Bonn Summer School
Advances in Empirical Macroeconomics

Karel Mertens

Cornell, NBER, CEPR

Bonn, June 2015

Overview

1.2 Identification Strategies

We have a dynamic model for $E[z_t \mid \mathcal{I}_{t-1}]$ and we can measure $\upsilon_t = z_t - E[z_t \mid \mathcal{I}_{t-1}]$.

The structural **impulse response** (IR) associated with e_t are the coefficients in the MA representation

$$z_t \quad = \quad M(L)\upsilon_t = \sum_{i=0}^{\infty} \mathcal{M}_i \upsilon_{t-i} \quad , \quad \upsilon_t = \mathcal{D} e_t, \; \mathcal{M}_0 = I$$

For shock j: $\frac{\partial E[z_{t+h}|\mathcal{I}_t]}{\partial e_{jt}} = \mathcal{M}_h \mathcal{D}_j$

We can back out $\mathcal{M}_h$, $h > 0$ from any of the reduced form models.

But to estimate the (average) dynamic causal effects $\partial E[z_{t+h} \mid \mathcal{I}_t]/\partial e_{jt}$ we also need to know $\mathcal{D}_j$, i.e. column j of the **impact matrix** $\mathcal{D}$.

We have estimates of υ_t and Σ, and we know that

$$\Sigma = \mathrm{Var}(\mathcal{D}e_t) = \mathcal{D}\mathrm{Var}(e_t)\mathcal{D}' = \mathcal{D}\mathcal{D}'$$

Symmetric positive semi-definite Σ provides $n \times (n+1)/2$ restrictions on the n^2 elements of $\mathcal{D}$

Not sufficient to uncover any of the columns of $\mathcal{D}$: the **identification problem**.

We need additional identifying restrictions.

In exactly identified systems, these restrictions are not testable.

In overidentified systems, these restrictions are testable.

Common are (combinations) of equality restrictions on

- the **impact matrix** $\mathcal{D}$,
  i.e. the contemporaneous response to shocks
- the **inverse impact matrix** $\mathcal{D}^{-1}$,
  i.e. the linear contemporaneous relationship between z_t.
- the **horizon** h**-impulse response coefficients** $\mathcal{M}_h\mathcal{D}$,
  i.e. the response after h period
- the **infinite horizon cumulative impulse responses** $M(1)\mathcal{D}$,
  i.e. the long run cumulative response to the shock

Subject to order and rank conditions for (local/global) identification (See Rubio-Ramirez, Wagonner and Zha, 2010)

Note, this generally involves solving a system of nonlinear equations.

Recursive Identification Scheme

Zero restrictions on the impact matrix, lower triangular $\mathcal{D}$:

$$\mathcal{D} = \begin{bmatrix} d_{11} & 0 & \dots & 0 \\ d_{21} & d_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \dots \\ d_{n1} & \dots & \dots & d_{nn} \end{bmatrix}$$

Adds $\frac{n\times(n-1)}{2}$ restrictions such that all n^2 elements of $\mathcal{D}$ are identified.

Easy computation through **Cholesky decomposition** of Σ, which factors a positive semi-definite matrix P into the product of a lower triangular matrices and its transpose $\Sigma = \mathcal{D}\mathcal{D}'$.

Partial Identification with Block-Recursive Scheme

Partition $z_t = [z_{1t},\ z_{2t},\ z_{3t}]'$ and $e_t = [e_{1t},\ e_{2t},\ e_{3t}]$ and consider the lower **block** triangular matrix

$$\mathcal{D} = \begin{bmatrix} \underset{n_1\times n_1}{d_{11}} & \underset{n_1\times 1}{0} & \underset{n_1\times n_2}{0} \\ \underset{1\times n_1}{d_{21}} & \underset{1\times 1}{d_{22}} & \underset{1\times n_2}{0} \\ \underset{n_2\times n_1}{d_{31}} & \underset{n_2\times 1}{d_{32}} & \underset{n_2\times n_2}{d_{33}} \end{bmatrix}$$

Christiano, Eichenbaum & Evans (1999) show that

1. Many matrices $\mathcal{D}$ of the above form, one of which is the lower triangular matrix, that satisfy $\Sigma = DD'$.
2. Each of these has the same $\mathcal{D}_2$ and IR to e_{2t}.
3. Using the Cholesky-identified $\mathcal{D}$, column $\mathcal{D}_2$ and the IR to e_{2t} are invariant to the ordering of variables within z_{1t} and z_{3t}.

Long Run Restrictions

Suppose z_t is in growth rates, then the long-run impact of e_t on levels is

$$\sum_{h=0}^{\infty} \mathcal{M}_h \mathcal{D} = M(1)\mathcal{D}$$

Common are zero restrictions on $M(1)\mathcal{D}$,

e.g. Blanchard and Quah (1989), Shapiro and Watson (1988), King, Plosser, Stock and Watson (1991), Gali (1999), Fisher (2006), Beaudry and Portier (2006).

Easily implemented by lower triangularization $M(1)\Sigma M(1)' = LL'$ and

$$\mathcal{D} = M(1)^{-1} L$$

Christiano, Eichenbaum & Evans (1999) results apply here as well.

Examples of Other Restrictions

- **Sign restrictions**, i.e. inequality instead of equality restrictions.

  Faust (1998), Uhlig (2005), Canova and De Nicoló (2002)

- **Medium run restrictions**, i.e. on $\mathcal{M}_h\mathcal{D}$

  Uhlig (2004)

- **Maximization of the FEV contribution**.

  Barsky and Sims (2006), Francis, Owyang, Roush, and DeCecio (2014)

- **Heteroskedastic covariance restrictions**

  Rigobon (2000), Sentana and Fiorentini JE (2001)

Instrumental Variables Approach

Identifying restrictions generate 'instruments'.

The elements of $\mathcal{D}$ can also be obtained by IV methods.

IV estimation:

$$y_t = \beta x_t + u_t \quad , \quad E[x_t u_t] \neq 0$$

Let w_t be an 'instrument' for x_t satisfying

$$E[w_t x_t] \neq 0 \qquad \text{(relevance)}$$
$$E[w_t u_t] = 0 \qquad \text{(exogeneity)}$$

Two Stage Least Squares (2SLS):

1. First Stage: Regress x_t on w_t and obtain $\hat{x}_t$
2. Second Stage: Regress y_t on $\hat{x}_t$ to obtain consistent estimate of β

Example with Recursive Identification

Suppose

$$\begin{aligned} \upsilon_{1t} &= d_{11} e_{1t} \\ \upsilon_{2t} &= d_{21} e_{1t} + d_{22} e_{2t} \\ \upsilon_{3t} &= d_{31} e_{1t} + d_{32} e_{2t} + d_{33} e_{3t} \end{aligned}$$

IV implementation:

1. Obtain d_{11} as the square root of the first diagonal element of Σ. Calculate $e_{1t} = d_{11}^{-1} \upsilon_{1t}$.
2. Regress υ_{2t} on υ_{1t} using e_{1t} as instrument to obtain d_{21}. Obtain d_{22} from std(residual) and calculate $e_{2t} = d_{22}^{-2} (\upsilon_{2t} - d_{21} e_{1t})$
3. Regress υ_{3t} on υ_{1t} and υ_{2t} using e_{1t} and e_{2t} as instruments to obtain d_{31} and d_{32}. Obtain d_{33} from std(residual).

Analogous for block recursive partial identification.

Analogous for long run zero restrictions, just replace v_t by $\tilde{v}_t = M(1)v_t$.
See Shapiro and Watson (1988)

Generally there is an equivalent IV implementation,
See Hausman and Taylor (1983).

For IV methods with inequalities, see Nevo and Rosen (2012)

Some Criticisms of Typical Identification Restrictions

- Short-run restrictions based on timing assumptions that are often hard to defend a priori.

  See Rudebusch (1998), Stock and Watson (2001)

- Long-run restrictions can be theoretically more appealing, but are unreliable in realistic samples.

  See Faust and Leeper (1997), Chari, Kehoe and McGrattan (2007), Kascha and Mertens (2010)

- Identified shocks seem often unrelated to known historical events.

  See Rudebusch (1998)

- Estimated innovations often based on insufficient information.

  See Reichlin and Lippi (1994), Romer and Romer (2004), Ramey (2011), Leeper, Walker and Yang (2013)

Event Study/Natural Experiment/Narrative Approach

A different approach to identification of causal effects is based on analyzing historical events that are

(1) unexpected by economic decision makers

(2) unrelated by other disturbances affecting economic decisions

These properties of events are established using ‘narrative’ methods.

Potentially addresses the concerns with traditional restrictions.

Examples:

- Monetary policy changes: Friedman and Schwartz (1963), Romer and Romer (1989)
- Oil price changes: Hamilton (1983), Hoover and Perez (1994)
- Military spending changes: Ramey and Shapiro (1998), Edelberg, Eichenbaum and Fisher (1999)
- Tax reforms: Romer and Romer (2010), Cloyne (2012)

Narrative Identification in Time Series Models

Let m_t be a scalar variable capturing the 'events' that satisfies

$$E[m_t e_{jt}] = \phi \neq 0 \qquad (A1)$$

$$E[m_t e_{-jt}] = 0 \qquad (A2)$$

$$E[m_t \mid \mathcal{I}_{t-1}] = 0 \qquad (A3)$$

This means m_t is assumed to be

correlated with the contemporaneous shock of interest (A1)
uncorrelated with other contemporaneous shocks (A2)
uncorrelated with any past shocks (A3).

Note: m_t may be a discrete variable (e.g. dummies), may be censored, ...

ϕ is unknown ex ante.

Common specifications for uncovering IR's to e_{jt} up to a scale λ:

Distributed Lag Specification (motivated by MA representation)

$$z_t = \delta(L)m_t + u_t \ , \ \delta(L) = \delta_0 + \delta_1 L + \delta_2 L^2 + \delta_3 L^3 + ...$$

VAR-X (motivated by VARMA representation)

$$B(L)z_t = \delta(L)m_t + u_t$$

Augmented SVAR (simply treats m_t as an observable)

$$B(L)\begin{bmatrix} m_t \\ z_t \end{bmatrix} = u_t$$

and identify IR to e_{1t} block-recursively using Cholesky of $Var(u_t)$.

Some important remarks:

- A1 is testable
- A2 is not testable
- A3 is testable
- For the DL and VARX specifications, it is crucial that the ‘events’ are unpredictable, i.e. A3 must hold (but is testable).
- For an Augmented SVAR with ‘adequate’ z_t, A3 is not required.
- Identification only up to a scale ϕ

  In practice m_t may satisfy A1-A3 but may be mismeasured $\Rightarrow$ innovations to m_t are uninterpretable without further assumptions.

  Instead scale IR’s according to one of the outcome variables in z_t (for which measurement error is assumed to be small).

Proxy/External Instruments Approach

Mertens and Ravn (2013)

Interpret m_t as a proxy measure of latent variable e_{jt}.

Estimate conventional SVAR

$$B(L)z_t = \upsilon_t$$

and, assuming $A1$, impose covariance restrictions $A2$ to solve for $\mathcal{D}^j$.

Easy to implement since A1 and A2 imply

$$\mathcal{D}_j = E[\upsilon_t m_t]/\phi$$

and solve for the scale that is consistent with $\Sigma = \mathcal{D}\mathcal{D}'$.

Some advantages over previous specifications

- No need to assume A3.
- Parsimonious, no need to estimate VAR or DL coefficients on m_t which often has many missing observations.
- Automatic scale adjustment, and robust to many types of measurement error in m_t.
- Easy comparison with alternative identification restrictions since the reduced form is the same.
- We can use as proxy m_t also the projection of narrative variables on observables. This nest the augmented VAR case, but we may also include observables not in z_t.

Generalization to Multiple Shocks

Partition $\upsilon_t = \begin{bmatrix} \underset{k\times 1}{\upsilon_{1t}} \\ \underset{n-k\times 1}{\upsilon_{2t}} \end{bmatrix}$, $e_t = \begin{bmatrix} \underset{k\times 1}{e_{1t}} \\ \underset{n-k\times 1}{e_{2t}} \end{bmatrix}$,

e_{1t} are the shocks of interest.

Suppose we have $k \times 1$ vector of proxy variables m_t

Identification assumptions:

$$E[m_t e_{1t}'] = \Phi \qquad \text{(A1)}$$

$$E[m_t e_{2t}'] = 0 \qquad \text{(A2)}$$

where Φ is $k \times k$, unknown and nonsingular, but not necessarily diagonal.

Partition $\mathcal{D} = \begin{bmatrix} \underset{k\times k}{\mathcal{D}_{11}} & \underset{k\times n-k}{\mathcal{D}_{12}} \\ \underset{n-k\times k}{\mathcal{D}_{21}} & \underset{n-k\times n-k}{\mathcal{D}_{22}} \end{bmatrix}$, $\mathcal{D}_1 = \begin{bmatrix} \underset{k\times k}{\mathcal{D}_{11}} \\ \underset{n-k\times k}{\mathcal{D}_{21}} \end{bmatrix}$

Assumptions A1/A2 imply $n \times k$ conditions

$$\Phi \mathcal{D}_1' = E[m_t \upsilon_t']$$

from which we extract $(n-k) \times k$ covariance restrictions

$$\mathcal{D}_{21} = (E[m_t \upsilon_{1t}']^{-1} E[m_t \upsilon_{2t}'])' \mathcal{D}_{11}$$

that can be used for identifying the first k columns.

These restrictions identify ${\mathcal{D}_{21}\mathcal{D}_{11}}^{-1}$.

An additional $k(k-1)/2$ restrictions are needed to fully identify $\mathcal{D}_1$.

${\mathcal{D}_{21}\mathcal{D}_{11}}^{-1}$ provides the impact matrix of e_{1t} up to a rotation.

Implementation with IV

Stock and Watson (2008, 2012) develop the equivalent IV approach that views m_t as 'external instruments'.

The RHS in

$$\mathcal{D}_{21}\mathcal{D}_{11}{}^{-1} = (E[m_t \upsilon'_{1t}]^{-1} E[m_t \upsilon'_{2t}])'$$

replaced with sample moments is just the 2SLS estimator of regression of υ_{2t} on υ_{1t} using m_t as instruments for υ_{1t}.

Generalizing further : if available we can use more than k instruments for k shocks and test for exogeneity of the instruments.

Implementation with IV

The approach is also equivalent to IV with observables directly:

1. First Stage: Regress z_{1t} on m_t and p lags of z_t and obtain $\hat{z}_{1t}$
2. Second Stage: Regress z_{2t} on $\hat{z}_{1t}$ and p lags of z_t

If $k = 1$, the IV estimates are the IR's to e_{1t} causing a unit innovation in z_{1t}.

If $k > 1$, combine with additional restrictions to obtain IR's.

The first stage here can be used for diagnostics of instrument relevance (assumption A1).

See also Stock and Montiel-Olea (WIP 2012).

Some Applications of Proxy/External Instrument VARs

- **Variety of Empirical Shock Measures**: Stock and Watson (2012)
- **Model based shocks (wedges)**: Evans and Marshall (2009) (same idea as Proxy SVAR)
- **Government Spending News**: Ramey (2010) (Augm. SVAR)
- **Austerity Packages**: Guajardo, Leigh and Pescatori (Augm. SVAR/IV)
- **Tax Reforms**: Romer and Romer (2010) (Augm. SVAR), Mertens and Ravn (2013, 2014) and Mertens (2013)
- **High Freq. Monetary Shocks**: Gertler and Karadi (2015), Passari and Rey (2015)
- **Financial News/Variables**: Brutti and Sauré (2015), Cesa-Bianchi, Cespedes, and Rebucci (2015), Bahaj (2013), Davis (2014)
- **Oil Shocks**: Stevens (2014)

Local Projections

So far, IR's were obtained from the MA representation, for instance by inverting a VAR.

IR's are nonlinear functions of parameters, which complicates inference:

Delta method, Monte Carlo/bootstrap methods, Bayesian methods.

An alternative way to estimate IRFs is by local projections (Jordà 2004):

$$z_{t+h} = C_h(L)z_t + u_{ht} \quad , \quad \text{for } h > 0$$

where $C_h(L) = C_{0h} + C_{1h}L + C_{2h}L^2 + ... + C_{ph}L^p$.

Suppose we know $\mathcal{D}_j = \partial z_t / \partial e_{jt}$, then $\partial E[z_{t+h} \mid \mathcal{I}_t] / \partial e_{jt} = C_{0h}\mathcal{D}_j$

If we know $\mathcal{D}_j$, we also know e_{jt}, so equivalently

$$z_{t+h} = D_{jh} e_{jt} + u_{ht} \quad , \quad \text{for } h > 0$$

where $\partial E[z_{t+h} \mid \mathcal{I}_t] / \partial e_{jt} = D_{jh}$.

Iteratively estimates the coefficients of the MA representation.

This approach is

- simple, univariate OLS regressions using HAC standard errors (Just type '`newey z e`' in Stata)
- but inefficient compared to (inverting) a linear system that generates the correct υ_t needed for identification
- more robust to misspecification (lag length), but only conditional on having the correct $\mathcal{D}_j$ or e_{jt}.

Local Projections-IV

Suppose e_{jt} is the shock of interest, but now unobserved.

Suppose we have variables m_t satisfying

$$E[m_t e_{jt}] \neq 0 \quad \text{(A1)}$$
$$E[m_t e_{-jt}] = 0 \quad \text{(A2)}$$
$$E[m_t \mid \mathcal{I}_{t-1}] = 0 \quad \text{(A3)}$$

First Stage:

$$z_{jt} = \delta m_t + u_{j0t} \quad , \quad \hat{z}_{jt} = \delta m_t$$

Second Stage:

$$\begin{aligned} z_{-jt} &= D_{-j0}\hat{z}_{jt} + u_{-j0t} \\ z_{t+h} &= D_{jh}\hat{z}_{jt} + u_{ht} \quad , \quad h > 0 \end{aligned}$$

D_{-j0} and D_{jh} are the IR's to e_{jt} causing a unit innovation in z_{1t}.

Local Projections-IV

We can drop assumption A3 if we have controls such that $E_t[m_t u_{j0t}] = 0$.

First Stage:

$$z_{jt} = \delta m_t + \text{controls}_t + u_{j0t} \quad , \quad \hat{z}_{jt} = \delta m_t$$

Second Stage:

$$\begin{aligned} z_{-jt} &= D_{-j0}\hat{z}_{jt} + \text{controls}_t + u_{-j0t} \\ z_{t+h} &= D_{jh}\hat{z}_{jt} + \text{controls}_t + u_{ht} \quad , \quad h > 0 \end{aligned}$$

D_{-j0} and D_{jh} are the IR's to e_{jt} causing a unit innovation in z_{1t}.

Naturally, use the same controls required for $\upsilon_t = z_t - E[z_t \mid \mathcal{I}_{t-1}]$.

With same controls, impact IR's will be identical to proxy/external instrument approach in linear system.

IR's for $h > 0$ will be different.

LP is simple but inefficient.
But estimating linear systems is also not hard, so why do LP?

Simplicity becomes a major advantage in case of **nonlinear** models and identification with (external) **instruments**.

Imagine general nonlinear impulse response function

$$z_t = f(e_t, e_{t-1}, e_{t-2}, ...)$$

and (locally) approximate using your favorite expansion

$$z_t \approx \text{linear terms} + \text{non-linear terms}$$

It is straightforward to add nonlinear terms in e_t or z_t to the LP.

Using IV, we never need to estimate the full nonlinear system to identify the contemporaneous impact.

Of course in practice we need to worry about parameter proliferation.

Some Applications of LP-IV

- **Spending Shocks**: Ramey and Zubairy (2014), Auerbach and Gorodnichenko (2014), Bernardini and Peersman (2015)
- **High Freq. Monetary Shocks**: Ramey (2015)
- **Tax Reforms**: Ramey (2015)
- **Austerity**: Jordà and Taylor (2015)

Some comments

- External instruments greatly expands options for identification in VARs and other models in combination with existing IV machinery.
- Worry about assumption A1: **instrument relevance**
- Worry about assumption A2: **contemporaneous exogeneity**

  'Narrative' variables do not automatically buy you identification.
- Worry about dropping assumption A3: **adequate controlling**

  Use a proper conditioning set to capture innovations to expectations.
- LP-IV is an easy way to allow for (some) nonlinearities.
- We should probably use more state space modeling.

Some general references on identification:

- Christiano, Eichenbaum and Evans (1999), 'Monetary Policy Shocks: What have we learned and to what end?'
- Stock and Watson, 2001, 'Vector Autoregressions'
- Luetkepohl, 2005, 'A New Introduction to Time Series Analysis'
- Ramey, 2015, 'Macroeconomic Shocks and Their Propagation'