Forecasts and VARs

A. Intro to VARs
B. Nonorthogonalized IRF
C. Standard errors
D. Jordà local projections
E. Unit roots
F. Instability

A. Intro to VARs
Suppose we want to forecast $y_{1t}$ based on:

$$y_{1,t-1}, y_{1,t-2}, \ldots, y_{1,t-p}$$
$$y_{2,t-1}, y_{2,t-2}, \ldots, y_{2,t-p}$$
$$\vdots$$
$$y_{n,t-1}, y_{n,t-2}, \ldots, y_{n,t-p}$$

deterministic functions of $t$

$(1, t, \cos(\pi t/6), \text{seasonal dummies})$

Let $y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$

$$(n \times 1)$$

$x_t = (1, y'_{t-1}, y'_{t-2}, \ldots, y'_{t-p})'$

$$(k \times 1)$$

$k = np + 1$

Suppose we consider linear forecast

$$\hat{y}_{1|t-1} = \gamma'_1 x_t$$

Best forecast within linear class:
value of $\gamma_1$ that minimizes

$$E(y_{1t} - \gamma'_1 x_t)^2$$

Proposition: If $y_t$ is covariance-stationary and $E(x_t x'_t)$ is nonsingular, then optimal forecast uses

$$\gamma_1^* = E(x_t x'_t)^{-1} E(x_t y_t)$$

Definition: The optimal linear forecast,

$$\hat{y}_{1|t-1} = \gamma'_1 x_t,$$

is called the “population linear projection” of $y_{1t}$ on $x_t$. 
Definition: Ordinary least squares (OLS) estimate is given by
\[
\hat{\gamma}_1 = \left( \sum_{t=1}^{T} x_t x_t' \right)^{-1} \left( \sum_{t=1}^{T} x_t y_{1t} \right)
\]

Proposition: If \( y_t \) is stationary and ergodic, then
\[
\hat{\gamma}_1 \xrightarrow{p} \gamma^*_1
\]

Proof: (Law of Large Numbers)
\[
\hat{\gamma}_1 = \left( T^{-1} \sum_{t=1}^{T} x_t x_t' \right)^{-1} \left( T^{-1} \sum_{t=1}^{T} x_t y_{1t} \right)
\]
\[
\xrightarrow{p} E(x_t x_t')^{-1} E(x_t y_{1t})
\]

If form separate forecasting equation for each element of \( y_t \) and collect in vector,
\[
y_{1t} = \gamma'_1 x_t + \epsilon_{1t}
\]
\[
\vdots
\]
\[
y_{nt} = \gamma'_n x_t + \epsilon_{nt}
\]
\[
y_t = \Gamma x_t + \epsilon_t
\]
result is called vector autoregression:
\[
y_t = \mathbf{c} + \mathbf{\Phi}_1 y_{t-1} + \mathbf{\Phi}_2 y_{t-2} + \cdots + \mathbf{\Phi}_p y_{t-p} + \epsilon_t
\]

Above results imply we can consistently estimate coefficients for VAR by OLS equation by equation
\[
\hat{\gamma}'_1 = \left( \sum_{t=1}^{T} y_{1t} x_t' \right) \left( \sum_{t=1}^{T} x_t x_t' \right)^{-1}
\]
\[
\vdots
\]
\[
\hat{\gamma}'_n = \left( \sum_{t=1}^{T} y_{nt} x_t' \right) \left( \sum_{t=1}^{T} x_t x_t' \right)^{-1}
\]
\[
\hat{\gamma}' = \left[ \hat{\mathbf{c}} \ \hat{\mathbf{\Phi}}_1 \ \hat{\mathbf{\Phi}}_2 \ \cdots \ \hat{\mathbf{\Phi}}_p \right]
\]
Example

\[ y_1t = 400 \times \text{quarterly log change in real GDP} \]
\[ y_2t = 400 \times \text{quarterly log change in PCE deflator} \]
\[ y_3t = \text{average fed funds rate over quarter} \]

- Estimate with 4 lags on each variable for 1960:Q1 to 1990:Q4.
- Data and code to replicate provided at course webpage.
- Sample code shows how to compare 4 versus 5 lags using hypothesis tests, AIC, or BIC.
- See Lütkepohl, Section 4.3 for description.

### VAR/System - Estimation by Least Squares

Quarterly Data From 1960:Q1 To 1990:Q4

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>GDPCH</th>
<th>Std Error</th>
<th>T-Stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDPCH(1)</td>
<td>0.0146839259</td>
<td>0.0663945316</td>
<td>2.2227</td>
</tr>
<tr>
<td>GDPCH(2)</td>
<td>-0.0276234562</td>
<td>0.0094523564</td>
<td>-3.0202</td>
</tr>
<tr>
<td>GDPCH(3)</td>
<td>0.1284856954</td>
<td>0.0563847562</td>
<td>2.5234</td>
</tr>
<tr>
<td>GDPCH(4)</td>
<td>0.5146823452</td>
<td>0.0139201234</td>
<td>4.5223</td>
</tr>
<tr>
<td>GDPCH(5)</td>
<td>0.0023456789</td>
<td>0.0345678901</td>
<td>0.6778</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>INFLATION</th>
<th>Std Error</th>
<th>T-Stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFLATION(1)</td>
<td>0.6545678901</td>
<td>0.0789012345</td>
<td>8.5234</td>
</tr>
<tr>
<td>INFLATION(2)</td>
<td>0.2145678901</td>
<td>0.0345678901</td>
<td>6.2234</td>
</tr>
<tr>
<td>INFLATION(3)</td>
<td>0.0234567890</td>
<td>0.0123456789</td>
<td>2.3234</td>
</tr>
<tr>
<td>INFLATION(4)</td>
<td>0.1234567890</td>
<td>0.0345678901</td>
<td>3.2323</td>
</tr>
<tr>
<td>INFLATION(5)</td>
<td>0.0345678901</td>
<td>0.0123456789</td>
<td>3.4567</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>FEDFUNDS</th>
<th>Std Error</th>
<th>T-Stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEDFUNDS(1)</td>
<td>0.1234567890</td>
<td>0.0345678901</td>
<td>3.4567</td>
</tr>
<tr>
<td>FEDFUNDS(2)</td>
<td>0.0345678901</td>
<td>0.0123456789</td>
<td>3.4567</td>
</tr>
<tr>
<td>FEDFUNDS(3)</td>
<td>0.0123456789</td>
<td>0.0345678901</td>
<td>3.4567</td>
</tr>
<tr>
<td>FEDFUNDS(4)</td>
<td>0.0345678901</td>
<td>0.0123456789</td>
<td>3.4567</td>
</tr>
<tr>
<td>FEDFUNDS(5)</td>
<td>0.0123456789</td>
<td>0.0345678901</td>
<td>3.4567</td>
</tr>
</tbody>
</table>

**Dependent Variable: GDPCH**

Mean of Dependent Variable: 4.12345678901

**Dependent Variable: INFLATION**

Mean of Dependent Variable: 2.12345678901

**Dependent Variable: FEDFUNDS**

Mean of Dependent Variable: 7.12345678901
B. Nonorthogonalized IRF

\[ y_{t+1} = c + \Phi_1 y_t + \Phi_2 y_{t-1} + \cdots + \Phi_p y_{t-p+1} + \varepsilon_{t+1} \]
\[ y_{t+2} = c + \Phi_1 y_{t+1} + \Phi_2 y_t + \cdots + \Phi_p y_{t-p+2} + \varepsilon_{t+2} \]
\[ y_{t+2} = c + \Phi_1 [c + \Phi_2 y_t + \cdots + \Phi_p y_{t-p+1} + \varepsilon_{t+1}] + \Phi_2 y_t + \cdots + \Phi_p y_{t-p+2} + \varepsilon_{t+2} \]
\[ = c_2 + \Psi_0 y_{t+2} + \Psi_1 y_{t+1} + \Psi_2 y_t + H_22 y_{t-1} + \cdots + H_2 p y_{t-p+1} \]

\[ \Psi_0 = I_n \]
\[ \Psi_1 = \Phi_1 \]
\[ \Psi_2 = \Phi_1^2 + \Phi_2 \]

\[ \Psi_s = \Phi_1 \Psi_{s-1} + \Phi_2 \Psi_{s-2} + \cdots + \Phi_p \Psi_{s-p} \]

Column \( j \) of \( \Psi_s \) is answer to the question: How does my forecast of \( y_{t+g} \) change if I increase \( y_t \) by one unit holding all other elements of \( y_t \) and all elements of \( y_{t-1}, \ldots, y_{t-p} \) constant.

The sequence of \( n \times n \) matrices \( \{\Psi_s\}_{s=0,1,2,...} \) is called the nonorthogonalized impulse-response function.

\[ y_{t+2} = c_2 + \Psi_0 y_{t+2} + \Psi_1 y_{t+1} + \Psi_2 y_t + H_22 y_{t-1} + \cdots + H_2 p y_{t-p+1} \]

We know that \( \varepsilon_{t+1} \) is uncorrelated with \( y_{t+1}, \ldots, y_{t-p+1} \) by definition of the plim.

If VAR has enough lags, \( \varepsilon_{t+2} \) is also uncorrelated with \( y_{t+1}, \ldots, y_{t-p+1} \).

\[ \frac{\zeta_{t+2}}{\zeta_t} = \Psi_2 \]
C. Standard errors

Generate standard errors using Bayesian posterior distribution based on diffuse priors.

\[ y_i = (y_{1i}, y_{2i}, \ldots, y_{ni})' \]
\[ x_i = (1, y_{i-1}, y_{i-2}, \ldots, y_{i-p})' \]
\[ k = np + 1 \]
\[ y_i = \Gamma' x_i + \varepsilon_i \]
\[ E(\varepsilon_i, \varepsilon_i) = \Omega \]

\[ \Omega^{-1} y_1, \ldots, y_T \sim \text{Wishart with } T-p \text{ degrees} \]

\[ \text{Wishart}(k, H) = z_1 z_1' + \cdots + z_k z_k' \]
\[ z_i \sim N(0, H^{-1}) \]

\[ \hat{\Gamma}(\text{col}) = \left( \sum_{j=1}^{T} y_{ij} x_j' \right) \left( \sum_{j=1}^{T} x_j x_j' \right)^{-1} \]
\[ \hat{\varepsilon}_i = y_i - \hat{\Gamma}' x_i \]
\[ \hat{\Omega} = T^{-1} \sum_{i=1}^{T} \hat{\varepsilon}_i \hat{\varepsilon}_i' \]

\[ \text{vec}(\Gamma) | \Omega, y_1, \ldots, y_T \sim \mathcal{N} \left( \text{vec}[\hat{\Gamma}], \Omega \otimes \left[ \sum_{j=1}^{T} x_j x_j' \right]^{-1} \right) \]

(1) Draw \( \Omega^{(m)} \) and \( \Gamma^{(m)} \) for this distribution
(2) For each \( m = 1, \ldots, 10^4 \), calculate \( \psi^{(m)} \)
(3) For each \( i, j, s \) find 95% interval for row \( i \) col \( j \) element of this matrix
D. Jordà local projections

As noted by Jordà (2005), we can also estimate forecast without imposing VAR structure.

\[ y_{t+s} = c_t + \Psi_1 y_t + H_{12} y_{t-1} + \cdots + H_{s1} y_{t-s+1} + u_{t+s} \]

Estimate by \( n \) different regressions separately for each \( s \).

Resulting \( \hat{\Psi}_s \) is direct estimate of nonorthogonalized IRF.

E. Unit roots

In exercises so far we took \( y_t \) to be growth rate of real GDP.

What if we had instead used the level of GDP without differencing?

<table>
<thead>
<tr>
<th>Growth rate regression</th>
<th>Levels regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDPCH(1)</td>
<td>0.14089</td>
</tr>
<tr>
<td>GDPCH(2)</td>
<td>0.37458</td>
</tr>
<tr>
<td>GDPCH(3)</td>
<td>0.03989</td>
</tr>
<tr>
<td>GDPCH(4)</td>
<td>0.02745</td>
</tr>
<tr>
<td>INFLATION(1)</td>
<td>-0.1362</td>
</tr>
<tr>
<td>INFLATION(2)</td>
<td>0.109624</td>
</tr>
<tr>
<td>INFLATION(3)</td>
<td>0.03572</td>
</tr>
<tr>
<td>INFLATION(4)</td>
<td>-0.00278</td>
</tr>
<tr>
<td>FEDFUNDS(1)</td>
<td>0.073564</td>
</tr>
<tr>
<td>FEDFUNDS(2)</td>
<td>-1.51527</td>
</tr>
<tr>
<td>FEDFUNDS(3)</td>
<td>1.13886</td>
</tr>
<tr>
<td>FEDFUNDS(4)</td>
<td>-0.00298</td>
</tr>
<tr>
<td>Constant</td>
<td>4.456602</td>
</tr>
</tbody>
</table>

- Local projections and VAR recursion should give similar broad picture.
- Local projections likely more volatile and may give worse forecasts (Marcellino, Stock and Watson, 2006)
Suppose the correct model would use growth rates
\[ \Delta y_{1t} = \zeta_1 \Delta y_{1,t-1} + \cdots + \zeta_p \Delta x_{1,p-1} + \beta' x_{1,t-1} + \varepsilon_{1t} \]
\[ y_{1t} = y_{1,t-1} = \xi_1 (y_{1,t-1} - y_{1,t-2}) + \cdots + \]
\[ \xi_p (y_{1,t-p} - y_{1,t-p-1}) + \beta x_{1,t-1} + \varepsilon_{1t} \]
This is a special case of regression in levels
\[ y_{1t} = \phi_1 y_{1,t-1} + \cdots + \phi_{p+1} y_{1,t-p-1} + \beta' x_{1,t-1} + \varepsilon_{1t} \]
\[ \phi_1 = 1 + \xi_1 \]
\[ \phi_2 = \xi_2 - \xi_1 \]
\[ \phi_3 = \xi_3 - \xi_2 \]
\[ \cdots \]
\[ \phi_p = \xi_p - \xi_{p-1} \]
\[ \phi_{p+1} = -\xi_p \]

This is a special case of regression in levels
\[ y_{1t} = \phi_1 y_{1,t-1} + \cdots + \phi_{p+1} y_{1,t-p-1} + \beta' x_{1,t-1} + \varepsilon_{1t} \]
OLS minimizes \( T^{-1} \sum_{t=1}^{T} \varepsilon_{1t}^2 \)
If \( y_{1t} \) has a unit root, this will be infinite unless we pick \( \phi_i \) consistent with the growth-rate specification.
In other words, OLS should force \( \phi_1 + \phi_2 + \phi_3 + \phi_4 \) close to one.
Actual OLS estimate of the sum is 1.004.

- For this example, levels regression and growth-rate regression are basically estimating the identical system
- If truth is growth-rate, when we estimate in levels we will force OLS to estimate the unit root for us
- But will have more efficient estimates if impose the unit root
- Also can avoid nonstandard distributions for hypothesis tests by using growth rates
Potential drawbacks to using growth rates
- GDP may not really have a unit root
- GDP and price level may be cointegrated
For this example, baseline specification seems sensible (growth rate of GDP, inflation rate, level of fed funds).

Other implications
- Does not usually make sense to throw in time trend (or quadratic time trend!) in levels regression because growth rates have no trend.
- A simple linear regression of level of a scalar on its own lagged levels is a robust, assumption-free way to remove unknown trend (much better than Hodrick-Prescott filter!)

Proposition: if $\Delta^d y_t$ is stationary for some $d$, then can write $y_{t+h}$ as a linear function of
$y_t, y_{t-1}, \ldots, y_{t-d+1}$ plus a stationary residual.

Example: $d = 1$
\begin{align*}
    u_t &= \Delta^2 y_t \sim I(0) \\
    y_{t+h} &= (h + 1)y_t - hy_{t-1} + u_{t+h} + 2u_{t+h-1} + \cdots + hu_{t+1} \\
    w_t^{(h)} &= u_{t+1} + u_{t+2} + \cdots + u_{t+h} \sim I(0)
\end{align*}

If $y_t \sim I(2)$, what happens if we regress $y_{t+h}$ on $(1, y_t, y_{t-1})'$?
- If coefficient on $y_t = h + 1$ and coefficient on $y_{t-1} = -h$, then average squared residual will tend to a finite number.
- For any other coefficients, average squared residual will tend to an infinite number.
- OLS will give a consistent estimate of parameters that characterize the trend.

Example: $d = 2$
\begin{align*}
    u_t &= \Delta^2 y_t \sim I(0) \\
    y_{t+h} &= (h + 1)y_t - hy_{t-1} + u_{t+h} + 2u_{t+h-1} + \cdots + hu_{t+1} \\
    w_t^{(h)} &= u_{t+1} + 2u_{t+h-1} + \cdots + hu_{t+1} \sim I(0)
\end{align*}
If $y_t \sim I(2)$, what happens if we regress $y_{t+h}$ on $(1, y_{t-1}, y_{t-2}, y_{t-3})$?

- Two of the coefficients will make the residuals stationary.
- Other two coefficients will then try to forecast stationary component.

Conclusion: we don’t need to know $d$.

If $y_t \sim I(d)$ for some unknown $d \leq 4$, the population linear projection of $y_{t+h}$ on $(1, y_{t-1}, y_{t-2}, y_{t-3})$ exists and can be consistently estimated by OLS regression.

Proposed definition: the cyclical component of $y_t$ is part we can’t predict 2 years ahead using linear regression.

For quarterly data estimate by OLS

$y_{t+8} = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + v_{t+8}$

interpret the residuals $v_{t+8}$ as the cyclical component.

F. Instability

- What happens if we estimate VAR over 1991.Q1 to 2007.Q4?

<table>
<thead>
<tr>
<th>Coef</th>
<th>Std error</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDPCH{1}</td>
<td>0.155077</td>
<td>0.143744</td>
</tr>
<tr>
<td>GDPCH{2}</td>
<td>0.185423</td>
<td>0.149189</td>
</tr>
<tr>
<td>GDPCH{3}</td>
<td>-0.10024</td>
<td>0.145688</td>
</tr>
<tr>
<td>GDPCH{4}</td>
<td>0.126178</td>
<td>0.145687</td>
</tr>
<tr>
<td>INFLATION{1}</td>
<td>-0.05773</td>
<td>0.295557</td>
</tr>
<tr>
<td>INFLATION{2}</td>
<td>-0.16592</td>
<td>0.274807</td>
</tr>
<tr>
<td>INFLATION{3}</td>
<td>-0.30131</td>
<td>0.27268</td>
</tr>
<tr>
<td>INFLATION{4}</td>
<td>0.057866</td>
<td>0.278822</td>
</tr>
<tr>
<td>FEDFUNDS{1}</td>
<td>-0.55891</td>
<td>0.948064</td>
</tr>
<tr>
<td>FEDFUNDS{2}</td>
<td>1.282767</td>
<td>1.700743</td>
</tr>
<tr>
<td>FEDFUNDS{3}</td>
<td>-1.46413</td>
<td>1.667997</td>
</tr>
<tr>
<td>FEDFUNDS{4}</td>
<td>0.670807</td>
<td>0.868767</td>
</tr>
<tr>
<td>Constant</td>
<td>3.253064</td>
<td>1.496104</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coef</th>
<th>Std error</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDPCH{1}</td>
<td>0.135977</td>
<td>0.143744</td>
</tr>
<tr>
<td>GDPCH{2}</td>
<td>0.164372</td>
<td>0.166859</td>
</tr>
<tr>
<td>GDPCH{3}</td>
<td>0.16824</td>
<td>0.161688</td>
</tr>
<tr>
<td>GDPCH{4}</td>
<td>0.13078</td>
<td>0.158897</td>
</tr>
<tr>
<td>INFLATION{1}</td>
<td>-0.05773</td>
<td>0.295557</td>
</tr>
<tr>
<td>INFLATION{2}</td>
<td>-0.16592</td>
<td>0.274807</td>
</tr>
<tr>
<td>INFLATION{3}</td>
<td>-0.30131</td>
<td>0.27268</td>
</tr>
<tr>
<td>INFLATION{4}</td>
<td>0.057866</td>
<td>0.278822</td>
</tr>
<tr>
<td>FEDFUNDS{1}</td>
<td>-0.55891</td>
<td>0.948064</td>
</tr>
<tr>
<td>FEDFUNDS{2}</td>
<td>1.282767</td>
<td>1.700743</td>
</tr>
<tr>
<td>FEDFUNDS{3}</td>
<td>-1.46413</td>
<td>1.667997</td>
</tr>
<tr>
<td>FEDFUNDS{4}</td>
<td>0.670807</td>
<td>0.868767</td>
</tr>
<tr>
<td>Constant</td>
<td>3.253064</td>
<td>1.496104</td>
</tr>
</tbody>
</table>
Options for dealing with instability

- Estimate allowing for GARCH to reduce impact of outliers (Hamilton, 2010)
- Find generalization of model that is stable
- Use Bayesian methods to bring in additional information
- Estimate system with time-varying parameters or changes in regime
- Use full sample as average summary (plim of regression)