Vector Autoregressions

Woong Yong Park

University of Hong Kong

Guest lecture for ECON6012

Empirical methods in macro

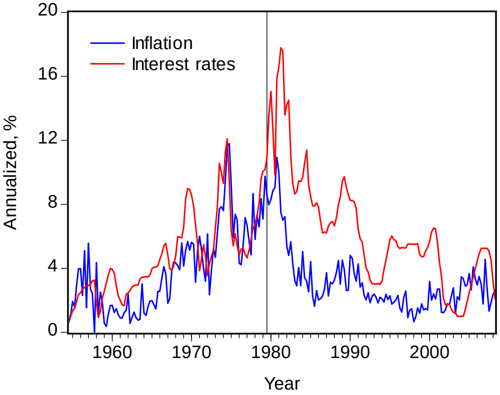
- Whether an economic model is explicitly used
- Whether interested in estimating a model (or a subset of parameters of a model) or estimating dynamic effects of macroeconomic shocks

1. Structural methods
   - Limited-information approach
     - Generalized method of moments
   - Full-information approach
     - Maximum-likelihood (ML) estimation (frequentist methods)
     - Bayesian methods
2. Non-structural (minimally-structural) methods
   - Vector autoregressions (VAR)

Vector autoregressions

- Vector autoregression (VAR) models are useful to summarize the information contained in the data.
  - VAR models are a natural starting point for empirical analysis.
  - The Wold theorem ensures that any vector of time series has a VAR representation under mild regularity conditions.
- VAR models are also useful to conduct certain types of economic experiments, among which policy experiments are most popular.
  - *Reduced-form* VAR models do not have economic interpretations.
  - To give an economic interpretation to a reduced-form VAR, we need to use economic theory. Compared with a macro models based on microfoundations, only a minimalist set of restrictions are employed.
  - This process is called *identification*.

Identification of monetary policy effect



"Monetary policy tries to control inflation, and as a result interest rates tend to be high when inflation is high. But we think that raising interest rates lowers inflation. The main contribution of this work is to provide a way to untangle the relationship between interest rates and inflation, so we can see what the effect of interest-rate policy changes are on the price level and inflation, and separate that from the reverse causality that makes central banks react to inflation by changing interest rates." - Christopher Sims, in an interview on Oct 10, 2011

VAR

- Based on Chapter 4 of Canova (2007) *Methods for Applied Macroeconomic Research*
- A good reference for time series techniques is Hamilton (1994) *Time Series Analysis*

The Wold Theorem

- The Wold theorem decomposes any vector stochastic process into two orthogonal components: one linearly predictable (based on the past information) and one linearly unpredictable (regular).
- We use a rather special case of the Wold theorem under two assumptions.
  - We consider a *linear* representation, that is we substitute conditional expectations with linear projections.
  - We consider covariance-stationary stochastic processes.

The Wold Theorem

- Then, any vector $m \times 1$ stochastic process $y_t^\dagger$ can be represented in the form
$$y_t^\dagger = ay_{-\infty} + \sum_{j=0}^{\infty} D_j e_{t-j},$$
where $y_{-\infty}$ is the initial value of $y_t^\dagger$ at the beginning of history and e_t is a white noise process ($E(e_t) = 0$, $E(e_t e_t') = \Sigma_e$, and $E\left(e_t e_{t-j}'\right) = 0$ if $j \neq 0$).
    - The term $ay_{-\infty}$ is the linearly deterministic component of $y_t^\dagger$.
    - The news or shock in period t, e_t, accumulates and produces the linearly regular component $\sum_{j=0}^{\infty} D_j e_{t-j}$.

MA representations

- Assuming that the data have zero mean and using the lag operator, we rewrite the decomposition as a moving-average (MA) representation

$$y_t = D(L)\, e_t,$$

where $y_t = y_t^{\dagger} - a y_{-\infty}$ and $D(L) = D_0 + D_1 L + D_2 L^2 + \cdots$.

  - L is a lag operator.

MA representations

- MA representations are not unique: There may exist more than one MA representations that imply the same variance-covariance structure (or for the same data y_t).
  - For example, consider $y_{1t} = e_t - 0.5e_{t-1}$ and $y_{2t} = \tilde{e}_t - 2\tilde{e}_{t-1}$ with $e_t \sim$ i.i.d. $\left(0, \sigma_1^2\right)$ and $\tilde{e}_t \sim$ i.i.d. $\left(0, \sigma_2^2\right)$. Then, the covariance generating function of y_{1t} and y_{2t} is $ACF_{y_1}(z) = (1 - 0.5z)\left(1 - 0.5z^{-1}\right)\sigma_1^2$ and $ACF_{y_2}(z) = (1 - 2z)\left(1 - 2z^{-1}\right)\sigma_2^2$. Therefore, if $\sigma_1^2 = 4\sigma_2^2$, the ACF of the two processes are the same.

Fundamental MA representation

- An MA representation is *fundamental* if all the roots of $D(z)$ are greater than 1 in modulus (outside the unit circle). Also it is called a Wold representation.
  - $\Longleftrightarrow \det\left(D_0 E\left(e_t e_t'\right) D_0'\right) > \det\left(D_j E\left(e_{t-j} e_{t-j}'\right) D_j'\right)$ for any $j \neq 0$.
- We are interested in a fundamental MA representation because a fundamental MA representation is *invertible*: Knowing y_t is the same as knowing e_t.
  - The space spanned by linear combinations of y_t and of e_t has the same information.
  - We can calculate (recover) the current value of e_t using $\{y_t, y_{t-1}, y_{t-2}, \cdots\}$.
- The white noise of a fundamental MA representation is called the fundamental innovation.

Fundamental MA representation

- Example. Consider

$$y_{1t} = e_t + \begin{pmatrix} 1 & 0 \\ 0 & 0.8 \end{pmatrix} e_{t-1} \quad \text{where} \quad Var\left(e_t\right) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \qquad (1)$$

$$y_{2t} = \epsilon_t + \begin{pmatrix} 1 & 0 \\ 0 & 1.25 \end{pmatrix} \epsilon_{t-1} \quad \text{where} \quad Var\left(\epsilon_t\right) = \begin{pmatrix} 2 & 0.8 \\ 0.8 & 0.64 \end{pmatrix} \qquad (2)$$

  - The ACF of the two processes is the same. But (2) is not fundamental.
- RBC and NK models usually have a fundamental MA representation for endogenous variables.
  - Therefore, if such a model is the true model of the economy, we can use a VAR to recover *structural* shocks.

Non-fundamental MA representation

- *The invertibility problem*: If a macro model implies a non-fundamental MA representation of the data vector, there is no hope to exactly recover the structural shocks of the model using a VAR. The model is said to be non-invertible.
- Example. Consider an RBC model which is standard except that the total factor productivity follows

$$\ln A_t = \ln A_{t-1} + 0.1\epsilon_{1t} + 0.2\epsilon_{1t-1} + 0.4\epsilon_{1t-2} + 0.2\epsilon_{1t-3} + 0.1\epsilon_{1t-4}.$$

  - Such a diffusion of technological innovations is appropriate when, for example, only the most advanced sectors employ the new technology and it takes some time for the innovation to spread to the economy.

Non-fundamental MA representation

- For a general condition for the invertibility of a macro model, see Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (AER, 2007).
  - When a model is not invertible, we need to use the model itself to recover/estimate structural shocks. Another alternative is suggested in Fernández-Villaverde et al (AER, 2007).
- We can reconstruct some of the structural shocks of a non-invertible macro model with good accuracy. See a discussion by Sims and Zha (Macroeconomic Dynamics, 2006).

VAR representation

- When an MA representation is invertible, we can “invert” the MA representations and express e_t as a linear combination of current and past y_t

  $$[D(L)]^{-1} y_t = e_t,$$

  which can be written using another lag polynomial $A(L) = A_1 + A_2 L + A_3 L^2 - \cdots$

  $$\begin{aligned} [A_0 - A(L) L] y_t &= e_t \\ y_t &= A(L) y_{t-1} + e_t, \end{aligned} \tag{3}$$

  where we normalize $A_0 = I$.
- This is what “Knowing y_t is the same as knowing e_t” means.

VAR representation

- Any vector of time series can be represented with a constant coefficients VAR(∞) under linearity, stationarity, and invertibility.
- If y_t has an invertible MA representation, it also has a stable VAR representation.
- In general $A(L)$ is of infinite length. With a finite stretch of data, only VAR(q) with q finite can be used. For a VAR(q) to approximate any y_t sufficiently well, we need D_j to converge to zero rapidly as j increases. Otherwise the approximation will be poor.

Identification

- One can reconstruct e_t from data. But e_t does not correspond to structural shocks of a macro model since it is contemporaneously correlated. That is, Σ_e is not diagonal.
  - e_t is called *reduced-form* shocks. We cannot attach a "name" to e_t.
  - Structural shocks of a macro model are assumed purely exogenous and orthogonal to each other.
- We need to *identify* structural shocks that are orthogonal to each other.
  - Identification needs economic theory.
  - Of course, if one derives the MA representation of a macro model and its VAR representation, then we have identification.

Identification

- A VAR with structural shocks not-identified is called a *reduced-form* VAR while a VAR with structural shocks identified is called a *structural* (*identified*) VAR.
- We will discuss identification in detail later.

Moments and parameter estimation

- The companion form representation transforms a VAR(q) model in a larger scale VAR(1) model, which is useful in computing moments or estimating parameters. Consider

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_q y_{t-q} + e_t$$

- Let

$$\underbrace{\mathbb{Y}_t}_{mq\times 1} = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-q+1} \end{bmatrix}, \quad \underbrace{\mathbb{E}_t}_{mq\times 1_t} = \begin{bmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \underbrace{\mathbb{A}}_{mq\times mq} = \begin{bmatrix} A_1 & A_2 & \cdots & A_q \\ I_m & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & & I_m & 0 \end{bmatrix}.$$

Then (3) is

$$\mathbb{Y}_t = \mathbb{A}\mathbb{Y}_{t-1} + \mathbb{E}_t, \quad \mathbb{E}_t \sim (0, \Sigma_{\mathbb{E}}).$$

Moments and parameter estimation

- Note that

$$E(\mathbb{Y}_t) = \left[(I - \mathbb{A}L)^{-1}\right] E(\mathbb{E}_t) = 0.$$

- Also,

$$\begin{aligned}
\Sigma_{\mathbb{Y}} &\equiv E\left[(\mathbb{Y}_t - E(\mathbb{Y}_t))(\mathbb{Y}_t - E(\mathbb{Y}_t))'\right] \\
&= \mathbb{A}E\left[(\mathbb{Y}_{t-1} - E(\mathbb{Y}_{t-1}))(\mathbb{Y}_{t-1} - E(\mathbb{Y}_{t-1}))'\right]\mathbb{A}' + \Sigma_{\mathbb{E}} \\
&= \mathbb{A}\Sigma_{\mathbb{Y}}\mathbb{A}' + \Sigma_{\mathbb{E}}.
\end{aligned}$$

Using $\text{vec}(TVR) = (R' \otimes T)\,\text{vec}(V)$,

$$\text{vec}(\Sigma_{\mathbb{Y}}) = \left[I_{(mq)^2} - (\mathbb{A} \otimes \mathbb{A})\right]^{-1} \text{vec}(\Sigma_{\mathbb{E}}).$$

- For the unconditional autocovariances,

$$ACF_{\mathbb{Y}}(\tau) = \mathbb{A}^{\tau}\Sigma_{\mathbb{Y}}.$$

Moments and parameter estimation

- Note that $y_t = \mathbb{A}_1 \mathbb{Y}_{t-1} + e_t$, where $\mathbb{A}_1$ is the first m rows of $\mathbb{A}$. Given the VAR structure and under the normality assumption on e_t,

$$y_t | y_{t-1}, \cdots, y_0, \cdots y_{-q+1} \sim \mathbb{N}\left(\mathbb{A}_1 \mathbb{Y}_{t-1}, \Sigma_e\right),$$

  where $y_0, \cdots y_{-q+1}$ are initial conditions.

- Therefore, the log likelihood of VAR(q) given the initial conditions is

$$\begin{aligned}
\mathcal{L}\left(\mathbb{A}_1, \Sigma_e | \{y_t\}\right) &= \frac{1}{2} T \left(m \ln(2\pi) - \ln \left|\Sigma_e^{-1}\right|\right) \\
&\quad - \frac{1}{2} \sum_{t=1}^{T} \left(y_t - \mathbb{A}_1 \mathbb{Y}_{t-1}\right)' \Sigma_e^{-1} \left(y_t - \mathbb{A}_1 \mathbb{Y}_{t-1}\right).
\end{aligned}$$

- MLE

$$\widehat{\mathbb{A}}'_{1,ML} = \left(\sum_{t=1}^{T} \mathbb{Y}_{t-1} \mathbb{Y}'_{t-1}\right)^{-1} \left(\sum_{t=1}^{T} \mathbb{Y}_{t-1} y'_t\right), \quad \widehat{\Sigma}_{e,ML} = \frac{1}{T} \sum_{t=1}^{T} \hat{e}_{t,ML} \hat{e}'_{t,ML}.$$

Moments and parameter estimation

- We did not impose any restrictions on the parameters to obtain the MLEs. When no restrictions are imposed and the initial conditions are given, ML and OLS estimators of $\mathbb{A}_1$ coincide.
- As long as all variables appear with the same lags in every equation, single-equation OLS estimation is sufficient. That is,

$$\widehat{\mathbb{A}}'_{1,j} = \left(\sum_{t=1}^{T} \mathbb{Y}_{t-1} \mathbb{Y}'_{t-1} \right)^{-1} \left(\sum_{t=1}^{T} \mathbb{Y}_{t-1} y'_{jt} \right)$$

  where $\widehat{\mathbb{A}}_{1,j}$ is the jth row of $\widehat{\mathbb{A}}_1$ and y_{jt} is the jth variable of y_t.
- For identification, we need to orthogonalize $\widehat{\Sigma}_e$. But the data do not contain the information for orthogonalization.

Reporting VAR results

- It is rare to report estimated VAR coefficients. Too many parameters. Also hard to interpret.
- To summarize information better, functions of the VAR coefficients are reported. Typically, impulse responses, variance and historical decompositions are reported.

Reporting VAR results

- Consider an MA representation

$$y_t = D(L)\, e_t = \sum_{j=0}^{\infty} D_j e_{t-j},$$

  and its VAR representation

$$y_t = A(L)\, y_{t-1} + e_t,$$

  with normalization $D_0 = A_0 = I$.
- We usually report those functions with respect to structural shocks $\tilde{e}_t$ rather than reduced-form shocks e_t.
- Let us suppose that we identify $\tilde{e}_t$ as $\tilde{e}_t = \tilde{\mathcal{P}}^{-1} e_t$ with some $\tilde{\mathcal{P}}$ such that $\Sigma_e = \tilde{\mathcal{P}}\tilde{\mathcal{P}}'$ and $E(\tilde{e}_t \tilde{e}_t') = I$.

Impulse responses

- Impulse responses trace out the MA coefficients of the system, that is, they describe how $y_{t+\tau}$ for $\tau \geq 0$ responds to a shock in e_t

$$\frac{\partial y_{t+\tau}}{\partial e_t'} = D_\tau,$$

and to a shock in $\tilde{e}_t$

$$\frac{\partial y_{t+\tau}}{\partial \tilde{e}_t'} = D_\tau \tilde{\mathcal{P}}.$$

Impulse responses

- Once one estimates a VAR(q), an approximation with finite lags, he can sequentially estimate D's:

$$D_\tau = \sum_{j=1}^{\max(\tau,q)} A_j D_{\tau-j},$$

  where $D_0 = I$ and $A_j = 0$ for any $j \geq q$.
- Example. Consider a VAR(2) with $y_t = A_0 + A_1 y_{t-1} + A_2 y_{t-2} + e_t$. Then, the impulse responses are $D_0 = I$, $D_1 = A_1 D_0$, $D_2 = A_1 D_1 + A_2 D_0$, ..., $D_\tau = A_1 D_{\tau-1} + A_2 D_{t-2}$. Note that the constant term does not matter.

Impulse responses

- Another way to compute impulse responses is to use forecast revisions of future y_t. That is, impulse responses measure how much forecast of future y_t is revised due to a shock.
- Let $\mathbb{Y}_t(\tau) \equiv E(\mathbb{Y}_{t+\tau}|y_t, y_{t-1}, ...) = \mathbb{A}^\tau \mathbb{Y}_t$, which is the forecast of $\mathbb{Y}_{t+\tau}$ obtained in information at time t. Then the τ-step-ahead forecast revision of $\mathbb{Y}_{t+\tau}$ due to a shock at time t is

$$\mathbb{Y}_t(\tau) - \mathbb{Y}_{t-1}(\tau+1) = \mathbb{A}^\tau(\mathbb{Y}_t - \mathbb{A}\mathbb{Y}_{t-1}) = \mathbb{A}^\tau \mathbb{E}_t,$$

  whose first m rows are impulse responses. One can use $\tilde{\mathcal{P}}$ to get impulse responses to $\tilde{e}_t$.
- For cumulative impulse responses, one computes $\sum_{j=0}^{\tau} D_j$. For long-run impulse responses, $\lim_{\tau\to\infty} \sum_{j=0}^{\tau} D_j$.

Variance decompositions

- The variance decomposition measures the contribution of a shock (e.g. $\tilde{e}_{jt}$, the jth shock in $\tilde{e}_t$) to the variability of the forecast error in $y_{t+\tau}$.
- Note that $\Sigma_e = \tilde{\mathcal{P}}\tilde{\mathcal{P}}' = \sum_{i=1}^{m}\left(\tilde{\mathcal{P}}_i\tilde{\mathcal{P}}_i'\right)$ where $\tilde{\mathcal{P}}_i$ is the ith column of $\tilde{\mathcal{P}}$. The ith summand $\tilde{\mathcal{P}}_i\tilde{\mathcal{P}}_i'$ is the contribution of the ith shock in $\tilde{e}_t$ to the variance of the contemporaneous shocks since

$$e_t = \tilde{\mathcal{P}}\tilde{e}_t = \tilde{\mathcal{P}}_1\tilde{e}_{1t} + \cdots + \tilde{\mathcal{P}}_m\tilde{e}_{mt}$$

  and $E\left(\tilde{e}_{kt}\tilde{e}_{\bar{k}t}\right) = 0$ for $k \neq \bar{k}$.
- Since the τ-step-ahead forecast error is $y_{t+\tau} - y_t(\tau) = \sum_{j=0}^{\tau-1} D_j e_{t+\tau-j}$, where $D_0 = I$, the mean squared error of the forecast is

$$MSE(\tau) = E\left[y_{t+\tau} - y_t(\tau)\right]^2 = \Sigma_e + D_1\Sigma_e D_1' + \cdots + D_{\tau-1}\Sigma_e D_{\tau-1}'.$$

Variance decompositions

- We can decompose the MSE as

$$MSE\left(\tau\right) = \sum_{i=1}^{m} \left(\tilde{\mathcal{P}}_i \tilde{\mathcal{P}}_i' + D_1 \tilde{\mathcal{P}}_i \tilde{\mathcal{P}}_i' D_1' + \cdots + D_{\tau-1} \tilde{\mathcal{P}}_i \tilde{\mathcal{P}}_i' D_{\tau-1}'\right).$$

- Let $MSE\left(\tau\right)^{(i)}$ denote the ith summand of $MSE\left(\tau\right)$. Then the fraction of the variance in $y_{i,t+\tau}$ due to $\tilde{e}_{\bar{i},t}$ conditional on information up to time $t-1$ is

$$VD_{i,\bar{i}}\left(\tau\right) \equiv \frac{MSE\left(\tau\right)_{i,i}^{(\bar{i})}}{MSE\left(\tau\right)_{i,i}}.$$

Historical decompositions

- The historical decomposition measures deviations from the baseline forecast of $y_{t+\tau}$ from t to $t+\tau$ due to a shock. The computations details can be found in any time series textbook.
- Impulse responses, variance and historical decompositions are computed using the same ingredients, which means that they simply package the same information in different ways.

Distributions of impulse responses

- To asses the statistical significance of impulse responses, variance and historical decompositions, we need to estimate their standard errors.
- Since they are a function of the VAR coefficient matrices and the covariance matrix of the shocks, their standard errors can be computed based on these parameters and their standard errors.
- There are three methods, one based on asymptotics and the other two which use the small-sample properties.
- We will discuss how to compute the distribution of impulse responses. Straightforward to extend to compute the distribution of variance and historical decompositions.
- Easy to compute with packaged codes (Eviews, RATS, STATA, ...)

Distributions of impulse responses

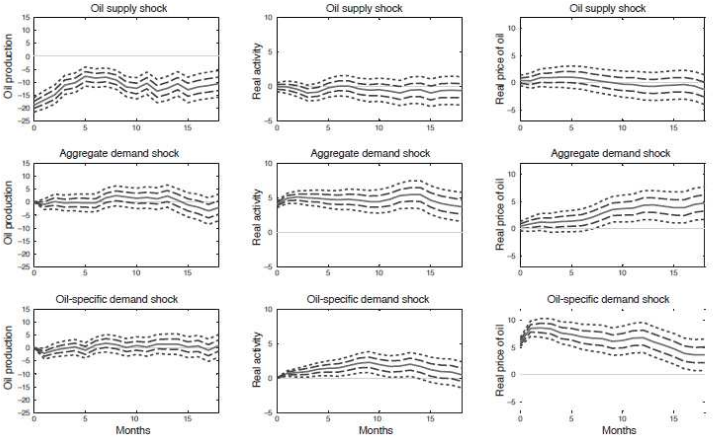


FIGURE 3. RESPONSES TO ONE-STANDARD-DEVIATION STRUCTURAL SHOCKS

(Point estimates with one- and two-standard error bands)

Kilian (AER 2009)

Distributions of impulse responses

The delta method

- The first method, called the delta method, uses the first-order approximation. Note that any differentiable function $f(\alpha)$ with a column vector α can be written as

$$f(\alpha) = f(\bar{\alpha}) + \frac{\partial f(\bar{\alpha})}{\partial \bar{\alpha}'}(\alpha - \bar{\alpha}) + O(|\alpha - \bar{\alpha}|).$$

- Now suppose that $\alpha \xrightarrow{D} \mathbb{N}(0, \Sigma_\alpha)$. Then,

$$f(\alpha) \xrightarrow{D} \mathbb{N}\left[f(0), \left(\frac{\partial f}{\partial \alpha'}\right)\Sigma_\alpha\left(\frac{\partial f}{\partial \alpha}\right)\right]$$

  if $\partial f / \partial \alpha \neq 0$.

- Let $\mathbb{S} = (I, 0, \ldots, 0)$ be an $m \times mq$ selection matrix so that $y_t = \mathbb{S}\mathbb{Y}_t$ and $\mathbb{E}_t = \mathbb{S}' e_t$. Then the τ-step-ahead forecast error is

$$\mathbb{S}\left[\mathbb{Y}_t(\tau) - \mathbb{Y}_{t-1}(\tau+1)\right] = \mathbb{S}\left(\mathbb{A}^\tau \mathbb{S}' e_t\right) \equiv \psi_\tau e_t.$$

Distributions of impulse responses

The delta method

- By matrix algebra, we can show that if $\mathbb{A}_1 \sim \mathbb{N}\left(0, \Sigma_{\mathbb{A}_1}\right)$, the impulse response at horizon τ

$$\psi_\tau \sim \mathbb{N}\left(0, \left[\frac{\partial \text{vec}\left(\psi_\tau\right)}{\partial \text{vec}\left(\mathbb{A}_1\right)}\right] \Sigma_{\mathbb{A}_1} \left[\frac{\partial \text{vec}\left(\psi_\tau\right)}{\partial \text{vec}\left(\mathbb{A}_1\right)}\right]'\right).$$

- For structural shocks $\tilde{e}_t$, the distribution of $\psi_\tau \tilde{\mathcal{P}}$ is straightforward.
- Caveats of the delta method
  - It relies on asymptotics. Poor small-sample properties with a typical sample size of 100-120 (quarterly data).
  - The asymptotics works poorly when near-unit-roots or near singularities are present.
  - Estimated VAR coefficients usually have large asymptotic standard errors, which lead to insignificant responses at all horizons.

Distributions of impulse responses

Bootstrap methods

- Bootstrapping is a method for estimating the sampling distribution of an estimator by resampling with replacement from the original sample.
- It produces an estimator with better small-sample properties.
- There are some caveats. In particular, the sampling distribution may suffer a bias due to the bias of the OLS estimator $\widehat{A}(L)$. One can try a bootstrap-after-the-bootstrap procedure by Kilian (1998) to correct the problem.

Distributions of impulse responses

Monte Carlo methods

- So far we used the frequentist approach.
- Using the Bayesian approach, one can simply sample from the posterior distribution of the VAR coefficient matrices and the covariance matrix, and construct the posterior distribution of impulse responses.
- No asymptotics. Generally applicable. The posterior distribution of impulse responses is exact under the distributional assumption on the shocks.

Reduced-form vs. structural VARs

- A VAR with the non-diagonal covariance matrix for the shocks is called a reduced-form VAR

$$y_t = A(L)\, y_{t-1} + e_t,$$

  where $e_t \sim \mathbb{N}(0, \Sigma_e)$. As we discussed, it is hard to interpret e_t as structural shocks to the economy.
- One can identify structural shocks $\tilde{e}_t$ as $\tilde{\mathcal{P}}\tilde{e}_t = e_t$ where $\tilde{\mathcal{P}}\tilde{\mathcal{P}}' = \Sigma_e$ and $E(\tilde{e}_t\tilde{e}_t') = I$.
- A VAR with identified structural shocks is called a structural VAR (SVAR)

$$y_t = A(L)\, y_{t-1} + \tilde{\mathcal{P}}\tilde{e}_t.$$

- The problem is that $\tilde{\mathcal{P}}$ is *not identified* in the data. For any orthonormal matrix $\mathcal{Q}$ such that $\mathcal{Q}\mathcal{Q}' = I$, $A_0 = \tilde{\mathcal{P}}\mathcal{Q}$ satisfies $A_0 A_0' = \Sigma_e$. $\tilde{\mathcal{P}}$ is not unique.

Identification

- *Economic theory* has to come into play to identify structural shocks and economically interpret the dynamics induced by them.
  - An economic model may or may not be explicitly considered.
- A DSGE model when approximated linearly or log-linearly around the steady state typically delivers finite-lag VAR solutions for the vector of endogenous variables of the model. One can find the mapping between DSGE parameters and VAR parameters and, in particular, $\tilde{\mathcal{P}}$.
- When doubts about the quality of a DSGE model exist, it is possible to conduct useful inference as long as a subset of the model restrictions are credible.
- Typical restrictions employed in the VAR literature include constraints on the short-run or the long-run impact of certain shocks on VAR variables or informational delays.

Short-run restrictions

- Consider a reduced-form VAR for an $m \times 1$ vector y_t

$$y_t = A(L)\, y_{t-1} + e_t, \quad e_t \sim \text{i.i.d. } (0, \Sigma_e)\,. \tag{4}$$

- We assume that there exists a macro model whose solution is of the form

$$y_t = \mathcal{A}(L)\, y_{t-1} + \mathcal{A}_0 \epsilon_t, \quad \epsilon_t \sim \text{i.i.d. } \left(0, \Sigma_\epsilon = \text{diag}\left\{\sigma_{\epsilon_i}^2\right\}\right). \tag{5}$$

  - The diagonal entries of $\mathcal{A}_0$ are all 1. We can drop Σ_ϵ and allow different diagonal entries.
  - There may exist the explicit mapping between the parameters of the macro model and $\mathcal{A}(L)$ and $\mathcal{A}_0$. But we don't want to impose all of the restrictions.

- We estimate (4), and want to identify (5) from (4).
  - In particular, identify $\mathcal{A}_0$ and Σ_ϵ from Σ_e. $\mathcal{A}_0$ is $\tilde{\mathcal{P}}$ in the previous slides.

Short-run restrictions

- Matching contemporaneous coefficients in (4) and (5) implies $e_t = \mathcal{A}_0 \epsilon_t$, or
$$\mathcal{A}_0 \Sigma_\epsilon \mathcal{A}_0' = \Sigma_e. \tag{6}$$
- As in simultaneous equations models, there are necessary and sufficient conditions that need to be met for identification.
- *The order condition* is about the number of free parameters to be estimated.
    - On the LHS of (6), there are m^2 parameters.
    - On the RHS of (6), there are $m(m+1)/2$ parameters given the symmetry of Σ_e.
    - Therefore, (6) will have a solution if at least $m(m-1)/2$ restrictions are imposed.
- *The rank condition* requires that $\operatorname{rank}(\mathcal{A}_0 \Sigma_\epsilon \mathcal{A}_0') = \operatorname{rank}(\Sigma_e)$.

Short-run restrictions

- A fundamental reason for which we need identification is that we can only learn (estimate) Σ_e from the data and as a result there are infinitely many solutions for $\mathcal{A}_0$ to this equation.

Recursive identification

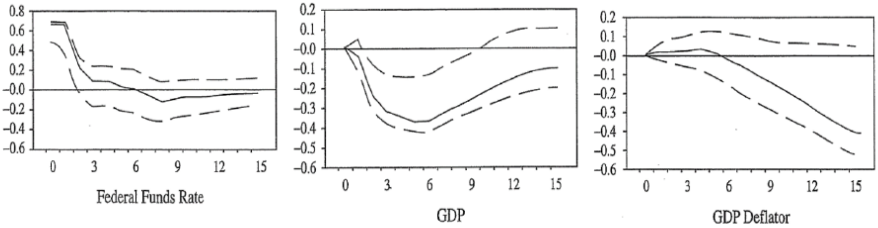
- Recursive identification. Sims (Econometrica, 1980)
  - Choleski decomposition: For any positive definite matrix B, there exists a lower triangular matrix A such that $B = AA'$. Unique with strictly positive diagonal entries.
  - $\mathcal{A}_0$ can be identified by using the *off*-diagonal entries of the Choleski factor of Σ_e.
- Example. $y_t =$(hours, productivity, interest rates)$'$, $e_t = \mathcal{A}_0 \epsilon_t$

$$\begin{pmatrix} e_t^{\text{hours}} \\ e_t^{\text{prod}} \\ e_t^{\text{int}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \times & 1 & 0 \\ \times & \times & 1 \end{pmatrix} \begin{pmatrix} \epsilon_t^{\text{hours}} \\ \epsilon_t^{\text{prod}} \\ \epsilon_t^{\text{int}} \end{pmatrix}$$

  - This could be the case if interest rate shocks take time to produce effects on hours and productivity and if hours are predetermined with respect to productivity.

Recursive identification

- Christiano, Eichenbaum and Evans (1999) about the effects of monetary policy shocks.
- In one benchmark specification, they use recursive identification in the order of Y_t, P_t, $PCOM_t$, FF_t, TR_t, NBR_t, and M_t. Note that the monetary authority is assumed to see Y_t, P_t and $PCOM_t$ when choosing FF_t.



Estimated dynamic response to a monetary policy shock (a one-standard deviation increase in the Fed funds rate)

Recursive identification

- Kilian (AER, 2009) $y_t = (\Delta prod_t, rea_t, rpo_t)^{'}$, monthly data
- Δprod is the percent change in global oil production, rea$_t$ denotes the index of real economic activity, and rpo_t is the real price of oil.

$$\begin{pmatrix} e_t^{\Delta prod} \\ e_t^{rea} \\ e_t^{rpo} \end{pmatrix} = \begin{pmatrix} \times & 0 & 0 \\ \times & \times & 0 \\ \times & \times & \times \end{pmatrix} \begin{pmatrix} \epsilon_t^{\text{oil supply shock}} \\ \epsilon_t^{\text{agg demand shock}} \\ \epsilon_t^{\text{oil-specific demand shock}} \end{pmatrix}$$

- Oil supply shocks are defined as unpredictable innovations to global oil production.
- Innovations to rea_t that cannot be explained based on oil supply shocks are referred to as shocks to the aggregate demand for industrial commodities. Production cannot respond immediately.
- Changes in rpo_t driven by shocks that are specific to the oil market will not affect rea_t immediately, but with a delay of at least a month.

Non-recursive identification

- Example. Consider a VAR with output growth, inflation, nominal interest rates, and money growth, and let $y_t = (\Delta GDP_t, \pi_t, i_t, \Delta M_t)'$.

$$\begin{pmatrix} e_t^{\Delta GDP} \\ e_t^{\pi} \\ e_t^{i} \\ e_t^{\Delta M} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \times & 1 & 0 & \times \\ 0 & 0 & 1 & \times \\ \times & \times & \times & 1 \end{pmatrix} \begin{pmatrix} \epsilon_t^{\Delta GDP} \\ \epsilon_t^{\pi} \\ \epsilon_t^{i} \\ \epsilon_t^{\Delta M} \end{pmatrix}$$

- There are six (zero restrictions) and ϵ_t shocks are identifiable from the VAR residuals.
- See Sims and Zha (IER, 1999) for a general discussion.

Long-run restrictions

- When y_t is integrated (has a unit root), there are additional restrictions which we can use for identification.
- Consider the MA representations of a VAR model and of a macro model

$$
\begin{aligned}
\Delta y_t &= D(L)\, e_t = D(1)\, e_t + \frac{D(L) - D(1)}{1 - L} \Delta e_t \\
\Delta y_t &= \mathcal{D}(L)\, \mathcal{A}_0 \epsilon_t = \mathcal{D}(1)\, \mathcal{A}_0 \epsilon_t + \frac{\mathcal{D}(L) - \mathcal{D}(1)}{1 - L} \mathcal{A}_0 \Delta \epsilon_t
\end{aligned}
$$

Long-run restrictions

- We decompose Δy_t into *permanent* and *transitory* components. Matching coefficients, we have additional restrictions

$$\mathcal{D}(1)\,\mathcal{A}_0\epsilon_t \quad = \quad D(1)\,e_t \tag{7}$$

  - When y_t is stationary, $\mathcal{D}(1) = D(1) = 0$ and the restrictions in (7) cannot be used.
  - For the transitory component, we use short-run restrictions.
- The long-run impacts of ϵ_t are $\mathcal{D}(1)\,\mathcal{A}_0$. Using $e_t = \mathcal{A}_0^{-1}\epsilon_t$,

$$\mathcal{D}(1)\,\mathcal{A}_0 = D(1)\,\mathcal{A}_0^{-1}.$$

- Then any restrictions on $\mathcal{D}(1)\,\mathcal{A}_0$ (as a whole) generate additional restrictions on $\mathcal{A}_0$ on the LHS.

Long-run restrictions

- Blanchard and Quah (AER, 1989) use a long-run restriction that demand shocks do not affect output in the long run to identify a bivariate VAR for output (GDP_t) and unemployment (U_t).
- They build a simple partial equilibrium model with overlapping labor contracts and derive

$$
\begin{aligned}
GDP_t &= GDP_{t-1} + (\epsilon_{2t} - \epsilon_{2t-1}) + a(\epsilon_{1t} - \epsilon_{1t-1}) + \epsilon_{1t} \\
U_t &= -\epsilon_{2t} - a\epsilon_{1t}
\end{aligned}
$$

  where ϵ_{1t} and ϵ_{2t} represent shocks to productivity (supply shocks) and shocks to money supply (demand shocks).
- Note that ϵ_{1t} affects GDP_t in the long-run, but ϵ_{2t} has only short-run effects on GDP_t. U_t is stationary.
- For a bivariate VAR, we need $1 = 2 \times (2-1)/2$ additional restriction for identification.
  - Blanchard and Quah use the implication of their model that one of the structural shocks does not have a long-run impact on output.

Long-run restrictions

- Chang, Gomes and Schorfheide (AER, 2002) suggest that skill accumulation through past work experience (learning-by-doing) can provide an important propagation mechanism in a DSGE model.
- They show that an RBC model with learning-by-doing mechanism can match impulse responses generated by a VAR while a standard RBC model cannot.
  - They use a VAR as a reference model to estimate impulse responses of structural shocks.

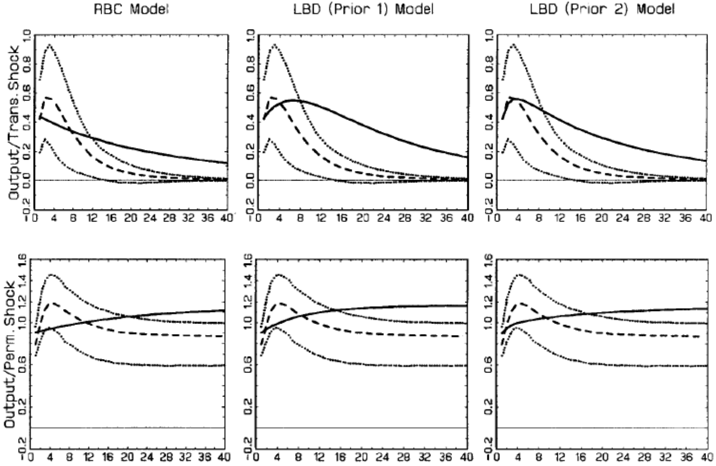
Long-run restrictions

- The question is how we can be sure that structural shocks of a VAR model are the same structural shocks of an RBC model.
  - For this, they use long-run restrictions similar to those used by Blanchard and Quah (1989).
- Their model has two shocks: a technology shock has a permanent effect on output while a demand shock has only a transitory effect on output:

$$
\begin{aligned}
\ln A_t &= \gamma + \ln A_{t-1} + \varepsilon_t^a, \\
\ln B_t &= (1-\rho)\ln B + \rho \ln B_{t-1} + \varepsilon_t^b
\end{aligned}
$$

Long-run restrictions

Chang, Gomes and Schorfheide (2002)



Long-run restrictions

- The long-run restrictions do not tell us anything about their behavioral content.
- Nevertheless, the literature has associated, somewhat arbitrary, permanent shocks with supply disturbances and transitory shocks with demand disturbances.

Sign restrictions

- Identification using short-run restrictions places zero restrictions on $\mathcal{A}_0$. There are criticisms of this identification scheme. In particular, standard DSGE models rarely provide such zero restrictions. Canova and Pina (2005).
- Similar criticisms on identification using long-run restrictions. Long-run restrictions may incompletely disentangle permanent and transitory shocks. Cooley and Dwyer (JoE, 1998).
- Alternative identification strategies that do not use zero restrictions emerged to produce a more solid bridge between economic theory and VARs. Among them, identification using sign restrictions.
- It achieves identification by restricting the sign (and/or shape) of impulse responses to structural shocks. Faust (1998), Canova and De Nicolo (JME, 2002) and Uhlig (JME, 2005).
- Economic theory / DSGE models contain a large number of sign restrictions usable for identification purposes.

Sign restrictions

- Example. Flexible price models usually have the feature that technology shocks increase output, consumption and investment either instantaneously or with a short lag, while prices and nominal interest rates decline.
  - Technology shocks can be identified by assuming that in impulse responses to positive shocks real variables increase and prices decrease, either contemporaneously or with a short lag.
- Example. Monetary models with nominal price rigidities usually have the feature that policy-driven increases in the nominal interest rate reduce real money balances instantaneously and induce a fall in inflation.
  - Contemporaneous (and lagged) comovements of real money balances, inflation and nominal interest rates in response to monetary policy shocks can be used to identify monetary disturbances.
- Though intuitively straightforward, implementation can be tricky. See the references.

Issues

- We considered covariance stationary stochastic processes. Even for nonstationary processes, a time-varying MA representation always exists and one can derive a VAR with time-varying coefficients.
- Unit roots?
- There are more problems that one should be aware of.
  - Time aggregation problems
  - The dimensionality of the VAR
  - Poor approximation of time series by a finite-order VAR
  - Nonfundamental MA representations

References

*** Canova. 2007. *Methods for Applied Macroeconomic Research.***

- Blanchard and Quah. 1989. The dynamic effects of aggregate demand and supply disturbances. *American Economic Review* 79: 655-673.

- Canova and De Nicolo. 2002. Monetary Disturbances Matter for Business Fluctuations in the G-7, *Journal of Monetary Economics* 49:1131-1159.

- Chang, Gomes and Schorfheide. 2002. Learning-by-Doing as a Propagation Mechanism. *American Economic Review* 92: 1498-1520..

- Christiano, Eichenbaum and Evans. 1999. Monetary policy shocks: what have we learned and to what end. In *Handbook of Macroeconomics*.

- Faust. 1998. The Robustness of Identified VAR Conclusions about Money, *Carnegie-Rochester Conference Series on Public Policy* 49:207-244.

- Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson. ABCs (and Ds) of understanding VARs. *American Economic Review* 97:1021-1026.

- Kilian. 2009. Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market. American Economic Review 99:1053-1069.

- Sims. 1980. Macroeconomics and Reality, *Econometrica* 48:1-48.

- Sims and Zha. 1998. Bayesian methods for dynamic multivariate models. *International Economic Review* 39:949-968.

- Sims and Zha. 2006. Does monetary policy generate recessions? *Macroeconomic Dynamics* 10:231-272.

- Uhlig. 2005. What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure. *Journal of Monetary Economics* 52:381-419.