Identifying Causal Effects in Time Series Models

Aaron Smith
UC Davis
July 27, 2015

Slides and references will be posted on my website
http://asmith.ucdavis.edu
Identifying Causal Effects in Time Series Models

If all the “Metrics” I Understand is in Mostly Harmless Econometrics, How Should I Think About Time Series Analysis?

Aaron Smith

UC Davis

July 27, 2015

Slides and references will be posted on my website http://asmith.ucdavis.edu
Outline

1. Potential Outcomes and Ideal Experiments
   ▶ Internal validity vs external validity
   ▶ Single discrete events
   ▶ Multiple discrete events

2. Potential events every time period
   ▶ Everything is autocorrelated and everything is endogenous
   ▶ The impulse response as an average treatment effect
   ▶ Identification

3. Conclusion
Effect of a Carbon Tax on the Demand for Gasoline?

\[ y_t = \alpha + \beta p_t + \varepsilon \]

Two requirements of a good estimate

1. **Internal validity**
   - Is \( \beta \) the predicted change in \( y_t \) if price had not changed the way it did in this sample?
   - Need “exogenous” or “random” variation in \( p_t \)

2. **External validity**
   - Is \( \beta \) the predicted change in \( y_t \) if we invoke a carbon tax?
   - Did we estimate \( \beta \) based on temporary or persistent price changes?
   - Does \( \beta \) represent the short-run or long-run response to a persistent price?
Effect of a Carbon Tax on the Demand for Gasoline?

\[ y_t = \alpha + \beta p_t + \epsilon \]

Two requirements of a good estimate

1. **Internal validity**
   - Is \( \beta \) the predicted change in \( y_t \) if price had not changed the way it did in this sample?
   - Need “exogenous” or “random” variation in \( p_t \)

2. **External validity**
   - Is \( \beta \) the predicted change in \( y_t \) if we invoke a carbon tax?
   - Did we estimate \( \beta \) based on temporary or persistent price changes?
   - Does \( \beta \) represent the short-run or long-run response to a persistent price?
Case 1: Single Discrete Event

\[ y_t(D_t) = \begin{cases} 
  y_t(0) & \text{if } D_t = 0 \quad \text{(untreated)} \\
  y_t(1) & \text{if } D_t = 1 \quad \text{(treated)} 
\end{cases} \]

- \( D_t \) is the treatment variable
- Observation \( t \) has two potential outcomes depending on whether the treatment is on or off.
- Average treatment effect is

\[ ATE = E[y_t(1) - y_t(0)] \]
StarLink is a variety of GMO corn that was not approved for human consumption or for export to some markets.

It was found in the food supply in 2000.

Carter and Smith (2007) estimate a 6.8% drop in corn price.
Estimating the ATE

Case 1: Single Discrete Event

- **Simplest Approach:** Use pre-treatment data to estimate the expected value in the absence of treatment

\[
\hat{ATE} = \frac{1}{\tau} \sum_{t=1}^{\tau} y_t - \frac{1}{T-\tau} \sum_{t=\tau+1}^{T} y_t
\]

- What if other explanatory variables differ before and after \( \tau \)?

\[
y_t = D_t \beta + X'_t \gamma + \varepsilon_t
\]

- \( D_t \in \{0, 1\} \) is treatment dummy
- \( X_t \) contains conditioning variables
- Need \( E[W_t \varepsilon_t | X_t] = 0 \) (White, 2006) (c.f. selection on observables)
- Parallels to event study and regression discontinuity literatures
Common Specification: Lagged $y_t$ as a Control
Case 1: Single Discrete Event

$$y_t = 0.63 - 0.02D_t + 0.84y_{t-1} + \varepsilon_t$$

- Why is $\beta$ coefficient $-0.02$ and not $-0.13$?
Common Specification: Lagged $y_t$ as a Control
Case 1: Single Discrete Event

\[ y_t = 0.63 - 0.02W_t + 0.84y_{t-1} + \varepsilon_t \]

- Why is $\beta$ coefficient $-0.02$ and not $-0.13$?
- single-period effect is $-0.02$
- mean effect is $-0.02/(1 - 0.84) = -0.13$
Case 2: Multiple Discrete Events

\[
y_t(D_t) = \begin{cases} 
y_t(0) & \text{if } D_t = 0 \quad \text{(untreated)} 
y_t(d_1) & \text{if } D_t = d_1 \quad \text{(treated with dose } d_1) 
\vdots 
y_t(d_J) & \text{if } D_t = d_J \quad \text{(treated with dose } d_J) 
\end{cases}
\]

- Observation \( t \) has \( J \) potential outcomes depending on dose
- Average treatment effect for dose \( j \) is \( ATE_j = E \left[ y_t(d_j) - y_t(0) \right] \)
- Overall average treatment effect is

\[
ATE = E_{d_j} \left[ \frac{ATE_j}{d_j} \right]
\]

- Angrist, Jorda, and Kuersteiner (2013) estimate average effects of discrete changes in Federal Funds Rate
Case 3: A Potential Event Every Period

- Example: estimating an elasticity of demand

- Potential outcomes in period $t$ are

$$y_t(p_t) = \begin{cases} y_t(\bar{p}) & \text{if } p_t = \bar{p} \\ y_t(\bar{p} + v_t) & \text{if } p_t = \bar{p} + v_t \end{cases}$$

- (untreated)

- (treated)

- Think of $v_t$ as this year’s yield shock

- Average treatment effect in period $t$ is

$$ATE_t = E \left[ \frac{y_t(\bar{p} + v_t) - y_t(\bar{p})}{v_t} \middle| v_t \right]$$

- We can’t estimate this because we have only one observation $t$
Average Treatment Effect Over Time

Case 3: A Potential Event Every Period

\[ ATE = E_v \left[ E \left[ \frac{y_t(\bar{p}_t + \nu_t) - y_t(\bar{p}_t)}{\nu_t} \right] \bigg| \nu_t \right] \]

▶ Demand example

\[ y_t = \alpha + \beta p_t + \varepsilon_t \]

▶ Suppose \( p_t = \bar{p} + \nu_t \). Then

\[ ATE = E_v \left[ \frac{(\alpha + \beta(\bar{p} + \nu_t)) - (\alpha + \beta \bar{p})}{\nu_t} \right] = \beta \]

▶ BUT not all price moves are supply shocks
Endogeneity
Case 3: A Potential Event Every Period

- Demand example

\[ y_t = \alpha + \beta p_t + \varepsilon_t \]
\[ p_t = \bar{p} + \nu_t + \gamma \varepsilon_t \]

- Could use an instrumental variable \( z_t \) with the properties

\[ E[z_t \varepsilon_t] = 0 \text{ and } E[z_t \nu_t] \neq 0 \]

- **BUT** instrument typically has persistent effect on \( \nu_t \)
  - \( \{z_t, z_{t-1}, z_{t-2}, ...\} \) all affect \( \nu_t \)
  - e.g., poor growing-season weather this year may affect prices for a few years

- **AND** instrument may be autocorrelated
  - We aren’t getting a new, clean experiment every period
Everything is Autocorrelated

Case 3: A Potential Event Every Period

- Even if we have a good instrument, we don’t have an independent experiment every period

- This raises questions
  - Are we estimating the demand response to short-lived or long-lived price changes?
  - Are we estimating the single-period or long-run demand response to price shocks?
  - How much of the variation in our data does the instrument explain?

- The “Mostly Harmless” answer: “I don’t care. I will correct my standard errors”

- A time series analyst’s answer: “I will try to answer these questions”
The Impulse Response as an Average Treatment Effect

Case 3: A Potential Event Every Period

- Estimate the average treatment effect of a forecast error
- Estimate this object at different horizons
- Demand example ($q_t$ is log quantity, $p_t$ is log price, $z_t$ is supply shifter (instrument))

\[
ATE(z_t \hookrightarrow z_{t+k}) = E_v \left[ E \left[ \frac{z_{t+k}(\bar{z}_t + v_{zt}) - z_{t+k}(\bar{z}_t)}{v_{zt}} \right] \right]
\]

\[
ATE(z_t \hookrightarrow p_{t+k}) = E_v \left[ E \left[ \frac{p_{t+k}(\bar{z}_t + v_{zt}) - p_{t+k}(\bar{z}_t)}{v_{zt}} \right] \right]
\]

\[
ATE(z_t \hookrightarrow y_{t+k}) = E_v \left[ E \left[ \frac{y_{t+k}(\bar{z}_t + v_{zt}) - y_{t+k}(\bar{z}_t)}{v_{zt}} \right] \right]
\]
Linear Vector Autoregression (VAR) Models

\[ X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + c_0 + c_1 t + \epsilon_t \]
\[ \epsilon_t \sim WN(0, \Omega) \]

- Multivariate system
  - \( X_t \) may include time series on production, consumption, inventory, demand and supply shifters, and futures
  - \( WN \equiv "white\ noise" \equiv no\ autocorrelation\)

- Everything may be endogenous

- How to identify causal effects?
  1. Estimate the average treatment effect of forecast errors
     - requires assumptions on meaning of contemporaneous correlations, i.e., off-diagonal elements of \( \Omega \)
     - If \( E[\epsilon_{1t}\epsilon_{2t}] \neq 0 \), did \( \epsilon_{1t} \) cause \( \epsilon_{2t} \) or the other way around?
  2. “Granger causality,” which is another word for predictability — requires assumption that effects occur after causes
VAR Identification

- Strip the problem down to its most basic form

\[ X_t = \varepsilon_t, \quad \varepsilon_t \sim WN(0, \Omega) \]

- Define structural errors \( \nu_t = A\varepsilon_t \), and write \( AX_t = \nu_t \)

- **Example:** commodity supply and demand with temperature

\[
\begin{pmatrix}
1 & \alpha_{12} & \alpha_{13} \\
\alpha_{21} & 1 & \alpha_{23} \\
\alpha_{31} & \alpha_{32} & 1
\end{pmatrix}
\begin{pmatrix}
p_t \\
q_t \\
w_t
\end{pmatrix}
= \nu_t
\]

\[
E[\nu_t\nu_t'] =
\begin{bmatrix}
\sigma_1^2 & 0 & 0 \\
0 & \sigma_2^2 & 0 \\
0 & 0 & \sigma_3^2
\end{bmatrix}
\]

- **Identification problem:** which variable causes which?

- With 3 variables, the data give us 3 covariances, so we can identify 3 parameters
VAR Identification: Triangular system

\[ AX_t = \nu_t, \quad \nu_t \sim WN(0, \Sigma) \]

- Impose zeros in lower triangle of A
- Direction of causation: temperature → quantity → price

\[
\begin{bmatrix}
1 & \alpha_{12} & \alpha_{13} \\
0 & 1 & \alpha_{23} \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
p_t \\
q_t \\
w_t
\end{bmatrix}
= 
\begin{bmatrix}
\nu_{1t} \\
\nu_{2t} \\
\nu_{3t}
\end{bmatrix}
\]

\[ E[\nu_t\nu_t'] = 
\begin{bmatrix}
\sigma_1^2 & 0 & 0 \\
0 & \sigma_2^2 & 0 \\
0 & 0 & \sigma_3^2
\end{bmatrix}
\]

- Identification assumption: Short-run supply perfectly inelastic
VAR Identification: Known Supply Elasticity

\[ AX_t = \nu_t, \quad \nu_t \sim WN(0, \Sigma) \]

- Impose a known value in A
- Could also get partial identification by specifying a range for \( \alpha_{21} \)

Demand equation

\[
\begin{bmatrix}
1 & \alpha_{12} & \alpha_{13} \\
0.1 & 1 & \alpha_{23} \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
p_t \\
q_t \\
w_t
\end{bmatrix} = \nu_t
\]

Supply equation

\[
E[\nu_t\nu_t'] =
\begin{bmatrix}
\sigma_1^2 & 0 & 0 \\
0 & \sigma_2^2 & 0 \\
0 & 0 & \sigma_3^2
\end{bmatrix}
\]

- Identification assumption: Short-run supply elasticity equals 0.1
VAR Identification: Instrumental Variables

\[ AX_t = \nu_t, \quad \nu_t \sim WN(0, \Sigma) \]

- Assume that weather does not shift demand
- Standard regression IV would allow \( \nu_{1t} \) to contain some supply shocks, i.e., \( \sigma_{12}^2 \neq 0 \)

\[
\begin{bmatrix}
1 & \alpha_{12} & 0 \\
\alpha_{21} & 1 & \alpha_{23} \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
p_t \\
q_t \\
w_t
\end{bmatrix}
= \nu_t
\]

\[
E[\nu_t \nu_t'] =
\begin{bmatrix}
\sigma_1^2 & \sigma_{12}^2 & 0 \\
\sigma_{12}^2 & \sigma_2^2 & 0 \\
0 & 0 & \sigma_3^2
\end{bmatrix}
\]

- System is not identified, even though \( \alpha_{12} \) is identified
VAR Identification: Instrumental Variables

\[ AX_t = \nu_t, \quad \nu_t \sim WN(0, \Sigma) \]

- Assume that weather does not shift demand
- Get system identification if assume either
  1. \( \sigma_{12}^2 = 0 \) — all supply shocks come from \( w_t \), or
  2. \( \alpha_{21} = 0 \) — supply is perfectly inelastic

\[
\begin{bmatrix}
1 & \alpha_{12} & 0 \\
\alpha_{21} & 1 & \alpha_{23} \\
0 & 0 & 1 \\
\end{bmatrix}
\begin{bmatrix}
p_t \\
q_t \\
w_t \\
\end{bmatrix}
= \nu_t
\]

\[
E[\nu_t \nu_t'] = 
\begin{bmatrix}
\sigma_1^2 & \sigma_{12}^2 & 0 \\
\sigma_{12}^2 & \sigma_2^2 & 0 \\
0 & 0 & \sigma_3^2 \\
\end{bmatrix}
\]

- Or identify system by adding an independent demand shifter
A Common Difference Between IV and VAR

**IV:** Exclusion restriction, but don’t label shocks

\[
\begin{align*}
\text{demand equation} & : \quad \begin{bmatrix} 1 & \alpha_{12} & 0 \end{bmatrix} \begin{bmatrix} p_t \\ q_t \\ w_t \end{bmatrix} = \nu_t \\
\text{supply equation} & : \quad \begin{bmatrix} 0 & 1 & \alpha_{23} \end{bmatrix} \begin{bmatrix} p_t \\ q_t \\ w_t \end{bmatrix} = \nu_t \\
\text{temperature equation} & : \quad \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_t \\ q_t \\ w_t \end{bmatrix} = \nu_t
\end{align*}
\]

\[
E[\nu_t \nu_t'] = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & 0 \\ \sigma_{12}^2 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}
\]

**VAR:** No exclusion restriction, but label shocks

\[
\begin{align*}
\text{demand equation} & : \quad \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} \end{bmatrix} \begin{bmatrix} p_t \\ q_t \\ w_t \end{bmatrix} = \nu_t \\
\text{supply equation} & : \quad \begin{bmatrix} 0 & 1 & \alpha_{23} \end{bmatrix} \begin{bmatrix} p_t \\ q_t \\ w_t \end{bmatrix} = \nu_t \\
\text{temperature equation} & : \quad \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_t \\ q_t \\ w_t \end{bmatrix} = \nu_t
\end{align*}
\]

\[
E[\nu_t \nu_t'] = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}
\]
Example: Effect of Aggregate Demand on Oil Prices
Kilian (AER 2009 and JAAE 2012)

Consider a VAR model based on monthly data for \( z_t = (\Delta \text{prod}_t, \text{rea}_t, \text{rpo}_t)' \), where \( \Delta \text{prod}_t \) is the percent change in global crude oil production, \( \text{rea}_t \) denotes the index of real economic activity constructed in Section I, and \( \text{rpo}_t \) defers to the real price of oil. The \( \text{rea}_t \) and \( \text{rpo}_t \) series are expressed in logs. The sample period is 1973:1–2007:12.\(^4\) The structural VAR representation is

\[
A_0 z_t = \alpha + \sum_{i=1}^{24} A_i z_{t-i} + \varepsilon_t, 
\]

where \( \varepsilon_t \) denotes the vector of serially and mutually uncorrelated structural innovations. I postulate that \( A^{-1}_0 \) has a recursive structure such that the reduced-form errors \( \varepsilon_t \) can be decomposed according to \( \varepsilon_t = A^{-1}_0 \varepsilon_t : \)

\[
e_t \equiv \begin{pmatrix} e_{\Delta \text{prod}}_t \\ e_{\text{rea}}_t \\ e_{\text{rpo}}_t \end{pmatrix} = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{pmatrix} \varepsilon_{\text{oil supply shock}}_t \\ \varepsilon_{\text{aggregate demand shock}}_t \\ \varepsilon_{\text{oil specific-demand shock}}_t \end{pmatrix}
\]

- Estimate the effect of production surprises on oil price
- Estimate the effect of REA surprises uncorrelated with production surprises
Example: Impulse Responses a.k.a. ATE

Kilian (AER 2009 and JAAE 2012)

- Demand has large persistent effect on prices
Example: Allowing SR Supply Response to be Positive
Kilian (AER 2009 and JAAE 2012)

(a) Model with sign restrictions only.

(b) Model with sign restrictions and oil supply elasticity bound.
Conclusion

- Time series analysts tend to ask questions that are less clean, perhaps because they have few clean experiments

- Rather than what is the effect of $X$ and $Y$, ask
  - What is the response of $Y$ to an unexpected change in $X$? How long does the response last?
  - How does $X$ vary?
  - How much of the variation in $Y$ does $X$ explain?

- Think of the impulse response as an average treatment effect
  - The treatment is a forecast error
  - Track the response at various horizons