Inference in Nearly Nonstationary SVAR Models with Long-Run Identifying Restrictions

Nikolay Gospodinov (Concordia University)

Second Forecasting Conference, Duke University, March 2007
1 Motivation

- disagreement about the response of hours worked to a positive technology shock (Gali, 1999; Chari, Kehoe and McGrattan, 2005; Christiano, Eichenbaum and Vigfusson, 2005)

- confidence bands are very wide and uninformative

- statistical properties of the IRF in this example have not been formally investigated
Response of hours worked to a 1% positive shock in technology estimated with US data for the period 1948Q1-2005Q3. **Left** graph: hours worked in *levels*; **Right** graph: hours worked in *first differences*. 
Based on a Monte Carlo simulation of 1,000 samples of length 180 generated from the RBC model (Christiano, Eichenbaum and Vigfusson, 2005)
2 Contributions of the Paper

- shows the inconsistency of the impulse responses and some structural parameters in a near-nonstationary SVAR with long-run restrictions
  - the inconsistency arises from the presence of weak instruments

- derives the limiting distributions of the structural parameters
  - nonstandard and fat-tailed for the parameters that determine the shape of the IRF

- proposes an improved (consistent and asymptotically normal) estimator of the IRF by incorporating additional statistical restrictions
Econometric Specification and Identifying Assumptions in Our Example

- labour productivity is an exact unit root process
  - the RBC model imposes a unit root on technology

- hours worked are highly persistent but not cointegrated with productivity
  - the RBC model implies that hours are stationary

- long-run identifying restriction: only technology has a permanent effect on labour productivity

- structural shocks are assumed to be orthogonal
Literature


- IV estimation of SVAR models
3  SVAR Models with Long-Run Identifying Restrictions

Reduced form VAR($p + 1$) model for $y_t = (l_t, h_t)'$

$$A(L)y_t = u_t,$$

where $A(L) = I - A_1L - \ldots - A_{p+1}L^{p+1}$.

Structural VAR model

$$B(L)y_t = \varepsilon_t,$$

where $B(L) = B_0A(L)$, $B_0 = \begin{bmatrix} 1 & -b_2^{(0)} \\ -b_1^{(0)} & 1 \end{bmatrix}$ and $\varepsilon_t = B_0u_t$ denote the structural shocks $(\varepsilon_t^z, \varepsilon_t^d)'$. 
Long-Run (LR) Identifying Restrictions

- Long-run $\Leftrightarrow L = I$

- LR identifying restrictions imposed on the matrix of long-run multipliers

$$B(I) = B_0 A(I)$$

- in our technology shock example, $B(I)$ is lower triangular
3.1 Nonstationary SVAR Model with No Cointegration

Due to the structure of our model, we let \( A(L) = \Psi(L)(I - \Phi L) \),

\[
\Psi(L)(I - \Phi L)y_t = u_t, \tag{1}
\]

where \( \Phi_T = \begin{bmatrix} 1 & 0 \\ 0 & 1 + c/T \end{bmatrix} \) for a fixed constant \( c < 0 \) (Elliott, 1998; Pesavento and Rossi, 2006) and \( \Psi(L) = I - \sum_{i=1}^{p} \Psi_i L^i \) = \[
\begin{bmatrix}
\psi_{11}(L) & \psi_{12}(L) \\
\psi_{21}(L) & \psi_{22}(L)
\end{bmatrix}.
\]

- diagonality of \( \Phi_T \) is crucial for the validity of our proposed estimator (this can be somewhat relaxed as we'll see later on)

- off-diagonal elements are set to zero to rule out near-I(2) behavior
Then, it follows that

\[
B(I) = \begin{bmatrix}
0 & -c/T \left( \psi_{12}(1) - b_{12}^{(0)} \psi_{22}(1) \right) \\
0 & -c/T \left( \psi_{22}(1) - b_{22}^{(0)} \psi_{12}(1) \right)
\end{bmatrix}.
\]

**Important Observation**

- imposing the restriction that \( B(I) \) is lower triangular constrains the parameters of \( B_0 \)
- in particular, \( b_{12}^{(0)} = \psi_{12}(1)/\psi_{22}(1) \)
4 Results

After imposing the LR restriction, the (Dickey-Fuller forms) of SVAR and VAR are

\[ \Delta l_t = b_{12}^{(0)} \Delta h_t + \text{lags}(\Delta l_t, \Delta h_t) + \varepsilon_t^z \]
\[ \Delta h_t = b_{21}^{(0)} \Delta l_t + b_{22}^{(1)} h_{t-1} + \text{lags}(\Delta l_t, \Delta h_t) + \varepsilon_t^d. \]

\[ \Delta l_t = \frac{c}{T} \psi_{12}(1) h_{t-1} + \text{lags}(\Delta l_t, \Delta h_t) + u_{1,t} \]
\[ \Delta h_t = \frac{c}{T} \psi_{22}(1) h_{t-1} + \text{lags}(\Delta l_t, \Delta h_t) + u_{2,t}. \]

Treat the higher-order model as a first-order model where the data matrices are residuals from projections on the predetermined variables (\( \Delta l_{t-1}, \Delta h_{t-1}, \ldots, \Delta l_{t-p}, \Delta h_{t-p} \)).
4.1 First-order model \((\Psi(L) = I)\)

**Structural form**

\[
\begin{align*}
\Delta l_t &= b_{12}^{(0)} \Delta h_t + \varepsilon_{zt} \\
 h_t &= b_{21}^{(0)} \Delta l_t + b_{22}^{(1)} h_{t-1} + \varepsilon_{td}
\end{align*}
\]

where \(\varepsilon_{zt}\) and \(\varepsilon_{td}\) are orthogonal.

**Reduced form**

\[
\begin{align*}
\Delta l_t &= u_{1,t} \\
\Delta h_t &= \frac{c}{T} h_{t-1} + u_{2,t}
\end{align*}
\]

**IV estimation:** instrument the endogenous variables in (2) \(\Delta h_t\) and \(\Delta l_t\) by \(h_{t-1}\) and \(\varepsilon_{zt}\), respectively.
Remarks:

- this IV estimator is identical to the OLS-based estimator which we typically use in applied work but provides important insights about the sampling behavior of the impulse response functions.

- local-to-zero correlation between the endogenous variable $\triangle h_t$ and instrument $h_{t-1}$ (Staiger and Stock, 1997). Interestingly, the weak instrument problem arises directly from the near-unit root parameterization of $h_t$.

- once we are in the IV framework, we can use inference procedures that are robust to weak instruments, search for better instruments, overidentifying restrictions, validity of instruments etc. (work in progress)
Theorem 1. Under the model assumptions,

\[ \frac{\hat{b}_{12}^{(0)} - b_{12}^{(0)}}{\gamma_1} \Rightarrow \frac{\gamma_2}{\gamma_1} \left[ \rho \int_0^1 J_c(s) dV_1(s) + \sqrt{1 - \rho^2} \int_0^1 J_c(s) dV_2(s) \right] \equiv \Lambda, \]

where \( V_1(r) \) and \( V_2(r) \) are independent standard Brownian motions, \( J_c(r) = \exp(cr) \int_0^r \exp(-cs) dV_1(s) \) is an Ornstein-Uhlenbeck process, \( \rho^2 = \frac{[b_{21}^{(0)}]^2 \sigma_1^2}{[b_{21}^{(0)}]^2 \sigma_1^2 + \sigma_2^2} \) is the squared correlation between \( \varepsilon_t \) and \( u_{2,t} \), and \( \frac{\gamma_2}{\gamma_1} \) is the ratio of the standard deviations of \( \varepsilon_t \) and \( u_{2,t} \), respectively.

Remark: In the presence of deterministic components, the asymptotic distribution has to be modified only slightly by replacing \( J_c(r) \) with its demeaned or detrended version.
Summary of results in Theorem 1:

- the IV estimator of $b_{12}^{(0)}$ is inconsistent. Cause of the inconsistency: $h_{t-1}$ is a weak instrument in a sense that it provides very little information about the endogenous variable $\triangle h_t$.

- the limiting distribution of $\hat{b}_{12}^{(0)} - b_{12}^{(0)}$ is nonstandard
  - the numerator is a mixture of a Gaussian random variable and a functional of an Ornstein-Uhlenbeck process
  - the denominator is also a random variable that involves functionals of the Ornstein-Uhlenbeck process
  - as $c \to 0$, it is a ratio of a standard normal and Dickey-Fuller random variables
\textbf{Theorem 2.} Under the model assumptions, 

\[ \left( \hat{b}_{21}^{(0)} - b_{21}^{(0)} \right) \Rightarrow \frac{-\Lambda M_2}{M_1 - \Lambda M_3} \]

\[ T \left( \hat{b}_{22}^{(1)} - b_{22}^{(1)} \right) \Rightarrow \frac{\sigma_2}{\gamma_1 \left( \int_0^1 J_c(s)^2 ds \right)^{1/2}} \left( \omega \frac{\int_0^1 J_c(s) dV_1(s)}{\left( \int_0^1 J_c(s)^2 ds \right)^{1/2}} + \sqrt{1 - \omega^2 \xi} \right), \]

where \( \Lambda \) is the limiting distribution of \( \left( \hat{b}_{12}^{(0)} - b_{12}^{(0)} \right) \) in Theorem 1, \( M_1 = E(u_1,t\varepsilon_t^z) = \frac{\sigma_1^2}{1 - b_{12}^{(0)} b_{21}^{(0)}} \), \( M_2 = E(u_2,t\varepsilon_t^d) = \frac{\sigma_2^2}{1 - b_{12}^{(0)} b_{21}^{(0)}} \), and \( M_3 = E(u_1,t u_2,t) = \frac{(b_{21}^{(0)} \sigma_1^2 + b_{12}^{(0)} \sigma_2^2)}{(1 - b_{12}^{(0)} b_{21}^{(0)})^2} \) are fixed matrices, \( \omega^2 \) is the squared correlation between \( \varepsilon_t^d \) and \( u_{2,t} \) and \( \xi \sim N(0,1) \).
Summary of results in Theorem 2:

- the IV estimator of $b_{21}^{(0)}$ is inconsistent and its limiting distribution is a function of the limiting distribution for $\hat{b}_{12}^{(0)}$

- the parameter on the persistent variable $b_{22}^{(1)} = (1 + c/T)$ can be consistently estimated (rate $T$-consistent) and its asymptotic distribution is a mixture of a Dickey-Fuller and standard normal distributions.
Summary of results (Theorem 3) for the structural IRF of hours worked at time \( t + l \) to a technology shock at time \( t \)

\[
\hat{\theta}_h(z)^{(l)} = \frac{\partial h_{t+l}}{\partial \zeta_t} = \frac{\hat{b}_{21}^{(0)} \left[ \hat{b}_{22}^{(1)} \right]^l}{1 - \hat{b}_{12}^{(0)} \hat{b}_{21}^{(0)}}
\]

- estimator of the impulse response function is inconsistent

- the inconsistency of the IRF estimators is driven by the presence of a weak instrument

- the weak instrument problem arises from the estimation of \( b_{12}^{(0)} \) and the inconsistency of this parameter contaminates the estimation of \( b_{21}^{(0)} \) and \( \theta_h(z)^{(l)} \).
Additional Remarks

• Theorems 1, 2 and 3 formalize and extend some of the results derived by Christiano, Eichenbaum and Vigfusson (2003) and Pagan and Fry (2005).

• In a different context, Hahn, Kuersteiner and Hausman (2001) and Han and Phillips (2005) study the weak identification problem that arises from instrumenting the first difference of a highly persistent variable with its lagged level for dynamic panel data models.
4.2 Improved Inference

- under the long-run identification scheme, $B(I)$ is lower triangular and $b_{12}^{(0)} = \psi_{12}(1)/\psi_{22}(1)$

- therefore, we can estimate $b_{12}^{(0)}$ as $\bar{b}_{12}^{(0)} = \hat{\psi}_{12}(1)/\hat{\psi}_{22}(1)$, where $\hat{\psi}_{12}(1)$ and $\hat{\psi}_{22}(1)$ are consistent estimators of $\psi_{12}(1)$ and $\psi_{22}(1)$ from the reduced form model.

- in the first-order model, $b_{12}^{(0)} = 0$. This restriction is testable (Theorem 4) using the $t$-test which is distributed as a mixture of a Dickey-Fuller and a normal random variables (similar to Cavanagh, Elliott and Stock, 1995; Hansen, 1995)

- by imposing $b_{12}^{(0)} = 0$, the source of the inconsistency of the IRF is removed
4.2.1 Main results for the new method

**Theorem 5.** If the restriction $b_{12}^{(0)} = 0$ is imposed,

$$\sqrt{T} \left( \tilde{b}_{21}^{(0)} - b_{21}^{(0)} \right) \xrightarrow{d} N(0, \sigma_1^{-2} \sigma_2^2),$$

$$T \left( \tilde{b}_{22}^{(1)} - b_{22}^{(1)} \right) \Rightarrow \Upsilon,$$

$$\sqrt{T} \left( \tilde{\theta}_{h2}^{(l)} - \theta_{h2}^{(l)} \right) \xrightarrow{d} N(0, \sigma_1^{-2} \sigma_2^2)$$

where $l$ is assumed fixed and $\Upsilon$ is the limiting distribution of $T \left( \tilde{b}_{22}^{(1)} - b_{22}^{(1)} \right)$ in Theorem 2.
Comments

- \( b_{21}^{(0)} \) is consistently estimable and normally distributed

- the estimator of \( b_{22}^{(1)} \) is still \( T \)-consistent with the same limiting distribution

- if \( l \) is fixed, the estimator of \( \theta_{h,z}^{(l)} \) is root-\( T \) consistent and normally distributed

- if \( l = [\lambda T] \), the distribution of the IRF estimator is similar to the one in Phillips (1998)
Extensions

- Higher-order models

- Bootstrap-based inference
  - bootstrap is inconsistent for the unrestricted estimator and consistent for the restricted estimator
  - in practice, we use Kilian’s (1998) bootstrap-after-bootstrap method

- Non-diagonal $\Phi_T$: Recall that if $\Phi_T = \begin{bmatrix} 1 & 0 \\ 0 & 1 + c/T \end{bmatrix}$, $b_{12}^{(0)} = \frac{\psi_{12}(1)}{\psi_{22}(1)}$.

Now let $\Phi_T = \begin{bmatrix} 1 & c/T \\ 0 & 1 + c/T \end{bmatrix}$. In this case, $b_{12}^{(0)} = \frac{\psi_{11}(1)+\psi_{12}(1)}{\psi_{21}(1)+\psi_{22}(1)}$. 
5 Simulation Experiment

- Model 3: data are generated from the VAR(2) model

\[
\begin{bmatrix}
I - \begin{pmatrix} -0.05 & 0.08 \\ 0.2 & 0.5 \end{pmatrix} \\
I - \begin{pmatrix} 1 & 0 \\ 0 & (1 + c/T) \end{pmatrix}
\end{bmatrix}
\begin{bmatrix}
L
\end{bmatrix}
\begin{bmatrix}
l_t \\
h_t
\end{bmatrix}
= 
\begin{bmatrix}
u_{1,t} \\
u_{2,t}
\end{bmatrix},
\]

where \( \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.78 & 0.1 \\ 0.1 & 0.55 \end{pmatrix} \right) \) with \( c=-5, \ T=250, \ 1000 \)

- Model 4: samples of 230 observations are simulated using the estimated coefficient and covariance matrices and the bootstrapped residuals from a VAR(4) model of U.S. labour productivity growth and hours worked

- estimation is performed on the demeaned \( \triangle l_t \) and \( h_t \) with 10,000 Monte Carlo replications
$c = -5; T = 250; \text{ solid: true IR; long dashes: median IR; short dashes: 95\% CI}$
$c = -5$; $T = 1000$; solid: true IR; long dashes: median IR; short dashes: 95% CI
Model 4; solid: true IR; long dashes: median IR; short dashes: 95% CI
Model 4; solid: true IRF; long dashes: IRF estimated with both variables in first differences; short dashes: 95% CI
Coverage of bootstrap 95% CIs for unrestricted and restricted IRF estimators
6 Empirical Example

- studies the effect of a positive technology shock on per capita hours worked

- U.S. data on labour productivity, hours worked in the non-farm business sector and population over the age of 16 from DRI Basic Economics

- period 1948Q1-2005Q3

- estimated VAR(4) model on the demeaned variables
Response of hours worked to a 1% positive shock in technology.
Left: conventional IRF estimator; Right: restricted IRF estimator.
7 Summary of Results

- formal results for the inconsistency of the IRF estimator in a SVAR model when one of the processes is local-to-unity

- the source of inconsistency (weak instrument) can be removed by imposing restrictions on the problematic parameter

- this is achieved by combining the long-run economic restrictions with additional statistical restrictions on the long-run reduced form VAR matrix

- the new identification scheme delivers a consistent IRF estimator which is asymptotically normally distributed and has excellent numerical properties
8 Concluding Remarks

- one has to proceed with caution when using long-run identifying restrictions in the presence of highly persistent variables

- short-run identification scheme does not suffer from the problems discussed in the paper

- imposing long-run restrictions through cointegrating relationships

- additional work is needed on optimal choice and validity of instruments, overidentifying restrictions and specification tests in SVAR models