{b}-\hat{\mathbf{b}}\right)'\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-1} \otimes \mathbb{X}'\mathbb{X}\right)\left(\mathbf{b}-\hat{\mathbf{b}}\right)\right\} \\
\times & \left|\boldsymbol{\Omega}_{\boldsymbol{\xi}}\right|^{-0.5(T-m)} \exp\left\{-\tfrac{1}{2}\mathrm{tr}\left(\left[\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbf{I}_T\right)\mathbf{y} - \left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbb{X}\right)\hat{\mathbf{b}}\right]'\right. \\
& \left.\left[\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbf{I}_T\right)\mathbf{y} - \left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-0.5} \otimes \mathbb{X}\right)\hat{\mathbf{b}}\right]\right)\Bigg\}
\end{aligned}$$

where $m = np + 1 \Longrightarrow$ the number of coefficients per regression and nm are the total number of coefficients in the VAR(p).

Building Blocks for Bayesian Confidence Bands for IRFs, V

- The term $\left|\boldsymbol{\Omega}_{\boldsymbol{\xi}}\right|^{-0.5m} \exp\left\{-0.5\left(\mathbf{b}-\hat{\mathbf{b}}\right)'\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-1}\bigotimes \mathbb{X}'\mathbb{X}\right)\left(\mathbf{b}-\hat{\mathbf{b}}\right)\right\}$ is the distribution of $\mathbf{b}$ conditional on $\hat{\mathbf{b}}$, $\boldsymbol{\Omega}_{\boldsymbol{\xi}}$, $\mathbf{y}$, and $\mathbb{X}$, which is normal.

- The second term is the moment matrix of the residuals $\hat{\boldsymbol{\xi}}$ $\Longrightarrow$ the covariance matrix $\boldsymbol{\Omega}_{\boldsymbol{\xi}} \times T$.

- The Wishart distribution is produced by the moment matrix $\mathbf{Z} = \mathbf{z}\mathbf{z}'$,
  1. where the $g \times s$ matrix $\mathbf{z} \sim \mathcal{N}\left(\mathbf{0}_{g\times s}, \boldsymbol{\Omega}_{\mathbf{z}}\right) \Longrightarrow \mathbf{Z} \sim \mathcal{W}\left(\boldsymbol{\Omega}_{\mathbf{z}}, g\right)$.
  2. $\mathbf{Z}$ has the Wishart distribution with mean $\boldsymbol{\Omega}_{\mathbf{z}}$ and g dfs.
  3. $\Longrightarrow$ A sample covariance matrix drawn from a multivariate normal distribution is distributed Wishart.
  4. The inverse Wishart distribution is $\mathbf{Z}^{-1} \sim \mathcal{W}^{-1}\left(\boldsymbol{\Omega}_{\mathbf{z}}^{-1}, g\right)$.
  5. $\Longrightarrow$ The conjugate prior of $\boldsymbol{\Omega}_{\boldsymbol{\xi}}$ is the inverse Wishart distribution.
  6. $\Longrightarrow$ The conjugate prior of $\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-1}$ is the Wishart distribution.

- These facts decompose the VAR(p)'s likelihood into

$$\mathcal{H}\left(\mathbf{b}, \boldsymbol{\Omega}_{\boldsymbol{\xi}} \middle| \mathbf{y}\right) \quad \propto \quad \mathcal{N}\left(\mathbf{b} \middle| \hat{\mathbf{b}}, \boldsymbol{\Omega}_{\boldsymbol{\xi}}, \mathbb{X}, \mathbf{y}\right) \times \mathcal{W}\left(\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-1} \middle| \mathbf{y}, \mathbb{X}, \hat{\mathbf{b}}, T-m\right).$$

Building Blocks for Bayesian Confidence Bands for IRFs, VI

- The likelihood $\mathcal{H}\left(\mathbf{b}, \boldsymbol{\Omega}_{\boldsymbol{\xi}} \middle| \mathbf{y}\right)$ of a VAR(p) is the product of the conditional normality of $\mathbf{b}$ and since the covariance matrix of the n reduced-form VAR errors, ε_t, are Gaussian, $\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-1}$ has the Wishart distribution with $T - m$ dfs.

- Decomposition yields a posterior of $\mathbf{b} \sim$ normal conditional on $\boldsymbol{\Omega}_{\boldsymbol{\xi}}$ and $\mathbf{y}$.
  1. Center posterior on OLS estimates, $\hat{\mathbf{b}}$, and its covariance matrix.
  2. The posterior and the likelihood share distributional properties $\Longrightarrow$ the prior $\times$ the likelihood is proportional to the posterior.
  3. The prior is the joint distribution of $\mathbf{b}$ and $\boldsymbol{\Omega}_{\boldsymbol{\xi}}$, which is proportional to $\boldsymbol{\Omega}_{\boldsymbol{\xi}}^{-1} \Longrightarrow$ Wishart because there is no information about these parameters beyond that embedded in the VAR(p).

- Analytic decomposition of $\mathcal{H}\left(\mathbf{b}, \boldsymbol{\Omega}_{\boldsymbol{\xi}} \middle| \mathbf{y}\right) \Longrightarrow$ careful choices of priors for $\mathbf{b}$ and $\boldsymbol{\Omega}_{\boldsymbol{\xi}}$ yield posterior distributions that are computed analytically.

- However, there are other priors that are consistent with analytic computation of posterior distributions of VAR parameters.

Bayesian Confidence Bands for IRFs

- Still, $\mathcal{H}\left(\mathbf{b}, \Omega_{\xi} \middle| \mathbf{y}\right) \propto \mathcal{N}\left(\mathbf{b} \middle| \hat{\mathbf{b}}, \Omega_{\xi}, \mathbb{X}, \mathbf{y}\right) \times \mathcal{W}\left(\Omega_{\xi}^{-1} \middle| \mathbf{y}, \mathbb{X}, \hat{\mathbf{b}}, T - m\right)$ suggests a simple Monte Carlo algorithm for generating posterior distributions of $\mathbf{IRF}(h)$ for just identified structural VARs.

- Given OLS estimates of a VAR(p)'s parameters, $\hat{\mathbf{b}}$, errors, $\hat{\Xi} = \mathbb{Y} - \mathbb{X}\hat{\mathbb{B}}$, and error covariance matrix, $\hat{\Omega}$, the multi-step Monte Carlo algorithm of IRF confidence bands consists of the following steps.
  1. Draw the covariance matrix $\Sigma_k \sim \mathcal{IW}\left(\left(\hat{\Xi}'\hat{\Xi}\right)^{-1}, T - m\right)$.
  2. Generate $\mathbf{b}_k = \hat{\mathbf{b}} + \upsilon_{k,mn}$, where $\upsilon_{k,mn}$ is the reshaped row vector of $\vartheta_k \sim \mathcal{N}\left(\mathbf{0}_{mn}, \left[\Sigma_k \otimes \left(\mathbb{X}'\mathbb{X}\right)^{-1}\right]\right)$.
  3. Compute $\mathbf{G}_{0,k} = \Omega_k^{0.5}$, where $\Omega_k = \left(\varepsilon_{k,t}\varepsilon_{k,t}'\right)/T$ and $\varepsilon_{k,t} = y_t - \Theta_k'\mathbf{X}_t$.
  4. Calculate $\mathbf{G}_{k,i,\ell}(h)$ using $\mathbf{G}_{0,k}$ and $\mathbf{b}_k$ (ignore intercepts), $h = 0, 1, 2, \ldots, H$.
  5. Repeat steps $1, \ldots, 4$ for $k = 1, 2, \ldots, \mathcal{K}$ iterations to report quantiles of $\left\{\left\{\mathbf{IRF}_{k,i,\ell}(h)\right\}_{h=0}^{H}\right\}_{k=1}^{K}$, w/r/t $y_{i,t}$ and the ℓth structural shock.

Notes on Bayesian Confidence Bands for IRFs

- This Bayesian Monte Carlo algorithm yields an exact posterior distribution of the likelihood, $\mathcal{H}\left(\mathbf{b}, \mathbf{\Omega}_{\xi} \middle| \mathbf{y}\right)$, conditional on the priors
  1. the VAR errors $\varepsilon_t \sim$ Gaussian,
  2. intercepts and lagged coefficients, $\mathbf{b} \sim$ conditionally normal,
  3. and the covariance matrix of $\mathbf{b}$, $\mathbf{\Omega}_{\xi}^{-1} \sim$ Wishart.

- Asymptotic approximate and bootstrapped confidence bands can be non-pivotal $\Longrightarrow$ these quantities depend on nuisance parameters, which are the VAR(p)'s OLS estimates $\Longrightarrow$ small sample bias matters.

- Sampling the empirical distribution of the IRFs from $\mathcal{H}\left(\mathbf{b}, \mathbf{\Omega}_{\xi} \middle| \mathbf{y}\right)$
  1. relies on the OLS estimates for all the information available in the data under the DGP of a VAR(p) $\Longrightarrow$ small sample bias in $\hat{\mathbf{b}}$ is not an issue.
  2. The algorithm is valid only for recursive identification schemes.
  3. There remain issues with the empirical distribution of the IRFs.

Notes on Bayesian Confidence Bands for IRFs, cont.

- Often report percentiles because the algorithm samples from $\mathcal{H}\left(\mathbf{b}, \boldsymbol{\Omega}_{\boldsymbol{\xi}} \middle| \mathbf{y}\right)$ $\Longrightarrow$ standard error bands are biased estimates of the coverage intervals of IRFs because a VAR's likelihood is not symmetric.

- An issue with inference using the coverage interval around $IRF(h)$
  1. is akin to a t-ratio $\Longrightarrow$ only the statistical significance of IRF(h).
  2. Coverage intervals of IRFs are not joint tests of IRF$(h), \ldots,$ IRF$(h+j)$ $\Longrightarrow$ not an F-, Lagrange multiplier, or Wald test.

- This is especially a problem because IRFs are often serially correlated across horizons $h = 0, 1, 2, \ldots, H$.

Serial Correlation and Bayesian Confidence Bands for IRFs

- Coverage intervals are biased if changes in uncertainty of an IRF at horizon h affects the uncertainty of nearby IRFs $\Longrightarrow$ serially correlated IRFs.
- The inference problem is that variation in $\mathbf{IRF}(h)$ is not independent of $\mathbf{IRF}(h+j)$ when there is serial correlation in IRFs.
- Sims and Zha (1999) assume that $\left\{\mathbf{G}_h\right\}_{h=0}^{H}$ are multivariate normal with covariance matrix $\mathbf{\Omega_G}$.
- Measure uncertainty or variability around $\left\{\mathbf{G}_h\right\}_{h=0}^{H}$ by projecting on the largest principal components of $\mathbf{\Omega_G}$.
  1. Principal components (PCs) are orthogonalized measures of the variability (*i.e.*, variance) of a covariance stationary multivariate time series.
  2. The largest PC of $\mathbf{\Omega_G}$ is the (unobserved) factor responsible for the largest share of variance in this symmetric (and positive definite) matrix.
  3. An eigenvalue decomposition is an estimator of PCs $\Longrightarrow$ the largest PC is a function of the largest eigenvalue.
- Computing PCs by eigenvalue decomposition adds two steps to the Monte Carlo algorithm for calculating IRF error bands.

Computing Bayesian Confidence Bands for IRFs, I

- Denote the $H+1 \times H+1$ covariance matrix of $\left\{\left\{\mathbf{G}_{k,i,\ell}(h)\right\}_{k=1}^{\mathcal{K}}\right\}_{h=0}^{H}$, $\mathbf{\Omega}_{\mathbf{G},i,\ell}$, which has a Jordan normal form $\mathbb{P}_{\mathbf{G}_{i,\ell}}\mathbf{\Lambda}_{\mathbf{G}_{i,\ell}}\mathbb{P}'_{\mathbf{G}_{i,\ell}}$, where the eigenvalues of $\mathbf{\Omega}_{\mathbf{G}_{i,\ell}}$ live on the diagonal of $\mathbf{\Lambda}_{\mathbf{G}_{i,\ell}}$ and the matrix of eigenvectors obeys $\mathbb{P}_{\mathbf{G}_{i,\ell}}\mathbb{P}'_{\mathbf{G}_{i,\ell}} = \mathbf{I}_{H+1\times H+1} \implies$ eigenvectors have unit length.

  1. $\lambda_{h,i,\ell}$ is the hth eigenvalue of $\mathbf{\Omega}_{\mathbf{G},i,\ell}$ and is the (h, h) element of $\mathbf{\Lambda}_{\mathbf{G}_{i,\ell}}$, where the eigenvector tied to $\lambda_{h,i,\ell}$ is $\mathbb{P}_{\mathbf{G}_{i,\ell},\cdot,h}$, $\implies$ the hth column of $\mathbb{P}_{\mathbf{G}_{i,\ell}}$.
  2. Define $\overline{\mathbf{G}}_{i,\ell}(h) = \mathcal{K}^{-1}\sum_{k=1}^{\mathcal{K}}\mathbf{G}_{i,\ell}^{k}(h)$, $h = 0, 1, 2, \ldots, H$, and $\overline{\mathbf{G}}_{i\ell}$ is the $H+1$ column vector of the Monte Carlo averages $\overline{\mathbf{G}}_{i,\ell}(h)$.
  3. $\implies$ Add a step to calculate $\mathbf{\Omega}_{\mathbf{G}_{i,\ell}}$, $\mathbf{\Lambda}_{\mathbf{G}_{i,\ell}}$, and $\mathbb{P}_{\mathbf{G}_{i,\ell}}$ and tabulate IRF error bands.

Computing Bayesian Confidence Bands for IRFs, II

- Error bands of **IRF**(h)s could be computed as $\overline{\mathbf{G}}_{i,\ell} \pm \alpha \sum_{h=1}^{H} \mathbb{P}_{\mathbf{G}_{i,\ell},\cdot,h} \sqrt{\lambda_{h,i,\ell}}$, where α denotes the significance level.
- However, there are often only a few λs that are "large" in absolute value.
  1. **Symmetric** standard error bands are $\overline{\mathbf{G}}_{i,\ell} \pm \sum_{m=f}^{sup} \mathbb{P}_{\mathbf{G}_{i,\ell},\cdot,m} \sqrt{\lambda_{m,i,\ell}}$ and $\overline{\mathbf{G}}_{i,\ell} \pm 1.96 \sum_{m=f}^{sup} \mathbb{P}_{\mathbf{G}_{i,\ell},\cdot,m} \sqrt{\lambda_{m,i,\ell}}$ for 1-standard deviation (*i.e.*, 68%) and 95% (*i.e.*, 2-standard deviation) confidence intervals, where f and sup denote the first non-negligible and largest eigenvalues (sup could $= f$) of $\mathbf{\Omega}_{\mathbf{G}_{i,\ell}}$ and the associated eigenvalue(s) $\Longrightarrow$ assume IRFs $\sim$ normal.
  2. **Asymmetric** percentile bands are $\overline{\mathbf{G}}_{i,\ell} + \gamma_{i,\ell}^{\alpha}$ and $\overline{\mathbf{G}}_{i,\ell} + \gamma_{i,\ell}^{1-\alpha}$, where $\gamma_{i,\ell}^{s} = \sum_{m=f}^{sup} \mathbb{P}_{\mathbf{G}_{k,i,\ell},m,\cdot} \mathbf{G}_{i,\ell}^{s}$ involves element by element multiplication of the $H+1$ elements of the row vector $\mathbb{P}_{\mathbf{G}_{i,\ell},m,\cdot}$ and the column vector $\mathbf{G}_{k,i,\ell}^{s}$, α is the significance level, $s = \alpha, 1-\alpha$, $\mathbb{P}_{\mathbf{G}_{i,\ell},m,\cdot}$ is the mth row of $\mathbb{P}_{\mathbf{G}_{i,\ell}}$, and $\mathbf{G}_{k,i,\ell}^{s}$ is the kth draw from the ensemble $\left\{\mathbf{G}_{k,i,\ell}\right\}_{k=1}^{\mathcal{K}}$ of IRFs.

Sup-t Confidence Bands

- An old statistics problem is conducting inference on a region of the parameter space.
- Suppose the hypothesis is a conditional moment (*i.e.* mean, median, etc.) that can take on many potential values.
  1. The test statistic is a confidence band of the conditional moment.
  2. The confidence band gives lower and upper bounds in which reside more than one potential realization of the moment, given a significance level.
  3. See Working and Hotelling (1929, "Application of the theory of error to the interpretation of trends," *Journal of the American Statistical Association* 24(Supplement), 73–85).
- Olea and Plagborg-Møller (2019) develop methods to construct uncertainty bands for IRFs in the tradition of Working and Hotelling (1929).
  1. Calculate quantiles to construct uncertainty bands that use all the IRF forecast horizons simultaneously.
  2. Show a *Sup-t* confidence band is best with respect to asymptotic coverage.
  3. The analytic result is supported by Monte Carlo results indicating the Bayesian version of the *Sup-t* confidence band dominates (in small sample).
  4. Differs from Sims and Zha (1998) $\Longrightarrow$ they rely only on the largest PCs to remove the sources of serial correlation in uncertainty bands.

Plug-in *Sup-t* Confidence Bands for IRFs

- Olea and Plagborg-Møller (2019) propose to construct Bayesian credible intervals for IRFs with one of two algorithms.
- Algorithm 1 needs the mean, $\overline{\mathbf{G}}_{i\ell}$, (or median) and the covariance matrix, $\mathbf{\Omega}_{\mathbf{G}_{i,\ell}}$, of the posterior distribution of the IRFs.
  1. Pull the standard deviations of the IRFs from $\mathbf{\Omega}_{\mathbf{G}_{i,\ell}}$.
  2. Eliminate standard deviations ≈ 0 (*i.e.*, $\sqrt{\texttt{machine error}}$).
  3. Adjust the IRF horizon for the deleted standard deviations, H_{adj}, to construct the correlation matrix, $\mathbf{\Omega}^{Corr}_{\mathbf{G}_{i,\ell}}$, of $\mathbf{\Omega}_{\mathbf{G}_{i,\ell}}$ using its eigenvalues.
  4. Draw $\mathcal{K}$ samples of t-stats $\sim \mathcal{N}\left(\mathbf{0}_{H_{adj}\times 1}, \mathbf{\Omega}^{Corr}_{\mathbf{G}_{i,\ell}}\right)$.
  5. The *Sup-t* credible intervals are the $1-\alpha$ percent quantiles of the $\mathcal{K}$ samples of the t-stats $\Longrightarrow \overline{\mathbf{G}}_{i\ell} \pm q_{1-\alpha}\sqrt{\texttt{Diag}\left(\mathbf{\Omega}_{\mathbf{G}_{i,\ell}}\right)}$.

Sup-t Bayesian Credible Intervals for IRFs

- Algorithm 2 produces a "quantile-based bootstrap or Bayes band" $\Longrightarrow$ for a confidence level, $1-\alpha$, calculate pointwise Bayesian credible intervals.
  1. Multiple 'equal tailed' credible intervals across all horizons $\Longrightarrow$ estimated credible intervals, $\widehat{C}$, are the Cartesian product $\widehat{C}_0 \times \widehat{C}_1 \times \widehat{C}_2 \times \ldots \times \widehat{C}_H$.
  2. Choose probabilities in the tails of the posterior of the IRFs to construct intervals with credibility of $1-\alpha$ percent simultaneously for all h.

- Only need the posterior distribution of the IRFs, $\left\{\left\{\mathbf{G}_{k,i,\ell}(h)\right\}_{k=1}^{\mathcal{K}}\right\}_{h=0}^{H}$, to compute α_x that achieves simultaneous credibility of $1-\alpha$ percent.
  1. Solve a nonlinear equation for α_x to find the fraction of $\mathcal{K}$ draws inside the credibility interval $[0.5\alpha, 1-0.5\alpha]$ for $h=0,\ldots,H \Longrightarrow$ quantiles.
  2. Credible intervals $\in [0.5\alpha_x\%,\ (1-0.5\alpha_x)\%]$ quantiles of $\left\{\mathbf{G}_{k,i,\ell}(h)\right\}_{k=1}^{\mathcal{K}}$.

- Olea and Plagborg-Møller show the plug-in *Sup-t* uncertainty bands and *Sup-t* Bayesian credible intervals are asymptotically equivalent.

- MatLab™ code for algorithms 1 and 2 are provided by Olea and Plagborg-Møller at `https://github.com/jm4474/Confidence_Bands`.

Measuring the Variability of Identified Shocks

- The forecast error variance decomposition (FEVD) measures the importance of an element of ε_t to forecast or explain the magnitude of fluctuations in $y_{i,t+h}$.
- This suggests that FEVDs reveals information about the impact of a change in $\varepsilon_{\ell,t}$ on the variability of fluctuations in $y_{i,t+h}$.
- A different piece of information about $y_{i,t+h}$ than found in IRFs.
- The IRF is only about the shape of the response of $y_{i,t+h}$ to a one unit change in $\varepsilon_{\ell,t}$.
- The FEVD explains *the importance of changes in* $\varepsilon_{\ell,t}$ *for future movements in* $y_{i,t+h}$.
- For FEVDs, the objects of interest revolve around the variance and MSEs.
- Nonetheless, IRFs and FEVDs employ the same information $\Longrightarrow$ the VAR(p)'s slope coefficients, errors, and covariance matrix of the errors.

Computing FEVDs, I

- When $y_t \sim$ Wold Decomposition, the forecast error at date $t+h$ is

$$y_{t+h} - \mathbf{E}_t \{y_{t+h}\} = \sum_{j=0}^{h-1} \mathbf{C}_j \varepsilon_{t+h-j}.$$

- At forecast horizon h, its MSE is

$$\mathbf{MSE}_y(h) \equiv \mathbf{E}\left\{[y_{t+h} - \mathbf{E}_t \{y_{t+h}\}][y_{t+h} - \mathbf{E}_t \{y_{t+h}\}]'\right\} = \sum_{j=0}^{h-1} \mathbf{C}_j \mathbf{\Omega}_{\boldsymbol{\varepsilon}} \mathbf{C}_j'.$$

- Switching to the VAR(1) companion form of a VAR(p) gives the forecast error at date $t+h$ as

$$z_{t+h} - \mathbf{E}_t \{z_{t+h}\} = \sum_{j=0}^{h-1} \mathbf{F}^j \nu_{t+h-j}.$$

- The MSE of this forecast error is

$$\mathbf{MSE}_z(h) \equiv \mathbf{E}\left\{[z_{t+h} - \mathbf{E}_t \{z_{t+h}\}][z_{t+h} - \mathbf{E}_t \{z_{t+h}\}]'\right\} = \sum_{j=0}^{h-1} \mathbf{F}^j \mathbf{Q} \mathbf{F}^{j\prime}.$$

Computing FEVDs, II

- Suppose the questions of interest concerns the h-step ahead FEVD of the ith element of y_t w/r/t to the ℓth shock of ε_t.
- The question asks about the contribution of a one unit change in $\varepsilon_{\ell,t}$ to the variability of $y_{i,t+h}$.
  1. $\Longrightarrow$ The MSE has to be normalized by the total MSE of $y_{i,t+h}$.
  2. The $\mathbf{FEVD}_{\ell,i}(h)$ measures the fraction or percentage of the variability of $y_{i,t+h}$ accounted for by $\varepsilon_{\ell,t}$.
- The $\mathbf{FEVD}_{y,\ell,i}(h) = \dfrac{\mathbf{MSE}_{y,\ell,i}(h)}{\mathbf{MSE}_{y}(h)}$ or $\mathbf{FEVD}_{z,\ell,i}(h) = \dfrac{\mathbf{MSE}_{z,\ell,i}(h)}{\mathbf{MSE}_{z}(h)}$.
- Thus, *a FEVD measures the relative contribution of $\varepsilon_{\ell,t}$ to fluctuations in $y_{i,t+h}$* using information in IRFs $\Longrightarrow$ the coefficient matrices of a VMA(∞) or the lag coefficient matrix of the associated companion VAR(1).

Identified FEVDs

- FEVDs face the same identification issues as IRFs $\Longrightarrow$ built on same VMA(∞).
- Define $\eta_t = \mathbf{D}^{-1}\varepsilon_t \Longrightarrow y_{t+h} - \mathbf{E}_t\{y_{t+h}\} = \sum_{j=0}^{h-1} \mathbf{G}_j \eta_{t+h-j}$, where $\mathbf{G}_j = \mathbf{C}_j\mathbf{D}$.
- The total h step-ahead MSE of the shocks identified by the recursive ordering is $\mathbf{MSE}_{y,\mathbf{G}}(h) = \sum_{j=0}^{h-1} \mathbf{G}_j\mathbf{G}_j'$.
- As shown in the previous slide, FEVDs are computed as the percentage contribution of ε_t to fluctuations in y_{t+h}.
- The relative contribution of a change in $\varepsilon_{\ell,t}$ to the variability of fluctuations in $y_{i,t+h}$ is measured by the (i, ℓ) element of the FEVDs at horizon h

$$\mathbf{FEVD}_{y,\mathbf{G},i,\ell}(h) = \frac{\mathbf{MSE}_{y,\mathbf{G},i,\ell}(h)}{\mathbf{MSE}_{y,\mathbf{G}}(h)}.$$

Notes on FEVDs

- The δ-method, bootstrap, and Bayesian Monte Carlo procedures can be adapted to generate confidence bands for FEVDs $\Longrightarrow \mathbf{B}_j$s, ε_t, and Ω.
- Confidence bands are not joint tests of the significance of several FEVDs in the same way the joint significance of IRFs cannot be assessed with confidence bands.
- FEVDs can suffer from serial correlation across horizons $h = 0, 1, 2, \ldots, H$.