SVARs & identification + Intro to continuous time stochastic processes

Tom Holden

http://www.tholden.org/

PhD Macroeconomics, Semester 2

Outline of today's talk

- SVARs.
- Identification methods.
- Continuous time stochastic processes.

Reading on SVARs and identification

- Canova: “Methods for applied macroeconomic research”.
  - Section 4.5 & 4.6 covers Identification and SVARs.
  - Section 10.3 covers this in a Bayesian context.
- Wikipedia, as needed for basic results in linear algebra.
  - Reading all of the pages in this category would be a good start: https://en.wikipedia.org/wiki/Category:Matrix_decompositions
- Christiano, Eichenbaum and Evans (2005):
  - A classic paper which you ought to be familiar with.
  - www.tau.ac.il/~yashiv/cee.pdf

Readings on continuous time processes etc.

- Cochrane (2012):
  - Nice review of continuous time stochastic processes, with a macro slant.
  - http://faculty.chicagobooth.edu/john.cochrane/research/papers/continuous_time_linear_models.pdf

- Any finance textbook for details (shouldn't really be needed).
  - E.g. Chapter 3 of Merton's "Continuous Time Finance".

- Wikipedia as needed…

Structural VARs: Motivation (1/2)

- We would like to know what the effects of (say) an unexpected increase in monetary policy is.
- But a change in monetary policy will produce changes in other variables within the same time period.
- Conversely, exogenous shocks to other variables will produce automatic reactions from monetary policy.
  - E.g. a Taylor Rule.
- Thus, if we see that interest rates were (say) tighter than was expected yesterday, we do not know if this was due to a change in policy or if it was an endogenous reaction to other changes in the economy.
  - A standard VAR tells us nothing about the effects of changes in policy!

Structural VARs: Motivation (2/2)

- Furthermore, even after contemporaneous responses of one variable to another have been taken into account, there may still be correlations in the shocks.
  - For example, an exogenous increase in rainfall may both decrease labour supply holding fixed the wage, and increase labour demand holding fixed the wage.
  - Thus in a VAR in which rainfall is omitted, it would show up as both a labour supply and a labour demand shock.
  - No way of knowing how much of this variance component due to rainfall should be assigned to supply, and how much should be assigned to demand.
  - However, this really reflects a failure of the model (omitting an observable variable).

  - Alternatively, some variables may respond directly to structural shocks to other variables.
  - In macroeconomic terms this is rather implausible, as shocks are generally not observed directly, and if they are observed, they're generally only observed by the agent that experiences the shock.
  - Nonetheless, in a few rare cases this may be justified.

Structural VARs: Definition

- This suggests the following structural representation:

$$x_t = c + a_0 x_t + a_1 x_{t-1} + \cdots + a_p x_{t-p} + u_t + b_0 u_t$$

  - where both a_0 and b_0 have a zero diagonal and where $u_t \sim \mathrm{WNIID}(0, \Sigma_u)$, with Σ_u diagonal.
- Then:

$$(I - a_0) x_t = c + a_1 x_{t-1} + \cdots + a_p x_{t-p} + (I + b_0) u_t$$

- Then if we define $A \coloneqq I - a_0$ and $B \coloneqq I + b_0$:

$$x_t = A^{-1} c + A^{-1} a_1 x_{t-1} + \cdots + A^{-1} a_p x_{t-p} + A^{-1} B u_t.$$

- Compare this to our previous reduced form:

$$x_t = \mu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{WNIID}(0, \Sigma_\varepsilon)$$

- Matching terms gives:

$$A\mu = c, \qquad A\phi_1 = a_1, \qquad \ldots, \qquad A\phi_p = a_p, \qquad A\varepsilon_t = B u_t,$$
$$A \Sigma_\varepsilon A' = B \Sigma_u B'$$

Structural VARs: Basic identification (1/2)

- Our hope is to be able to use some prior restrictions on A and B (or equivalently a_0 and b_0), in order to solve for Σ_u in the equation $A\Sigma_\varepsilon A' = B\Sigma_u B'$.
- We know A and B have a unit diagonal, and that Σ_u is zero everywhere except the diagonal.
- If we knew $A\Sigma_\varepsilon A'$ could we at least work out B and Σ_u?
  - No, not uniquely, without additional information.
  - By the Cholesky decomposition, there exists a lower triangular matrix L such that $A\Sigma_\varepsilon A' = LL'$. So one candidate solution is $B \coloneqq L(\operatorname{diag}\operatorname{diag} L)^{-1}$, and $\Sigma_u \coloneqq (\operatorname{diag}\operatorname{diag} L)^2$.
  - But let U be any real orthogonal matrix. Then $A\Sigma_\varepsilon A' = (LU)(LU)'$ too. Thus $B \coloneqq LU(\operatorname{diag}\operatorname{diag} LU)^{-1}$, and $\Sigma_u \coloneqq (\operatorname{diag}\operatorname{diag} LU)^2$ is another solution.
  - The space of all $n \times n$ orthogonal matrices is $\frac{n(n-1)}{2}$ dimensional, so this is the number of restriction we need on B if we already know $A\Sigma_\varepsilon A'$.
  - This may be seen directly from noting that in the equation $RR' = S$, with S symmetric, the equations above the diagonal are identical to those below.

Structural VARs: Basic identification (2/2)

- In practice B is almost always assumed to be equal to the identity matrix, for the reasons I gave previously.
  - If it's not, it reflects either strange informational assumptions, or omitted variables.
- So with B known, can we pin down A without additional assumptions?
  - No. Much as before, the equation $A\Sigma_\varepsilon A' = B\Sigma_u B'$ has $\frac{n(n-1)}{2}$ free parameters with B known.
  - So this is the number of assumptions we need to make on A.
- A common assumption is that A is lower triangular, which gives the required $\frac{n(n-1)}{2}$ restrictions.
  - This means a_0 is strictly lower triangular, implying a "causal ordering" on the variables.
  - The variable ordered first is assumed to have no contemporaneous response to later variables.
  - The one ordered second just responds to the first contemporaneously, but no others. Etc. etc. till…
  - The one ordered last responds contemporaneously to all variables.

Reporting SVAR results: Impulse responses

- Normally, we are primarily interested in the response of the economy to some shock.
  - So, suppose there's a shock of one standard deviation to variable i in period 0, and then from then on, no further shocks hit the economy.
  - I.e., suppose that $u_{0,i} = 1$ for some $i \in \{1, \dots, n\}$, but that $u_{t,j} = 0$ for all $t \in \mathbb{N}$ and all $j \in \{1, \dots, n\}$, unless $t = 0$ and $j = i$.
  - Given these assumptions on the shocks, we can simulate the SVAR model and plot the responses of the variables of interest.
  - The result of this is an "impulse response". Often plotted relative to the variable's mean.
- In non-linear models, there are various possible definitions of an impulse response.
  - One is $\mathbb{E}\left[x_t \middle| u_{0,i} = 1\right]$.
  - Another is $\mathbb{E}\left[x_t \middle| u_{0,i} = \tilde{u}_{0,i} + 1\right]$, where $\tilde{u}_{0,i}$ has the same distribution as that of $u_{t,i}$.
    - Some authors also condition on the initial state in these expectations. (Dynare does not.)

Other identification methods: AB restrictions

- Causal orderings are deeply implausible. Most variables have some contemporaneous effect on most other variables.
- Indeed, many variables have strong anticipatory effects on other variables.
  - If a shock to another variable is expected in future (and the econometricians data-set is insufficient to pick this up) then shocks (observed) tomorrow might have an effect on variables today.

- Other restrictions on the A and B matrices based on theory are often as bad, for basically the same reason.

- The Blanchard and Perotti (2002) approach uses micro data to estimate some parameters of the A matrix in a fiscal policy context.
  - Their argument is based upon government taking more than quarter to respond, so is little better than the causal ordering approach.

Other identification methods: Sign restrictions

- The sign restriction approach (Uhlig 2005) effectively places a flat prior over the space of all orthogonal rotation matrices, then truncates this prior to zero in areas where the model generates "the wrong results" in some sense.
  - "Flat" may not be as intuitive as it seems, see e.g. Baumeister & Hamilton (2017).
- "Wrong" is usually defined in terms of the sign of the impulse response to a certain shock at a certain point in time.
  - May end up assuming what it wants to prove. E.g. causal ordering identification of monetary policy shocks often produces "price puzzles", with increasing interest rates increasing inflation.
  - Assuming away price puzzles begs the question of whether these are real features or not.
- If interpreted classically, sign restrictions only produce set identification, not point identification. (see Moon, Schorfheide and Granziera 2013).
  - Following identification via sign restrictions, there is no such thing as "the" estimated impulse response.
  - Rather, the estimator provides a band of impulse responses, even with infinite data.
  - Finite sample parameter uncertainty produces even larger bands.

Other identification methods: Narrative evidence

- Pioneered by Romer and Romer (1989), who use the text of FOMC meetings to identify times when policy makers intended to use contractionary policy to bring down inflation.
  - Later work has tightened the definition of a monetary shock.
- In fiscal contexts, Ramey and Shapiro (1998) performed a similar analysis using military build-ups.

- One difficulty with this approach is that hand selected shocks will always "smell funny".
  - Some recent researchers have ameliorated this via using automated textual analysis.
- Another problem is that it's not always clear that the narrative analysis procedure really succeeds in purging all endogeneity.

Other identification methods: Long-run restrictions

- While there's a lot of debate about how the economy evolves in the short-run, there's a lot more consensus about the long-run effects of various shocks.
  - E.g. only a technology shock increases GDP per capita in the long-run. Monetary shocks are neutral for all variables in the long-run. Etc.

- Blanchard and Quah (1989) exploit this for identification.
  - It is a bit like a sign restriction at $t = \infty$, but since they are imposing exact coefficients for the long-run response they get point, not set, identification.
  - Can be imposed without simulation, either:
    - by deriving the limit of the IRF by diagonalization (see exercises), or,
    - by using the following Beveridge-Nelson type decomposition: $\Delta x_t = \Theta(L)\varepsilon_t = \Theta(1)\varepsilon_t + \frac{\Theta(L)-\Theta(1)}{I-L}\Delta\varepsilon_t$, so the permanent impact of a structural shock is $\Theta(1)A^{-1}Bu_t$.

SVARs and identification in practice

- There is a huge literature looking at the responses of monetary and fiscal shocks.
- Results vary wildly depending on which identification method is used, though there is more consensus about monetary shocks than fiscal ones.
  - For example, in a cross country study Iletzki, Mendoza and Vegh (2013) find basically zero fiscal multipliers in developed, open economies, and Ramey's narrative based work finds at most moderate multipliers, around one.
  - On the other hand Perotti continues to find large multipliers.
- The correct response is broad distrust of most VAR identification methods.
- In any case, it is unclear why we should care about fiscal multipliers.
  - The fact that government expenditure increases GDP more than one for one tells us nothing about whether this is good for welfare.
  - In fact, in most modern macro models that generate large multipliers, expansionary fiscal policy is unambiguously bad for welfare.

Continuous time stochastic processes

- The stochastic processes we looked at in the first lecture were random variables taking their value from the vector space of sequences (i.e. functions $\mathbb{Z} \to \mathbb{R}$).

- In some circumstances, it is easier to work in continuous time, i.e. with random variables taking their value from the vector space of functions $\mathbb{R} \to \mathbb{R}$.
  - This is the standard in finance.
  - It's also increasingly more common in macro, and we'll look at some continuous time DSGE models later in the course.

The Wiener process

- The Wiener process (aka “standard Brownian motion”) is the building blocks of most continuous time stochastic processes.
- You might find it helpful to think of the Wiener process as the continuous time analogue of a random walk.
- The process, W_t is characterised by the following properties:
  1. $W_0 = 0$.
  2. W_t is almost surely everywhere continuous.
  3. If $s_1, s_2, \dots, s_{n+1}, t_1, t_2, \dots, t_n \in \mathbb{R}^+$ satisfy $0 < s_i < t_i \leq s_{i+1}$ for all $i \in \{1, \dots, n\}$, then $W_{t_i} - W_{s_i}$ is independent of $W_{t_j} - W_{s_j}$ for all $i \neq j$.
  4. If $0 \leq s < t$, then $W_t - W_s \sim \mathrm{N}(0, t - s)$.

- The process $\mu t + \sigma W_t$ is called a Wiener process with drift μ and infinitesimal variance σ^2.

The Itō integral

- We would often like to work with processes with time varying drift and time varying infinitesimal variance.
  - Scaling the Wiener process by a time varying amount will not work, as this will change both the level of the process and its future infinitesimal variance.
  - In some loose sense then, we need to “differentiate” the process, scale it, and then integrate back.
  - However, the Wiener process is not differentiable.
- Itō defined a new integral (with different integration laws) in order to tackle this.
  - It allows us to integrate a function times a kind of “derivative” of the Wiener process.
- In particular, if:
  - for all $n \in \mathbb{N}$, π_n is an increasing sequence of length $n+1$, with $\pi_{n,0} = S$ and $\pi_{n,n} = T$, and $\lim_{n\to\infty} \max_{i\in\{1,\dots,n\}} \left|\pi_{n,i} - \pi_{n,i-1}\right| = 0$,
  - W_t is a Wiener process, and X_t is another continuous time stochastic process that is left-continuous and locally bounded,
  - then we define:

$$\int_S^T X_t \, dW_t \coloneqq \operatorname*{plim}_{n\to\infty} \sum_{i\in\{1,\dots,n\}} X_{\pi_{n,i-1}} \left(W_{\pi_{n,i}} - W_{\pi_{n,i-1}}\right).$$

Useful property: The Itō isometry

- Given $S_X < T_X$, $S_Y < T_Y$, with $\max\{S_X, S_Y\} < \min\{T_X, T_Y\}$ and given continuous stochastic processes X_t and Y_t:

$$\mathbb{E}\left[\left(\int_{S_X}^{T_X} X_t\, dW_t\right)\left(\int_{S_Y}^{T_Y} Y_t\, dW_t\right)\right] = \mathbb{E}\left[\int_{\max\{S_X,S_Y\}}^{\min\{T_X,T_Y\}} X_t Y_t\, dt\right]$$

- Informal "proof":
  - Note:

$$\mathbb{E}\left[\left(\int_{S_X}^{T_X} X_t\, dW_t\right)\left(\int_{S_Y}^{T_Y} Y_t\, dW_t\right)\right] = \mathbb{E}\left[\left(\int_{S_X}^{T_X}\int_{S_Y}^{T_Y} X_s Y_t\, dW_s\, dW_t\right)\right]$$

  - $\mathbb{E}\, dW_s\, dW_t$ "equals" $0\, dt$ if $s \neq t$ and "equals" $1\, dt$ if $s = t$.

Drift diffusion processes

- Processes used in finance (and continuous time macro) often take the form:
$$X_t = X_0 + \int_0^t \mu(X_u, u)\, du + \int_0^t \sigma(X_u, u)\, dW_u\,.$$
  - The first integral here is a standard integral, the second is an Itō one!
  - The function $\mu(X_t, t)$ controls the drift of the process at t.
  - The function $\sigma(X_t, t)$ controls the infinitesimal variance at t.

- In practice, this expression is usually written in the more compact "stochastic differential equation" form:
$$dX_t = \mu(X_t, t)\, dt + \sigma(X_t, t)\, dW_t.$$
- However, it is important to remember that the latter expression is just a shorthand for the former.

Itō's lemma

- Itō's lemma is an equivalent of the chain rule for continuous time stochastic processes.
- Suppose:

$$dX_t = \mu_t \, dt + \sigma_t \, dW_t.$$

- Then for any twice differentiable function $f\colon \mathbb{R}^2 \to \mathbb{R}$, Itō's lemma states:

$$df(t, X_t) = \left(f_1 + \mu_t f_2 + \frac{1}{2} \sigma_t^2 f_{22} \right) dt + \sigma_t f_2 \, dW_t$$

- For example, let $Y_t = \exp X_t$, then:

$$dY_t = d \exp X_t = \left(\mu_t Y_t + \frac{1}{2} \sigma_t^2 Y_t \right) dt + \sigma_t Y_t \, dW_t.$$

Ornstein-Uhlenbeck processes

- Ornstein-Uhlenbeck processes are the continuous time equivalent of $\mathrm{AR}(1)$ processes.
  - Recall for later that the $\mathrm{AR}(1)$ process $x_t = (1-\rho)\mu + \rho x_{t-1} + \sigma\varepsilon_t$ has an $\mathrm{MA}(\infty)$ representation $x_t = \mu + \sigma\sum_{s=0}^{\infty}\rho^s\varepsilon_{t-s}$.

- Ornstein-Uhlenbeck processes are solutions to the s.d.e.:
$$dX_t = \theta(\mu - X_t)\,dt + \sigma\,dW_t.$$
- To find their properties, first define $Y_t = e^{\theta t}X_t$, then, by Itō's lemma:
$$dY_t = d\left(e^{\theta t}X_t\right) = \left(\theta Y_t + \theta(\mu - X_t)e^{\theta t}\right)dt + \sigma e^{\theta t}\,dW_t$$
$$= \theta\mu e^{\theta t}\,dt + \sigma e^{\theta t}\,dW_t.$$
- I.e.:
$$Y_t = Y_0 + \int_0^t \theta\mu e^{\theta u}\,du + \int_0^t \sigma e^{\theta u}\,dW_u = (X_0 - \mu) + \mu e^{\theta t} + \sigma\int_0^t e^{\theta u}\,dW_u\,.$$
- Thus:
$$X_t = (X_0 - \mu)e^{-\theta t} + \mu + \sigma\int_0^t e^{\theta(u-t)}\,dW_u = (X_0 - \mu)e^{-\theta t} + \mu + \sigma\int_{s=0}^t e^{-\theta s}\,dW_{t-s}\,.$$
- Define $Z_t = X_{t+\tau}$ (i.e. Z_t is an Ornstein-Uhlenbeck process started at time $-\tau$). Then, in the limit as $\tau \to \infty$:
$$Z_t = \mu + \sigma\int_{s=0}^{\infty} e^{-\theta s}\,dW_{t-s}\,.$$

Conclusion and recap

- Reduced form VARs do not identify shocks.
- Identification is impossible without making strong prior assumptions.
- Continuous time stochastic processes are not so different to discrete time ones.