Advanced Macroeconomics: Structural Vector Autoregressive Analysis

Lectures 16-17

U. Michael Bergman
University of Copenhagen

Fall 2024

CURRICULUM:

KL 12.1-12.2, 12.4-12.6 and 12.13

These slides cover:

- Inference on impulse responses and forecast error variance decomposition.
- Focus: Bootstrap methods
  - Stationary VARs.
  - Non-stationary VARs.
  - Known pitfalls and finite-sample properties.
- Bias-adjusted estimators.
- Other practical issues.

How to construct confidence bands for impulse response functions and forecast error variance decompositions?

- Use asymptotic distributions (derived in Lütkepohl, 2005 and discussed in KL 12.1.1-12.1.2).
- Alternative is to use Bootstrap methods.
- Why not use asymptotic distribution?
  - Asymptotic distribution depends on whether or not the lag length is known and finite.
  - We need to choose between these two choices.
  - Bootstrap methods do not require us to choose.
  - Simulations suggest that confidence bands based on standard bootstrap methods are asymptotically valid for both cases (for VARMA if a suitable large lag length is chosen).
  - Small sample sizes in most applications, asymptotic distribution may not be a good approximation.
  - VAR residuals are not Gaussian in practice, an assumption used when deriving asymptotic distribution.
- We will use bootstrap methods to construct confidence intervals for impulse responses and variance decompositions.

BOOTSTRAP INTERVALS FOR STRUCTURAL IMPULSE RESPONSES

- Basic idea behind bootstrap methods:
  - Approximate the distribution of the statistic based on the DGP.
  - Bootstrap methods do not require residuals to be Gaussian iid.
  - Suitably constructed bootstrap confidence intervals have been shown to be more accurate in small samples than asymptotic approximations.
- Note: Standard caveats apply to bootstrap methods
  - May not be applicable at all.
  - Must be used carefully, can produce less accurate inference.
  - Only use bootstrap methods when they are asymptotically justified.
- All bootstrap methods to be discussed can be used to construct confidence intervals for impulse responses, cumulative impulse responses, linear combinations of impulse responses, and corresponding forecast error variance decompositions.
- First discuss bootstrap methods, then turn to the question how to construct confidence bands and finally we cover the issue of significance level.

Alternative bootstrap methods

- General idea: Simulate the VAR model using either estimated residuals or regressands and regressors.
- Several methods suggested in the literature
  - Standard residual-based recursive-design bootstrap.
  - Standard residual-based fixed-design bootstrap.
  - Residual-based wild bootstrap.
  - Residual-based block bootstrap.
  - Residual-based block bootstrap in conditionally heteroscedastic VAR models.
  - Bootstrapping tuples of regressands and regressors.
- We will suggest one particular approach!

- Runkle (1987) was the first study suggesting bootstrap method to construct confidence intervals for impulse responses.
- Asymptotic closed-form solution unknown at the time. First study to establish asymptotic validity of bootstrap came a year later (Bose, 1988).
- Assume that the DGP is a VAR (p) model (lag order p is known)

$$y_t = \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t$$

- Basic assumptions: $\mathbb{E}(u_t) = 0$ and that u_t has finite moments.
- Estimate the VAR(p) model to obtain estimates of ν, $A_1, \ldots, A_p$ ($\hat{\nu}$, $\hat{A}_1, \ldots, \hat{A}_p$) and Σ_u ($\hat{\Sigma}_u$). And compute the statistic of interest, for example the impulse response function for an exactly identified structural VAR (using any of the approaches discussed so far).

Standard residual-based recursive-design bootstrap

- To construct a bootstrap sample we:
  - Draw u_t^* with replacement from the estimated residuals $\hat{u}_t$.
  - If $\nu = 0$, demean u_t^*.
  - Given initial conditions $\left[y_{-p+1}^*, \ldots, y_0^*\right]$ generate recursively the bootstrap realizations $\{y_t^*\}_{t=-p+1}^{T}$.
- Re-estimate the VAR(p) model and compute the impulse response function (computing B_0^{-1} again). Save the impulse response function!
- Repeat the procedure R times (construct a new bootstrap sample, re-estimate the VAR(p) model, compute B_0^{-1} and compute the impulse response function).
- We have then generated R estimates of the impulse response function. These can then be used to construct confidence intervals.

STANDARD RESIDUAL-BASED RECURSIVE-DESIGN BOOTSTRAP

- How to treat initial conditions?
- Three alternatives:
  - Use fixed initial conditions (for example sample values) $\left[y_{-p+1}, \ldots, y_0\right]$ for all $r = 1, \ldots, R$ replications. Resulting sampling uncertainty will be understated but will be unimportant in stationary models as $T \to \infty$.
  - Draw initial conditions $\left[y^*_{-p+1}, \ldots, y^*_0\right]$ with replacement as a block of p consecutive vector valued observations from the observed data $\{y_t\}_{t=-p+1}^{T}$ for each bootstrap replication $r = 1, \ldots, R$. Ensures that the initial conditions are drawn from the same distribution as the remaining observations. This is the preferred approach!
  - Set initial conditions $\left[y^*_{-p+1}, \ldots, y^*_0\right]$ equal to an arbitrary value, for example the average. In this case we need to generate $\tau + T$ observations and discard the first τ observations to avoid that subsequent observations depend on the initial condition. Problem: How to determine τ?
- All other residual-based bootstrap methods are variants of this (standard residual-based recursive-design bootstrap) method.

Illustration

- KL example pp. 337 for univariate AR model: How to draw residuals with replacement.
- MATLAB example: Illustrate.m and IllustrateSimulation.m Python example: Illustrate.py and IllustrateSimulation.py

Random numbers from uniform distribution

- Draw Tbig numbers from a uniform distribution
  - Matlab: Draw Tbig random numbers from uniform distribution between 1 and Tbig: indexstar = randi([1 Tbig],1,Tbig)
  - Python: Draw Tbig random numbers from uniform distribution between 1 and Tbig: indexstar = np.random.randint(0, Tbig, size=Tbig)
- Draw rows of residuals with replacement.
  - Matlab: ur = uhat(indexstar,:)
  - Python: ur = uhat[indexstar,:]
- Illustrate.m and Illustrate.py

Prepare initial conditions

- We allow for two different approaches: Either use the first p observations as initial conditions, or draw p consecutive observations from the data.
  - Matlab: boot = 0: use actual initial conditions or boot = 1: use a sequence of p consecutive initial conditions drawn from a uniform distribution
    ```
    if boot==0
    y0p = y(1:p,1:K);
    else
    pos=randi([p+1,t]); % Draw the largest position for initial condition minimum=p+1 and maximum=t
    y0p = y(pos-p+1:pos,:); % Define initial conditions starting with pos and then add previous p-1 values
    end
    ```
  - Python:
    ```
    if boot == 0:
    y0p = y[:p-1, :K-1] # in Python, indexing starts from 0
    else:
    pos = random.randint(p+1, t) # in Python, randint's upper limit is inclusive
    y0p = y[(pos-p+1):(pos+1), :] # add 1 to pos for Python's exclusive upper limit in slicing
    ```

HOW TO GENERATE *yr*

- In the Matlab/Python codes we use a loop to generate the simulated data contained in *yr*.
- To do this efficiently we construct an index for the lags in *yr* and use the Matlab/Python instruction "flip" to turn the index upside down.
- Illustration: Assume that $K = 3$, $p = 2$ and $Tbig = 5$.
- This implies that our total sample contains 7 observations (note that we take the transpose of the time series vector). This implies that *yr* has K rows and 7 columns.
- Since we have estimated a VAR(2) model we have $\hat{A}_1$ and $\hat{A}_2$ and we then define $a = \begin{bmatrix} A_1 & A_2 \end{bmatrix}$.

HOW TO GENERATE *yr*

- To generate the simulated data given initial conditions we need to compute (abstracting from deterministic components)

$$yr_t = \begin{bmatrix} \hat{A}_1 & \hat{A}_2 \end{bmatrix} \begin{bmatrix} yr_{t-1} \\ yr_{t-2} \end{bmatrix} + ur_t$$

- This implies that

$$yr_3 = \begin{bmatrix} \hat{A}_1 & \hat{A}_2 \end{bmatrix} \begin{bmatrix} yr_2 \\ yr_1 \end{bmatrix} + ur_3$$

$$yr_4 = \begin{bmatrix} \hat{A}_1 & \hat{A}_2 \end{bmatrix} \begin{bmatrix} yr_3 \\ yr_2 \end{bmatrix} + ur_4$$

$$\vdots$$

$$yr_7 = \begin{bmatrix} \hat{A}_1 & \hat{A}_2 \end{bmatrix} \begin{bmatrix} yr_6 \\ yr_5 \end{bmatrix} + ur_7$$

- Therefore we need an index to generate the sequence

$$\begin{bmatrix} 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

How to generate *yr*

- To do this we use $j : j + p - 1$ for $j = 1, ..., 5$. For $p = 2$ we then obtain

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \end{bmatrix}$$

so we need to use the command “flip” to produce

$$\begin{bmatrix} 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

- In Matlab we do this using a range and then apply flip: index = flip((j:j+p-1));
- In Python we do this using a range and then apply flip: index = np.flip(np.arange(j, j+p))
- We can then use this index (representing rows in *yr*) to define the lags in the VAR (what we call *ylags* in the example code). In Matlab:
ylags = vec(yr(:,[index]));
and in Python:
ylags = vec(yr[:, index])
ylags = np.reshape(ylags, (K,p))
ylags = vec(ylags.T)

How to generate *yr*

- First we use the index to extract the corresponding observations in yr. To compute yr_3 we need observations for $t-1$ (row 2) and $t-2$ (row 1) corresponding to column 1 in the index, i.e.,

$$\begin{bmatrix} yr_{1,t-1} & yr_{1,t-2} \\ yr_{2,t-1} & yr_{2,t-2} \\ yr_{3,t-1} & yr_{3,t-2} \end{bmatrix}$$

and then we vectorize (column-wise using the vec function)

$$\begin{bmatrix} yr_{1,t-1} \\ yr_{2,t-1} \\ yr_{3,t-1} \\ yr_{1,t-2} \\ yr_{2,t-2} \\ yr_{3,t-2} \end{bmatrix}$$

and finally we premultiply with a (the corresponding coefficients)

$$\begin{bmatrix} a_{1,11} & a_{1,12} & a_{1,13} & a_{2,11} & a_{2,12} & a_{2,13} \\ a_{1,21} & a_{1,22} & a_{1,23} & a_{2,21} & a_{2,22} & a_{2,23} \\ a_{1,31} & a_{1,32} & a_{1,33} & a_{2,31} & a_{2,32} & a_{2,33} \end{bmatrix}$$

to obtain

$$yr_{1,t} = a_{1,11} \times yr_{1,t-1} + a_{1,12} \times yr_{2,t-1} + \ldots + a_{2,13} \times yr_{3,t-2}$$

and similar expressions for the remaining two variables.

How to generate *yr*

- IllustrateSimulation.m and IllustrateSimulation.py

Standard residual-based fixed-design bootstrap

- Same method as above except that we are holding the regressor matrix (the independent variables) fixed.
- Draw u_t^* with replacement from the estimated residuals $\hat{u}_t$ as above.
- If $\nu = 0$, demean u_t^*.
- Given initial conditions $\left[y_{-p+1}^*, \ldots, y_0^*\right]$ and the fixed regressor matrix generate the bootstrap realizations $\{y_t^*\}_{t=-p+1}^{T}$.
- Then continue as above.
- This approach is less accurate!

Residual-based wild bootstrap

- Basic assumption underlying the two previous methods is that u_t is iid.
- Very often conditional heteroscedasticity in the residuals.
- To handle this, use the same approach as above, but multiply each element of u_t^* with a scalar drawn from an auxiliary distribution, for example the normal distribution with mean zero and variance equal to 1.
- Other distributions may also be considered (no differences in practice).
- This approach is preferred to modeling the heteroscedasticity!
- Confidence bands will be too narrow for short horizons.

Allow for both types of bootstrap

- Matlab: Define BS_type = 0 if standard bootstrap and BS_type == 1 for wild bootstrap:

```
if BS_type==0
% Non-parametric BS
yr(:,i)=V*determ(i,:)'+a*ylags+ur(:,i);
else
% If Wild Gaussian BS, multiply residuals with random number
yr(:,i)=V*determ(i,:)'+a*ylags+randn*ur(:,i);
end
```

- Python: Define BS_type = 0 if standard bootstrap and BS_type == 1 for wild bootstrap:

```
if BS_type == 0
# Non-parametric BS
yr[:, i] = (V @ determ[i, :].reshape(-1, 1) + a @ ylags + ur[:, i].reshape(-1,1)).ravel()
else:
# If Wild Gaussian BS, multiply residuals with random number
yr[:, i] = (V @ determ[i, :].reshape(-1, 1) + a @ ylags + np.random.randn() * ur[:,
i].reshape(-1,1)).ravel() # np.random.randn() generates a random number from a
standard normal distribution
```

Residual-based block bootstrap

- In the residual-based recursive-design bootstrap method we draw one residual at a time.
- Heteroscedasticity and autocorrelation in the uniformly drawn residuals will disappear even if it appears in the estimated residuals.
- To handle this, we can draw, with replacement, blocks of residuals.
- Draw l consecutive residuals with replacement and possibly overlapping and concatenate these. The number of blocks $s = T - l + 1$.
- Proceed as above and recursively generate the variables in the VAR model.
- Note: Uncertain properties, no derived asymptotic properties.

Residual-based block bootstrap in conditionally heteroscedastic VAR models but no autocorrelation

- Combine the residual-block bootstrap with a transformation of the residuals.
- Use the residual-block bootstrap as above to generate bootstrap innovations $\tilde{u}_t^*$.
- De-mean these and construct the bootstrap innovations

$$u_{jl+i}^* = \tilde{u}_{jl+i} - \mathbb{E}^*\left(\tilde{u}_{jl+i}^*\right) = \tilde{u}_{jl+i} - \frac{1}{T-l+i}\sum_{r=0}^{T-l}\hat{u}_{i+r}$$

  where $\hat{u}_t$ is the vector of estimated residuals.
- Coverage accuracy is high in fairly large samples.

Bootstrapping tuples of regressands and regressors

- Instead of drawing residuals with replacement, we can bootstap regressors and regressands.
- Draw tuples of the y_t of size $1 \times (Kp + K + 1)$ for $t = 1, \ldots, T$ and stacking them into a matrix X, i.e., draw with replacement from $x_t \equiv \left(y_t', 1, y_{t-1}', \ldots, y_{t-p}'\right)$ to form $x_t^{*r} \equiv \left(y_t^{*r\prime}, 1, y_{t-1}^{*r\prime}, \ldots, y_{t-p}^{*r\prime}\right)$.
- Re-estimate the VAR using x_t^{*r}, compute B_0^{-1} and IRFs as above.
- Problem: When drawing with replacement we destroy time dependency in the data.
- This approach is robust to heteroscedasticity of unknown form but works only if residuals are uncorrelated.

Blocks-of-Blocks bootstrap (Naive block bootstrap)

- Solve problem with losing time dependency by drawing blocks of data.
- Draw blocks of consecutive observations from x_t above. For example, for a VAR(2) we draw $m = 4$ consecutive observations.
- The size of m and T together determine the number of blocks l.
- Stacking the l blocks of data in a matrix, we can then re-estimate the model and proceed as above.
- There are no good rule of thumbs to determine l. Need to conduct robustness tests with different l.
- Does not improve accuracy in VARs and may worsen it in small samples.

IDENTIFY THE STRUCTURAL VAR AND COMPUTE IRF'S

- Identify the structural model using any of the approaches discussed in the course and compute IRF's.
- Then prepare the bootstrap.
- Define matrix to hold results. Need to decide on number of trials (ntrials).
  Matlab:
  IRFmat=zeros(ntrials,(K*K)*(h+1));
  Python:
  IRFmat=np.zeros((ntrials,(K*K)*(h+1)), dtype=np.float64)
- Then, for each replication compute IRF's and save in IRFmat.
  Matlab:
  IRFmat(trial,:)=vec(IRFr');
- Python:
  IRFmat[trial-1,:] = np.reshape(vec(irfr.T), (1,-1))
- Then, repeat the procedure until trial=ntrials.
- Important to check that the model satisfies stability (if VAR model) and reduced rank (if VEC model).

What to do if a bootstrapped VAR model is unstable?

- We need to test whether the VAR model is stable before computing the IRF's!
- To do this:
  - Compute the companion matrix A.
  - Compute the eigenvalues of A.
  - If the largest eigenvalue (modulus) > 0.999999999999, then discard this bootstrap trial, if not, then compute IRF's
- If we have a VEC model.
  - Compute the companion matrix A.
  - Compute the eigenvalues of A.
  - If the largest eigenvalue (modulus) > 1.000000000001, then discard this bootstrap trial, if not, then compute IRF's.

Bootstrap confidence intervals

- We have now generated ntrial IRF's kept in the matrix IRFmat.
- How to compute confidence bands?
- Delta method: Assume that we have generated R bootstrap estimates of the impulse response function $\hat{\theta}^*$. We can then compute the standard deviation of these and form confidence intervals centered on the point estimate of the impulse response function

$$\hat{\theta} \pm z_{\gamma/2}\hat{\sigma}(\hat{\theta}^*)$$

  where $z_{\gamma/2}$ denotes the $\gamma/2$ quantile of the standard normal distribution.
- Matlab:

```
IRFrstd=reshape((std(IRFmat)'),K*K,h+1);
CI1LO=IRF-sign*IRFrstd';
CI1UP=IRF+sign*IRFrstd';
```

  Python:

```
irfstd = np.std(IRFmat.T, axis=1)
IRFrstd = np.reshape(irfstd, (K*K,h+1))
CI1LO=irf-sign*IRFrstd.T
CI1UP=irf+sign*IRFrstd.T
```

- Illustration: Figure 12.5 in KL VARbootTestDelta.m and VARBootTestDelta.py

Illustration: KL Figure 12.5

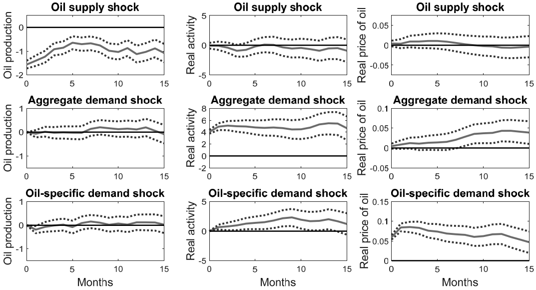


Figure 12.5: 95% delta method confidence intervals based on bootstrap standard error estimates.
Notes: All estimates were generated using the data in Figure 12.4 and the recursively identified structural VAR model in Kilian (2009).

BOOTSTRAP CONFIDENCE INTERVALS

- Efron's percentile interval: Let $\hat{\theta}^*_{\gamma/2}$ and $\hat{\theta}^*_{1-\gamma/2}$ be the critical points defined by the $\gamma/2$ and the $1-\gamma/2$ quantiles of the distribution of $\hat{\theta}^*$, then the interval is

$$\mathrm{CI}^{\mathrm{Efron}}_{\mathrm{PER}} = \left[\hat{\theta}^*_{\gamma/2}, \hat{\theta}^*_{1-\gamma/2}\right]$$

  Note: This percentile interval is valid even if the distribution of the impulse response function is not normal but in empirical applications impulse responses are far from normal and there may be small-sample bias questioning whether this approach is valid for impulse responses. Use with caution! Bias corrected Efron percentile intervals do not solve this problem.

- We will use Efron's percentile method to compute confidence bands for FEVD:s. Reason is that it is range respecting, FEVD:s cannot be negative and cannot exceed 1 (100).

Efron's percentile interval

Same code as before, but compute confidence bands using Efron's percentile method.

- Matlab:
  ```
  % Construct upper and lower confidence bands using Efron 10% level
  CI=prctile(IRFmat,[5 95]);
  CI1LO=reshape(CI(1,:),K*K,h+1)';
  CI1UP=reshape(CI(2,:),K*K,h+1)';
  ```
  Python:
  ```
  # Construct upper and lower confidence bands using Efron 10% level
  CI = np.percentile(IRFmat, [5, 95], axis=0, method='nearest')
  CI1LO = np.reshape(CI[0,:], [K*K,h+1]).T
  CI1UP = np.reshape(CI[1,:], [K*K,h+1]).T
  ```
- Illustration: VARbootTestEfron.m and VARbootTestEfron.py

- Equal-tailed percentile-*t* intervals: Let $\hat{t}^*_{\gamma/2}$ and $\hat{t}^*_{1-\gamma/2}$ be the critical points defined by the $\gamma/2$ and the $1-\gamma/2$ quantiles of the distribution of $\hat{t}^* \equiv \left(\hat{\theta}^* - \hat{\theta}\right)/\hat{\sigma}\left(\hat{\theta}^*\right)$, then

$$\mathrm{CI}_{\mathrm{PER}-t} = \left[\hat{\theta} - \hat{t}^*_{1-\gamma/2}\hat{\sigma}\left(\hat{\theta}\right), \hat{\theta} - \hat{t}^*_{\gamma/2}\hat{\sigma}\left(\hat{\theta}\right)\right]$$

  Note: Tends to be more accurate than the percentile interval, depends on the accuracy of $\hat{\sigma}(\hat{\theta})$, the estimated standard error of $\hat{\theta}$ which can be estimated by bootstrap simulation or computed analytically, this requires an additional bootstrap loop, small sample accuracy can be erratic.

BOOTSTRAP CONFIDENCE INTERVALS

- Symmetric Percentile-t intervals: Let $\hat{t}^*_{1-\gamma}$ be the critical point defined by the $1-\gamma$ quantile of the distribution of $\hat{t}^* \equiv \mid \hat{\theta}^* - \hat{\theta} \mid / \hat{\sigma}\left(\hat{\theta}^*\right)$, then

$$\mathrm{CI}_{\mathrm{PER}-t} = \left[\hat{\theta} - \hat{t}^*_{1-\gamma}\hat{\sigma}\left(\hat{\theta}\right), \hat{\theta} + \hat{t}^*_{1-\gamma}\hat{\sigma}\left(\hat{\theta}\right)\right]$$

  Note: More accurate than the equal-tailed percentile-t interval.

- Hall's percentile interval: Let $\hat{\theta}^*_{\gamma/2}$ and $\hat{\theta}^*_{1-\gamma/2}$ be the critical points defined by the $\gamma/2$ and the $1-\gamma/2$ quantiles of the distribution of $\hat{\theta}^*$, then the interval is

$$\mathrm{CI}^{\mathrm{Hall}}_{\mathrm{PER}} = \left[2\hat{\theta} - \hat{\theta}^*_{1-\gamma/2}, 2\hat{\theta} - \hat{\theta}^*_{\gamma/2}\right]$$

  Note: Poor small-sample accuracy.

BIAS-ADJUSTED BOOTSTRAP METHOD

- Suggested by Kilian in a number of papers. Idea: Correct for small-sample bias in the Runkle bootstrap.
- No consensus in the literature that IRFs should be estimated based on bias-adjusted slope parameters instead of the original LS/ML estimates.
- Replicate Figure 12.6 illustrating different bootstrap confidence bands. MATLAB example: Figure12_6.m

ILLUSTRATION: KL FIGURE 12.6

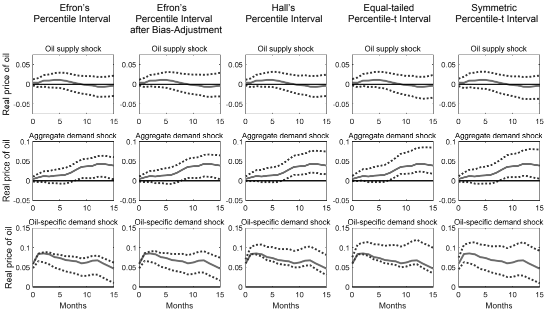


Figure 12.6: Alternative 95% bootstrap confidence intervals.
Notes: All estimates were generated using the data in Figure 12.4 and the recursively identified structural VAR model in Kilian (2009).

Bootstrap non-stationary VARs

- If $y_t \sim I(1)$, then estimate VAR in first difference and use any of the methods above.
- If $y_t \sim I(1)$ and cointegrated: Two cases depending on whether we know the cointegration vector or not.
  - If cointegration vector is known: Estimate the VEC model, rewrite as a VAR in levels and bootstrap using any method above re-estimating the VEC model using LS/ML.
  - If cointegration vector is unknown: Estimate the cointegration vector using ML in the first stage, then use the estimated cointegration vector in the second stage when estimating the VEC.
- The approaches above rest on the assumption that we know the cointegration rank. If cointegration rank is unknown, we first need to determine the rank, then condition the bootstrap on the obtained cointegration rank.
- May need to check that the bootstrap model (VAR or VEC) is stable. Discard any unstable VAR/VEC from bootstrap.

Bootstrap non-stationary VARs

Homework Assignment: Exchange rate model.

Practical issues/Lessons

- Use bootstrap methods, not asymptotic confidence bands.
- What confidence level? There is no consensus! Often use 68% confidence level ($\pm$ one standard error). Note: Need to increase the number of replications to estimate the 95% confidence band. How many replications?
- Use residual-based methods.
- Bias-adjustment: matter of taste but may not make a difference in practice.
- Do not use Hall or Efron for IRFs.
- Need to use range respecting methods when computing FEVD. Efron may be a good choice even though it by construction does not include 0 or 1.
- Next topic: Local projection KL 12.8

Advanced Macroeconomics: Structural Vector Autoregressive Analysis

Lectures 9-11

U. Michael Bergman
University of Copenhagen

Fall 2024
This version: October 24, 2024

Learning objective:

Curriculum: KL 4.1-4.2, 8.1-8.5, 9.1-9.2, Sims (1992), Eichenbaum (1992), Bergman and Worm (2021) and lecture note on stable VAR models

1. Structural VAR model.
2. Reduced form VAR to structural VAR.
3. How to impose restrictions to identify structural VAR model.
4. Recursive identification.
5. How to implement identification in Python/MATLAB.
6. Forecast Error Variance Decomposition.
7. Empirical applications: Effects of monetary policy and economic policy uncertainty.

Structural VAR models

- Consider the bivariate model

$$x_t = b_{10} z_t + b_{11,1} x_{t-1} + b_{12,1} z_{t-1} + \ldots + b_{11,p} x_{t-p} + b_{12,p} z_{t-p} + w_{1t}$$

$$z_t = b_{20} x_t + b_{21,1} x_{t-1} + b_{22,1} z_{t-1} + \ldots + b_{21,p} x_{t-p} + b_{22,p} z_{t-p} + w_{2t}.$$

  where $\mathbb{E}(w_t w_t') = I_K$.

- Note that we cannot estimate the parameters in this model directly since z_t is correlated with w_{1t} and x_t is correlated with w_{2t}.
- Or written as

$$B_0 y_t = B_1 y_{t-1} + \ldots + B_p y_{t-p} + w_t$$

- Rewrite the model in reduced form

$$y_t = \underbrace{B_0^{-1} B_1}_{A_1} y_{t-1} + \ldots + \underbrace{B_0^{-1} B_p}_{A_p} y_{t-p} + \underbrace{B_0^{-1} w_t}_{u_t}$$

  where $\mathbb{E}(u_t u_t') \equiv \Sigma_u = B_0^{-1}(B_0^{-1})'$.

- We can estimate the reduced form VAR to obtain estimates of $A_1, \ldots, A_p$ and Σ_u.
- Is it possible to recover estimates in the structural VAR $(B_0, B_1, \ldots, B_p)$?

Structural VAR models

- Consider again the residual covariance matrix

$$\Sigma_u = \mathbb{E}(u_t u_t') = B_0^{-1}\mathbb{E}(w_t w_t')(B_0^{-1})' = B_0^{-1}\Sigma_w(B_0^{-1})' = B_0^{-1}(B_0^{-1})'$$

- $\Sigma_u = B_0^{-1}(B_0^{-1})'$ is a system of nonlinear equations in the unknown parameters of B_0^{-1}.
- Let $K = 2$ for simplicity, then B_0 be written as

$$\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

with inverse: $\begin{bmatrix} \frac{b_{22}}{b_{11}b_{22}-b_{12}b_{21}} & -\frac{b_{12}}{b_{11}b_{22}-b_{12}b_{21}} \\ -\frac{b_{21}}{b_{11}b_{22}-b_{12}b_{21}} & \frac{b_{11}}{b_{11}b_{22}-b_{12}b_{21}} \end{bmatrix}$

- $B_0^{-1}(B_0^{-1})'$ is equal to

$$\frac{1}{(\det(B_0))^2}\begin{bmatrix} b_{22}^2 + b_{12}^2 & b_{22}b_{21} + b_{12}b_{11} \\ b_{22}b_{21} + b_{12}b_{11} & b_{21}^2 + b_{11}^2 \end{bmatrix}$$

where we have four unknowns.
- But in Σ_u we have three independent equations since this matrix is symmetric.
- We need to impose one restriction on B_0 in order to identify the unknowns.

STRUCTURAL VAR MODELS

- How to impose restrictions?
  - Exclusion restrictions
  - Proportionality restrictions
  - Equality restrictions
- In general we have $K(K+1)/2$ independent equations in Σ_u but K^2 unknowns in B_0^{-1} so therefore we need $K(K-1)/2$ restrictions.
- How to motivate restrictions?
  - Information delays (observe financial data without delay but GDP with delay)
  - Physical constraints (investment decisions and actual investment)
  - Institutional knowledge (suppliers cannot respond immediately to demand)
  - Assumptions about market structure (no feedback from small open economies to the rest of the world)
  - Homogeneity of demand functions (money holdings rise in proportion to nominal income implying that the response to a change in real income is the same as a response to a change in prices)
  - Extraneous parameter estimates (predetermined parameter values)
  - High-frequency data

Structural VAR models

- Lesson: No generally accepted approach to motivate restrictions. But, rely on specific economic models, for example DSGE models or RBC models. If economic models are not available, then use one or more of the examples above.

Recursively identified models

- A simple and popular approach is to use a Cholesky decomposition. (Also called orthogonalization.)
- Define a lower-triangular $K \times K$ matrix P with positive diagonal such that $\Sigma_u = PP'$.
- Example: Assume $K = 2$, then

$$P = \begin{bmatrix} a & 0 \\ b & c \end{bmatrix}$$

  and the Cholesky decomposition solves the problem

$$\begin{bmatrix} a & 0 \\ b & c \end{bmatrix} \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} = \begin{bmatrix} \sigma_{u1}^2 & \sigma_{u1,u2} \\ \sigma_{u1,u2} & \sigma_{u2}^2 \end{bmatrix}$$

  Solution is: $a = \sigma_{u1}$, $b = \frac{\sigma_{u1,u2}}{\sigma_{u1}}$ and $c = \sqrt{\sigma_{u2}^2 - (\frac{\sigma_{u1,u2}}{\sigma_{u1}})^2}$
- From the definition that $\Sigma_u = B_0^{-1}(B_0^{-1})'$ we have that $B_0^{-1} = P$.
- Since P is lower triangular, it has $K(K-1)/2$ zero parameters, thus implying exact identification.
- Cholesky decomposition implies a recursive structure. Need to motivate the recursive structure chosen!

More on Cholesky decomposition

- Consider again the bivariate VAR model above written as

$$\underbrace{\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}}_{B_0} y_t = B_1 y_{t-1} + \ldots + B_p y_{t-p} + w_t$$

- To identify the system we need to impose $K(K-1)/2 = 1$ restriction on B_0. Since we maintain the assumption that $\Sigma_w = I_2$, the diagonal in B_0 must be nonzero. We have two alternatives, either $b_{12} = 0$ or $b_{21} = 0$.
- If $b_{12} = 0$, then z_t has no contemporaneous effect on x_t and if $b_{21} = 0$ then x_t has no contemporaneous effect on z_t.

More on Cholesky decomposition

- If $b_{12} = 0$, then

$$B_0 = \begin{bmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{bmatrix} \quad \text{and} \quad B_0^{-1} = \begin{bmatrix} \frac{1}{b_{11}} & 0 \\ -\frac{b_{21}}{b_{11} b_{22}} & \frac{1}{b_{22}} \end{bmatrix}$$

- If $b_{21} = 0$, then

$$B_0 = \begin{bmatrix} b_{11} & b_{12} \\ 0 & b_{22} \end{bmatrix} \quad \text{and} \quad B_0^{-1} = \begin{bmatrix} \frac{1}{b_{11}} & -\frac{b_{12}}{b_{11} b_{22}} \\ 0 & \frac{1}{b_{22}} \end{bmatrix}$$

Recursively identified models

- Consider again the Cholesky decomposition $PP' = \Sigma_u$ where P is a lower triangular matrix.
- We know that $\Sigma_u = B_0^{-1}(B_0^{-1})'$ so $B_0^{-1} = P$.
- Consequences:
  - Structural model is recursive, we impose a causal chain.
  - Ordering of the K variables matter! Obtain different P for each ordering.
- Sensitivity analysis testing different orderings is not the solution!
  - The argument that the system is recursive but with no knowledge about the recursiveness is not credible.
  - Many alternative ordering of the K variables: If $K = 4$ we have 24 different orderings!
  - Even if we find no difference, this finding does not support the identification.

Alternative normalization

- We have implicitly assumed that $\Sigma_w = I_K$ leaving the diagonal elements of B_0 unrestricted.
- Identification is achieved by imposing restrictions on B_0^{-1}.
- This implies that: A unit innovation in the structural shocks is an innovation of a magnitude of one standard deviation.
- Alternative is to let diagonal elements of Σ_w to be unrestricted and set diagonal elements in B_0 to unity. In this case we need $K(K-1)/2$ restrictions on B_0 not counting the diagonal.
- We can combine these as $B_0 u_t = C w_t$.
- If $\Sigma_w = I_K$ then $\Sigma_u = B_0^{-1} C C' (B_0^{-1})'$.
- Cholesky decomposition (restrictions on B_0^{-1}) is the case when $C = I_K$.
- We can now allow $B_0 = I_K$ and cases when the diagonal of both B_0 and C are normalized to unity but neither B_0 nor C are diagonal.

Examples of recursively identified models

- Macroeconomic models
- A model of the global market for crude oil
- Oil price shocks and stock returns
- Transmission of energy price shocks
- Monetary policy (three to nine variables)
- Monetary policy shocks and exchange rates

Estimation of stable VAR with short-run restrictions: KL 9.1

- We are interested in the response of the US economy (inflation and GDP growth) to oil price shocks.
- Assume that the joint behavior of inflation, GDP growth and real oil price changes is given by a structural VAR model.
- We have the 3-dimensional VAR(4) model

$$B_0 y_t = B_1 y_{t-1} + B_2 y_{t-2} + B_3 y_{t-3} + B_4 y_{t-4} + w_t$$

  where

$$y_t = \begin{pmatrix} \Delta rpoil_t & \Delta p_t & \Delta gdp_t \end{pmatrix}'$$

  and

$$w_t \sim (0, I_3)$$

  and define $B \equiv [B_1, \ldots, B_4]$.
- Pre-multiply the VAR by B_0^{-1} such that

$$y_t = A_1 y_{t-1} + \ldots, A_4 y_{t-4} + u_t$$

  where $A \equiv [A_1, \ldots, A_4] = B_0^{-1} B$ and $u_t = B_0^{-1} w_t \sim (0, \Sigma_u)$ and $\Sigma_u = B_0^{-1} \Sigma_w \left(B_0^{-1}\right)'$.

ESTIMATION OF STABLE VAR WITH SHORT-RUN RESTRICTIONS: KL 9.1

- How do we estimate B_0^{-1}?
- We have $K(K+1)/2 = 3(3+1)/2 = 6$ independent equations in B_0^{-1} but $K^2 = 3^2 = 9$ parameters.
- Then we need to impose $K(K-1)/2 = 3(3-1)/2 = 3$ restrictions.
- We know that we can use a Cholesky (a lower triangular structure) decomposition of B_0^{-1}.
- Then we need to define a recursive structure.
- We will assume that inflation and GDP growth shocks cannot affect the change in oil prices in the first period. This implies that oil price changes must be put first in y_t.
- GDP growth shocks cannot affect inflation in the first period. This implies that GDP growth must be the last variable in y_t.

ESTIMATION OF STABLE VAR WITH SHORT-RUN RESTRICTIONS: KL 9.1

- Data: Percentage change in the real WTI price of crude oil, GDP deflator inflation and real GDP growth (US data) covering the sample 1987q1-2013q2.
- Estimates of reduced form model above in KL 9.2.1 pp. 243.
- Replicate these in Figure9_1_Estimates.m for Matlab or Figure9_1.py for Python.

Estimation of stable VAR with short-run restrictions: KL 9.1

- Use reduced form estimates to compute $B_0^{-1} = \text{chol}(\Sigma_u)$.
- In MATLAB use the command chol(Σ_u) or Python command np.linalg.cholesky(Σ_u)
- Upper triangular matrix in Matlab! Need to take the transpose. Lower triangular matrix in Python.
- Use the reduced form estimates of $[A_1, \ldots, A_4]$ to compute the companion matrix A.

Structural VMA representation

- Use the VAR(4) to compute the VMA representation

$$y_t = A_1 y_{t-1} + \ldots, A_4 y_{t-4} + u_t$$

  rewritten as

$$A(L) y_t = u_t$$

  such that

$$y_t = A(L)^{-1} u_t = \Phi(L) u_t$$

  where we know that $\Phi_i = JA^i J'$.

- Since $B_0^{-1} B_0 = I_3$ we find that

$$y_t = A(L)^{-1} u_t = \Phi(L) B_0^{-1} B_0 u_t = \Phi(L) B_0^{-1} w_t = \Theta(L) w_t$$

  since $u_t = B_0^{-1} w_t$.

- $\Theta(L)$ is the structural impulse response functions.
- What is the response of current and future values of each of the variables to a one-unit increase in the current value of one of the structural shocks, assuming that all other shocks are equal to zero?

Structural VMA representation

- To compute $\Theta(L)$ we need an estimate of $\Phi(L)$ and B_0^{-1}.
- Workflow: From reduced form estimates we compute the companion matrix A and Φ, then given our estimate of B_0^{-1} we compute Θ. We know that $\Phi_i = JA^iJ'$ and therefore we find that $\Theta_i = \Phi_i B_0^{-1}$.

Function to compute structural IRF in Matlab

```
function [irf]=irfstruc(A,B0inv,K,p,h)
% Process to invert VAR to VMA
% Note: A is the companion matrix
% K = number of var's
% p = number of lags
% h = horizon
%
% Output: irf organized such that the first K columns contain the effects
% of shock 1 on all variables, next K columns contain the effects of shock
% 2 on all variables and so on
if p==1;
jmat = eye(K);
end;
if p>1;
jmat = [eye(K,K) zeros(K,K*(p-1))];
end;
irf = reshape( jmat*(A^0)*jmat'*B0inv, [1,K*K] );
for i=1:h
irf = [ irf; reshape( jmat*(A^i)*jmat'*B0inv, [1,K*K] )];
end
end
```

FUNCTION TO COMPUTE STRUCTURAL IRF IN PYTHON

```
def IRF(A,K,p,horizon,B0inv):
”””
Process to invert VAR to VMA
Note: A is the companion matrix
K = number of variables
p = number of lags
horizon = horizon to compute MA representation
B0inv = Identifying matrix
Output: IRF’s are organized such that the first K columns contain
the effects of all shocks on the first variable, next K columns contain the
effects of all shocks on the second variable and so on
”””
jmat = Jmatrix(K,p)
irf = np.dot(np.dot(np.dot(jmat,np.linalg.matrix_power(A,0)),jmat.T),B0inv)
irf = np.reshape(irf, [ 1, K*K ])
i=1
while i ¡= horizon:
help = np.dot(np.dot(np.dot(jmat,np.linalg.matrix_power(A,i)),jmat.T),B0inv)
irf = np.concatenate((irf,np.reshape(help, [ 1, K*K ])), axis=0)
i=i+1
return irf
```

Example: KL Figure 9.1

- MATLAB Example: Figure9_1.m Python example: Figure9_1.py

Example: KL Figure 9.1

- Is there an alternative way to compute B_0^{-1}?
- Use the Python/MATLAB solver!
- We know that
$$\Sigma_u = B_0^{-1}(B_0^{-1})'$$
- Reformulate this as a system of equations (taking into account the fact that both sides of the equation are symmetric matrices).
- The vech operator selects the unique elements of $B_0^{-1}(B_0^{-1})' - \Sigma_u = 0$ (all elements on and below the diagonal).
- Then we need to add our restrictions imposed on B_0^{-1}, that is $b_0^{12} = b_0^{13} = b_0^{23} = 0$.
- Then solve the system of equations.
- This can be done by Python/MATLAB.

THE IMPLIED SYSTEM OF EQUATIONS

- Imposing the restrictions on B_0^{-1} we have

$$B_0^{-1} = \begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

- This implies that the system of equations $B_0^{-1}(B_0^{-1})' - \Sigma_u = 0$ can be written as

$$\begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{21} & b_{31} \\ 0 & b_{22} & b_{32} \\ 0 & 0 & b_{33} \end{bmatrix} - \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22}^2 & \sigma_{32} \\ \sigma_{13} & \sigma_{32} & \sigma_{33}^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

which in turn implies the following equations

$$\begin{bmatrix} b_{11}^2 - \sigma_{11}^2 & b_{11}b_{21} - \sigma_{12} & b_{11}b_{31} - \sigma_{13} \\ b_{11}b_{21} - \sigma_{12} & b_{21}^2 + b_{22}^2 - \sigma_{22}^2 & b_{21}b_{31} + b_{22}b_{32} - \sigma_{32} \\ b_{11}b_{31} - \sigma_{13} & b_{21}b_{31} + b_{22}b_{32} - \sigma_{32} & b_{31}^2 + b_{32}^2 + b_{33}^2 - \sigma_{33}^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

THE IMPLIED SYSTEM OF EQUATIONS

- We then need to solve the following system of equations for the unknowns in B_0^{-1}

$$\begin{bmatrix} b_{11}^2 - \sigma_{11}^2 & b_{11}b_{21} - \sigma_{12} & b_{11}b_{31} - \sigma_{13} \\ b_{11}b_{21} - \sigma_{12} & b_{21}^2 + b_{22}^2 - \sigma_{22}^2 & b_{21}b_{31} + b_{22}b_{32} - \sigma_{32} \\ b_{11}b_{31} - \sigma_{13} & b_{21}b_{31} + b_{22}b_{32} - \sigma_{32} & b_{31}^2 + b_{32}^2 + b_{33}^2 - \sigma_{33}^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- But, only six equations are unique

$$\begin{aligned} b_{11}^2 - \sigma_{11}^2 &= 0 \\ b_{11}b_{21} - \sigma_{12} &= 0 \\ b_{21}^2 + b_{22}^2 - \sigma_{22}^2 &= 0 \\ b_{11}b_{31} - \sigma_{13} &= 0 \\ b_{21}b_{31} + b_{22}b_{32} - \sigma_{32} &= 0 \\ b_{31}^2 + b_{32}^2 + b_{33}^2 - \sigma_{33}^2 &= 0 \end{aligned}$$

- These are the equations that the vech or vec functions extract in Python/Matlab!

Example: KL Figure 9.1 using Matlab solver

- The system of equations can now be written as

$$\begin{bmatrix} \mathrm{vech}(B_0^{-1}(B_0^{-1})' - \Sigma_u) \\ b_0^{12} \\ b_0^{13} \\ b_0^{23} \end{bmatrix} = 0$$

- We need to add the following to our program
  % Set some options for fsolve
  warning off
  options=optimset('TolX',1e-10,'TolFun',1e-10,'MaxFunEvals',1e+10);
  % Standard identification based on recursive short-run restrictions
  B0inv=fsolve('versionNorm',randn(K,K),options);
- and
  global so
- Then we need to setup the function ”versionNorm”.

Example: KL Figure 9.1 using Matlab solver

```
% VersionNorm.m
% Solving equation (9.2.2) on page 245
% Normalization: Sigma_w = I
function q=versionNorm(guess)
global so
K = size(guess,1);
B0inv=guess;
F=vec(B0inv*B0inv'-so);
% Using recursive short-run restrictions (like in chol-decomposition)
q=[F; B0inv(1,2); B0inv(1,3); B0inv(2,3)];
q'+1
end
```

MATLAB example: Figure9_1_Estimates_Solver.m

EXAMPLE: KL FIGURE 9.1 USING PYTHON SOLVER

- The system of equations is as above.
- We need to add the following to our program
  from statsmodels.tsa.tsatools import duplication_matrix, unvec, vec, **vech**
  from scipy import optimize
  from scipy.optimize import fsolve
- First we need to define the matrix B0inv
  B0invs = np.zeros([K,K]) where K=3 in our example.
- Then we need to define the parameters to be constrained and freely estimated in a function

```
def func2(vars):
# Define restrictions imposed on B0invs
# B0invs = [ * 0 0
# * * 0
# * * *]
B0invs[0,1] = 0
B0invs[0,2] = 0
B0invs[1,2] = 0
# Define all parameters to be estimated
B0invs[0,0], B0invs[1,0], B0invs[2,0], B0invs[1,1], B0invs[2,1], B0invs[2,2] = vars
# Then define the system of equations to be solved
fun1 = (vech(np.dot(B0invs,B0invs.T) - so)).flatten()
return (fun1)
```

- And finally we start the solver using the command
  root2 = fsolve(func2, [1, 1, 1, 1, 1, 1])
- And then we print B0invs and compare to the original estimate.

NEXT TOPICS

- Forecast error variance decomposition.
- Empirical analysis using Cholesky decompositions
  - Effects of monetary policy.
  - Uncertainty and consumer confidence.
  - Uncertainty and exchange rate expectations.

Forecast error variance decomposition

- Forecast error variance decomposition: decomposition of the forecast error variance of a variable in the system into the effects of each structural shock.
- Note that this decomposition is based on an identified structural model where the structural shocks are orthogonal.
- Consider the bivariate VMA model

$$\begin{bmatrix} x_t \\ z_t \end{bmatrix} = \sum_{i=0}^{\infty} \begin{bmatrix} \Theta_{11}(i) & \Theta_{12}(i) \\ \Theta_{21}(i) & \Theta_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{xt-i} \\ \varepsilon_{zt-i} \end{bmatrix}$$

$$y_t = \sum_{i=0}^{\infty} \Theta_i \varepsilon_{t-i}.$$

- Use this model to compute the n–step forecast and then compute the forecast error. The forecast is given by

$$y_{t+n} - E[y_{t+n}|t] = \sum_{i=0}^{n-1} \Theta_i \varepsilon_{t+n-i}$$

which implies that the forecast error depends on the structural shocks affecting the system from time $t+1$ until $t+n$.

Forecast error variance decomposition

- Let us only consider forecasts of the first variable in the system x_t. The forecast error for this variable is given by

$$x_{t+h} - E\left[x_{t+h}|t\right] = \Theta_{11}(0)\,\varepsilon_{xt+h} + \Theta_{12}(0)\,\varepsilon_{zt+h} + \Theta_{11}(1)\,\varepsilon_{xt+h-1} + \Theta_{12}(1)\,\varepsilon_{zt+h-1} + \ldots + \Theta_{11}(h-1)\,\varepsilon_{xt+1} + \Theta_{12}(h-1)\,\varepsilon_{zt+1}.$$

- Compute the variance of the forecast error

$$Var\left(x_{t+h} - E\left[x_{t+h}|t\right]\right) = \sigma_x^2(n) = \sigma_{\varepsilon x}^2\left(\Theta_{11}(0)^2 + \Theta_{11}(1)^2 + \ldots + \Theta_{11}(h-1)^2\right) + \sigma_{\varepsilon z}^2\left(\Theta_{12}(0)^2 + \Theta_{12}(1)^2 + \ldots + \Theta_{12}(h-1)^2\right).$$

- Since the structural shocks $\sigma_{\varepsilon x}$ and $\sigma_{\varepsilon z}$ have unit variance we find that

$$1 = \frac{\Theta_{11}(0)^2 + \Theta_{11}(1)^2 + \ldots + \Theta_{11}(h-1)^2}{\sigma_x^2(h)} + \frac{\Theta_{12}(0)^2 + \Theta_{12}(1)^2 + \ldots + \Theta_{12}(h-1)^2}{\sigma_x^2(h)}$$

  where

$$\sigma_x^2(h) = \Theta_{11}(0)^2 + \Theta_{11}(1)^2 + \ldots + \Theta_{11}(h-1)^2 + \Theta_{12}(0)^2 + \Theta_{12}(1)^2 + \ldots + \Theta_{12}(h-1)^2$$

  In a similar way we can compute the forecast error variance for all other variables in our system and also extending the number of variables.

- MATLAB Example: Table4_1.m Python example: Table4_1.py

MATLAB FUNCTION

```
function [VC]=fevd(A,B0inv,K,p,h);
% Structural forecast error variance Decomposition
% using B0inv for a K dimensional VAR and horizon h
% Input:
% A: companion matrix
% B0inv: Identifying matrix
% K: number of variables
% p: number of lags
% h: horizon
%
% Output: fevd organized as
% Each row is for horizon 1, 2, ..., h
% First K columns contain effects of shock 1,
% next K columns contain effects of shock 2,
% ....
% last K columns contain effect of shock K
%
```

Python function

```
def FEVD(A,B0inv,K,p,h):
"""
Structural forecast error variance Decomposition
using B0inv for a K dimensional VAR and horizon h
Input:
A: companion matrix
B0inv: Identifying matrix
K: number of variables
p: number of lags
h: horizon
Output: fevd organized as
Each row is for horizon 1, 2, ..., h
First K columns contain effects of shock 1,
next K columns contain effects of shock 2,
....
last K columns contain effect of shock K
```

SIMS (1992) AND EICHENBAUM (1992) ON MONETARY POLICYT

- Sims had published previous papers on money-income causality and how to use VAR models to study monetary policy in the US.
- Sims (1992) extended this work by also looking at other countries.
- Empirical puzzle in previous work: Restrictive monetary policy tends to increase inflation in the short-term in the US economy.
- Controversy among economists (and practitioners):
  - Some argue that monetary policy is effective, monetary authorities can control short-term interest rates and also aggregate income. But evidence suggest that restrictive monetary policy preceded almost all US postwar recessions.
  - Real Business Cycle (RBC) theory suggests that real shocks dominate while monetary policy is relatively unimportant.

Sims on the effects of monetary policy

- VAR models for five countries (France, Germany, Japan, the United Kingdom and the US). Monthly data covering the postwar period until the beginning of the 1990s.
- Two models, one 6-variable VAR and one 4-variable VAR.
- Time series vector: Short-term interest rate, money supply M1, consumer price index, industrial production (and exchange rate and commodity price index). Logs of all variables except interest rates. 14 lags in all models. Cholesky decomposition.
- Identification: All shocks affect interest rate contemporaneously, industrial production shocks only affect industrial production contemporaneously. Ordering (assuming lower triangular B_0^{-1} matrix): Y, P, M, R.
- Main conclusion: Weaker empirical evidence for both strands of the literature. Models should include both nominal and real variables.

IMPULSE RESPONSES: US

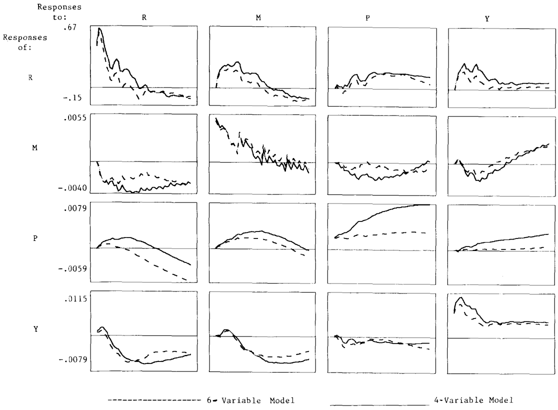


Fig. 10. United States 1958:4–1991:2.

EICHENBAUM COMMENTS

- Inference based on the assumption that residuals in the interest rate equation reflect monetary policy shocks.
- Sims argue that expansionary monetary policy should move output up and money and interest rates down. Interest rate increases in response to positive M1 shocks.
- Many different measures of monetary policy: direct measures of open market operations (non borrowed reserves), M0, M1, M2 and so on.
- Prize puzzle: restrictive monetary policy increases the price level. Questions the choice of monetary policy measure.
- Sims explanation: Information about inflation pressure not captured by data. But this implies a mis-specification! Using Non Borrowed Reserves, these shocks explain only small portions of output.
- VAR models: Federal Funds rate, industrial production, consumer price index and three measures of money stock (Non Borrowed Reserves, M0 and M1). Same sample and 14 lags in VAR. Same identification.
- Results: Figures 1, 2 and 3.

EICHENBAUM COMMENTS

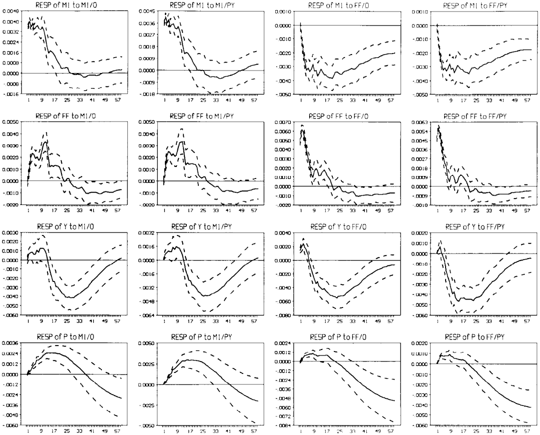


Fig. 1

Note: Columns 1 and 3 (ordering: M, P, Y, FF), columns 2 and 4 (P, Y, M, FF)

Eichenbaum comments

- Same results for M0 and Non Borrowed Reserves.
- Still a prize puzzle and uncertainty concerning measures of monetary policy.
- Alternative approaches: Narrative approach, alternative identification (sign restrictions), alternative econometric approach (local projections).
- Next another empirical example: Economic policy uncertainty and consumer confidence. Bergman & Worm (2021)

Bergman & Worm "Economic Policy Uncertainty and Consumer Perceptions: The Danish Case"

- Empirical example: Granger non-causality tests and structural VAR analysis using Cholesky decomposition.

Purpose and background

- Test whether newspaper reporting on economic policy in Denmark affects household perceptions about their own future financial situation and the future of the Danish economy.
- Need to construct a newspaper based measure of economic policy uncertainty in Denmark.
- Measure household expectations using consumer confidence surveys conducted each month.
- VAR analysis in reduced form and in structural form using Cholesky decomposition.
- Background
  - Many papers studying the relationships between consumer confidence and consumption (Carroll, Fuhrer and Wilcox, 1994; Cotsomitis and Kwan, 2006; Gausden et.al., 2020)
  - Literature on the factors determining consumer confidence (Acemoglu and Scott, 1994; Bram and Ludvigson, 1998; De Boef and Kellstedt, 2004; Ramalho et.al., 2011)
  - Literature on measuring uncertainty (Jurado et.al., 2015; Castelnuovo and Tran, 2017; Bloom et.al., 2018; Baker et.al., 2016).
  - Literature on the effects of uncertainty on real economy and on financial markets (Jansen and Nahuis, 2003; Bordo et.al., 2016; Krol, 2014; Beckmann and Czuday, 2017).
- We combine: measuring economic policy uncertainty using the newspaper based approach and the literature on the determinants of consumer confidence.

Economic policy uncertainty in Denmark

- We develop an index of economic policy uncertainty in Denmark following the approach suggested by Baker et.al. (2016).
- Method
  - Select keywords representing “economic policy”.
  - Search newspaper articles that include these keywords. We use web scraping through Infomedia homepage. Five nation-wide newspapers.
  - Count the number of articles each month.
  - Construct an index

KEYWORDS

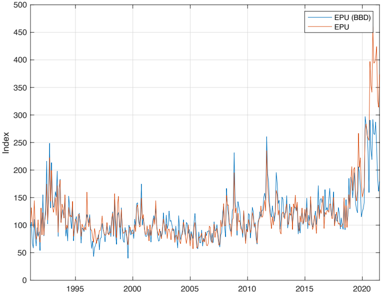
Table 1: Categories and keywords used to construct the Danish economic policy uncertainty index.

Category	Keywords
Economy (økonomi)	economy, economic (økonomi, økonomisk, økonomiske)
Uncertainty (usikkerhed)	Uncertainty, uncertain, turmoil (usikkerhed, usikker, usikkert, uro, uroligheder)
Geographic location (geografi)	Denmark (Danmark)
Politics (politik)	Nationalbanken, central bank, ministry of finance, government, regulation, legislation, parliament, Christiansborg, budget deficit (Nationalbanken, centralbank, finansministeriet, regering, regulering, lovgivning, Folketinget, Christiansborg, underskud)
Other political terms (politisk udvidelse)	Fiscal policy, budget deficit, availability for work, unemployment, general election, reform, tax, tax reform (finanspolitik, budgetunderskud, ledighed, arbejdsløshed, folketingsval, reform, skat, skattereform)

Note: The Danish language keywords used in web scraping are shown within parentheses.

Danish EPU index

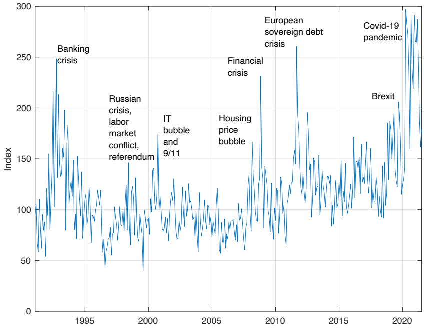
Figure 1: Comparison of EPU index with and without extended set of keywords.



Note: EPU(BBD) is the index constructed using the standard keywords suggested by Baker et.al. (2016) whereas EPU is the index constructed using the extended set of keywords stated in Table 1.

Danish EPU index

Figure 2: Danish economic policy uncertainty index January 1991 to June 2021.



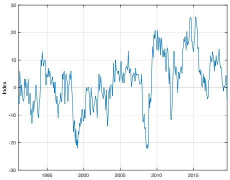
Consumer confidence measures

- Compiled on a monthly frequency by Statistics Denmark.
- Omnibus survey, representative sample of about 1500 persons each month from the population of persons of ages between 16 and 74 residing in Denmark.
- Several different survey based measures: We focus on “household expectation concerning their own financial situation 12 months ahead” and “household expectation about the Danish economy 12 months ahead”.
- Five possible answers: get a lot better; get a little bit better; stay the same; get a little bit worse; get a lot worse. These answers are weighted and aggregated to an index.

CONSUMER CONFIDENCE MEASURES

Figure B.3: Consumer confidence indices, household expectation about the Danish economy 12 months ahead (CCC) and their expectation about their own economy 12 months ahead (CCH).

(a) CCC (b) CCH



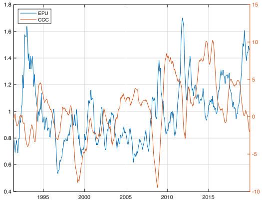
Control variable

- Composite leading indicator (CLI) to measure future state of the economy. Forward looking variable.
- CLI signals future turning points in the business cycle using a wide range of underlying data.

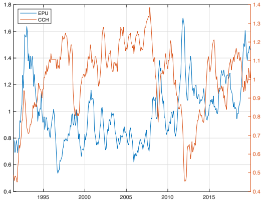
Data

Figure 3: EPU, consumer confidence and composite leading indicator, 5 months symmetric moving average filtered data.

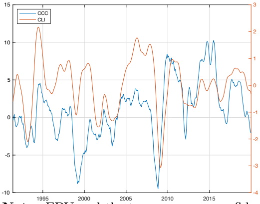
(a) EPU & CCC



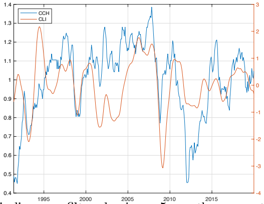
(b) EPU & CCH



(c) CCC & CLI



(d) CCH & CLI



Note: EPU and the two consumer confidence indices are filtered using a 5 months symmetric moving average filter. The composite leading indicator is unfiltered.

DESCRIPTIVE STATISTICS

Table 3: Cross-correlations and first order autocorrelations.

	EPU	CCC	CCH	CLI	DCLI	ρ_1	[95 percent Conf. Interval]
EPU	1					0.557	[.446 .667]
CCC	-0.015	1				0.901	[.852 .950]
CCH	-0.320*	0.223*	1			0.690	[.615 .764]
CLI	-0.205*	0.281*	0.301*	1		0.985	[.965 1.005]
DCLI	-0.240*	0.294*	0.122*	0.078	1	0.963	[.928 .998]

Note: * denotes significance at the 10 percent level using Bonferroni-adjusted significance levels.

Empirical model

- Let $x_t = \begin{bmatrix} CLI_t & EPU_t & CCH_t \end{bmatrix}'$

$$x_t = \nu + \sum_{j=1}^{\infty} \Phi_j \varepsilon_{t-j} \tag{1}$$

or, equivalently given by the Vector Autoregressive (VAR) model

$$A(L)x_t = \mu + \varepsilon_t \tag{2}$$

- Granger non-causality hypotheses tested on $A(L)$ and $\Phi(L)$.
- Identify structural VAR using Cholesky decomposition.

$$x_t = \nu + \sum_{j=0}^{\infty} \Phi_j \varepsilon_{t-j} = \nu + \sum_{j=0}^{\infty} \Phi_j P^{-1} P \varepsilon_{t-j} = \nu + \sum_{j=0}^{\infty} \Theta_j \omega_{t-j} \tag{3}$$

where the structural shocks $\omega_{t-j} = P\varepsilon_{t-j}$ and the structural impulse responses $\Theta_j = \Phi_j P^{-1}$.

Reduced form model

- Lag length (given in Table B.1)
- Stability. Tests for cointegration

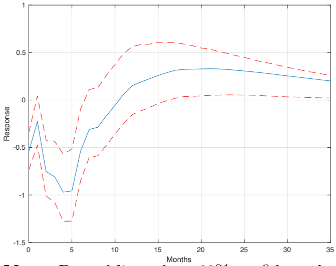
Table 4: Johansen trace test in bivariate and trivariate VAR systems.

Variables	Lag length	Modulus	Johansen trace test		
			$r = 0$	$r = 1$	$r = 2$
CCC & EPU	7	0.946	23.44***	7.28***	
CCH & EPU	6	0.956	35.60***	6.05**	
CCC, EPU & CLI	7	0.948	72.99***	24.12***	7.28***
CCH, EPU & CLI	7	0.959	86.57***	35.25***	5.64**

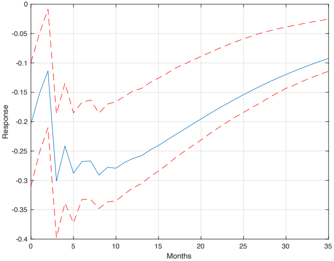
Note: Eigenvalue denotes the largest modulus in a levels VAR model. *** indicates significance at the 1 percent level and ** at the 5 percent level.

Figure 4: Impulse response of consumer confidence to a one standard deviation shock to EPU in a bivariate VAR model.

(a) EPU shock on CCC.



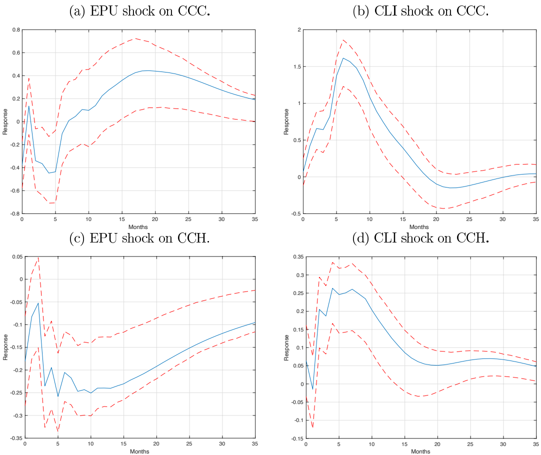
(b) EPU shock on CCH.



Note: Dotted lines show 68% confidence bands computed using non-parametric bootstrap with 1000 trials.

STRUCTURAL FORM MODEL

Figure 5: Impulse response of consumer confidence to a one standard deviation shock to EPU and to a one standard deviation shock to CLI.



(a) EPU shock on CCC.
(b) CLI shock on CCC.
(c) EPU shock on CCH.
(d) CLI shock on CCH.

Note: Dotted lines show 68% confidence bands computed using non-parametric bootstrap with 1000 trials.

STRUCTURAL FORM MODEL

Table 6: Forecast error variance decomposition of consumer confidence. The share explained by EPU shocks.

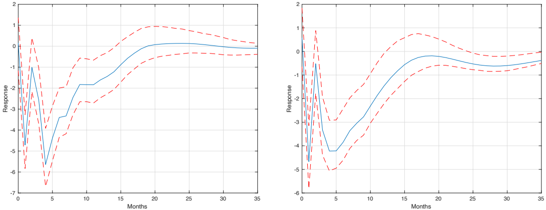
	Bivariate VAR		Trivariate VAR	
Horizon	CCC	CCH	CCC	CCH
1	0.019	0.012	0.01	0.01
	[0.001 0.061]	[0.000 0.043]	[0.000 0.044]	[0.000 0.043]
2	0.014	0.016	0.007	0.009
	[0.002 0.054]	[0.001 0.061]	[0.002 0.034]	[0.001 0.049]
6	0.061	0.059	0.017	0.028
	[0.010 0.173]	[0.019 0.151]	[0.005 0.086]	[0.008 0.096]
12	0.052	0.12	0.011	0.092
	[0.016 0.185]	[0.027 0.264]	[0.006 0.093]	[0.026 0.224]
18	0.052	0.159	0.019	0.132
	[0.021 0.191]	[0.031 0.332]	[0.008 0.144]	[0.030 0.291]
24	0.058	0.181	0.031	0.154
	[0.023 0.213]	[0.034 0.368]	[0.009 0.187]	[0.032 0.324]
30	0.063	0.193	0.039	0.166
	[0.023 0.232]	[0.035 0.387]	[0.009 0.207]	[0.032 0.342]
36	0.066	0.199	0.042	0.173
	[0.024 0.237]	[0.035 0.401]	[0.009 0.217]	[0.032 0.356]

Note: 95 percent confidence bands computed using non-parametric bootstrap with 1000 trials are shown within parentheses below each estimate.

Structural form model

Figure 6: Impulse response of EPU to a one standard deviation shock to CLI.

(a) CLI shock on EPU (VAR with CCC). (b) CLI shock on EPU (VAR with CCH).



Note: Dotted lines show 68% confidence bands computed using non-parametric bootstrap with 1000 trials.

Table 7: Forecast error variance decomposition of consumer confidence and EPU in trivariate VAR models. The share explained by shocks to CLI.

Horizon	CCC	EPU	CCH	EPU
1	0.002	0	0.001	0.001
	[0.000 0.024]	[0.000 0.019]	[0.000 0.022]	[0.000 0.017]
2	0.002	0.031	0.001	0.03
	[0.000 0.030]	[0.007 0.083]	[0.000 0.024]	[0.007 0.077]
6	0.073	0.091	0.041	0.075
	[0.026 0.167]	[0.039 0.181]	[0.010 0.114]	[0.031 0.160]
12	0.118	0.126	0.082	0.114
	[0.039 0.233]	[0.052 0.251]	[0.016 0.211]	[0.043 0.246]
18	0.127	0.131	0.084	0.114
	[0.041 0.270]	[0.052 0.273]	[0.019 0.229]	[0.044 0.256]
24	0.127	0.131	0.082	0.112
	[0.041 0.275]	[0.054 0.277]	[0.019 0.232]	[0.044 0.251]
30	0.130	0.131	0.083	0.112
	[0.043 0.279]	[0.054 0.277]	[0.019 0.241]	[0.044 0.252]
36	0.133	0.133	0.084	0.113
	[0.044 0.285]	[0.054 0.279]	[0.019 0.245]	[0.045 0.254]

Note: 95 percent confidence bands computed using non-parametric bootstrap with 1000 trials are shown within parentheses below each estimate.

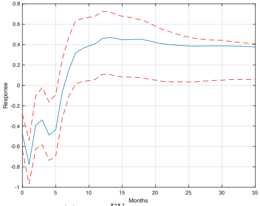
STRUCTURAL FORM MODEL: SPILLOVER EFFECTS FROM EU

- Denmark is a small open economy.
- Fixed exchange rate versus the euro.
- Add EU economic policy (EPU-EU) uncertainty index to VAR model.

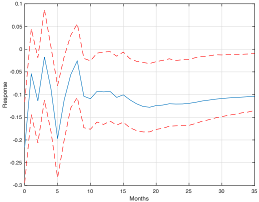
Structural form model: Spillover effects from EU

Figure 7: Impulse response of consumer confidence to a one standard deviation shock to economic policy uncertainty.

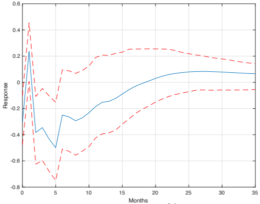
(a) EPU^{EU} shock on CCC.



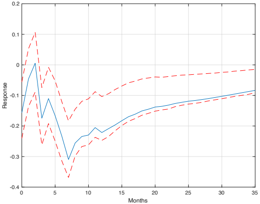
(b) EPU shock on CCC.



(c) EPU^{EU} shock on CCH.



(d) EPU shock on CCH.

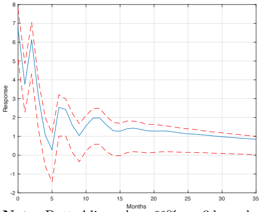


Note: Dotted lines show 68% confidence bands computed using non-parametric bootstrap with 1000 trials.

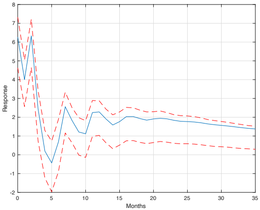
Structural form model: Spillover effects from EU

Figure 8: Impulse response of EPU to a one standard deviation shock to EPU-EU.

(a) EPU^{EU} shock on EPU (CCC).



(b) EPU^{EU} shock on EPU (CCH)



Note: Dotted lines show 68% confidence bands computed using non-parametric bootstrap with 1000 trials.

Structural form model: Spillover effects from EU

Table 8: Forecast error variance decomposition of consumer confidence and Danish economic policy uncertainty in four-variable VAR models. The share explained by shocks to economic policy uncertainty.

Horizon	(a) Shares of CCC		(b) Shares of EPU	(c) Shares of CCH		(d) Shares of EPU
	EPU^{EU}	EPU	EPU^{EU}	EPU^{EU}	EPU	EPU^{EU}
1	0.016	0.007	0.074	0.014	0.007	0.065
	[0.001 0.055]	[0.000 0.035]	[0.029 0.143]	[0.000 0.052]	[0.000 0.038]	[0.020 0.129]
2	0.036	0.007	0.087	0.014	0.007	0.083
	[0.006 0.099]	[0.001 0.029]	[0.039 0.164]	[0.001 0.057]	[0.000 0.042]	[0.031 0.153]
6	0.033	0.018	0.119	0.024	0.021	0.118
	[0.007 0.120]	[0.005 0.088]	[0.061 0.211]	[0.007 0.090]	[0.008 0.084]	[0.059 0.209]
12	0.030	0.018	0.125	0.027	0.079	0.120
	[0.014 0.112]	[0.005 0.124]	[0.063 0.235]	[0.010 0.114]	[0.022 0.185]	[0.062 0.221]
18	0.044	0.017	0.132	0.034	0.104	0.135
	[0.015 0.165]	[0.006 0.143]	[0.063 0.257]	[0.012 0.145]	[0.023 0.229]	[0.064 0.253]
24	0.056	0.017	0.138	0.046	0.115	0.151
	[0.017 0.205]	[0.007 0.148]	[0.063 0.269]	[0.014 0.183]	[0.023 0.248]	[0.067 0.275]
30	0.065	0.017	0.143	0.055	0.122	0.163
	[0.019 0.227]	[0.007 0.155]	[0.063 0.281]	[0.014 0.210]	[0.024 0.256]	[0.070 0.299]
36	0.074	0.017	0.146	0.063	0.125	0.171
	[0.020 0.245]	[0.007 0.157]	[0.063 0.284]	[0.015 0.233]	[0.025 0.266]	[0.070 0.314]

Note: 95 percent confidence bands computed using non-parametric bootstrap with 1000 trials are shown within parentheses below each estimate.

CONCLUSIONS

- EPU significantly affects consumer confidence, in particular household perceptions about their own economy.
- Other factors are as important, CLI equally important as EPU for CCH.
- EPU-EU also determines consumer confidence but to a lesser degree than EPU. Sizable effect from EPU-EU to EPU (significant spillover effects).
- Lesson: Households read newspapers and articles concerning economic policy. Readings affect their expectations about their own economy.
- Additional panel data regression results using disaggregated consumer confidence data (not included in the curriculum).

BECKMANN AND CZUDAJ (2017)

- Idea: Uncertainty affects financial markets, for example the foreign exchange market.
- Purpose: Study the effects of economic policy uncertainty on exchange rate expectations and forecast errors of market participants.
- Approach related to the ‘news’ approach and also to the microstructure literature.
- Potential problem: Need survey data on exchange rate expectations and data on policy uncertainty.

Beckmann and Czudaj (2017)

- Exchange rate expectations: Survey data obtained from FX4casts (formerly known as the Financial Times Currency Forecaster). Forecast horizons: 3, 6 and 12 months.

Table 1
Contributors to FX4casts consensus forecasts.

Allied Irish Bank, ANZ Bank, Bank of America/Merrill Lynch, Bank of New York Mellon, Barclays Capital, Bayerische Landesbank, BNP Paribas, Canadian Imperial Bank of Commerce, Credit-Agricola, Citigroup, Commerzbank, Credit Suisse - First Boston, Danske Bank, Deka, Deutsche Bank, DnBNOR, Economist Intelligence Unit, Goldman Sachs, Handels Banken, HSBC, IHS Global Insight, ING Bank, Intesa Sanpaolo, JP Morgan Chase, Julius Baer, Lloyds TSB, Macquarie Capital Securities, Moody's Economy.com, Morgan Stanley, National Australia Bank, Nomura, Nordea, Rabobank, Royal Bank of Canada, Royal Bank of Scotland, Scotia Bank, SEB, Societe Generale, Standard Chartered, Suntrust, Swedbank, Tokyo-Mitsubishi UFJ, Toronto Dominion, UBS Warburg, UniCreditHVB, Vontobel, Wachovia, Westpac.

- Economic policy uncertainty: Baker et al. (2013). Three different US newspaper based measures of uncertainty: Economic policy uncertainty, monetary policy uncertainty, and fiscal policy uncertainty. (Based on text-searching in articles published in the 10 largest US newspapers: the USA Today, Miami Herald, Chicago Tribune, Washington Post, Los Angeles Times, Boston Globe, San Francisco Chronicle, Dallas Morning News, New York Times, and the Wall Street Journal.)

Table 2
Full list of terms for the monetary and fiscal policy uncertainty index.

Monetary policy	Federal Reserve, the fed, money supply, open market operations, quantitative easing, monetary policy, fed funds rate, overnight lending rate, Bernanke, Volker, Greenspan, central bank, interest rates, fed chairman, fed chair, lender of last resort, discount window, European Central Bank, ECB, Bank of England, Bank of Japan, BOJ, Bank of China, Bundesbank, Bank of France, Bank of Italy
Fiscal policy	Government spending, federal budget, budget battle, balanced budget, defense spending, military spending, entitlement spending, fiscal stimulus, budget deficit, federal debt, national debt, Gramm-Rudman, debt ceiling, fiscal footing, government deficits, balance the budget

Beckmann and Czudaj (2017)

- Empirical framework: Vector Autoregressive (VAR) Model.
- Variables: Policy uncertainty (monetary and fiscal), difference between actual and expected exchange rate (change in exchange rate expectation), interest rate differential, relative money supply and relative output.
- Four major currencies vs. the dollar, three forecast horizons, sample: August 1986 until December 2014.
- Focus on the effects of policy uncertainty on the forecast error (and the change in expectations).

BECKMANN AND CZUDAJ (2017)

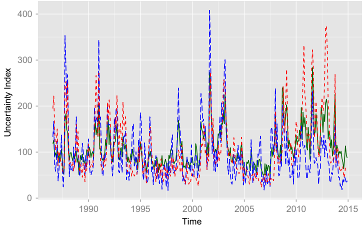


Fig. 1. US economic, monetary and fiscal policy uncertainty. The plot shows three measures of uncertainty over a sample period running from 1986:08 until 2014:12 on a monthly frequency: Economic policy uncertainty (green line), monetary policy uncertainty (blue line), and fiscal policy uncertainty (red line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Correlation matrix of the uncertainty measures.

	EPU	MPU	FPU
EPU	1.0000000	0.6941259	0.8205214
MPU	0.6941259	1.0000000	0.5919167
FPU	0.8205214	0.5919167	1.0000000

Note: This table reports the correlation matrix of uncertainty measures, i.e. US economic policy uncertainty (EPU), US monetary policy uncertainty (MPU), and US fiscal policy uncertainty (FPU).

Beckmann and Czudaj (2017)

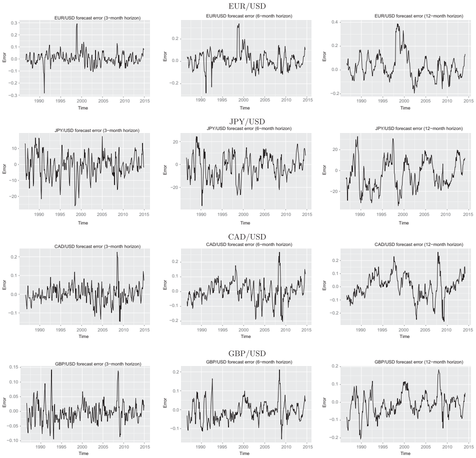


Fig. 3. Forecast errors. The plots show the exchange rate forecast errors computed as the difference between the actual and the expected exchange rate for all currencies (EUR/USD, JPY/USD, CAD/USD, GBP/USD) and all horizons (h=3,6,12).

- Time series vector

$$Y_t = \begin{bmatrix} EPU_t \\ \varepsilon_t \\ \tilde{r}_t \\ \tilde{m}_t \\ \tilde{y}_t \end{bmatrix}$$

  where EPU_t is either US economic policy uncertainty or US monetary policy uncertainty or US fiscal policy uncertainty; ε_t is either the forecast error (difference between actual and expected exchange rate) or the change in exchange rate expectations.
- Fundamentals: $\tilde{m}$ is the relative money supply, $\tilde{r}_t$ is the interest differential and $\tilde{y}_t$ is the relative income.
- VAR model

$$Y_t = A_0 + A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + \eta_t$$

- VAR is assumed to be stationary and use Cholesky decomposition to identify shocks. Need to impose recursive structure.
- In this case, EPU is only affected by own shocks in the first period and expectations are affected only by EPU and own shocks in the first period.
- Since focus is on the effects of EPU on exchange rate expectations, then the ordering of the remaining variables is irrelevant.
- Next topic: Alternative ways to identify structural VARs. Long-run restrictions.

Other examples of SVAR analyses focusing on uncertainty and using Cholesky decomposition

- Hardouvelis, G., G. Karalas, D. Karanastasis, and P. Samartzis (2024), "Economic policy uncertainty and the Greek economic crisis", Journal of Economic Studies, 51:1199-1215.
- Forni, M., L. Gambetti and L. Sala (2023), "Macroeconomic uncertainty and vector autoregressions", Econometrics and Statistics, in press.
- Albrecht, P., S. Kapounek, and Ku?erová, Z. (2023), "Economic policy uncertainty and stock markets' co-movements", International Journal of Finance & Economics, 28:3471?3487.
- Ludvigson, S. C., Ma, S. and and Ng. S. (2021), "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?" American Economic Journal: Macroeconomics, 13: 369?410.

Advanced Macroeconomics: Structural Vector Autoregressive Analysis

Lectures 6 to 8

U. Michael Bergman
University of Copenhagen

Fall 2024

Non-Stationary VAR models

Learning objective: Vector Error Correction (VEC) models (estimation, specification and interpretation), testing for cointegration, testing for unit roots and interpretation of estimated cointegration vectors in a VEC. Implementation of these methods in Python/Matlab.
Curriculum: KL 3.1, 3.2 (LS estimation with known cointegration vector), 3.3-3.4, Johansen (1991), Johansen and Juselius (1990,1992) and Python/ Matlab examples.

- Vector Error Correction (VEC) model.
- Estimation of VEC with known cointegration vector.
- Maximum Likelihood estimation of cointegration vectors.
- Tests for cointegration.
- Tests for specific cointegration vectors.
- Applications: Money demand, PPP and UIP.

THE VEC MODEL

- Assume that the K–dimensional time series vector y_t is generated by a VAR(p) model

$$y_t = \rho + A_1 y_{t-1} + \ldots + A_p y_{t-p} + \varepsilon_t.$$

- This is a stable VAR if y_t is stationary.
- If all components of the time series vector are integrated of order 1, but not cointegrated we can simply take the first difference and rewrite the VAR model in first difference terms.
- What are the consequences if y_t is integrated of order 1 and cointegrated?

Definition of co-integration

- Engle and Granger definition of cointegration

Definition
The components of the vector x_t are said to be co-integrated of order d, b, denoted $x_t \sim CI(d, b)$, if (i) all components of x_t, are $I(d)$; (ii) there exists a vector $\beta(\neq 0)$ so that $z_t = \beta' x_t \sim I(d - b)$, $b > 0$. The vector β is called the co-integration vector.

- Johansen (1991) definition of cointegration.

Definition
If the K–dimensional time series vector y_t is integrated of order 1 and, for any $K \times r$ matrix β, $\beta' y_t - E[\beta' y_t]$ is stationary, then y_t is cointegrated with r cointegration vectors.

- Note: We allow for both non-stationary and stationary variables in the time series vector y_t but not in x_t!
- Engle and Granger definition requires that all individual time series x_t are integrated of order 1, Johansen requires that at least 1 of the elements of y_t is I(1).

The VEC model

- If all components of x_t are I(1), then the stable vector ARMA model (or the stable VMA in first differences) is written as

$$(1 - L)x_t = C(L)\varepsilon_t \tag{1}$$

- Definition (Engle and Granger, 1987)

Definition

A vector time series x_t has an error correction representation if it can be expressed as:

$$A(L)(1 - L)x_t = -\alpha z_{t-1} + u_t$$

where u_t is a stationary multivariate disturbance, with $A(0) = I_K$, $A(1)$ has all elements finite, $z_t = \beta' x_t$, and $\beta \neq 0$.

GRANGERS REPRESENTATION THEOREM (ENGLE AND GRANGER, 1987)

Theorem

If the $K \times 1$ vector x_t, given in (1) is co-integrated with $d = 1$, $b = 1$ and with co-integration rank r, then:

1. *$C(1)$ is of rank $K - r$.*
2. *There exists a vector ARMA representation*
$$A(L)x_t = d(L)\varepsilon_t \tag{2}$$
with the properties that $A(1)$ has rank r and $d(L)$ is a scalar lag polynomial with $d(1)$ finite, and $A(0) = I_K$. When $d(L) = 1$, this is a vector autoregression.
3. *There exists $K \times r$ matrices, β, α, of rank r such that*
$$\beta' C(1) = 0,$$
$$C(1)\alpha = 0,$$
$$A(1) = \alpha\beta'$$
4. *There exists an error correction representation with $z_t = \beta' x_t$, and $r \times 1$ vector of stationary random variables:*
$$A^*(L)(1 - L)x_t = -\alpha z_{t-1} + d(L)\varepsilon_t \tag{3}$$
with $A^*(0) = I_K$
5. *The vector z_t is given by*
$$z_t = K(L)\varepsilon_t,$$
$$(1 - L)z_t = -\beta'\alpha z_{t-1} + J(L)\varepsilon_t,$$
where $K(L)$ is an $r \times K$ matrix of lag polynomials given by $\beta' C^*(L)$ with all elements of $K(1)$ finite with rank r, and let $\det(\beta'\alpha) > 0$.
6. *If a finite vector autoregressive representation is possible, it will have the form given by (2) and (3) above with $d(L) = 1$ and both $A(L)$ and $A^*(L)$ as matrices of finite polynomials.*

- Consider again the VAR

$$y_t = \rho + A_1 y_{t-1} + \ldots + A_p y_{t-p} + \varepsilon_t. \tag{4}$$

- subtract y_{t-1} from both sides, and (ii) add and subtract $(A_{i+1})\, y_{t-i}$ for $i = 1, \ldots, p-1$ to the RHS so that we obtain

$$\Delta y_t = \rho + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + \varepsilon_t \tag{5}$$

  where $\Gamma_i = -(A_{i+1} + \ldots + A_p)$ for $i = 1, \ldots, p-1$, and $\Pi = -(I_K - A_1 - \ldots - A_p) = -A(1)$.

- Let $\Pi = \alpha\beta'$ where α is the adjustment coefficient and β is the cointegration vector. Note that α and β are $K \times r$ matrices where r is the number of cointegration vectors.

Applying Grangers Representation Theorem: Number of cointegration vectors

- The rank, i.e., the number of eigenvalues that are different from zero, of the matrix Π determines how many cointegration vectors there exist in the VAR model.
- Given our new definition of cointegration above we can distinguish between the following three cases.
  1. $\text{Rank}[\Pi] = K$, implying that $A(1)$ has full rank and $y_t \sim \text{I}(0)$.
  2. $\text{Rank}[\Pi] = 0$ implying that $A(1)$ is a null matrix and $y_t \sim \text{I}(1)$ and $\Delta y_t \sim \text{I}(0)$.
  3. $\text{Rank}[\Pi] = r < K$ implying that $A(1)$ has reduced rank such that the time series vector contains r cointegration relations.

Example 1

- Assume that $y_t = \begin{bmatrix} x_t & z_t \end{bmatrix}'$ where x_t is non–stationary whereas z_t is stationary. Under these assumptions we know that the cointegration vector $\beta = \begin{bmatrix} 0 & 1 \end{bmatrix}'$ such that

$$\beta' y_t = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ z_t \end{bmatrix} = z_t$$

  is a stationary linear combination since we have assumed that z_t is stationary.
- Note: No other linear combination between the two variables can be stationary.
- Assume that $\alpha = \begin{bmatrix} 0 & 1 \end{bmatrix}'$ such that

$$\Pi = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

  Compute eigenvalues of the matrix Π, the number of cointegration vectors is equal to the number of non–zero eigenvalues.
- We have the eigenvalue problem:

$$\det\left(\lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) = \lambda(\lambda - 1)$$

  which has the solutions $\lambda_1 = 0$ and $\lambda_2 = 1$.
- Conclusion: only one eigenvalue is different from zero and there must be one cointegration vector.

Example 2

- Assume that the two variables variables x_t and z_t are cointegrated with cointegration vector $\beta = \begin{bmatrix} 1 & -1 \end{bmatrix}'$ and that the adjustment parameter $\alpha = \begin{bmatrix} 0 & 1 \end{bmatrix}'$.
- Then

$$\Pi = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}$$

- To test for the number of cointegration vectors in this system we compute eigenvalues of this matrix.
- We have the following eigenvalue problem

$$\det\left(\lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}\right) = \lambda(\lambda + 1)$$

with the solution $\lambda_1 = 0$, and $\lambda_2 = -1$.

- There is, thus, only one eigenvalue different from zero implying that there must be only 1 cointegration vector in the system (as we also assumed in the beginning).

General case

- In the general case
$$\Pi = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} \beta_1 & 1 \end{bmatrix}$$
- In this case we obtain one eigenvalue equal to zero while the second eigenvalue is $\beta_1 \alpha_1 + \alpha_2$.
- Conclusion: As long as any of the adjustment parameters are different from zero, one eigenvalue will be different from zero. This implies that in order for the two variables to be cointegrated, one of the adjustment parameters must be different from zero.

General case

- We also reach a similar conclusion if we assume that the second variable z_t is stationary while x_t is integrated of order 1. In this case we have

$$\Pi = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix}$$

  where one eigenvalue is equal to zero while the second eigenvalue is equal to α_2.
- The conclusion is that if $\alpha_2 \neq 0$, then the rank of Π is equal to 1 implying that there is 1 cointegration vector in the system. What would happen if $\alpha_2 = 0$? Then Π will be a null matrix and both variables must be integrated of order 1 and not cointegrated.

THE NON-UNIQUENESS OF COINTEGRATION VECTORS

- The cointegration vectors are not unique. This implies that the adjustment coefficients in α cannot be unique either.
- We can always normalize the cointegration vector by dividing or multiplying with any nonzero number.
- Assume that $K = 3$ $p = 2$ and that $r = 1$.
- In this case, $\alpha\beta' y_{t-1}$ must be given by

$$\alpha\beta' y_{t-1} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} [\beta_1 y_{1t-1} + \beta_2 y_{2t-1} + \beta_3 y_{3t-1}] .$$

- It is often useful to normalize the coefficients in the cointegration vector on one of the variables, for example y_{1t-1}, by multiplying β with $1/\beta_1$ and α with β_1. We then obtain

$$\begin{bmatrix} \alpha_1\beta_1 \\ \alpha_2\beta_1 \\ \alpha_3\beta_1 \end{bmatrix} \left[y_{1t-1} + \frac{\beta_2}{\beta_1} y_{2t-1} + \frac{\beta_3}{\beta_1} y_{3t-1} \right]$$

so that the first equation in the VEC model can be written as

$$\Delta y_{1t} = \Gamma_{11}\Delta y_{1t-1} + \Gamma_{12}\Delta y_{2t-1} + \Gamma_{13}\Delta y_{3t-1} + \tilde{\alpha}_1 \left[y_{1t-1} + \frac{\beta_2}{\beta_1} y_{2t-1} + \frac{\beta_3}{\beta_1} y_{3t-1} \right] + \varepsilon_{1t}$$

where $\tilde{\alpha}_1 = \alpha\beta_1$.

The non-uniqueness of cointegration vectors

- Assume now that $r = 2$. In this case $\alpha\beta' y_{t-1}$ is given by

$$\alpha\beta' y_{t-1} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix} =$$

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} \beta_{11} y_{1t-1} + \beta_{21} y_{2t-1} + \beta_{31} y_{3t-1} \\ \beta_{12} y_{1t-1} + \beta_{22} y_{2t-1} + \beta_{32} y_{3t-1} \end{bmatrix}.$$

- Let us normalize the first cointegration vector using β_{11} and the second vector using β_{12}. The first equation in the VEC model can then be written as

$$\Delta y_{1t} = \Gamma_{11}\Delta y_{1t-1} + \Gamma_{12}\Delta y_{2t-1} + \Gamma_{13}\Delta y_{3t-1} + \tilde{\alpha}_{11}\left[y_{1t-1} + \frac{\beta_{21}}{\beta_{11}} y_{2t-1} + \frac{\beta_{31}}{\beta_{11}} y_{3t-1}\right] +$$
$$\tilde{\alpha}_{12}\left[y_{1t-1} + \frac{\beta_{22}}{\beta_{12}} y_{2t-1} + \frac{\beta_{32}}{\beta_{12}} y_{3t-1}\right] + \varepsilon_{1t}$$

where $\tilde{\alpha}_{11} = \alpha_{11}\beta_{11}$ and $\tilde{\alpha}_{12} = \alpha_{12}\beta_{12}$.

The non-uniqueness of cointegration vectors

- How should we identify the cointegration vectors?
- One approach is to use economic theory as a guideline. Examples often used in the literature are:
  - Purchasing Power Parity which imposes a certain restriction on the parameters in the cointegration vector and
  - stationarity of consumption and investment shares of output as suggested by a simple Real Business Cycle model, or
  - money demand functions.
- Otherwise, we can normalize the cointegration vectors. One approach is to use the so–called Phillips normalization.

PHILLIPS NORMALIZATION

- Assume that we have estimated the cointegration vector β which is a $K \times r$ matrix. Partition this matrix in the following way

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

  where β_1 is an $r \times r$ matrix and β_2 is a $(K - r) \times r$ matrix. Under the assumption that β_1 has full rank, we can post–multiply β by β_1^{-1} such that

$$\beta = \begin{bmatrix} I_r \\ \beta_2 \beta_1^{-1} \end{bmatrix} = \begin{bmatrix} I_r \\ -B \end{bmatrix}.$$

- Note: When using this normalization, it may be important to specify the system in such a way that interpretation of the cointegration vectors is simplified, i.e., the ordering of the variables could simplify the interpretation of the vectors.

Example: Phillips normalization

- Assume that $K = 4$ and $r = 2$, then the cointegration vector is

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \beta_{13} & \beta_{23} \\ \beta_{14} & \beta_{24} \end{bmatrix}$$

- The $r \times r$ upper matrix

$$\beta_1 = \begin{bmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \end{bmatrix}$$

$$\beta_1^{-1} = \begin{bmatrix} \frac{\beta_{22}}{\beta_{11}\beta_{22}-\beta_{21}\beta_{12}} & -\frac{\beta_{21}}{\beta_{11}\beta_{22}-\beta_{21}\beta_{12}} \\ -\frac{\beta_{12}}{\beta_{11}\beta_{22}-\beta_{21}\beta_{12}} & \frac{\beta_{11}}{\beta_{11}\beta_{22}-\beta_{21}\beta_{12}} \end{bmatrix}$$

- Post–multiplying β by β_1^{-1}:

$$\begin{bmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \beta_{13} & \beta_{23} \\ \beta_{14} & \beta_{24} \end{bmatrix} \begin{bmatrix} \frac{\beta_{22}}{\beta_{11}\beta_{22}-\beta_{21}\beta_{12}} & -\frac{\beta_{21}}{\beta_{11}\beta_{22}-\beta_{21}\beta_{12}} \\ -\frac{\beta_{12}}{\beta_{11}\beta_{22}-\beta_{21}\beta_{12}} & \frac{\beta_{11}}{\beta_{11}\beta_{22}-\beta_{21}\beta_{12}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -\frac{-\beta_{13}\beta_{22}+\beta_{23}\beta_{12}}{\beta_{11}\beta_{22}-\beta_{21}\beta_{12}} & \frac{-\beta_{13}\beta_{21}+\beta_{23}\beta_{11}}{\beta_{11}\beta_{22}-\beta_{21}\beta_{12}} \\ -\frac{-\beta_{14}\beta_{22}+\beta_{24}\beta_{12}}{\beta_{11}\beta_{22}-\beta_{21}\beta_{12}} & \frac{-\beta_{14}\beta_{21}+\beta_{24}\beta_{11}}{\beta_{11}\beta_{22}-\beta_{21}\beta_{12}} \end{bmatrix}$$

- Note 2: This is automatically implemented in Python! More on this later.

Deterministic terms in cointegrated processes

- Examples given in KL 3.1.2.
- Consider the VEC model

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + v + \delta t + \varepsilon_t$$

- Let $v = \alpha\mu + \gamma$ and $\delta t = \alpha\rho t + \tau t$
- Then we can write the VEC model as

$$\Delta y_t = \alpha\left(\beta' y_{t-1} + \mu + \rho t\right) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \gamma + \tau t + \varepsilon_t$$

DETERMINISTIC TERMS IN COINTEGRATED PROCESSES

- VEC model from previous page is

$$\Delta y_t = \alpha\left(\beta' y_{t-1} + \mu + \rho t\right) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \gamma + \tau t + \varepsilon_t$$

- We can now distinguish between five cases.
  - Case 1. $v = 0$ and $\delta = 0$: No deterministic components implying that all cointegration vectors have mean zero and that the mean of first differences is zero.

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t$$

  - Case 2. $\tau = \rho = \gamma = 0$: Mean of cointegration vectors is μ and no trends in either levels ($\gamma = 0$) or first differences ($\tau = 0$) and variables are stationary around a constant mean.

$$\Delta y_t = \alpha\left(\beta' y_{t-1} + \mu\right) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t$$

  - Case 3. $\tau = \rho = 0$: Unrestricted constant, no quadratic trend in levels, linear trend in levels since $\gamma \neq 0$ and variables are stationary around a constant mean.

$$\Delta y_t = \alpha\left(\beta' y_{t-1} + \mu\right) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \gamma + \varepsilon_t$$

Deterministic terms in cointegrated processes

- VEC model from previous page is

$$\Delta y_t = \alpha\left(\beta' y_{t-1} + \mu + \rho t\right) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \gamma + \tau t + \varepsilon_t$$

  - Case 4. $\tau = 0$: No quadratic trends in levels, linear trend in levels and variables are trend stationary.

$$\Delta y_t = \alpha\left(\beta' y_{t-1} + \rho t\right) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \gamma + \varepsilon_t$$

  - Case 5. No restrictions on deterministic components, linear trend in first differences, quadratic trend in levels and variables are trend stationary.

$$\Delta y_t = \alpha\left(\beta' y_{t-1} + \rho t\right) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \gamma + \tau t + \varepsilon_t$$

- Default should be case 4! But case 3 is often used as the default! But, what is the definition of the cointegration vector in your economic model?

Matlab and Python syntax

- In Matlab we have the following syntax:
  - Case 1: ”H2”
  - Case 2: ”H1*”
  - Case 3: ”H1”
  - Case 4: ”H*”
  - Case 5: ”H”
- In Python we use the following syntax:
  - Case 1: ”n”
  - Case 2: ”co”
  - Case 3: ”ci”
  - Case 4: ”cili”
  - Case 5: ”colo”
- Note: In some Python illustrations and in the function to perform the Johansen cointegration test (to be discussed below) I also use numbers to refer to cases (for case = 3 I use model=3). See, JJ1992replication.py for an example.

Estimation of VARs with integrated variables

- Depends on assumptions about cointegration vectors
  - Known cointegration matrix: LS, ML, GLS
  - Unknown cointegration vectors: ML
- ML estimation known β. Example KL pp. 91, ML estimator in KL 3.2.2.
- LS estimation known β.
- ML estimation unknown β and cointegration tests. KL 3.2.2
- Matlab examples: LSML_knownBeta.m and VECMLcoint.m Python examples: VECknowncoint.py and VECMLcoint.py making use of own function LSKnownBeta included in the package you can download from the course homepage.

Estimating VEC models

- Let us reformulate the VEC model from the previous slides

$$\Delta y_t = \alpha\left(\beta' y_{t-1} + c_0 + d_0 t\right) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + c_1 + d_1 t + \varepsilon_t$$

- Consider then the 5 cases:
    - Case 1: No deterministic components $c_0 = c_1 = d_0 = d_1 = 0$.
    - Case 2: Constant in the cointegration vector $c_0 \neq 0$, $c_1 = d_0 = d_1 = 0$.
    - Case 3: Constant in the cointegration vector and constant in first differences (linear trend in levels) $c_0 \neq 0$, $c_1 \neq 0$, $d_0 = d_1 = 0$.
    - Case 4: Constant and linear trend in the cointegration vector and constant in first differences (linear trend in levels) $c_0 \neq 0$, $c_1 \neq 0$, $d_0 \neq 0$, $d_1 = 0$
    - Case 5: Constant and linear trend in the cointegration vector, constant in first differences (linear trend in levels) and linear trend in first differences (quadratic trend in levels) $c_0 \neq 0$, $c_1 \neq 0$, $d_0 \neq 0$, $d_1 \neq 0$
- When estimating VEC we need to define what case we aim to estimate!

ESTIMATING VEC MODELS

- We do this in Matlab and in Python in the following way:
- In Matlab we have the following syntax:
  - Case 1: ”H2”
  - Case 2: ”H1*”
  - Case 3: ”H1”
  - Case 4: ”H*”
  - Case 5: ”H”
- Python is using the following syntax:
  - Case 1: ”n”
  - Case 2: ”co”
  - Case 3: ”ci”
  - Case 4: ”cili”
  - Case 5: ”colo”

A NOTE ON STATSMODELS IN PYTHON

- Do not use Statsmodels (function VECM) to estimate VEC for unknown cointegration vector!
- It's impossible to verify the code for all model specifications (different assumptions about deterministic components), the Phillips normalization (to be illustrated later) is not always correct and the standard errors of normalized cointegration vector is incorrect.
- For later reference, the implementation of Johansen's cointegration test is incorrect.
- Instead, use the function jcitest included in the TraceTestNew.py file to estimate VEC for unknown cointegration vectors and the function LSknownBeta to estimate the VEC for known cointegration vector.

Estimating VEC models

- Illustration in Matlab (IllustrateMLandLS.m) and in Python (IllustrateMLandLS.py). Note that the functions LSknownBeta use the same syntax in Matlab and in Python.
- We first load some data.
- Determine the model including deterministic components.
- Estimate the VEC using ML for unknown cointegration vector.
- Print results.
- Then, given a specified rank, compute the cointegraiton vector depending on the model specification.
- Use this cointegration vector as a known vector and estimate VEC model using LS.

Illustrating Phillips normalization

- Illustration in Matlab (LMUnknownBetaNorm.m) and in Python (LMUnknownBetaNorm.py).
- The codes are equivalent and include the following steps.
  - We first load some data.
  - Determine the model including deterministic component.
  - Then we test for cointegration using the Johansen trace test. jcitest in Python and either jcitestOwn or the built-in function jcitest in Matlab.
  - Determine the rank, the number of cointegration vectors in the system.
  - For a certain rank, compute the cointegration vectors and use the LSknownBeta function to estimate the VEC for known cointegration vector.
  - Use the function VECMLHelp (same name and syntax in both Matlab and Python) to produce some matrices needed to compute standard errors.
  - Use the function PhillipsNorm to compute normalized cointegration vector, the implied speed of adjustment coefficients and the Π matrix together with standard errors.
  - In case you need to estimate the VEC model using the normalized cointegration vector, use LSknownBeta function again.

Testing for cointegration in a VAR model

- This section describes Johansen test based on the maximum likelihood estimation of the VEC model in (5) discussed in Johansen (1991), Johansen and Juselius (1990,1992) and in KL chapter 3.2.2.
- Consider the VEC model in equation (5):

$$\Delta y_t = \rho + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} - A(1) y_{t-1} + \varepsilon_t \tag{6}$$

  where as before $\Gamma_i = -(A_{i+1} + \ldots + A_p)$ for $i = 1, \ldots, p-1$ and $-A(1) = \alpha\beta' = \Pi$.
- To test for the number of cointegration vectors in the system we need an estimate of $A(1)$.
- Then we have to compute the rank of this matrix since rank$[A(1)]$ is equal to the number of nonzero eigenvalues which in turn is equal to the number of cointegration vectors r.
- In addition we need an estimate of the adjustment coefficients and the parameters in the cointegration vector.

Testing for cointegration in a VAR model

- Let $Z_{0,t} = \Delta y_t$, $Z_{1,t} = \left[\Delta y'_{t-1}, \ldots, \Delta y'_{t-p+1}, 1\right]'$ and $Z_{p,t} = y_{t-p}$. Compute the moment matrices

$$M_{ij} = T^{-1} \sum_{t=1}^{T} Z_{i,t} Z'_{j,t} \qquad i,j = 0, 1, p\ ,$$

residuals

$$R_{i,t} = Z_{i,t} - M_{i1} M_{11}^{-1} Z_{1,t} \qquad i = 0, p$$

and the sum of squares

$$S_{ij} = M_{ij} - M_{i1} M_{11}^{-1} M_{1j} \qquad i,j = 0, p.$$

Testing for cointegration in a VAR model

- To estimate the cointegration vectors in β we only have to solve the following eigenvalue problem
$$\mid \lambda S_{pp} - S_{p0} S_{00}^{-1} S_{0p} \mid = 0$$
and order the resulting eigenvalues such that $\hat{\lambda}_1 > \ldots > \hat{\lambda}_n$.
- A likelihood ratio test for the number of cointegration vectors is given by
$$-2 \ln Q(r \mid n) = -T \sum_{i=r+1}^{n} \ln(1 - \hat{\lambda}_i)$$
which is called the (Johansen's) Likelihood ratio trace test.

Testing for cointegration in a VAR model

- The next step is to compute the parameters in the cointegration vector β. This can be done if we solve for the eigenvectors V, i.e., we solve for V from

$$\left|\lambda S_{pp} - S_{p0} S_{00}^{-1} S_{0p}\right| V = 0.$$

- The eigenvectors $\hat{V} = (\hat{v}_1, \ldots, \hat{v}_n)$ can then be normalized such that $\hat{V}' S_{pp} \hat{V} = I$.
- The Maximum likelihood estimate of β is finally given by

$$\hat{\beta} = (\hat{v}_1, \ldots, \hat{v}_n)$$

  and the estimate of α is given by

$$\hat{\alpha} = S_{0p}\hat{\beta}.$$

- We have now a test for the number of cointegration vectors in the time series vector y_t (The LR-trace test), a procedure to estimate the adjustment coefficients in α and the parameters in the cointegration vectors β.
- Johansen also shows how to estimate the matrices $\hat{\Gamma}_i$.

Testing for cointegration in a VAR model

- How should the LR–trace test be used?
- If the null hypothesis that $\hat{\lambda}_{r+1} = \hat{\lambda}_{r+2} = \ldots = \lambda_K = 0$ cannot be rejected, then the system contains $K - r$ unit roots and therefore r cointegration vectors.
- To determine the number of vectors we make use of he following sequence of tests:
  - start by testing the hypothesis that there exists K unit roots. If this hypothesis is rejected it implies that $\lambda_1 > 0$ and we continue to the next null hypothesis which is that $\lambda_2 = \lambda_3 = \cdots = \lambda_K = 0$.
  - If this hypothesis is rejected we continue with the next null hypothesis $\lambda_3 = \cdots = \lambda_K = 0$ and so on.
  - A non–rejection of the null hypothesis determines the number of unit roots.
- Use correct critical values! Keep in mind that the asymptotic distribution of the trace statistic depends on the deterministic components in the model including dummy variables and dummy–type variables.
- An exception is centered seasonal dummies since they by construction sum to zero.
- Broken trends or shifts in the deterministic trend and the inclusion of weakly exogenous variables may also affect the asymptotic distribution.
- Many cases where standard critical values do not apply. Use judgment!

Hypotheses testing

- Consider the K-dimensional VEC model

$$\Delta y_t = \alpha \left(\beta' y_{t-1} + c_0 + d_0 t\right) + c_1 + d_1 t + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + u_t$$

- Hypotheses on α:
  - Variable i does not adjust to deviations from cointegration vector j.
  - Test whether $\alpha_{ij} = 0$ for $i = 1, \ldots, K$ and $j = 1, \ldots, r$.
  - Weak exogeneity test (equation i can be excluded from VEC model)
- Hypotheses on β:
  - Test whether one variable is stationarity around a constant mean (or linear trend).
  - Test whether one variable can be excluded from cointegration space.
  - Test whether a linear relation is stationary.
- Note 1: All these tests are asymptotically distributed as χ^2 as shown in Johansen (1991).
- Note 2: All tests are based on known r.

How to test for cointegration in Python/Matlab?

- Workflow:
  - Decide on the number of lags (always use one of the suggested methods in KL) and deterministic components, default model is restricted trend (a linear trend in the cointegration vector but no quadratic trends in levels). This corresponds to model 'H*' in Matlab or model "cili" in Python.
  - Use jcitest (or jcitestOwn in Matlab) to produce trace statistics and jcontest (or johcontest in Python) to test for specific cointegration vectors.

Workflow: Testing for cointegration

Step 1 Collect data and set up a VAR model.

(a) No need to pretest to asses whether all variables are integrated of order 1.

(b) Plot the data to see if deterministic trends should be included in the model. Are there structural breaks?

(c) Determine the lag length.

(d) Diagnostic tests, test for autocorrelation in the residuals, heteroscedasticity and normality. Both univariate and multivariate tests can be used.

Step 2 Estimate the model and compute the trace test to determine the rank of Π, i.e., the number of cointegration vectors in the system. It is useful to look at the estimated eigenvalues as they can provide additional information.

WORKFLOW: TESTING FOR COINTEGRATION

STEP 3 Test for stationarity, exclusion and weak exogeneity.

STEP 4 Normalize the estimated cointegration vectors and estimate the adjustment coefficients. Interpret the cointegration vectors, i.e., put a name on the vectors such for example PPP, a money demand function and so on.

STEP 5 We have now a fully specified VEC model. The next step is to analyze the model, for example test for Granger non–causality, identify a structural VAR model and estimate impulse responses and variance decomposition, the sources of the common stochastic trends and so on. Options are endless.

How to test for cointegration in Matlab?

- Use Matlab built in function!
  - "jcitest" to test for cointegration using Johansen ML estimator (or you can use the jcitestOwn.m function)
  - "jcontest" to test restrictions on α and on β
  - "estimate" to estimate VEC for unknown β but known r.
- Use own function (LSKnownBeta.m) to estimate VEC for known cointegration vectors.
- Johansen cointegration test on data matrix Y, for model 'H*' (restricted trend implying trend stationarity and no quadratic trends in levels) and p lags in the VEC (note: this is $p-1$):
  [h,pValue,stat,cValue,mles] = jcitest(Y,'model','H*','lags',p);
- To display full results, results for all possible r
  [h,pValue,stat,cValue,mles] = jcitest(Yl,'model','H*','lags',p,'display','full');
- Or use the jcitestOwn.m function:
  [Jres,c,c0,c1,d,d0,d1,beta,alpha]=jcitestOwn(x,p,model)

How to test for cointegration in Python?

- Use Python built in function to estimate VEC for unknown cointegratioon vector. Illustration: VECMLcoint.py
- Do not use the function to test for cointegration distributed with statsmodels! Use the function jcitest included in TracetestNew.py
- Johansen cointegration test on data matrix Y, for model 'H*' (restricted trend implying trend stationarity and no quadratic trends in levels) and p lags in the VEC (note: this is $p-1$):
  (lr1, cval5, cval10, pval, l, beta, alpha, c, c0, c1, d, d0, d1) = jcitest(data,p,model)
  Illustrated in VECMLcoint.py
- Note: Model is a number referring to the case, in this case model=4!

How to test restrictions on β and α in Python/Matlab?

- Restrictions on adjustment parameters α and cointegration vectors β can be implemented in Matlab using the function 'jcontest' and in Python johcontest (the same syntax works for both functions!)
- To test for specific cointegration vectors in Matlab: Use the built in function jcontest (illustrated below when replicating papers by Johansen and Juselius)
- To test for specific cointegration vectors in Python:
  (LRtest, pval, dof, Arest, Brest, c0, c1, d0, d1) =
  johcontest(Y,r,test,cons,model,lags,alpha)
  which is a Python function calling the built-in Matlab function jcontest and we use similar syntax! (illustrated below when replicating papers by Johansen and Juselius)
- Weak exogeneity (zero restrictions on α): use option 'Acon'.
- Exclusion from coint vec (zero restrictions on β): use option 'Bcon'
- Stationarity: use option 'Bvec'
- Specific cointegration vectors: either option 'Bvec' or 'Bcon'

ILLUSTRATIONS: JOHANSEN AND JUSELIUS (1990,1992)

Learning objective: Cointegration and unit root tests in practice.
Curriculum: Johansen (1991), Johansen and Juselius (1990,1992) and Python/Matlab examples.

- Johansen and Juselius (1990) estimating the Danish money demand function.
- Johansen and Juselius (1992) testing PPP and UIP.

Johansen and Juselius (1990)

- Estimation of money demand functions for Denmark and Finland.
- Illustrate trace test, tests on cointegration vectors and estimation under restrictions.

Johansen and Juselius (1990)

- A standard money demand function

$$\frac{M}{P} = Y^{\phi} \exp(-\psi i)$$

take logs (and add time sub-script)

$$m_t = \phi y_t - \psi i_i$$

- This is an equilibrium relation (a cointegration vector).
- We can add a constant term to capture time-invariant real money balance demand and add interest rates.
- Danish data: M2, real GDP (nominal GDP divided by GDP deflator), cost of holding money measured as the difference between bank deposit rates and the bond rate. Sample is 1974:1-1987:3.
- Transform the data into: $m =$ the log of real money balance, $y =$ logarithm of nominal GDP.
- VAR(2) model, constant in the cointegration vector (corresponding to model H1 in Matlab).

Johansen and Juselius (1990)

- Tests for autocorrelation and normality in Table 1.

TABLE 1

Some Test Statistics for the niid Assumption for the Residuals in the Model (1.2) with $k=2$

	The Danish data				*The Finnish data*			
	$\Delta m2$	Δy	Δi^b	Δi^d	$\Delta m1$	Δy	Δi^m	Δp
τ_1	7.15	11.48	10.57	7.34	11.30	19.21	4.30	6.99
τ_2	2.12	1.93	1.06	1.61	1.61	1.88	10.86	28.02
$\hat{\sigma}_\varepsilon$	0.019	0.019	0.007	0.005	0.045	0.029	0.034	0.011

where $\tau_1 = T\Sigma r_i^2 (i=1,\ldots,10) \sim \chi^2(10)$,

$$\tau_2 = \frac{T-m}{6}\left(SK^2 + \frac{EK^2}{4}\right) \overset{as}{\sim} \chi^2(2),$$

m is the number of regressors, SK is skewness and EK is excess kurtosis.
$\hat{\sigma}_\varepsilon$ is the standard error of regression estimate.

We cannot reject the null of no autocorrelation and normality for Danish data.

Johansen and Juselius (1990)

- Next step is to test for cointegration. Table 3 reports results.

TABLE 3

Test statistics for the hypothesis H_2^* and H_2 for various values of r versus r + 1 (λ_{max}) and versus the general alternative H_1 (trace) for the Danish and Finnish data. The 95% quantiles are taken from Table A2 (H_2) and A3 (H_2^*)

	The Danish data					*The Finnish data*			
H_2^*	*trace*	*trace (0.95)*	λ_{max}	λ_{max} *(0.95)*	H_2	*trace*	*trace (0.95)*	λ_{max}	λ_{max} *(0.95)*
$r \leq 3$	2.35	9.09	2.35	9.09	$r \leq 3$	3.11	8.08	3.11	8.08
$r \leq 2$	8.69	20.17	6.34	15.75	$r \leq 2$	11.01	17.84	7.90	14.60
$r \leq 1$	19.06	35.07	10.37	21.89	$r \leq 1$	37.65	31.26	26.64	21.28
$r = 0$	49.14	53.35	30.08	28.17	$r = 0$	76.14	48.42	38.49	27.34

- Use trace test!
- Null hypothesis that $r = 0$ cannot be rejected at the 5 percent level ($49.14 <$ critical value 53.35) also slightly below the critical value at the 10 percent level ($49.14 < 49.92$).
- Two alternatives: Either conclude that all variables are non-stationary or that there is one cointegration vector.

JOHANSEN AND JUSELIUS (1990)

- Always look at the eigenvalues!
- Table 2 reports eigenvalues.

TABLE 2

The Eigenvalues $\hat{\lambda}$ and Eigenvectors $\hat{V}$ as well as the Weights $\hat{W}$ for the Danish and Finnish Data

The Danish data

Eigenvalues $\hat{\lambda}^*$	(0.4332	0.1776	0.1128	0.0434	0)
Eigenvectors $\hat{V}^*$					
$m2$	1.00	1.00	1.00	1.00	1.00
y	−1.03	−1.37	−3.23	−1.88	−0.63
i^b	5.21	0.24	0.54	24.40	1.70
i^d	−4.22	6.84	−5.65	−14.30	−1.90
1	−6.06	−4.27	7.90	−2.26	−8.03
Weights $\hat{W}^*$					
$m2$	−0.21	0.00	0.04	0.00	0.00
y	0.12	0.02	0.05	0.00	0.00
i^b	0.02	−0.01	0.00	0.00	0.00
i^d	0.03	−0.03	0.00	0.00	0.00

The finnish data

Eigenvalues $\hat{\lambda}$	(0.3093	0.2260	0.0731	0.0295)
Eigenvectors $\hat{V}$				
$m1$	1.00	0.04	1.00	1.00
y	−0.98	−0.06	−0.92	1.61
i^m	−7.09	−0.09	−0.26	−1.38
Δp	−7.02	1.00	−1.82	−15.69
Weights $\hat{W}$				
$m1$	0.03	−0.46	−0.13	0.00
y	0.02	−0.13	0.01	−0.01
i^m	0.05	1.07	−0.01	0.00
Δp	0.01	−0.39	0.01	0.00

First eigenvalue is relatively large (0.4332) compared to the second eigenvalue (0.1776). Can be interpreted as suggesting 1 cointegration vector.

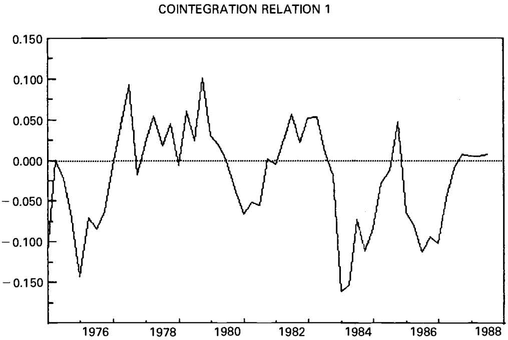
- Decision: There is one cointegration vector in the Danish system.

Johansen and Juselius (1990)

- The estimated unrestricted cointegration vector (β) is the first eigenvector in Table 2.
$$m2 = 1.03y - 5.21i^{b} + 4.22i^{d} + 6.06$$
- Adjustment coefficients (α) are equal to the weights associated with the first cointegration vector reported in Table 2.
$$\hat{\alpha}' = (-0.213, 0.115, 0.023, 0.029)$$

JOHANSEN AND JUSELIUS (1990)

- Cointegration vector shown in Figure 3.



(e)

Johansen and Juselius (1990)

- Test hypotheses on cointegration vector.

$$0 = \beta_{1,1} m2 + \beta_{1,2} y + \beta_{1,3} i^b + \beta_{1,4} i^d + \beta_{1,0}$$

  - The coefficients associated with money and output are equal but with opposite signs, $\beta_{1,1} = -\beta_{1,2}$.
  - The coefficients associated with the bond rate and the deposit rate are equal but with opposite signs, $\beta_{1,3} = -\beta_{1,4}$
- In addition, we may want to test hypotheses on α.

Johansen and Juselius (1990)

- Results in Table 6.

TABLE 6

Estimates for the Danish Data of the First Eigenvalue and the Corresponding $\boldsymbol{\beta}$- and $\boldsymbol{\alpha}$-vector under various hypotheses about $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$

	H_2^*	$H_{3.1}^*$	$H_{3.2}^*$	$H_{5.1}^*$	$H_{5.2}^*$	$H_{5.3}^*$	$H_{4.1}^*$
β-restrictions	—	$\beta_1=-\beta_2$	$\beta_1=-\beta_2$ and $\beta_3=-\beta_4$	$\beta_1=-\beta_2$ and $\beta_3=-\beta_4$	$\beta_1=-\beta_2$ and $\beta_3=-\beta_4$	$\beta_1=-\beta_2$ and $\beta_3=-\beta_4$	—
α-restrictions	—	—	—	$\alpha_3=0$	$\alpha_3=\alpha_4=0$	$\alpha_3=\alpha_4=\alpha_2=0$	$\alpha_3=\alpha_4=\alpha_2=0$
$\hat{\lambda}_1^*$	0.433	0.432	0.423	0.410	0.356	0.286	0.357
$-T\ \ln(1-\hat{\lambda}_1^*)$	30.09	30.04	29.15	27.96	23.34	17.91	23.42
β_1	1.00	1.00	1.00	1.00	1.00	1.00	1.00
β_2	−1.03	−1.00	−1.00	−1.00	−1.00	−1.00	−0.96
β_3	5.21	5.30	5.88	5.95	5.81	5.88	4.76
β_4	−4.22	−4.29	−5.88	−5.95	−5.81	−5.88	−2.57
β_5	−6.06	−6.26	−6.21	−6.22	−6.21	−6.21	−6.58
α_1	−0.213	−0.212	−0.177	−0.152	−0.137	−0.197	−0.254
α_2	0.115	0.108	0.095	0.099	0.139	0	0
α_3	0.023	0.022	0.023	0	0	0	0
α_4	0.029	0.030	0.032	0.029	0	0	0

Johansen and Juselius (1990)

- Replication in Matlab: JJ1990replication.m Replication in Python: JJ1990replication.py

Johansen and Juselius (1992)

- Focus on PPP and UIP. Also summarize ML estimation and tests.
- PPP states that

$$e_{12} = p_1 - p_2$$

  wheres UIP states that

$$e^e_{12} - e_{12} = i_1 - i_2$$

  Note: variables in logs.
- Two equilibrium relations conditional on no transaction costs, risk neutrality and rational expectations.
- In a system comprised of 5 variables, we expect to find two cointegration vectors, PPP and UIP.

Johansen and Juselius (1992)

- Potential problem: PPP in levels form includes the nominal exchange rate whereas UIP includes the (expected) change in the nominal exchange rate.
- JJ solves this by stating that under the assumption of constant inflation differential, then there is no change in the nominal exchange rate and interest differential is unchanged.
- It then follows that there should be two cointegration vectors in the time series vector

$$X_t = \begin{bmatrix} p_1 & p_2 & e_{12} & i_1 & i_2 \end{bmatrix}$$

defined as

$$\beta_1 = \begin{bmatrix} 1 & -1 & -1 & 0 & 0 \end{bmatrix}$$

and

$$\beta_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

Johansen and Juselius (1992)

- Note: If we combine PPP in its relative form and UIP we find that the real interest differentials in two countries must be equal in the long-run. This is called Real Interest Rate Parity. Allows for another way to jointly test PPP and UIP.
- UK data: Wholesale price, trade-weighted foreign wholesale price, effective nominal exchange rate, 3-month T-bill rate. Foreign interest rate is the 3-month eurodollar interest rate. All variables are in logarithms and the sample is 1972:1-1987:2.
- Add oil price changes as an exogenous variable to the model and assume that the lag length is 2.

Johansen and Juselius (1992)

- Residual tests in Table 1

Table 1

Residual misspecification tests in model (13).

Eq.	Standard deviation	Skewness	Excess kurtosis	Normality test $\chi^2(2)$	Autocorr. test $\chi^2(20)$
1	0.007	0.29	1.27	4.84	6.09
2	0.007	0.28	2.16	12.44	9.59
3	0.030	0.30	0.17	0.95	13.54
4	0.011	0.58	0.25	3.55	9.11
5	0.013	−0.51	3.76	37.95	16.41

Excess Kurtosis in equations 2 (foreign price level) and 5 (eurodollar interest rate).

Johansen and Juselius (1992)

- Cointegration tests in Table 2.

Table 2
Tests of the cointegration rank.

i	$\hat{\lambda}_i$	$-T\ln(1-\hat{\lambda}_i)$	$\hat{\lambda}_{max}(0.95)$	$-T\sum\ln(1-\hat{\lambda}_i)$	$\hat{\lambda}_{trace}(0.95)$
1	0.407	31.33	33.18	80.75	68.91
2	0.285	20.16	27.17	49.42	47.18
3	0.254	17.59	20.78	29.26	29.51
4	0.102	6.48	14.03	11.67	15.20
5	0.083	5.19	3.96	5.19	3.96

Results suggest at least 2 possibly 3 cointegration vectors.

- JJ decides on 2 cointegration vectors guided by economic theory, eigenvalues and the trace test.

Johansen and Juselius (1992)

- Unrestricted cointegration vectors in Table 3.

Table 3

The estimated eigenvectors of (7) partitioned into the stationary components $(\beta_{0.1}, \beta_{0.2})$ and their weights $(\hat{\alpha}_{0.1}, \hat{\alpha}_{0.2})$ together with the remaining eigenvectors $(\hat{v}_3, \hat{v}_4, \hat{v}_5)$ and the corresponding weights $(\hat{w}_3, \hat{w}_4, \hat{w}_5)$ defined as $\hat{w}_i = S_{0k} v_i$; see (9).

Eigenvectors				
$\hat{\beta}_{0.1}$	$\hat{\beta}_{0.2}$	$\hat{v}_3$	$\hat{v}_4$	$\hat{v}_5$
1.00	0.03	0.36	1.00	1.00
−0.91	−0.03	−0.46	−2.40	−1.45
−0.93	−0.10	0.41	1.12	−0.48
−3.38	1.00	1.00	−0.41	2.28
−1.89	−0.93	−1.03	2.98	0.76

Weights				
$\hat{\alpha}_{0.1}$	$\hat{\alpha}_{0.2}$	$\hat{w}_3$	$\hat{w}_4$	$\hat{w}_5$
−0.07	0.04	−0.01	0.00	−0.01
−0.02	0.00	−0.04	0.01	0.01
0.10	−0.01	−0.15	−0.04	−0.05
0.03	−0.15	0.03	0.01	−0.02
0.06	0.29	0.01	0.03	−0.01

Note: PPP present in first eigenvector and stationary interest rate differential in second eigenvector.

Johansen and Juselius (1992)

- Estimates of $\Pi = \alpha\beta'$, a measure of the importance of the cointegration vectors in each equation, in Table 4.

Table 4

The estimates of $\Pi = \alpha\beta'$ for $\mathscr{H}_1(2)$.

Eq.	p_1	p_2	e_{12}	i_1	i_2
1	−0.067	0.061	0.060	0.272	0.090
2	−0.018	0.016	0.016	0.064	0.030
3	0.101	−0.091	−0.093	−0.345	−0.186
4	0.030	−0.026	−0.018	−0.263	0.072
5	0.066	−0.062	−0.082	0.097	−0.382

  Note: Signs consistent with PPP and strong effects in exchange rate equation. UIP in equations 4 and 5 (the two interest rate equations).

- Proceed to test theoretical restriction on the cointegration space.

Johansen and Juselius (1992)

- Tests for weak exogeneity (section 4). Find that the foreign price is weakly exogenous but the eurodollar interest rate is not weakly exogenous.
- Can reject PPP but not UIP. Find that a linear stationary relation between prices and exchange rates cannot be rejected.
- Replication using updated data. Matlab example: JJ1992replication.m Python example: JJ1992replication.py
- Next topic: Identification of structural VARs, KL 4.1-4.3.

Advanced Macroeconomics: Structural Vector Autoregressive Analysis

Lectures 2 to 5

U. Michael Bergman
University of Copenhagen

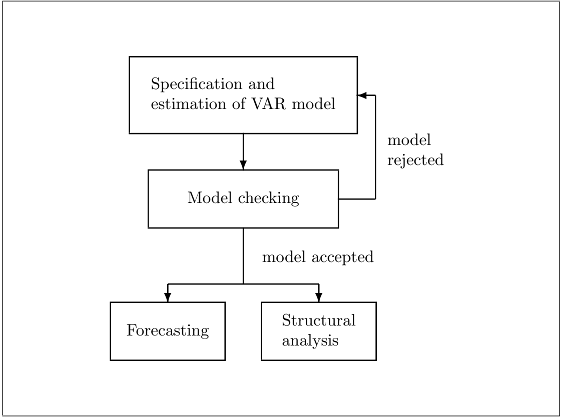
Fall 2024

Learning objective:

Curriculum: KL ch. 2.2-2.3 (except 2.3.3-2.3.6), 2.5-2.7, and lecture note on stable VAR models

1. LS: from sums to vectors/matrices.
2. What is a VAR model?
3. Properties of stationary VAR models.
4. Moving Average Representation.
5. Estimation (MATLAB/Python).
6. Simulate VAR models and use suggested estimators.
7. Determining the lag length.
8. Granger non-causality hypothesis.
9. Empirical Example: Sims (1972) Granger non-causality tests.

VAR analysis



Linear VAR processes

- Consider the following bivariate Vector Auto Regressive (VAR) system with one lag

$$x_t = b_{10} - b_{12} z_t + \gamma_{11} x_{t-1} + \gamma_{12} z_{t-1} + \epsilon_{xt} \tag{1}$$

$$z_t = b_{20} - b_{21} x_t + \gamma_{21} x_{t-1} + \gamma_{22} z_{t-1} + \epsilon_{zt} \tag{2}$$

- Note that we in this model have contemporaneous effect from z_t to x_t in equation (1)) and an opposite effect in equation (2).
- Rewrite (1) and (2) in the following way

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} x_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{xt} \\ \epsilon_{zt} \end{bmatrix} \tag{3}$$

LINEAR VAR PROCESSES

- Premultiply both sides with the matrix

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}^{-1} \tag{4}$$

and let $y_t = \begin{bmatrix} x_t & z_t \end{bmatrix}'$ so that we obtain

$$y_t = B^{-1}\Gamma_0 + B^{-1}\Gamma_1 y_{t-1} + u_t \tag{5}$$

where $\Gamma_0 = \begin{bmatrix} b_{10} & b_{20} \end{bmatrix}'$ and

$$\Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$$

and

$$u_t = \begin{bmatrix} u_{xt} \\ u_{zt} \end{bmatrix} = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \epsilon_{xt} \\ \epsilon_{zt} \end{bmatrix}$$

- which can also be written as

$$y_t = A_0 + A_1 y_{t-1} + u_t.$$

This is the linear VAR(1) model!

- Note 1: Since the residuals in equations (1) and (2) are assumed to be white noise, it follows that the transformed residuals in u_t also are white noise.
- To show this, we have that

$$u_{xt} = \frac{\epsilon_{xt} - b_{12}\epsilon_{zt}}{1 - b_{12}b_{21}}$$

and

$$u_{zt} = \frac{\epsilon_{zt} - b_{21}\epsilon_{xt}}{1 - b_{12}b_{21}}$$

so that $\mathrm{E}\left[u_{xt}\right] = \mathrm{E}\left[u_{zt}\right] = 0$,

$$\mathrm{E}\left[u_{xt}^2\right] = \left(\frac{1}{1 - b_{12}b_{21}}\right)^2 \left(\sigma_{\epsilon_x}^2 + b_{12}^2\sigma_{\epsilon_z}^2\right),$$

$$\mathrm{E}\left[u_{zt}^2\right] = \left(\frac{1}{1 - b_{12}b_{21}}\right)^2 \left(\sigma_{\epsilon_z}^2 + b_{21}^2\sigma_{\epsilon_x}^2\right)$$

and $\mathrm{E}\left[u_{xt}u_{xt-i}\right] = \mathrm{E}\left[u_{zt}u_{zt-i}\right] = 0$ for all $i \neq 0$, and

$$\mathrm{E}\left[u_{xt}u_{zt}\right] = \left(\frac{1}{1 - b_{12}b_{21}}\right)^2 \left(-b_{21}\sigma_{\epsilon_x}^2 - b_{12}\sigma_{\epsilon_z}^2\right) \neq 0.$$

Linear VAR processes

- Note 2: Note that $\mathrm{E}\left[u_{xt}u_{zt}\right] = 0$ when $b_{12} = b_{21} = 0$, i.e., when there is no contemporaneous effect between the variables x_t and z_t in the VAR model.

Linear VAR processes

- Assume now that y_t is a K dimensional time series vector and that the lag length is p.
- The VAR model can now be written as

$$y_t = \nu + A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + u_t \tag{6}$$

where $\nu = A_0$,

- and where

$$\begin{array}{rcll} y_t & = & \begin{bmatrix} y_{1,t} & \cdots & y_{Kt} \end{bmatrix}' & (K \times 1) \\ A_i & = & \begin{bmatrix} a_{11,i} & \cdots & a_{1K,i} \\ \vdots & & \vdots \\ a_{K1,i} & \cdots & a_{KK,i} \end{bmatrix} & (K \times K) \\ \nu & = & \begin{bmatrix} \nu_1 & \cdots & \nu_K \end{bmatrix}' & (K \times 1) \\ u_t & = & \begin{bmatrix} u_1 & \cdots & u_K \end{bmatrix}' & (K \times 1) \end{array}$$

- and the residuals satisfies

$$\begin{array}{rcl} \mathrm{E}[u_t] & = & 0 \\ \mathrm{E}[u_t u_t'] & = & \Sigma_u \\ \mathrm{E}[u_t u_s'] & = & 0 \quad \text{for} \quad s \neq t \end{array}$$

where Σ_u is non–singular.

- Define the lag operator L in the following way:

$$L^i y_t \equiv y_{t-i}$$

- Define the matrix polynomial in the lag operator

$$A(L) = I_K - A_1 L - \cdots - A_p L^p$$

such that the VAR model can be written as

$$A(L) y_t = u_t$$

Definition

A VAR(p) process is stable if the reverse characteristic polynomial

$$\det\left(I_K - Az\right) = \det\left(I_K - A_1 z - \ldots - A_p z^p\right) \neq 0 \quad \mathit{for} \quad | z | \leq 1$$

does not have any roots inside or on the complex unit circle.

- The VAR(p) model can be written as the VAR(1)

$$Y_t = \nu + \mathbf{A} Y_{t-1} + U_t$$

where

$$\underset{Kp \times 1}{\nu = \begin{bmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{bmatrix}}, \quad \underset{Kp \times Kp}{\mathbf{A} = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_K & 0 & \dots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_K & 0 \end{bmatrix}}, \text{ and } \quad \underset{Kp \times 1}{U_t = \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}} \tag{7}$$

- Note: The matrix $\mathbf{A}$ is called the *companion matrix*.
- Matlab function: comp.m Python function: companion in Ownfunctions.py

Definition

Let $|\lambda_{Kp}| \geq |\lambda_{Kp-1}| \geq \ldots \geq |\lambda_1| \geq 0$ *be the eigenvalues of the matrix A defined in (7). The VAR(p) model is stable and invertible if and only if* $|\lambda_{Kp}| < 1$*.*

These definitions allow us to define stationarity in the following way:

Definition
A vector time series y_t *is stationary if (i)* $E[y_t] = \mu_t = \mu$*, and (ii)* $E[(y_t - \mu)(y_{t-k} - \mu)'] = \Upsilon_k$ *for all* t *and* k*.*

In order to avoid misunderstandings we also introduce the definition:

Definition
A vector time series y_t *is trend–stationary if* $\tilde{y}_t \equiv y_t - \mu_t$ *is stationary.*

These definitions also lead us to the following definition:

Definition
A stable VAR(p) process y_t *is stationary.*

Example: Checking stability condition

- Consider the following 3–dimensional VAR(1) model

$$y_t = \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} y_{t-1} + \varepsilon_t.$$

- The reverse characteristic polynomial is

$$\det\left[\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{bmatrix} z\right] = \det\left(\begin{bmatrix} 1-0.5z & 0 & 0 \\ -0.1z & 1-0.1z & -0.3z \\ 0 & -0.2z & 1-0.3z \end{bmatrix}\right) =$$

$$(1-0.5z)(1-0.1z)(1-0.3z) - (1-0.5z)(-0.3z)(-0.2z) = (1-0.5z)\left(1-0.4z-0.03z^2\right)$$

  implying that the roots are: $z_1 = 2$, $z_2 = 2.1525$, and $z_3 = -15.486$.

- Since the absolute value of all roots are greater than unity (no root inside or on the unit circle), the process is stable.

Example: Checking stability condition

- Consider the following VAR(2) process

$$y_t = \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} y_{t-2} + \varepsilon_t.$$

  The reverse characteristic polynomial is

$$\det \left[\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} z - \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} z^2 \right] = 0.21z^2 - z - 0.025z^3 + 1$$

  which implies the roots: $z_1 = 1.3, z_2 = 3.55 - 4.2623i, z_3 = 3.55 + 4.2623i$.
- Modulus for z_2 and z_3 are: $|z_2| = |z_3| = \sqrt{3.55^2 + 4.2623^2} = 5.547$. This process is also stable since all roots lie outside the unit circle.

EXAMPLE: CHECKING STABILITY CONDITION

- An alternative to use the reverse characteristic polynomial is the compute the eigenvalues of the companion matrix

$$\begin{bmatrix} 0.5 & 0.1 & 0 & 0 \\ 0.4 & 0.5 & 0.25 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

  which gives us the solutions $\lambda_1 = -0.76926$,$\lambda_2 = -0.11537 + 0.13852i$,$\lambda_3 = -0.11537 - 0.13852i$ and $\lambda_4 = 0$. Modulus for λ_2 and λ_3 is $\sqrt{0.11537^2 + 0.13852^2} = 0.1803$. The absolute values of all eigenvalues are less than one and the system is stable.
- Check stability using MATLAB: StabilityVAR1.m or use function stabVAR.m. Python: StabilityVAR1.py or use function stabVar included in Ownfunctions.py

Vector Moving Average representation

- Under the assumption that the VAR model is stable, we can rewrite our model as a moving average model (VMA)

$$
\begin{aligned}
A(L)y_t &= \nu + u_t \\
y_t &= A(L)^{-1}(\nu + u_t) \\
y_t &= A(1)^{-1}\nu + A(L)^{-1}u_t \\
y_t &= \delta + \Phi(L)u_t
\end{aligned}
\tag{8}
$$

  where $\Phi(L) = I_K + \sum_{i=1}^{\infty} \Phi_i L^i$ and $\Phi(L)A(L) = I_K$.
- Note: Since $A(L)$ is non–singular we find that $\Phi(L)$ is absolutely summable such that $\Phi(L)u_t$ is well–defined.
- How can we compute $\Phi(L)$?

Vector Moving Average representation

- Use the relationship $\Phi(L)A(L) = I_K$, such that

$$I_K = (\Phi_0 + \Phi_1 L + \Phi_2 L^2 + \ldots)(I_K - A_1 L - A_2 L^2 - \ldots - A_p L^p)$$

  which implies that

$$I_K = \Phi_0 + (\Phi_1 - \Phi_0 A_1)L + (\Phi_2 - \Phi_1 A_1 - \Phi_0 A_2)L^2 + \cdots$$

  which further implies that

$$\begin{aligned} I_K &= \Phi_0 \\ 0 &= \Phi_1 - \Phi_0 A_1 \\ 0 &= \Phi_2 - \Phi_1 A_1 - \Phi_0 A_2 \\ &\vdots \end{aligned}$$

  where $A_j = 0$ for $j > p$.

- Φ_i can be computed recursively using the relationship

$$\begin{aligned} \Phi_0 &= I_K \\ \Phi_i &= \textstyle\sum_{j=1}^{i} \Phi_{i-j} A_j \quad \text{for} \quad i = 1, 2, \ldots \end{aligned}$$

Vector Moving Average representation

- The mean of y_t, i.e., μ, can be computed using the following expression

$$\mu = \Phi(1)\nu = A(1)^{-1}\nu.$$

Vector Moving Average representation

- An alternative is to use the companion matrix **A** defined in equation (7).
- Let

$$J = \begin{bmatrix} I_K & 0 & \cdots & 0 \end{bmatrix}$$

  which has the dimension $K \times Kp$. Matlab function: jmatrix.m Φ. Python function: Jmatrix in Ownfunctions
- The VMA representation can now be computed using the following expression

$$\Phi_i = JA^i J' \tag{9}$$

  where the matrices Φ_i have the dimension $K \times K$.
- Matlab illustration: VARtoVMA3.m Python: VARtoVMA3.py

ESTIMATION OF VAR MODELS

- Multivariate Least Squares estimator.
- Different ways to formulate the LS estimator, either based on VAR(1) model, as standard OLS, based on the VAR(1) representation or using the companion form.

Estimation of VAR models: Based on VAR(1)

$$\hat{A} = YZ'(ZZ')^{-1}$$

where

$$Y \equiv \begin{bmatrix} y_{1,p} & y_{1,p+1} & \cdots & y_{1,T} \\ y_{2,p} & y_{2,p+1} & \cdots & y_{2,T} \\ \vdots & \vdots & \vdots & \vdots \\ y_{K,p} & y_{K,p+1} & \cdots & y_{K,T} \end{bmatrix},$$

$$\hat{A} \equiv \begin{bmatrix} \hat{\nu} & \hat{A}_1 & \hat{A}_2 & \ldots & \hat{A}_p \end{bmatrix}$$

and

$$Z \equiv \begin{bmatrix} 1 & 1 & \ldots & 1 \\ y_{1,p-1} & y_{1,p} & \cdots & y_{1,T-1} \\ y_{2,p-1} & y_{2,p} & \cdots & y_{2,T-1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{K,p-1} & y_{K,p} & \cdots & y_{K,T-1} \\ y_{1,p-2} & y_{1,p-1} & \cdots & y_{1,T-2} \\ y_{2,p-2} & y_{2,p-1} & \cdots & y_{2,T-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{K,p-2} & y_{K,p-1} & \cdots & y_{K,T-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1,0} & y_{1,1} & \cdots & y_{1,T-p} \\ y_{2,0} & y_{2,1} & \cdots & y_{2,T-p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{K,0} & y_{K,1} & \cdots & y_{K,T-p} \end{bmatrix}$$

$$\hat{U} = Y - \hat{A}Z$$

$$\widehat{\Sigma}_u = \frac{\hat{U}\hat{U}'}{T - Kp - 1}$$

where

$$U \equiv \begin{bmatrix} u_{1,p} & u_{1,p+1} & \cdots & u_{1,T} \\ u_{2,p} & u_{2,p+1} & \cdots & u_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ u_{K,p} & u_{K,p+1} & \cdots & u_{K,T} \end{bmatrix}$$

and covariance matrix of A

$$(ZZ')^{-1} \otimes \hat{\Sigma}_u$$

Estimation of VAR models: Standard Multivariate LS

Let $z = Z'$ and $y = Y'$, then

$$\hat{A} = (z'z)^{-1}z'y$$

$$\hat{u} = y - z\hat{A}$$

$$\hat{\Sigma}_u = \frac{\hat{u}'\hat{u}}{T - Kp - 1}$$

and covariance of A is

$$\hat{\Sigma}_u \otimes (z'z)^{-1}$$

Matlab illustration: LSestimators.m and Python LSestimators.py

Matlab users: Use function VARls.m as illustrated in LSestimators.m to estimate the models or built-in function estimate. Python users: use model.fit available in statsmodels (but do not use the built-in tests) illustrated in VARGCtest.py.

Simulate VAR models

- We will now simulate a VAR model and then estimate that model using the estimators suggested previously.
- Main question: How good are these estimators?
- A controlled experiment where we specify parameters in the VAR model and then we estimate the parameters and check whether we obtain correct estimates.
- To do this, we need to find out how to simulate the VAR model.
- We will use the Standard residual-based recursive-design approach (also used later in the course).

STANDARD RESIDUAL-BASED RECURSIVE-DESIGN BOOTSTRAP

- Assume that the DGP is a VAR (p) model (lag order p is known)

$$y_t = \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t$$

- Basic assumptions: $\mathbb{E}(u_t) = 0$ and that u_t has finite moments.
- We need to specify the lag length p, the parameters ν, $A_1, \ldots, A_p$ and Σ_u.

STANDARD RESIDUAL-BASED RECURSIVE-DESIGN BOOTSTRAP

- Draw residuals u_t (based on a distribution).
- If $\nu = 0$, demean u_t^*.
- Given initial conditions $\left[y^*_{-p+1}, \ldots, y^*_0\right]$ we can generate recursively the realizations $\{y_t^*\}_{t=-p+1}^{T}$. We assume that all initial conditions are equal to zero.
- Estimate the VAR(p) model using the simulated data and compare to the specified model.

- Assume that we want to simulate the following VAR model

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.2 & -0.1 & 0.5 \\ -0.4 & 0.2 & 0 \\ -0.1 & 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + u_t$$

where

$$\Sigma_u = u_t' u_t = \begin{bmatrix} 1 & 0.2 & 0.4 \\ 0.2 & 1 & 0.3 \\ 0.4 & 0.3 & 1 \end{bmatrix}$$

for the sample $t = 1, \ldots, T$. The dimension $K = 3$.

Standard residual-based recursive-design bootstrap

- How to generate Σ_u?
- First we draw $T \times K$ i.i.d. random numbers, this is u_t.
- Then we transform these residuals such that $\hat{\Sigma_u} = \Sigma_u$.
- To do this, use the following procedure:
  - Subtract the mean from u, $\hat{u} = u - mean(u)$ and then standardize such that $\tilde{u} = \hat{u} * (Chol(cov(\hat{u})))^{-1}$.
  - Final step is to compute $\tilde{u} = \tilde{u} * Chol(\Sigma_u)$.
  - To check that this procedure produces residuals with the assumed properties, use the Python/Matlab examples.
- We now have the residuals and we assume that initial conditions (the first observations) are equal to zero. This allows us to compute the time series vector at $t = 1$.
- Then we can recursively generate the time series vector for all time periods up until T.

Standard residual-based recursive-design bootstrap: Python/Matlab

- MATLAB example: IllustrateSimulation.m Python example will be provided during the exercise.
- Use the simulated time series vector and estimate the VAR and compare.

Specification of VAR models

- Lag length p?
- Check stability!
- Unit roots and cointegration?
- Properties of the residuals? (Autocorrelation, heteroscedasticity, normality)

DETERMINE LAG LENGTH p

- Possible procedures
  - Top-down sequential testing (general-to-specific): Start with a maximum number of lags p_{max} testing a sequence of null hypotheses: $\mathbb{H}_0$: $A_{p_{\max}} = 0$ vs. $\mathbb{H}_1$: $A_{p_{\max}} \neq 0$, $\mathbb{H}_0$: $A_{p_{\max}-1} = 0$ vs. $\mathbb{H}_1$: $A_{p_{\max}-1} \neq 0$, ..., $\mathbb{H}_0$: $A_1 = 0$ vs. $\mathbb{H}_1$: $A_1 \neq 0$. Process terminates when there is a rejection. Use Wald or LR tests.
  - Bottom-up sequential testing (specific-to-general): Start with $p_{\min}$ testing for autocorrelation in the residuals (using for example a multivariate test). Add lags until there is no significant autocorrelation.
  - Information criteria (alternative to sequential testing): The three most commonly used criteria are; Akaike Information Criterion (AIC), Hannan-Quinn Criterion (HQC) and Schwarz Information Criterion (SIC). Idea: Trade-off between model fit and parsimony.

Top-down sequential testing

- Use either a Wald test to test for parameter restrictions or an LR test testing a VAR(m) model against a VAR($m+1$). (Discussed in section 2.6.1)
- MATLAB illustration: TopDownSequence.m Python illustration: TopDownSequence.py
- Use Portmanteu or LM (Part of HW1) tests for residual autocorrelation (discussed in section 2.6.2 and part of HW1).
- Multivariate non-normality tests (Part of HW1).

Information Criteria

Akaike Information Criterion:

$$\mathrm{AIC}(m) = \ln(det(\tilde{\Sigma}_u(m))) + \frac{2}{T}(mK^2 + K)$$

Hannan-Quinn Criterion:

$$\mathrm{HQC}(m) = \ln(det(\tilde{\Sigma}_u(m))) + \frac{2\ln(\ln(T))}{T}(mK^2 + K)$$

Schwarz Information Criterion:

$$\mathrm{SIC}(m) = \ln(det(\tilde{\Sigma}_u(m))) + \frac{\ln(T)}{T}(mK^2 + K)$$

where m is the lag order.

Note: $\hat{p}^{\mathrm{SIC}} \leq \hat{p}^{\mathrm{HQC}} \leq \hat{p}^{\mathrm{AIC}}$

Part of HW1

GRANGER NON-CAUSALITY

Curriculum: KL section 2.5, 7.2-7.4 and lecture note on stable VAR models.

- Definition
- Standard formulation in bivariate VAR and MATLAB example.
- What happens in higher dimensional VARs?

Granger non-causality

- Granger causality $\neq$ causality in the usual sense!
- Granger causality is about prediction! Can z_t be predicted more efficiently when also taking into account the variable x_t?
- Then x_t Granger causes z_t!

- Let Ω_t be the information set that includes all available and relevant information at time t.
- Let $z_t(h|\Omega_t)$ be the optimal h-steps predictor of the time series process z_t at time t based on all information in Ω_t.
- The expected value of the squared forecast error (MSE) is denoted $\Sigma_z(h|\Omega_t)$.
- Then, we can define Granger causality as follows.

Theorem

(Granger, 1969) *The process x_t is said to cause z_t in Granger's sense if*

$$\Sigma_z(h \mid \Omega_t) < \Sigma_z(h \mid \Omega_t \setminus \{x_s \mid s \leq t\}) \quad \textit{for at least one} \quad h = 1, 2, \ldots$$

where $\Omega_t \setminus \{x_s \mid s \leq t\}$ is the information set excluding historical and current information about the process x_t.

- If z_t can be predicted more efficiently when we also take into account the variable x_t, then x_t Granger causes z_t.

- Start with a bivariate VAR(p) process written in its VMA form

$$y_t = \delta + \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t-i} = \delta + \Phi(L)\varepsilon_t \tag{10}$$

where $\Phi_0 = I_n$ and $\mathrm{E}[\varepsilon_t \varepsilon_t'] = \Sigma_\varepsilon$ and $y_t = \begin{bmatrix} z_t & x_t \end{bmatrix}'$.

- The VMA model can also be written as

$$\begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}. \tag{11}$$

- The optimal 1–step predictor of z_t based on historical values of both z_t and x_t is

$$z_t(1 \mid \{y_s \mid s \le t\}) = \delta_1 + \sum_{i=1}^{\infty} \Phi_{11,i}\varepsilon_{1,t+1-i} + \sum_{i=1}^{\infty} \Phi_{12,i}\varepsilon_{2,t+1-i}.$$

- The optimal 1–step predictor of z_t based only on historical values of z_t is

$$z_t(1 \mid \{z_s \mid s \le t\}) = \delta_1 + \sum_{i=1}^{\infty} \Phi_{11,i}\varepsilon_{1,t+1-i}.$$

Characterization of Granger causality

- If
$$z_t(1 \mid \{y_s \mid s \leq t\}) = z_t(1 \mid \{z_s \mid s \leq t\})$$
then this implies that
$$\Phi_{12,i} = 0 \qquad \text{for} \quad i = 1, 2, \ldots.$$
- Interpretation: x_t does not improve the prediction of z_t. Therefore, x_t does not Granger cause z_t.
- A test of the null hypothesis that x_t does not Granger cause z_t can then be performed by testing whether the moving average parameters $\Phi_{12,i}$ are significantly different from zero.
- Potential problem 1: How many lags i to include in the test?
- Potential problem 2: How to compute covariance matrix of $\Phi_{12,i}$?

Characterization of Granger causality

- Number of lags i: Lütkepohl suggests the following result.

  Theorem Lütkepohl (2005)
  If y_t is a stable K dimensional VAR(p) process, then for $j \neq k$,

  $$\Phi_{jk,i} = 0 \qquad \textit{for} \;\; i = 1, 2, \ldots$$

  is equivalent to $\Phi_{jk,i} = 0 \;\; \textit{for} \;\; i = 1, 2, \ldots, p(K-1)$.
- Need to compute variance-covariance matrix of Φ. Lütkepohl (1991) derives expressions that can be used (also found in lecture note).

Characterization of Granger causality

- Alternative approach to test for Granger causality in bivariate models.
- The bivariate VMA model (with infinite lags) can be written as the VAR(p) model

$$\begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} + \begin{bmatrix} A_{11,1} & A_{12,1} \\ A_{21,1} & A_{22,1} \end{bmatrix} \begin{bmatrix} z_{t-1} \\ x_{t-1} \end{bmatrix} + \ldots + \begin{bmatrix} A_{11,p} & A_{12,p} \\ A_{21,p} & A_{22,p} \end{bmatrix} \begin{bmatrix} z_{t-p} \\ x_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

- If $\Phi_{12}(L) = 0$ then

$$\Phi(L) = \begin{bmatrix} \Phi_{11}(L) & 0 \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix}$$

  and if we invert $\Phi(L)$ we find that

$$\Phi(L)^{-1} = \begin{bmatrix} \Phi_{11}(L)^{-1} & 0 \\ -\Phi_{22}(L)^{-1}\Phi_{21}(L)\Phi_{11}(L)^{-1} & \Phi_{22}(L)^{-1} \end{bmatrix} = A(L).$$

- Then we find that if

$$z_t(1 \mid \{y_s \mid s \leq t\}) = z_t(1 \mid \{z_s \mid s \leq t\})$$

  then

$$A_{12,i} = 0 \qquad \text{for} \quad i = 1, \ldots, p.$$

  This hypothesis is simple to test once we have estimated the VAR model.
- MATLAB Example: BivariateGC.m Python Example: VARGCtest.py

Granger causality in higher dimensional VAR models

- Discussion above only applies to bivariate systems or systems where we test for Granger causality between blocks of variables.
- The procedure above *does not* apply to higher dimensional systems.
- To illustrate: Consider the following 3-dimensional VAR model

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix} + \begin{bmatrix} A_{11,1} & A_{12,1} & A_{13,1} \\ A_{21,1} & A_{22,1} & A_{23,1} \\ A_{31,1} & A_{32,1} & A_{33,1} \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} + \ldots + \begin{bmatrix} A_{11,p} & A_{12,p} & A_{13,p} \\ A_{21,p} & A_{22,p} & A_{23,p} \\ A_{31,p} & A_{32,p} & A_{33,p} \end{bmatrix} \begin{bmatrix} x_{1t-p} \\ x_{2t-p} \\ x_{3t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}.$$

- Test the null hypothesis that x_{3t} does not Granger cause x_{1t}.
- x_{1t} is affected by x_{3t} through the parameters $A_{13,i}$.
- It is tempting to test whether $A_{13,i}(L) = 0$ for $i = 1, \ldots, p$. If this hypothesis is not rejected, we conclude that the null hypothesis that $x_{3,t}$ does not Granger cause x_{1t} cannot be rejected.
- But, x_{3t} affects x_{2t} which in turn may affect x_{1t}. This relationship between the third and the first variable is not captured by the parameters in $A_{13,i}$.

Granger causality in higher dimensional VAR models

- To illustrate: Start by writing down the first two equations of the VAR model

$$\begin{aligned} A_{11}(L)\,x_{1t} &= \rho_1 - A_{12}(L)\,x_{2t} - A_{13}(L)\,x_{3t} + \varepsilon_{1t} \\ A_{22}(L)\,x_{2t} &= \rho_2 - A_{21}(L)\,x_{1t} - A_{23}(L)\,x_{3t} + \varepsilon_{2t}. \end{aligned}$$

- Premultiply the first equation with $A_{22}(L)$ and substitute $A_{22}(L)\,x_{2t}$ from the second equation into the first equation such that

$$\begin{aligned} \left(A_{22}(L)\,A_{11}(L) - A_{12}(L)\,A_{21}(L)\right)x_{1t} &= A_{22}(L)\left(\rho_1 + \varepsilon_{1t}\right) - A_{12}(L)\left(\rho_2 + \varepsilon_{2t}\right) + \\ &\quad \left(A_{12}(L)\,A_{23}(L) - A_{22}(L)\,A_{13}(L)\right)x_{3t}. \end{aligned}$$

- This equation describes how x_{3t} affects x_{1t} within the 3–dimensional system. We find that x_{3t} does not Granger cause x_{1t} if $A_{12}(L)\,A_{23}(L) - A_{22}(L)\,A_{13}(L) = 0$.
- Note that this is different from $A_{13,i}(L) = 0$!
- It is clearly the case that this latter hypothesis does not imply that $A_{12}(L)\,A_{23}(L) - A_{22}(L)\,A_{13}(L) = 0$ or the opposite.
- Conclusion: A test of the null hypothesis that $A_{13,i}(L) = 0$ only captures one possible channel of influence from $x_{3,t}$ to $x_{1,t}$, *the direct channel*. It may be the case that $x_{3,t}$ Granger causes $x_{1,t}$ even if $A_{13,i}(L) = 0$.

- How to test Granger causality hypotheses in higher dimensional systems?
- Apply method suggested by Lütkepohl based on VMA representation.
- Rewrite the VAR model in its VMA form

$$x_t = A(1)^{-1}\rho + A(L)^{-1}\varepsilon_t = \Phi(1)\rho + \Phi(L)\varepsilon_t.$$

- Partition the matrix $A(L)$ in the following way

$$A(L) = \begin{bmatrix} A_{11}(L) & A_{12}(L) & A_{13}(L) \\ A_{21}(L) & A_{22}(L) & A_{23}(L) \\ A_{31}(L) & A_{32}(L) & A_{33}(L) \end{bmatrix}$$

- Compute its inverse (use the relation that $\Phi(L) = A(L)^{-1}$)

$$A(L)^{-1} = \frac{1}{\det}\begin{bmatrix} A_{22}(L)A_{33}(L) - A_{23}(L)A_{32}(L) & A_{12}(L)A_{33}(L) - A_{13}(L)A_{32}(L) & A_{12}(L)A_{23}(L) - A_{13}(L)A_{22}(L) \\ A_{21}(L)A_{33}(L) - A_{23}(L)A_{31}(L) & A_{11}(L)A_{33}(L) - A_{13}(L)A_{31}(L) & A_{11}(L)A_{23}(L) - A_{13}(L)A_{21}(L) \\ A_{21}(L)A_{32}(L) - A_{22}(L)A_{31}(L) & A_{11}(L)A_{32}(L) - A_{12}(L)A_{31}(L) & A_{11}(L)A_{22}(L) - A_{12}(L)A_{21}(L) \end{bmatrix}$$

- $\Phi_{13}(L) = (A_{12}(L)A_{23}(L) - A_{13}(L)A_{22}(L))/\det(L)$ implying that if $\Phi_{13}(L) = 0$ then

$$A_{12}(L)A_{23}(L) - A_{22}(L)A_{13}(L) = 0.$$

Granger causality in higher dimensional VAR models

- To test for Granger non-causality in higher dimensional systems we need to compute the VMA representation.
- Compute the standard error of these estimates (Lütkepohl (2005) provides expressions to use).
- Then use Theorem 2 above.

Granger causality

- Consider the following bivariate VAR model

$$\begin{bmatrix} \Delta m_t \\ \Delta n_t \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} + \begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix} \begin{bmatrix} \Delta m_{t-1} \\ \Delta n_{t-1} \end{bmatrix} + \dots$$

$$+ \begin{bmatrix} a_{11,p} & a_{12,p} \\ a_{21,p} & a_{22,p} \end{bmatrix} \begin{bmatrix} \Delta m_{t-p} \\ \Delta n_{t-p} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

- $\mathbb{H}_0$: $a_{12,1} = a_{12,2} = \ldots = a_{12,p} = 0$ means that Δn_t does not Granger cause Δm_t.
- $\mathbb{H}_0$: $a_{21,1} = a_{21,2} = \ldots = a_{21,p} = 0$ means that Δm_t does not Granger cause Δn_t.
- Possible results:
  - Granger noncausality cannot be rejected in either direction
  - Unidirectional Granger causality
  - Bidirectional Granger causality
- Examples: Money-income causality, roosters internal clock (rooster crow Granger causes sunrise), oil price increases and recessions.
- Caution warranted!

Next topic

- Sims (1972)

Learning objective: Replicating analysis in Sims (1972)

Curriculum: Sims (1972)

- Purpose: Is money exogenous in the money-income relationship?
- Also show how time-series analysis can be used to shed light on money-income relations.
- Conduct tests for unidirectional causality.
- Main findings:
  - Money Granger causes income.
  - Income does not Granger cause money.
- Extensive literature on the relationship between money and income.
- Granger (1969): Definition of Granger non-causality.

Testing for the direction of causality

- Standard Granger non-causality tests based on single equation OLS estimates.
- Regression equation

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}.$$

- x_t does not Granger cause y_t if $\Phi_{12}(L) = 0$
- We know that the bivariate VMA can be written as a VAR(p) model

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} + \begin{bmatrix} A_{11,1} & A_{12,1} \\ A_{21,1} & A_{22,1} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \ldots + \begin{bmatrix} A_{11,p} & A_{12,p} \\ A_{21,p} & A_{22,p} \end{bmatrix} \begin{bmatrix} y_{t-p} \\ x_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

- $\Phi_{12}(L) = 0$ is equivalent to $A_{12,j} = 0$ for all $j = 1, ..., p$.
- Sims estimate each equation in the VAR using OLS.

Testing for the direction of causality

- Data: Monetary Base and M1, and GNP (in current dollars) covering the sample 1947-69 on a quarterly basis.
- Transformation of variables: Natural logarithm of GNP and Money.
- Prefiltered: replace each time series by $x_t - 1.5x_{t-1} + 0.5625x_{t-2}$ in order to make the regression residual white noise.
- No robustness tests!

Results

TABLE 1—SUMMARY OF *OLS* REGRESSIONS[a]

	F for Independent Variables	$\bar{R}^2$	Standard Error of Estimate	Degrees of Freedom
$GNP = f(M1$, 8 past lags)	1.89*	0.7927	0.01018	64
$GNP = f(M1$, 4 future, 8 past lags)	1.37	0.7840	0.01040	60
$GNP = f(MB$, 8 past lags)	2.24**	0.8004	0.00999	64
$GNP = f(MB$, 4 future, 8 past lags)	1.61	0.7924	0.01019	60
$M1 = f(GNP$, 4 future, 8 past lags)	11.25**	0.8385	0.00403	60
$MB = f(GNP$, 4 future, 8 past lags)	5.89**	0.8735	0.00410	60

* Significant at 0.10 level.
** Significant at 0.05 level.
[a] All regressions were fit to the period 1949III-1968IV. $M1$ is currency plus demand deposits. MB is monetary base as prepared by the Federal Reserve Bank of St. Louis. The F-tests shown are for the null hypothesis that all right-hand side variables except trend and seasonal dummies had zero coefficients. See also notes to Table 4.

RESULTS

TABLE 2—F'S FOR COMPARISONS OF SUBPERIODS 1948III–1957III vs. 1957IV–1968IV[a]

Regression Equation	F	Degrees of Freedom
$GNP = f(M1$, 8 past lags)	1.44	(14, 50)
$GNP = f\ (MB$, 8 past lags)	0.64	(14, 50)
$M1 = g(GNP$, 4 future, 8 past lags)	0.88	(18, 46)
$MB = f(GNP$, 4 future, 8 past lags)	1.01	(18, 46)

[a] Tests are for the null hypothesis that all coefficients (including trend and seasonals) remained the same in both subsamples.

Results

TABLE 3—F-TESTS ON FOUR FUTURE QUARTERS' COEFFICIENTS[a]

Regression Equation	F
GNP on $M1$	0.36
GNP on MB	0.39
$M1$ on GNP	4.29**
MB on GNP	5.89**

** Significant at 0.05 level

[a] All tests apply to regressions run over the full sample and are assumed distributed as $F(4, 60)$.

RESULTS

TABLE 4—LAG DISTRIBUTIONS FROM TIME-DOMAIN REGRESSIONS[a]

Coefficient on lag of:	*GNP* on *MB* past only	*GNP* on *MB* with future	*MB* on *GNP*	*GNP* on *M*1 past only	*GNP* on *M*1 with future	*M*1 on *GNP*
−4		−0.65	.162		−.300	.050
−3		.290	−.013		.120	.117
−2		−.088	.105		.126	.069
−1		−.110	.179		.105	.125
0	.603	.532	.171	.570	.484	.181
1	.593	.507	.015	.370	.412	.089
2	.509	.515	.052	−.034	−.017	.116
3	−.029	.080	.264	.543	.582	.107
4	−.011	.023	.107	−.242	−.363	.027
5	−.865	−.822	−.009	−.178	−.147	.027
6	−.037	−.053	.016	−.180	−.136	.025
7	−.296	−.282	.147	−.157	−.139	.123
8	.072	.039	.130	−.326	−.405	.112
Standard errors of coefficients:						
Largest s.e.	.313	.338	.052	.293	.318	.051
Smallest s.e.	.272	.276	.045	.274	.294	.044
Sum of coefficients	.540	—	—	.365	—	—
Standard error of sum	.442	—	—	.523	—	—

[a] Regressions were on *logs* of variables, prefiltered as explained in the text. Each regression included, in addition to the leading and lagging values of the independent variable for which coefficients are shown, a constant term, a linear trend term, and three seasonal dummies. Trends were in all cases significant. Seasonal dummies were insignificant. (The data were seasonally adjusted.)

Conclusions

- Reject the null that Money does not Granger cause GNP.
- Reject the null that GNP does not Granger cause M1.
- No significant difference in sub-samples.
- Future Money explains GNP but future GNP cannot predict Money. Suggests that money is not passive.
- Robustness tests using real GNP. Results unchanged.

Replicating Sims results

- Download new updated data from FRED. GNP Gross National Product, Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate, MANMM101USQ189S M1 for the United States, National Currency, Quarterly, Seasonally Adjusted
- Matlab illustration: SimsGranger.m (also need sdummy.m a procedure to define seasonal dummies and the data file Sims1972D.txt) Python: No example file

Comparison

Table: Comparison Sims full sample with 1959-68.

	Sims		Replication		
	F-test	$\bar{R}^2$	Wald	F-test	$\bar{R}^2$
GNP=f(M1, 8 lags)	1.89*	0.79	15.67**	1.95	0.74
GNP=f(M1, 4 future, 8 past lags)	1.37	0.78	18.55	1.43	0.90
M1=f(GNP, 4 future, 8 past lags)	11.25**	0.78	7.41	0.57	0.81
GNP on M1	0.36		4.59	1.15	
M1 on GNP	4.29		7.55	1.88	

Table: Comparison Sims full sample with 1959-68, no filtering of data (first log differences).

	Sims		Replication		
	F-test	$\bar{R}^2$	Wald	F-test	$\bar{R}^2$
GNP=f(M1, 8 lags)	1.89*	0.79	8.10	1.01	0.00
GNP=f(M1, 4 future, 8 past lags)	1.37	0.78	24.54 **	1.89**	0.17
M1=f(GNP, 4 future, 8 past lags)	11.25**	0.78	13.68	1.05	0.10
GNP on M1	0.36		2.57	0.64	
M1 on GNP	4.29		4.58	1.15	

POTENTIAL PITFALLS?

- Sims estimates a regression in levels: GNP and Money are non-stationary. Tests illustrated in unitroottest.m
- No VAR estimates! Will results change?
- Lag length dependence?
- Does it matter if we formulate tests on autoregressive or moving average coefficients?

VAR or VMA based tests?

- Lütkepohl (2005) derives the asymptotic distribution of VMA parameters.

Theorem Lütkepohl Proposition 3.6
Suppose

$$\sqrt{T}\begin{bmatrix} \hat{\alpha}-\alpha \\ \hat{\sigma}-\sigma \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} \Sigma_{\hat{\alpha}} & 0 \\ 0 & \Sigma_{\hat{\sigma}} \end{bmatrix}\right)$$

then

$$\sqrt{T}\mathrm{vec}\left(\hat{\Phi}_i - \Phi_i\right) \xrightarrow{d} \mathcal{N}\left(G_i \Sigma_{\hat{\alpha}} G_i'\right) \quad i = 1, 2, ...$$

where

$$G_i \equiv \frac{\partial vec\left(\Phi_i\right)}{\partial \alpha'} = \sum_{m=0}^{i-1} J\left(A'\right)^{i-1-m} \otimes \Phi_m$$

How does it work in practice?

- We have estimated the following
  - $\alpha = \text{vec}(A_1, ..., A_p)$; $K^2p \times 1$.
  - The companion matrix A; $Kp \times Kp$. Remember that J is a $K \times Kp$ matrix.
  - $\sigma \equiv \text{vech}(\Sigma_u)$; $\frac{1}{2}K(K+1) \times 1$.
- This implies that G_i is a $K^2 \times K^2p$ matrix.
- Construct G_i and compute $G_i \Sigma_{\hat{\alpha}} G_i'$ and select the row corresponding to the element i, j in Φ.
- Note that: $G_1 = J \otimes I_K$, $G_2 = JA' \otimes I_K + J \otimes \Phi_1$ and so on.

How to conduct Granger non-causality tests in VAR models

- Determine the lag length using, for example, general-to-specific method.
- Check stability.
- Test for autocorrelation, ARCH and normality using multivariate tests.
- Estimate VAR using the preferred lag length.
- If bivariate VAR, test null hypotheses on autoregressive parameters, otherwise test null hypotheses on moving average parameters. (Check for bi-directional causality.)
- Conclude.
- Next topic: Integrated VAR models, tests for unit roots and cointegration. Curriculum: KL ch. 3.1, 3.2 (LS estimation with known cointegration vector), 3.3-3.4 and Johansen (1991).

Advanced Macroeconomics: Structural Vector Autoregressive Analysis

Lectures 12-15

U. Michael Bergman
University of Copenhagen

Fall 2024

CURRICULUM:

KL 10-11, Blanchard and Quah (1989), Bergman (1996), Galí (1992, 1999), King, Plosser, Stock and Watson (1991), Warne (1992), Bergman, Cheung & Lai 2010.

- Identification using long-run restrictions
  - The Blanchard and Quah approach.
  - Approach based on VEC model.
  - Approach based on Common Trends model.
  - Applications: supply-demand shocks in an AD-AS models (Blanchard & Quah, 1989; Bergman, 1996), Productivity and hours worked (Galí, 1999); and real exchange rate fluctuations (Bergman, Cheung & Lai, 2010).
- Identification using both long-run and short-run restrictions
  - Closed-form solution (Warne, 1992)
  - King, Plosser, Stock and Watson (1991) macro-model
  - Galí (1992) IS-LM model.
- The Galí–Christiano, Eichenbaum and Vigfusson–Chari, Kehoe and McGrattan debate.

Long-run restrictions using the stationary representation

Curriculum: KL 10.1, 11.2, Blanchard and Quah (1989).

- Long-run restrictions: A restriction on the long-run response of a variable to shocks.
- Examples: Demand shocks can have no long-run effect on output; nominal shocks do not affect real variables in the long-run.
- First suggested by Blanchard and Quah (1989) and then used extensively in the literature.
- Restriction is motivated using an AD-AS model.

- Macro model:
  - A version of Fischer (1977) model (the role of monetary policy in a model with long-term labor contracts).
  - Solution of the model is:

$$\Delta Y = e_d(t) - e_d(t-1) + a\,(e_s(t) - e_s(t-1)) + e_s(t)$$

$$U = -e_d(t) - ae_s(t)$$

  - If $e_d(1) = 1$, then $Y(1) = 1$, $Y(2) = Y(1) - 1 = 0$, and $U(1) = -1$, $U(2) = 0$ and if $e_s(1) = 1$, then $Y(1) = a + 1$, $Y(2) = Y(3) = Y(4) = \ldots = Y(\infty) = 1$ and $U(1) = -a$, $U(2) = 0$.
  - Model implies that the supply shock has permanent effects on output and transitory effects on unemployment whereas demand shocks have transitory effects on both output and unemployment.
  - Another version of the Fischer model for a bivariate system comprised of output and inflation can be found in Bergman(1996).

Blanchard and Quah empirical model

- Starting point:
  - GNP is an I(1) process, a shock to GNP has a long-run effect on the level.
  - GNP is affected by more than one type of shocks, problem distinguishing between the effects of all these shocks using univariate time series analysis.
  - Solution: impose restrictions allowing some shocks to have permanent effects whereas other shocks only have temporary (or transitory) effects.
  - Restrictions are derived using a macro model.

Blanchard and Quah empirical model

- Empirical model:
  - Two structural shocks: aggregate supply shock w_t^{AS} and aggregate demand shock w_t^{AD}.
  - $z_t = \begin{bmatrix} \Delta gdp_t & ur_t \end{bmatrix}' \sim I(0)$.
  - $gdp_t \sim I(1)$.
  - Structural VAR model:
  $$B(L)z_t = w_t$$
  where $B(L) = B_0 - B_1 L - B_2 L^2 - \ldots - B_p L^p$, $B(1) = B_0 - B_1 - \ldots - B_p$ and $w_t = \begin{bmatrix} w_t^{AS} & w_t^{AD} \end{bmatrix}' \sim (0, I_2)$.
  - Structural VMA representation
  $$z_t = B(L)^{-1} w_t = \Theta(L) w_t$$
- Note that $z_t \sim I(0)$ implying that long-run effects of structural shocks on z_t is zero.
- But, level of GDP does not necessarily return to its initial level in the long-term.
- The effect of a structural shock on the level of GDP is the cumulative sum of its effects on Δgdp_t.
- The cumulative effect is $\Theta(1) = \sum_{i=1}^{\infty} \Theta_i = B(1)^{-1}$.

- Empirical model continued:
- The restriction that aggregate demand shocks have no long-run effects on the level of GDP implies that

$$\Theta(1) = \begin{bmatrix} \theta_{11}(1) & 0 \\ \theta_{21}(1) & \theta_{22}(1) \end{bmatrix}$$

  where $\theta_{11}(1)$ is unrestricted and measures the long-run effect of aggregate supply shocks on the level of GDP.
- There are no restrictions on the second row of $\Theta(1)$
- Is it possible to recover the parameters of the structural form using reduced form estimates?

Blanchard and Quah empirical model

- Assume that the time series vector z_t is generated by

$$A(L)z_t = u_t$$

  where $A(L) = I_2 - A_1 L - \ldots - A_p L^p$, $A(1) = I_2 - A_1 - \ldots - A_p$ and $u_t \sim (0, \Sigma_u)$.
- Compare reduced form and structural form:
  - Remember that $A_i = B_0^{-1} B_i$ or, which is equivalent $B_i = B_0 A_i$!
  - This implies that $B(L) = B_0 A(L)$ and that $B(1) = B_0 A(1)$
  - $w_t = B_0 u_t$ or $u_t = B_0^{-1} w_t$ implying that $\Sigma_u = B_0^{-1}(B_0^{-1})'$
- We know that $\Theta(1) = B(1)^{-1} = A(1)^{-1} B_0^{-1}$.
- A restriction on Θ is a restriction on B_0.

- We also know that

$$\Sigma_u = B_0^{-1}(B_0^{-1})' = \left[A(1)B(1)^{-1}\right]\left[A(1)B(1)^{-1}\right]'$$

- To see this, pre-multiply $B(1) = B_0 A(1)$ with B_0^{-1} and post-multiply with $B(1)^{-1}$.
- Pre-multiply both sides by $A(1)^{-1}$ and post-multiply both sides by $(A(1)^{-1})'$

$$A(1)^{-1}\Sigma_u(A(1)^{-1})' = A(1)^{-1}A(1)B(1)^{-1}\left[A(1)B_0^{-1}\right]'(A(1)^{-1})'$$

and simplify

$$A(1)^{-1}\Sigma_u(A(1)^{-1})' = B(1)^{-1}(B(1)^{-1})' = \Theta(1)\Theta(1)'$$

Blanchard and Quah empirical model

- From previous slide

$$A(1)^{-1}\Sigma_u(A(1)^{-1})' = B(1)^{-1}(B(1)^{-1})' = \Theta(1)\Theta(1)'$$

- We know that Σ_u is symmetric so we have three equations. But, $\Theta(1)$ has four parameters. We need to impose one restriction on $\Theta(1)$ to identify the system.
- Apply a Cholesky decomposition on

$$A(1)^{-1}\Sigma_u(A(1)^{-1})'$$

such that

$$\hat{\Theta}(1) = \text{chol}\left(\hat{A}(1)^{-1}\hat{\Sigma}_u(\hat{A}(1)^{-1})'\right)$$

and then we find that

$$\hat{B}_0^{-1} = \hat{A}(1)\hat{\Theta}(1)$$

- Once we have estimated $\hat{B}_0^{-1}$, we can proceed as usual and compute impulse responses and variance decompositions.

Replication: Blanchard and Quah analysis

- VAR system comprised of the first log difference of real GDP and unemployment rate with 8 lags, quarterly data.
- Sample: 1950:2-1987:4.
- Four specifications: detrending and allowing for breaks. Small differences across these models.
- MATLAB example: BQDecompNew.m Python example: BQDecompNew.py

Blanchard-Quah results

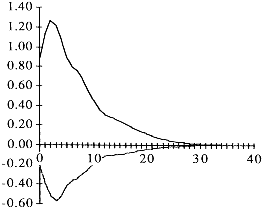


Figure 1. Response to Demand, — = Output, ~~ = Unemployment

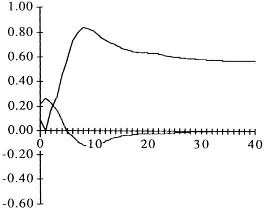


Figure 2. Response to Supply, — = Output, ~~ = Unemployment

Blanchard-Quah results: Historical decomposition

- BQ also reports forecast error variance decomposition answering the question: What portion of the variance of the forecast error in predicting, say, output is due to supply shocks and demand shocks?
- That is, the relative importance of each structural shock! See Table 2.
- GDP: Demand shocks dominate at short horizons but supply shocks are more important after 12 quarters. Unemployment: Demand shocks dominate for all horizons.
- Figures 7-10 (and Table 1) report historical decompositions. Question: What portion of the deviation of GDP growth and unemployment from its unconditional mean is due to each structural shock?
- In other words: What structural shock at what time contributed to the fluctuations in GDP growth and unemployment during the sample?

Historical decomposition

- Note that each observation of our data can be re-written as the cumulative sum of the structural shocks.
- Assume for simplicity that the structural VAR(1) model is given by

$$B_0 x_t = B_1 x_{t-1} + \varepsilon_t$$

which can be written as the reduced form model

$$x_t = \underbrace{B_0^{-1} B_1}_{A_1} x_{t-1} + B_0^{-1} \varepsilon_t$$

- Use this model to compute the data at each point in time ($t > 1$)

$$x_2 = A_1 x_1 + B_0^{-1} \varepsilon_2$$

$$x_3 = A_1 x_2 + B_0^{-1} \varepsilon_3 = A_1^2 x_1 + A_1 B_0^{-1} \varepsilon_2 + B_0^{-1} \varepsilon_3$$

and putting it all together we have that

$$x_t = A_1^{t-1} x_1 + \sum_{j=0}^{t-2} A_1^j B_0^{-1} \varepsilon_{t-j}$$

Historical decomposition

- Example: Compute the historical decomposition of the third observation in a bivariate VAR(1) model. Let $x_t = \begin{bmatrix} x_{1,t} & x_{2,t} \end{bmatrix}$.
- Using the relations from previous slide we find that

$$x_3 = \underbrace{A_1^2}_{init_3} + \underbrace{A_1 B_0^{-1}}_{\Lambda} \varepsilon_3$$

- We can re-write this as

$$\begin{bmatrix} x_{1,3} \\ x_{2,3} \end{bmatrix} = \begin{bmatrix} init_{1,3} \\ init_{2,3} \end{bmatrix} + \begin{bmatrix} \lambda_{11}^1 & \lambda_{12}^1 \\ \lambda_{21}^1 & \lambda_{22}^1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,2} \\ \varepsilon_{2,2} \end{bmatrix} + \begin{bmatrix} \lambda_{11}^0 & \lambda_{12}^0 \\ \lambda_{21}^0 & \lambda_{22}^0 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,3} \\ \varepsilon_{2,3} \end{bmatrix}$$

- Implying that

$$x_{1,3} = init_{1,3} + \lambda_{11}^1 \varepsilon_{1,2} + \lambda_{12}^1 \varepsilon_{2,2} + \lambda_{11}^0 \varepsilon_{1,3} + \lambda_{12}^0 \varepsilon_{2,3}$$

$$x_{2,3} = init_{2,3} + \lambda_{21}^1 \varepsilon_{1,2} + \lambda_{22}^1 \varepsilon_{2,2} + \lambda_{21}^0 \varepsilon_{1,3} + \lambda_{22}^0 \varepsilon_{2,3}$$

HISTORICAL DECOMPOSITION

- The historical decomposition is then given by

$$\underbrace{\begin{cases} \mathcal{HD}_{1,3}^{init} = init_{1,3} \\ \mathcal{HD}_{1,3}^{\varepsilon 1} = \lambda_{11}^{1}\varepsilon_{1,2} + \lambda_{11}^{0}\varepsilon_{1,3} \\ \mathcal{HD}_{1,3}^{\varepsilon 2} = \lambda_{12}^{1}\varepsilon_{2,2} + \lambda_{12}^{0}\varepsilon_{2,3} \end{cases}}_{\text{This sums to } x_{1,3}} \quad \underbrace{\begin{cases} \mathcal{HD}_{2,3}^{init} = init_{2,3} \\ \mathcal{HD}_{2,3}^{\varepsilon 1} = \lambda_{21}^{1}\varepsilon_{1,2} + \lambda_{21}^{0}\varepsilon_{1,3} \\ \mathcal{HD}_{2,3}^{\varepsilon 2} = \lambda_{22}^{1}\varepsilon_{2,2} + \lambda_{22}^{0}\varepsilon_{2,3} \end{cases}}_{\text{This sums to } x_{2,3}}$$

- Matlab example: BQhdecomp.m Python example: BQhdecomp.py

Historical decomposition: Matlab function

```
function [HDinit,HDconst,HDshock,HDendo]=hdecomp(A,mu,What,B0inv,q,p,indep)
% Input
% A = Companion matrix A
% mu = constant
% What = structural shocks
% B0inv = inv(B0)
% q = # variables
% p = # lags
% indep = dependent variables excluding constant
%
% Output
% HDinit = initial conditions
% HDconst = constant
% HDshocks(obs,shock,variable) = (nobs x shock & variable, i.e.,
% HDshock(:,1,1) contains the effect of shock 1 on variable 1;
% HDshock(:,2,1) contains the effect of shock 2 on variable 1 and so on.
%
```

```
def hdecomp(A,mu,What,B0inv,K,p,indep):def hdecomp(A,mu,What,B0inv,K,p,indep):
"""
# Input
# A = Companion matrix A
# mu = constant
# What = structural shocks
# B0inv = inv(B0)
# K = # variables
# p = # lags
# indep = dependent variables excluding constant
# Output
# HDinit = initial conditions
# HDconst = constant
# HDshocks(obs,shock,variable) = (nobs x shock & variable, i.e.,
# HDshock(:,1,1) contains the effect of shock 1 on variable 1;
# HDshock(:,2,1) contains the effect of shock 2 on variable 1 and so on.
```

An alternative approach for imposing long-run restrictions

Curriculum: KL 10.2 pp. 268-271

- In the Blanchard-Quah decomposition we impose restrictions on the cumulative responses within a stationary VAR representation $z_t = \begin{bmatrix} \Delta gnp_t & ur_t \end{bmatrix}'$.
- An alternative approach is to base identification on the VEC representation

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_t + \ldots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t$$

- We also know that this model has a vector ARMA (or VMA representation in first differences)

$$\Delta y_t = \varepsilon_t + \Phi_1 \varepsilon_{t-1} + \Phi_2 \varepsilon_{t-2} + \ldots = C(L)\,\varepsilon_t$$

- The VMA representation for the level of y_t can be obtained if we add y_{t-1} on both sides

$$y_t = C(L)\,\varepsilon_t + y_{t-1}$$

and replace y_{t-1} $(= C(L)\,\varepsilon_{t-1} + y_{t-2})$ recursively so that

$$\begin{aligned}
y_t &= C(L)\,\varepsilon_t + y_{t-1} \\
y_t &= C(L)\,\varepsilon_t + C(L)\,\varepsilon_{t-1} + y_{t-2} \\
&\vdots \\
y_t &= C(L)\,(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_0) + y_0
\end{aligned}$$

- Rewrite this equation by adding and subtracting

$$C(1)\,(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_0)$$

such that we obtain

$$y_t = C(1)\,(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_0) + (C(L) - C(1))\,(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_0) + y_0$$

- From previous slide

$$y_t = C(1)(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_0) + (C(L) - C(1))(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_0) + y_0$$

- Assume now that $\varepsilon_s = 0$ for $s < 0$ such that the second term on the RHS can be written as

$$\left(I + \Phi_1 L + \Phi_2 L^2 + \ldots - C(1)\right)(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_0) =$$

$$\left(I + \Phi_1 L + \Phi_2 L^2 + \ldots - I - \Phi_1 - \Phi_2 - \ldots\right)(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_0) =$$

$$(-\Phi_1 - \Phi_2 - \Phi_3 - \ldots)\varepsilon_t + (-\Phi_2 - \Phi_3 - \ldots)\varepsilon_{t-1} + (-\Phi_3 - \ldots)\varepsilon_{t-2} + \ldots$$

- Let $C_s^* = -\sum_{j=s+1}^{\infty} C_j$ and define a random walk $\xi_t = \xi_{t-1} + \varepsilon_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_0$ so that we can write our model as

$$y_t = y_0 + C(1)\xi_t + C^*(L)\varepsilon_t$$

- From previous slide
$$y_t = y_0 + C(1)\,\xi_t + C^*(L)\,\varepsilon_t$$
- This is the reduced form Common Trends (CT) model. Note that we have decomposed the time series vector y_t into two components, a permanent component reflecting the stochastic trends
$$y_t^p = C(1)\sum_{i=1}^{t}\varepsilon_i$$
and a transitory component reflecting the stationary part
$$y_t^t = C^*(L)\,\varepsilon_t$$
This decomposition is known as the multivariate Beveridge–Nelson decomposition.

- Let $\Xi = C(1)$, $\Xi^*(L) = C^*(L)$, $y_0^* = y_0$ and $u_t = \varepsilon_t$, then the CT model can be written as equation (10.2.2) in KL, i.e.,

$$y_t = \Xi \sum_{i=1}^{t} u_i + \Xi^*(L)u_t + y_0^*$$

- From Grangers Representation Theorem (Theorem 4.1 in Johansen, 1991) we also have that

$$\Xi = \beta_\perp \left[\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp \right]^{-1} \alpha'_\perp$$

  where $\beta_\perp$ and $\alpha_\perp$ are orthogonal complements of β and α ($\beta'\beta_\perp = 0$).
- As above we know that $\Xi^*(L)u_t = \sum_{j=0}^{\infty} \Xi^* u_{t-j}$ is an $I(0)$ process, y_0^* is the initial condition and $\Xi \sum_{i=1}^{t} u_i$ contains the common stochastic trends.
- The matrix Ξ has rank $K - r$ (we have r stationary relations and $K - r$ common trends).

- We can define a *structural* CT model in the same way as we have done before, by replacing u_t by $B_0^{-1} w_t$. The structural CT model can then be written as

$$y_t = \Xi \sum_{i=1}^{t} B_0^{-1} w_i + \Xi^*(L) B_0^{-1} w_t + y_0^*$$

$$y_t = \underbrace{\Xi B_0^{-1}}_{\Upsilon} \sum_{i=1}^{t} w_i + \Xi^*(L) B_0^{-1} w_t + y_0^* = \Upsilon \sum_{i=1}^{t} w_i + \Xi^*(L) B_0^{-1} w_t + y_0^*$$

- The matrix Υ is the matrix of long-run multipliers, it measures the long-run effect of the common trends (or the permanent shocks). Note that the long-run effects of the stationary part $\Xi^*(L) B_0^{-1} w_t$ goes to zero as $j \to \infty$. The rank of Υ is the same as the rank of Ξ, i.e., rank $K - r$.
- Long-run restrictions can be imposed directly on Υ, if the long-run effect of a shock is zero on all variables, then the corresponding column of Υ is restricted to zero.

- The Blanchard-Quah model can then be written as

$$y_t = \begin{pmatrix} gdp_t \\ ur_t \end{pmatrix} = \begin{bmatrix} \zeta_{11} & 0 \\ \zeta_{21} & \zeta_{22} \end{bmatrix} \sum_{i=1}^{t} \begin{pmatrix} w_i^{AS} \\ w_i^{AD} \end{pmatrix} + \dots$$

- We also note that the second row of Υ must be equal to zero, no shock can have permanent effects on unemployment (we have assumed that $ur_t \sim I(0)$). Therefore

$$\Upsilon = \begin{bmatrix} \zeta_{11} & 0 \\ 0 & 0 \end{bmatrix}$$

- The element $\zeta_{11} \neq 0$, otherwise the rank of Υ is not equal to one!

An alternative approach for imposing long-run restrictions

- We can apply this decomposition in both Matlab and Python.
- To do this, we need to solve a system of non-linear equations!
- Remember that a restriction on B_0^{-1} is also a restriction on Θ since $\Theta = \Xi B_0^{-1}$.

An alternative approach for imposing long-run restrictions: Matlab

- Remember, again, that a restriction on B_0^{-1} is also a restriction on Θ since $\Theta = \Xi B_0^{-1}$.
- In Matlab we need to define the equations and the restrictions imposed. As before, we use fsolve and define the system of equations and restrictions in the following way.

```
function q=restrictions(B0inv)
global SIGMA Xi
K=size(B0inv,1);
THETA1=Xi*B0inv;
F=vec(B0inv*B0inv'-SIGMA(1:K,1:K));
% Long run restriction
q=[F; THETA1(1,2)];
q'+1;
```

- The estimation is illustrated in BQsolver.m

An alternative approach for imposing long-run restrictions: Python

- Remember, once again, that a restriction on B_0^{-1} is also a restriction on Θ since $\Theta = \Xi B_0^{-1}$.
- In Python we define the system of equations and restrictions in a similar way. But, in Python we need to state the equations explicitly.
- There are 5 parameters to be estimated, four in $B0^{-1}$ and one parameter in Θ. Therefore we need five equations.

An alternative approach for imposing long-run restrictions: Python

- We have three unique equations in $B_0^{-1}\left(B_0^{-1}\right)' - \Sigma_u$.
- Remember that $\Theta = \Xi B_0^{-1}$ implying that the RHS can be written as

$$\Xi B_0^{-1} = \left[\begin{array}{cc} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{array}\right]\left[\begin{array}{cc} b_{11} & b_{12} \\ b_{21} & b_{22} \end{array}\right] = \left[\begin{array}{cc} \xi_{11}B_{11} + \xi_{12}b_{21} & \xi_{11}b_{12} + \xi_{12}b_{22} \\ \xi_{21}B_{11} + \xi_{22}b_{21} & \xi_{21}b_{12} + \xi_{22}b_{22} \end{array}\right]$$

- We want to estimate one parameter in this matrix $\xi_{11}B_{11} + \xi_{12}b_{21} = \theta_{11}$ and we want to restrict $\theta_{12} = \xi_{11}b_{12} + \xi_{12}b_{22} = 0$.
- We now have five equations and five unknowns.

An alternative approach for imposing long-run restrictions: Python

- The system of equations to be solved can then be written in the following way:

```
def func(vars):
# Define restrictions imposed on Theta1
# Theta1 = [ * 0
# 0 0]
#Theta1s = np.dot(Xi,B0invs)
#Theta1s[0,1] = 0
#and note that Theta1s[1,0] = Theta1s[1,1] = 0 by definition
# Define variables to be estimated
B0invs[0,0], B0invs[0,1], B0invs[1,0], B0invs[1,1], Theta1s[0,0] = vars
# Then define the system of equations to be solved
eq1 = B0invs[0,0]**2 + B0invs[0,1]**2 - so[0,0]
eq2 = B0invs[0,0]*B0invs[1,0]+B0invs[0,1]*B0invs[1,1] - so[0,1]
eq3 = B0invs[1,0]**2+B0invs[1,1]**2-so[1,1]
eq4 = Xi[0,0]*B0invs[0,1]+Xi[0,1]*B0invs[1,1]
eq5 = Xi[0,0]*B0invs[0,0]+Xi[0,1]*B0invs[1,0]-Theta1s[0,0]
return ([eq1, eq2, eq3, eq4, eq5])
```

- The estimation is illustrated in BQsolver.py (Code also include two alternative ways to formulate and solve the system of equations.)

An alternative approach for imposing long-run restrictions: Python using LSQ

- An alternative way to solve the system of equations using scipy.optimize.least_squares:

```
def func(vars):
# Define restrictions imposed on Theta1
# Theta1 = [ * 0
# * *]
#Theta1s[0,1] = 0
# variables
B0invs[0,0], B0invs[0,1], B0invs[1,0], B0invs[1,1] = vars
# Then define the system of equations to be solved
# We only need to define the vec(B0inv*B0inv'-Sigma_u) equations
eq1 = (vec(np.dot(B0invs,B0invs.T) - SIGMA)).flatten()
# Then we need to compute Theta1s = Asum*B0inv
Theta1s = np.dot(Asum,B0invs)
# Then, finally define the restrictions imposed on Theta1
eq2 = np.array(Theta1s[0,1]).flatten()
eq = np.hstack((eq1,eq2))
return (eq)
root5 = scipy.optimize.least_squares(func, [1, 1, 1, 1])
```

- The estimation is illustrated in IllustratingSolversinPython.py

Potential issues when imposing long-run restrictions

- Note 1: If $y_t \sim I(0)$, then $\Upsilon = 0_{K \times K}$ and shocks cannot have permanent effects.
- Note 2: If $y_t \sim I(1)$ and there is no cointegration relation, $r = 0$, then $\alpha = \beta = 0$ the orthogonal complements of α and β are $K \times K$ identity matrices implying that

$$\Xi = \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right)^{-1}$$

and

$$\Upsilon = \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right)^{-1} B_0^{-1}$$

Potential issues when imposing long-run restrictions

- If $y_t \sim I(1)$, then we can rewrite the VEC model as the first difference VAR(p-1) model
$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t$$
- The structural form representation is
$$B(L)\Delta y_t = w_t$$
where $B(L) = B_0 \Gamma(L) = B_0(I_K - \Gamma_1 L - \ldots - \Gamma_{p-1} L^{p-1})$.
- The structural VMA representations is
$$\Delta y_t = B(L)^{-1} w_t = \Gamma(L)^{-1} B_0^{-1} w_t$$
and the long-run effects on the level of y_t is $\Upsilon = \Gamma(1)^{-1} B_0^{-1}$.
- We can now proceed and estimate the Blanchard-Quah model again using the solver.

Combining long- and short-run restrictions

- The approaches discussed previously will work as long as we are only interested in identifying long-run shocks.
- If we also are interested in identifying the transitory shocks (in case there are more than 1 transitory shock) we need to add restrictions on B_0^{-1}, i.e., restrictions on the contemporaneous effects of the transitory shocks.
- To identify the full system, it is suggested in KL to use the non-linear solver. Homework Assignment 2! And examples later on.
- Alternatively, one may use and extend the CT model representation approach to allow us to identify both permanent and transitory shocks without using the non-linear solver.

King, Plosser, Stock and Watson (1991)

- Purpose of paper and economic model

Identification in the three-variable KPSW model

Curriculum: KL chapter 10 pp. 272-276, KL section 11.3.2 pp. 311-315.

- KPSW apply the approaches to implement long-run restrictions, an extension of the Blanchard-Quah method.
- Consider the three-variable VAR model for $y_t = \begin{pmatrix} gnp_t & c_t & inv_t \end{pmatrix}'$ where we assume that $c_t - gnp_t \sim I(0)$ and that $inv_t - gnp_t \sim I(0)$.
- Cointegration rank $r = 2$, the cointegration vector is known, there is one shock (balanced growth shock) representing productivity, and there are two transitory shocks (shocks having only short-run effects).
- We can therefore write the VAR model as the VEC model

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_i \Delta y_{t-1} + \cdots + \Gamma_p \Delta y_{t-p+1} + u_t$$

where $\beta' = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$ is known.

- We know that this VEC can be written as a CT model

$$y_t = \Xi \sum_{i=1}^{t} u_i + \Xi^*(L)u_t + y_0^*$$

where

$$\Xi = \beta_\perp \left[\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp \right]^{-1} \alpha'_\perp$$

- The structural CT model is

$$y_t = \Upsilon \sum_{i=1}^{t} w_i + \Xi^*(L) B_0^{-1} w_t + y_0^*$$

  where $\Upsilon = \Xi B_0^{-1}$.
- In our structural model

$$\Upsilon = \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \end{bmatrix}$$

  if we put the balanced growth shock as the first element in w_t. The balanced growth shock has a long-run effect on all three variables. The remaining two transitory shocks have no long-run effects on any of the variables.
- How to identify the transitory shocks?
- We need to combine the long-run restriction with restrictions on the contemporaneous effects of the transitory shocks.
- This can be done by placing restrictions on the last two columns of B_0^{-1}.

- How many restrictions do we need?
- To just identify the system we need $K(K-1)/2 = 3$ restrictions. The two zero columns of Υ stand for 2 restrictions. We therefore need one additional restriction. Another way to see this is that we have two transitory shocks and therefore we need one restriction to identify these two shocks.
- Note: The identification of permanent and transitory shocks are independent implying that the identification of the permanent shock is independent on how we identify the two transitory shocks. If we only are interested in the effects of the long-run shock, we can impose an arbitrary restriction on the transitory shocks.

IDENTIFICATION IN THE THREE-VARIABLE KPSW MODEL

- One could use, for example, the restriction suggested in section 10.3.1:

$$B_0^{-1} = \begin{bmatrix} * & * & * \\ * & * & 0 \\ * & * & * \end{bmatrix}$$

- How to implement these restrictions? Use the non-linear solver as suggested in 11.3.2.
- Let ζ_{ij} denote the *ij*:th element in Υ. This implies that $\zeta_{12} = \zeta_{13} = \zeta_{22} = \zeta_{23} = \zeta_{32} = \zeta_{33} = 0$.
- We also have the restriction that $b_0^{23} = 0$ and that $B_0^{-1}(B_0^{-1})' = \Sigma_u$.

- Combining these we have the following restrictions

$$\begin{bmatrix} \text{vech}\left(B_0^{-1}(B_0^{-1})' - \Sigma_u\right) \\ b_0^{23} \\ \zeta_{12} \\ \zeta_{13} \\ \zeta_{22} \\ \zeta_{23} \\ \zeta_{32} \\ \zeta_{33} \end{bmatrix} = 0$$

- Given an initial guess for B_0^{-1} such that $\Upsilon = \Xi B_0^{-1}$ where Ξ is given above (and given that we can estimate the VEC model and compute the orthogonal complements of α and β), we can use the MATLAB/Python non-linear solver to find a solution.

KPSW: MATLAB

```
% restrictions.M % Normalization: SIGMA_w=I
function q=restrictions(B0inv)
global GAMMA SIGMA alpha beta alpha_perp beta_perp Xi p
K=size(B0inv,1);
THETA1=Xi*B0inv;
F=vec(B0inv*B0inv'-SIGMA(1:K,1:K));
% Long run and short run restrictions q=[F; B0inv(2,3); THETA1(1,2); THETA1(1,3);
THETA1(2,2); THETA1(2,3); THETA1(3,2); THETA1(3,3)];
q'+1;
warning off options=optimset('TolX',1e-10,'TolFun',1e-10,'MaxFunEvals',1e+10);
% Compute B_0^1 B0inv=fsolve('restrictions',randn(q,q),options);
```

KPSW: Python

def ident3(b0_invflat,Xi,Sigma_u): # b0_flat = vectorized B0inv # Xi = estimated Xi matrix # Sigma_u = estimated var-cov matrix of residuals b = b0_invflat.reshape(3,3) Upsilon = Xi@b block_free = vech(b@b.T-Sigma_u) block_short = b[2,1] block_long = Upsilon[:,1:].flatten() return np.hstack([block_free,block_long,block_short]) # LS solver guess = np.ones(9) obj = lambda x: ident3(x,Xi,sigma) result = least_squares(obj,guess) B0inv3 = result.x.reshape(3,3)

Identification in the three-variable KPSW model

- Once we have found the solution of B_0^{-1} we can compute the impulse responses $C(L)$ in the first difference VAR model using the VEC estimates.
- For this, we need the companion matrix based on VEC estimates to compute the impulse responses

$$\begin{bmatrix} \Pi + I_K + \Gamma_1 & \Gamma_2 - \Gamma_1 & \cdots & \Gamma_{p-1} - \Gamma_{p-2} & -\Gamma_{p-1} \\ I_K & 0 & \ldots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & I_K & 0 \end{bmatrix}$$

- Functions: Matlab function vectovar.m and Python vec2var.py

VECTOVAR FUNCTIONS

MATLAB

```
function [B]=vectovar(Gamma,Pi);
% This function converts VEC estimates into companion matrix for VAR in
% levels
%
% Michael Bergman
% Checked October 2018
%
% Input: KxKp Gamma matrix
% KxK Pi = alpha*beta' matrix
% Output: KpxKp Companion matrix B
```

PYTHON

```
def vectovar(Gamma,Pi):
# This function converts VEC estimates into companion matrix for VAR in
# levels.
#
# Michael Bergman
# Checked October 2023
#
# Input: Gamma = KxKp coefficient matrix
# Pi = alpha*beta' a KxK matrix
# Output: KpxKp Companion matrix B
```

Identification in the 6-variable KPSW model

- KPSW also consider a 6-variable model, with 3 cointegration vectors and therefore 3 common trends
- This implies that we need 3 restrictions to identify the permanent shocks and 3 restrictions to identify the transitory shocks.
- They suggest the following long-run multiplier with three zero restrictions imposed

$$\Theta = \begin{bmatrix} * & 0 & 0 & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 & 0 \end{bmatrix}$$

- We also need to impose three restrictions on B_0^{-1}, but they are irrelevant since they are not interested in these shocks.

KPSW: MATLAB

```
% restrictions.M % Normalization: SIGMA_w=I
function q=restrictions(B0inv)
global GAMMA SIGMA alpha beta alpha_perp beta_perp Xi p
K=size(B0inv,1);
THETA1=Xi*B0inv;
F=vec(B0inv*B0inv'-SIGMA(1:K,1:K));
% Long run and short run restrictions q=[F; B0inv(2,3); THETA1(1,2); THETA1(1,3);
THETA1(2,2); THETA1(2,3); THETA1(3,2); THETA1(3,3)];
q'+1;
warning off options=optimset('TolX',1e-10,'TolFun',1e-10,'MaxFunEvals',1e+10);
% Compute B_0^-1 B0inv=fsolve('restrictions',randn(q,q),options);
```

```
def ident6(b0_invflat,Xi,Sigma_u):
b = b0_invflat.reshape(6,6)
Upsilon = Xi@b
block_free = vech(b@b.T-Sigma_u)
block_short = [b[5,3],b[5,4],b[4,3]]
block_long1 = [Upsilon[0,1],Upsilon[0,2],Upsilon[5,2]]
block_long2 = Upsilon[:,3:].flatten()
return np.hstack([block_free,block_short,block_long1,block_long2])
# Solver guess = np.ones(K*K) obj6 = lambda x: ident6(x,Xi,sigma) result6 =
least_squares(obj6,guess,xtol=1e-15,ftol=1e-15,gtol=1e-15) # Impose tolerance,
needed because of many variables B0inv6 = result6.x.reshape(6,6)
```

Galí (1992) IS-LM model

- Test the implications of an IS-LM-Phillips curve textbook model using US data.
- Structural VAR model using both long-run and short-run restrictions.

GALÍ (1992) IS-LM MODEL

- Four equation model: IS-equation

$$y = \alpha + u_s - \sigma\left(i - E\Delta p_{+1}\right) + u_{is}$$

  LM equation

$$m - p = \phi y - \lambda i + u_{md}$$

  money supply process

$$\Delta m = u_{ms}$$

  and Philips curve

$$\Delta p = \Delta p_{-1} + \beta\left(y - u_s\right)$$

  where u_i is the structural shock to the Phillips curve $i = s$, money supply $i = ms$, money demand $i = md$ and IS $i = is$.
- Galí considers a quarterly model for $z_t = \begin{pmatrix} \Delta gnp_t & \Delta i_t & i_t - \Delta p_t & \Delta m_t - \Delta p_t \end{pmatrix}'$.
- From reduced form estimates and using the restrictions listed in Table I, he then estimates the structural model.
- Results:
  - Supply shocks explain 70-80% of output.
  - Nominal shocks explain nominal variables.
  - IS shocks explain inflation.

GALÍ (1992) IDENTIFICATION

- KL section 10.4.1 discusses the identification used by Galí (1992). The underlying IS-LM model is used as a theoretical background to the model specification.
- Galí considers a quarterly model comprising the growth rate of GNP (Δgnp_t), the change in the interest rate (Δi_t), the real interest rate ($i_t - \Delta p_t$) and the change in the real money balance ($\Delta m_t - \Delta p_t$).
- The time series vector is defined as

$$z_t = \begin{bmatrix} \Delta gnp_t & \Delta i_t & i_t - \Delta p_t & \Delta m_t - \Delta p_t \end{bmatrix}'$$

determined by four structural shocks: an aggregate supply shock (w_t^{AS}), a money supply shock (w_t^{MS}), a money demand shock (w_t^{MD}) and a shock to the IS-curve (w_t^{IS}). The vector of structural shocks is given by

$$w_t = \begin{bmatrix} w_t^{AS} & w_t^{MS} & w_t^{MD} & w_t^{IS} \end{bmatrix}'$$

GALÍ (1992) IS-LM MODEL

- To identify the structural VAR, Galí imposes 6 restrictions divided into three groups of restrictions. These are all stated in Table 1.
- The first three restrictions (R1, R2 and R3) imposes long-run restrictions, the next two imposes short-run restrictions (R4 and R5). This provides 5 identifying restrictions.
- In addition, Galí discusses three additional restrictions (R6, R7 and R8) that can be used together with the long-run and short-run restrictions. These restrictions imply non-linear restrictions on the short-term (contemporaneous) effects.
- These restrictions are implemented on $\Theta(1)$ (the long-run restrictions R1, R2 and R3), on B_0^{-1} (short-run restrictions R4 and R5) and on B_0 (short-run restrictions R6, R7 and R8). Note that comparing our standard notation with the one used by Galí we have that $S = B_0^{-1}$, $A(0) = B_0$ and $C(1) = \Theta(1)$.

- We can now implement the long-run restrictions R1 ($\Theta_{12}(1) = 0$), R2 ($\Theta_{13}(1) = 0$) and R3 ($\Theta_{14}(1) = 0$) suggested by Galí implying that

$$\Theta(1) = \begin{bmatrix} * & 0 & 0 & 0 \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}$$

Restrictions R4 ($b^{-1}_{12,0} = 0$) and R5 ($b^{-1}_{13,0} = 0$) implies that

$$B_0^{-1} = \begin{bmatrix} * & 0 & 0 & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}$$

GALÍ (1992) IS-LM MODEL

- The remaining three restrictions are imposed on B_0. Restriction R6 which is discussed in KL implies that

$$B_0 = \begin{bmatrix} * & * & * & * \\ * & * & b_{23} & -b_{23} \\ * & * & * & * \\ * & * & * & * \end{bmatrix}$$

  Restriction R7 implies that

$$B_0 = \begin{bmatrix} * & * & * & * \\ * & 0 & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}$$

  whereas restriction R8 implies that

$$B_0 = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & 0 & * \\ * & * & * & * \end{bmatrix}$$

Galí (1992) IS-LM model

- We can now use the Matlab solver to estimate $\Theta(1)$, B_0^{-1} and B_0 conditional on restrictions R1 to R5 and then one of the restrictions R6, R7 or R8.
- To implement these restrictions we can make use of the following Matlab code (the restrictions.m file)

```
% restrictions.M
% Implements Gali (1992) restrictions R1-R5 and R6
function q=restrictions(guess)
global GAMMA SIGMA p
K=size(guess,1);
B0inv = guess(1:K,1:K);
B0 = inv(B0inv);
THETA1=Xi*B0inv;
F=vec(B0inv*B0inv'-SIGMA(1:K,1:K));
% Long run and short run restrictions
q=[F; B0inv(1,2); B0inv(1,3); B0(2,3)+B0(2,4); THETA1(1,2);
THETA1(1,3); THETA1(1,4)];
q'+1;
```

Galí (1992) IS-LM model

- It is then straightforward to make changes to this code to implement restrictions R7 and R8. MATLAB illustration: Gali1992.m.
- Potential problem: Using Galí's own replication code, the original data and his own implementation of the identification scheme (using restrictions R1-R6) we find that all results except the impulse response for the real interest rate to be identical or very close. But, impulse responses of the real interest rate is not!

A third alternative approach

- Closed form solution!

Identification based on CT model: Closed form solution

Curriculum: Warne (1992), sections 2.2, 3.2

- KL suggestion is to use the MATLAB/Python non-linear solver to compute B_0^{-1}.
- As an alternative we may use the closed form solutions suggested by Warne (1993) discussed in sections 2.2 and 3.2.
- Consider again the reduced form CT model (KL notation)

$$y_t = \Xi \sum_{i=1}^{t} u_i + \Xi^*(L) u_t + y_0^*$$

- We also have the structural form CT model

$$y_t = \underbrace{\Xi B_{0-1}}_{\Upsilon} \sum_{i=1}^{t} w_i + \Xi^*(L) B_0^{-1} w_t + y_0^* = \Upsilon \sum_{i=1}^{t} w_i + \Xi^*(L) B_0^{-1} w_t + y_0^*$$

- Note that we assume that the deterministic components defined in Warne (1992) equations (4) and (10) are assumed to be zero (they play no role when identifying the CT model).

Identification based on CT model: Closed form solution

- Let
$$B_0^{-1} = \begin{bmatrix} F_k \\ F_r \end{bmatrix}^{-1}$$
where F_k identifies the permanent shocks whereas F_r identifies the transitory shocks under the normalization that $\Sigma_w = I_K$.
- Is it possible to derive expressions for F_k and F_r such that we can compute B_0^{-1}?

Identification based on CT model: Closed form solution

- Comparing reduced and structural forms we find that

$$\Upsilon w_t = \Xi u_t, \text{and} \quad \Upsilon\Upsilon' = \Xi\Sigma_u\Xi'$$

- Given the definition of Ξ we also have that

$$\beta'\Upsilon = 0$$

- To identify the permanent shocks, Warne suggest the following method:
    - When there are $K - r = k$ stochastic trends

    $$\Upsilon = \Upsilon_0\pi$$

    where Υ_0 is a $K \times k$ matrix with known parameters.
    - Υ_0 is chosen such that $\beta'\Upsilon_0 = 0$.
    - This implies that the freely estimated parameters in Υ are lumped into the $k \times k$ matrix π.
    - Given $\Upsilon\Upsilon' = \Xi\Sigma_u\Xi'$ we find that

    $$\pi\pi' = \left(\Upsilon_0'\Upsilon_0\right)^{-1}\Upsilon_0'\Xi\Sigma_u\Xi'\Upsilon_0\left(\Upsilon_0'\Upsilon_0\right)^{-1}$$

    where we can uniquely identify $k(k+1)/2$ parameters using the Cholesky decomposition.

- Since Υ_0, Ξ and Σ_u are known (we use the VEC estimates), we can compute π. Given π, we can then compute Υ.
- Since $\Upsilon w_t = \Xi u_t$ and if we decompose $w_t = \begin{pmatrix} \phi_t & \psi_t \end{pmatrix}'$ we find

$$\phi_t = \left(\Upsilon'\Upsilon\right)^{-1} \Upsilon'\Xi u_t$$

  where ϕ_t contain the k permanent shocks and ψ_t are the r transitory shocks.
- We have now identified the permanent shocks. The top $k \times K$ matrix F_k in B_0^{-1} is equal to $\left(\Upsilon'\Upsilon\right)^{-1} \Upsilon'\Xi$.

Identification based on CT model: Closed form solution

- It remains to identify the transitory shock. Warne suggest that the $r \times K$ matrix F_r in B_0^{-1} is determined by

$$\psi_t = Q^{-1}\xi'\Sigma_u^{-1}u_t$$

where Q is an $r \times r$ matrix, $\xi = \alpha\left(U\alpha\right)^{-1}$ and U is an $r \times K$ matrix.
- The covariance matrix for the transitory shocks is:

$$\mathrm{E}\left(\psi_t\psi_t'\right) = Q^{-1}\xi'\Sigma_u^{-1}\xi\left(Q'\right)^{-1}$$

implying that Q must be chosen such that $\xi'\Sigma_u^{-1}\xi$ is diagonalized.
- U-matrix: This is the main matrix to be determined. Note: $U\alpha$ must be invertible.
- If we know U, then we can compute ξ (given the estimates of α). Then we can use a Cholesky decomposition of $\xi'\Sigma_u^{-1}\xi$ to determine Q, i.e.

$$QQ' = \xi'\Sigma_u^{-1}\xi$$

and then finally we can compute $F_r = Q^{-1}\xi'\Sigma_u^{-1}$ to be inserted into B_0^{-1}.

- How to find the U matrix?
- It is an $r \times K$ selection matrix. We can either design this automatically in case we are not interested in identifying the transitory shocks or implement specific restrictions.
- A simple choice is to use

$$\begin{bmatrix} 0_{r \times K-r} & I_r^+ \end{bmatrix}$$

  where I_r^+ is the $r \times r$ anti-diagonal matrix (exchange matrix).
- Examples using Blanchard-Quah and King-Plosser-Stock-Watson models below.

Potential issues with Blanchard & Quah identification

- Works if time series vector is $I(1)$ and there are r cointegration vectors.
- If cointegration vector is not of the form $\begin{pmatrix} 0 & 1 \end{pmatrix}'$, then we need to adjust the method.
- Use the closed form solution.

Blanchard-Quah: Analytical solution

- We assume, as in Blanchard-Quah, that the bivariate time series vector x_t is generated by the reduced form VMA model (using notation and equation numbers from Bergman, 1996)
$$\Delta x_t = \mu + C(L)\varepsilon_t \tag{3}$$
where $x_t = \begin{pmatrix} y_t & \pi_t \end{pmatrix}'$.
- We assume that $y_t \sim I(1)$ whereas $\pi_t \sim I(0)$ such that the cointegration vector is $\beta' = \begin{bmatrix} 0 & 1 \end{bmatrix}$. It follows that: $\mathrm{rank}\,[C(1)] = 1$, $\beta' C(1) = 0$, $\beta' x_t \sim I(0)$
- We know that the VMA model is the solution of the reduced form VAR in levels
$$A(L)x_t = \rho + \varepsilon_t \tag{4}$$
and that the reduced form VAR can be written as the VEC model
$$\Delta x_t = \sum_{i=1}^{p-1} \Lambda_i \Delta x_{t-i} + \alpha\beta' x_{t-1} + \rho + \varepsilon_t \tag{5}$$
and that the model can be written as the reduced form CT model
$$x_t = C(1)\left[\rho t + \sum_{i=1}^{t} \varepsilon_i\right] + x_0 + C^*(L)\varepsilon_t \tag{6}$$

Blanchard-Quah: Analytical solution

- The structural form VMA is
$$\Delta x_t = \mu + R(L)\nu_t \tag{1}$$
where $\nu_t = \begin{pmatrix} \phi_t & \psi_t \end{pmatrix}'$, $\mathrm{E}\left[\nu_t \nu_t'\right] = I_2$ and $R(L) = C(L)\Gamma^{-1}$.
- Problem: How to compute Γ^{-1} given reduced form estimates.
- The idea in Warne is based on the transformed reduced form model (called restricted VAR in Warne).
- Define some matrices:
$$D(L) \equiv \begin{bmatrix} I_k & 0 \\ 0 & (1-L)I_r \end{bmatrix}, \qquad D_\perp(L) \equiv \begin{bmatrix} (1-L)I_k & 0 \\ 0 & I_r \end{bmatrix}$$
and
$$M \equiv \begin{bmatrix} S_k' & \beta \end{bmatrix}'$$
where the rows of the $k \times K$ matrix S_k satisfy $S_{i,k}C(1) \neq 0$.

Blanchard-Quah: Analytical solution

- For the Blanchard-Quah model we find that

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad D_{\perp} = \begin{bmatrix} (1-L) & 0 \\ 0 & 1 \end{bmatrix}$$

- According to Warne, $z_t = D_{\perp} M x_t$. Using the matrices above we find

$$\begin{bmatrix} (1-L) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = \begin{bmatrix} \Delta y_t \\ \pi_t \end{bmatrix} = z_t$$

- Warne shows that we now can represent our reduced forms VAR, VMA, VEC and CT models as the Restricted VAR

$$B(L) z_t = \theta + \eta_t$$

  where $\theta = M\rho$ and $\eta_t = M\varepsilon_t$.
- Warne also provides us with expression allowing us to translate estimates of the Restricted VAR into estimates of the the VAR, VMA and CT models.

Blanchard-Quah: Analytical solution

- Using the expressions $A(L) = M^{-1}B(L)D_{\perp}(L)M$ and $C(L) = M^{-1}D(L)B(L)^{-1}M$, we find

$$A(L) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{bmatrix} \begin{bmatrix} (1-L) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} (1-L)B_{11}(L) & B_{12}(L) \\ (1-L)B_{21}(L) & B_{22}(L) \end{bmatrix}$$

  and

$$C(L) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & (1-L) \end{bmatrix} \begin{bmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} =$$

$$\frac{1}{\det[B(L)]} \begin{bmatrix} B_{22}(L) & -B_{12}(L) \\ -(1-L)B_{21}(L) & (1-L)B_{11}(L) \end{bmatrix}$$

Blanchard-Quah: Analytical solution

- The adjustment coefficients α (γ in Warne) is given by $\alpha = M^{-1}B(1)P_r$ where $P_r = \begin{bmatrix} 0 & I_r \end{bmatrix}'$, see equation (26).

$$\alpha = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} B_{11}(1) & B_{12}(1) \\ B_{21}(1) & B_{22}(1) \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} B_{12}(1) \\ B_{22}(1) \end{bmatrix}$$

- Finally,

$$C(1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & (1-L) \end{bmatrix} \begin{bmatrix} B_{11}(1) & B_{12}(1) \\ B_{21}(1) & B_{22}(1) \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} =$$

$$\frac{1}{\det\left[B(1)\right]} \begin{bmatrix} B_{22}(1) & -B_{12}(1) \\ 0 & 0 \end{bmatrix}$$

Blanchard-Quah: Analytical solution

- We can now proceed and identify the structural shocks as discussed above. Note that $\Upsilon_0 = \begin{bmatrix} 1 & 0 \end{bmatrix}'$, ϕ (or π in Warne) is a scalar given by $\left(\Upsilon_0'\Upsilon_0\right)^{-1}\Upsilon_0'\Xi\Sigma_u\Xi'\Upsilon_0\left(\Upsilon_0'\Upsilon_0\right)^{-1}$ which reduces to the solution

$$\varphi_t = \frac{1}{\phi}\frac{1}{\det\left[B(1)\right]}\begin{bmatrix} B_{22}(1) & -B_{12}(1) \end{bmatrix}\varepsilon_t = \Gamma_k\varepsilon_t$$

Blanchard-Quah: Analytical solution

- The transitory shock is given by ($U = \begin{bmatrix} 0 & 1 \end{bmatrix}$)

$$\psi_t = \left(\alpha' \Sigma^{-1} \alpha\right)^{-0.5} \alpha' \Sigma^{-1} \varepsilon_t = \Gamma_r \varepsilon_t$$

- We have now identified the structural shocks and can proceed as usual, either by using the VEC estimates or by using the restricted VAR estimates using the expressions in Warne.
- Note: Warne also derives the asymptotic distribution of IRFs and FEVDs, we will use bootstrap techniques instead.
- Matlab example: BQBergman.m Python example: BQBergman.py

Blanchard-Quah: Analytical solution

- Results: Permanent shocks important for output, transitory shocks more important for inflation. Mixed results on cointegration tests.

Blanchard-Quah: Analytical solution

Table 2

Likelihood ratio tests for the number of cointegration vectors in the unrestricted VAR system [a]

Country	Estimated coint. vec. $\hat{\beta}$		LR_{tr} test $r=0$	$r \leq 1$	Stationarity tests $H_s(y)$	$H_s(\pi)$
Germany	0.013 1.000	−0.026 1.000	16.37 *	1.85	9.06 **	5.11 *
Japan	0.026 1.000	−0.022 1.000	23.08 **	5.72 *	8.68 **	9.36 **
Sweden	0.029 1.000	−0.290 1.000	14.45 +	0.24	13.42 **	2.75 +
U.K.	−0.017 1.000	−0.327 1.000	12.76	0.31	11.78 **	0.20
U.S.	−0.003 1.000	−0.040 1.000	19.26 *	0.33	15.66 **	0.53

[a] All tests are based on VAR models with a constant and seasonal dummies. The quantiles for the trace statistic, LR_{tr}, are taken from Johansen and Juselius (1990, Table A1). Stationarity tests refer to a likelihood ratio test of the null hypothesis that each variable is stationary given that one cointegration vector is present in the VAR model. This test is χ^2 distributed with 1 degree of freedom.

** Statistically significant at the 1 percent level,

* Statistically significant at the 5 percent level

\+ Statistically significant at the 10 percent level.

Blanchard-Quah: Analytical solution

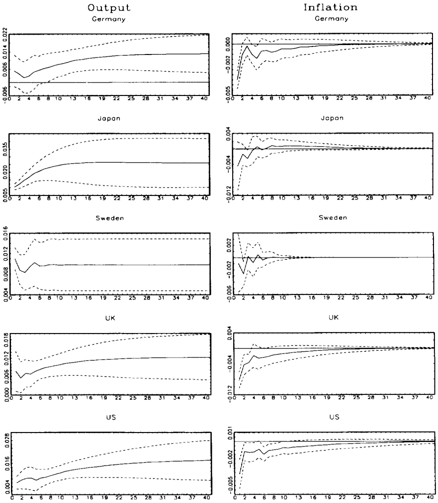


Fig. 1. Impulse response of the level of output and inflation to a one standard deviation permanent shock and the two standard deviation confidence bands.

Blanchard-Quah: Analytical solution

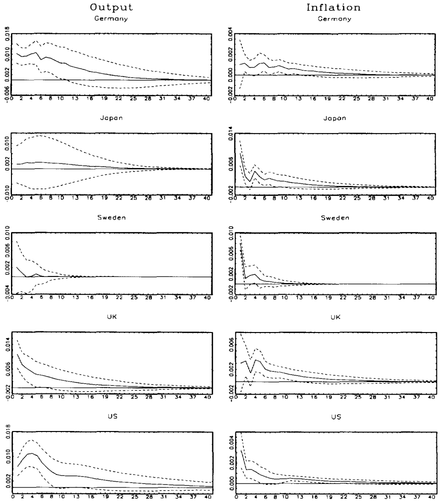


Fig. 2. Impulse response of the level of output and inflation to a one standard deviation transitory shock and the two standard deviation confidence bands.

Blanchard-Quah: Analytical solution

Table 3
Variance decomposition analysis. The fraction of the level of output and the rate of inflation explained by permanent shocks [a]

	Output					Inflation				
k	Germany	Japan	Sweden	U.K.	U.S.	Germany	Japan	Sweden	U.K.	U.S.
1	0.206	0.973	0.964	0.332	0.454	0.919	0.202	0.013	0.883	0.564
	(0.280)	(0.111)	(0.100)	(0.274)	(0.237)	(0.184)	(0.284)	(0.067)	(0.182)	(0.234)
2	0.184	0.975	0.973	0.341	0.426	0.846	0.202	0.108	0.807	0.584
	(0.268)	(0.106)	(0.082)	(0.274)	(0.232)	(0.202)	(0.283)	(0.103)	(0.201)	(0.232)
3	0.147	0.973	0.980	0.396	0.403	0.819	0.243	0.104	0.816	0.536
	(0.236)	(0.105)	0.060)	(0.273)	(0.228)	(0.196)	(0.286)	0.096)	(0.196)	(0.230)
4	0.128	0.976	0.984	0.436	0.392	0.803	0.212	0.107	0.746	0.536
	(0.223)	(0.096)	(0.047)	(0.269)	(0.226)	(0.204)	(0.256)	(0.104)	(0.192)	(0.230)
8	0.214	0.984	0.992	0.617	0.472	0.702	0.191	0.108	0.685	0.571
	(0.264)	(0.067)	(0.024)	(0.208)	(0.205)	(0.230)	(0.201)	(0.099)	(0.203)	(0.227)
12	0.350	0.988	0.995	0.736	0.618	0.661	0.194	0.108	0.672	0.581
	(0.277)	(0.047)	(0.016)	(0.150)	(0.162)	(0.235)	(0.173)	(0.099)	(0.203)	(0.228)
16	0.485	0.992	0.996	0.806	0.716	0.646	0.201	0.108	0.667	0.584
	(0.254)	(0.035)	(0.012)	(0.109)	(0.127)	(0.235)	(0.164)	(0.099)	(0.202)	(0.230)
20	0.593	0.994	0.997	0.851	0.783	0.640	0.206	0.108	0.666	0.586
	(0.215)	(0.027)	(0.010)	(0.081)	(0.099)	(0.232)	(0.161)	(0.099)	(0.202)	(0.231)
24	0.672	0.995	0.997	0.881	0.829	0.639	0.208	0.108	0.665	0.587
	(0.177)	(0.022)	(0.008)	(0.063)	(0.078)	(0.231)	(0.160)	(0.099)	(0.201)	(0.232)
32	0.770	0.996	0.998	0.916	0.885	0.637	0.209	0.108	0.665	0.588
	(0.122)	(0.015)	(0.006)	(0.042)	(0.051)	(0.230)	(0.159)	(0.099)	(0.201)	(0.233)
40	0.825	0.997	0.998	0.936	0.916	0.638	0.209	0.108	0.665	0.588
	(0.091)	(0.012)	(0.005)	(0.030)	(0.036)	(0.230)	(0.159)	(0.099)	(0.201)	(0.233)

[a] Asymptotic standard errors are reported in the parentheses below each estimate. These standard errors are computed using the theory developed by Warne (1993) which is described in Appendix A.

KPSW model: Analytical solution

- In the King, Plosser, Stock and Watson (1991) trivariate model we have the following cointegration vector

$$\beta = \begin{pmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}'$$

- Υ_0 satisfying the condition that $\beta' \Upsilon_0 = 0$ is given by

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

- It is then straightforward to compute

$$\pi = \text{chol}\left(\left(\Upsilon_0' \Upsilon_0 \right)^{-1} \Upsilon_0' \Xi \Sigma_u \Xi' \Upsilon_0 \left(\Upsilon_0' \Upsilon_0 \right)^{-1} \right)$$

- Which in turn will give us an estimate of $F_k = \left(\Upsilon' \Upsilon \right)^{-1} \Upsilon' \Xi$ in

$$B_0^{-1} = \begin{bmatrix} F_k \\ F_r \end{bmatrix}$$

KPSW model: Analytical solution

- It remains to identify the transitory shock. Warne suggest that the $r \times K$ matrix F_r in B_0^{-1} is determined by

$$\psi_t = Q^{-1}\xi'\Sigma_u^{-1}u_t$$

where Q is an $r \times r$ matrix, $\xi = \alpha\left(U\alpha\right)^{-1}$ and U is an $r \times K$ matrix.

- The covariance matrix for the transitory shocks is:

$$\mathrm{E}\left(\psi_t\psi_t'\right) = Q^{-1}\xi'\Sigma_u^{-1}\xi\left(Q'\right)^{-1}$$

implying that Q must be chosen such that $\xi'\Sigma_u^{-1}\xi$ is diagonalized.

- U-matrix: This is the main matrix to be determined. Note: $U\alpha$ must be invertible.
- If we know U, then we can compute ξ (given the estimates of α). Then we can use a Cholesky decomposition of $\xi'\Sigma_u^{-1}\xi$ to determine Q, i.e.

$$QQ' = \xi'\Sigma_u^{-1}\xi$$

and then finally we can compute $F_r = Q^{-1}\xi'\Sigma_u^{-1}$ to be inserted into B_0^{-1}.

- How to find the U matrix?
- It is an $r \times K$ selection matrix. We can either design this automatically in case we are not interested in identifying the transitory shocks or implement specific restrictions.
- A simple choice is to use

$$\begin{bmatrix} 0_{r \times K-r} & I_r^{+} \end{bmatrix}$$

where I_r^{+} is the $r \times r$ anti-diagonal matrix (exchange matrix).

KPSW model: Analytical solution

- Alternatively, we can pre-set U to reflect the specific restrictions we want to impose on the contemporaneous effects of a shock on a specific variable. In the KPSW trivariate model, to impose a restriction that $b_0^{22} = 0$ we use the following U:

$$U = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

  and if we want to impose the restriction that $b_0^{12} = 0$, we set

$$U = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

  and finally, to implement the restriction that $b_0^{32} = 0$, we set

$$U = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Note: We can always re-order the variables such that we impose a zero restriction on the first column associated with a transitory shock (could help our interpretation of the shocks).

- KPSW also consider a 6-variable case with 3 cointegration vectors.
- This case can also be analyzed using either the solver or the analytical approach.
- We need to define Υ_0 and U.
- Y_0 should be the orthogonal complement to the cointegration vectors as suggested above. KPSW suggest that the matrix A in equation (8) on page 831 is a suitable choice.
- Since KPSW are only interested in the three common trends (permanent shocks) the choice of U is irrelevant. One could for example define U as

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- With these matrices and the estimated VEC model we can compute the B_0^{-1} matrix and then the implied IRF's.

Application: Galí (1999)

- Starting points:
  - RBC models predict high positive correlation between hours worked and labor productivity.
  - This positive correlation reflects the effects of technology shocks that shift labor demand schedule combined with an upward sloping labor supply curve.
  - This prediction is inconsistent with most empirical evidence.
  - As a result, RBC models are augmented with additional shocks (non-technology shocks) that shift the labor supply leading to negative correlation between productivity and hours worked.
  - Main question: What are the effects of technology and non-technology shocks on productivity and hours worked?
  - If technology shocks are the only source of fluctuations, sticky price (RBC) models predict a negative (positive) correlation between hours and labor productivity. For non-technology shocks, the correlation should be positive (negative) in the sticky price (RBC) model.
- Estimate the comovements: Galí suggests that the impulse response function from a bivariate VAR using long-run restrictions can reveal the comovements.

Application: Galí (1999)

- Let $y_t = \begin{pmatrix} gdp_t - h_t & h_t \end{pmatrix}'$ where gdp_t is the log of real GDP and h_t is the log of total employee hours in non-agricultural sector.
- Assume that both variables are $I(1)$ but not cointegrated.
- Estimate a VAR(5) model and impose long-run restriction that the second shock (identified as a non-technology shock) can have no long-run effect on productivity.

Application: Galí (1999)

- The VAR(5) model in levels can be rewritten as the VAR(4) model in first differences

$$z_t = \nu + A_1 z_{t-1} + \ldots + A_4 z_{t-4} + u_t$$

- To impose the long-run restriction we have

$$\Theta(1) = \left(I_2 - \sum_{i=1}^{4} A_i \right)^{-1} B_0^{-1} = \begin{bmatrix} \theta_{11}(1) & 0 \\ \theta_{12}(1) & \theta_{22}(1) \end{bmatrix}$$

  which is the same structure as in Blanchard-Quah decomposition.

- It is then straightforward to use the Cholesky decomposition.
- Remember that $\Theta(1)(\Theta(1))' = A(1)^{-1}\Sigma_u(A(1)^{-1})'$
- Cholesky decomposition implies

$$\hat{\Theta}(1) = \mathrm{chol}(\hat{A}(1)^{-1}\hat{\Sigma}_u(\hat{A}(1)^{-1})') = \begin{bmatrix} \hat{\theta}_{11}(1) & 0 \\ \hat{\theta}_{12}(1) & \hat{\theta}_{22}(1) \end{bmatrix}$$

- The structural impact multiplier matrix

$$\hat{B}_0^{-1} = \hat{A}(1)\hat{\Theta}(1)$$

  and the structural shocks $\hat{w}_t = \hat{B}_0 u_t$

Application: Galí (1999)

- Since $r = 0$ the VEC reduces to a VAR in first differences and we can apply the method above.

$$\hat{\Upsilon} = \mathrm{chol}(\Upsilon\Upsilon') = \mathrm{chol}(\hat{\Gamma}(1)^{-1}\hat{\Sigma}_u(\hat{\Gamma}(1)^{-1})')$$

  and

$$\hat{B}_0^{-1} = \left[\hat{\Gamma}(1)\mathrm{chol}(\Upsilon\hat{\Upsilon}')\right]$$

- MATLAB example: figure11_1_VAR.m, figure11_1_VEC.m and figure11_1_solve.m
  Python example: figure11_1_VAR.py, figure11_1_VEC.py and figure11_1_solve.py
- Galí (1999) replication: Gali1999replication.m and Gali1999replication.py

GALÍ RESULT

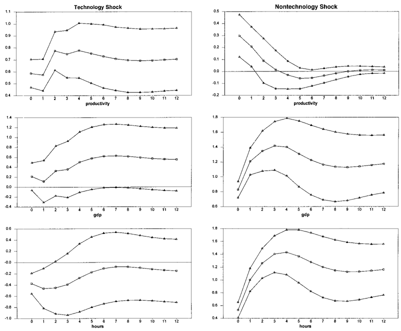


FIGURE 2. ESTIMATED IMPULSE RESPONSES FROM A BIVARIATE MODEL: U.S. DATA, FIRST-DIFFERENCED HOURS (POINT ESTIMATES AND ±2 STANDARD ERROR CONFIDENCE INTERVALS)

Results: Negative correlation between productivity and hours following a positive productivity shock and a positive correlation following a non-technology shock. Inconsistent with RBC.

Assumptions made by Galí

- The time series vector $y_t = \begin{pmatrix} gdp_t - h_t & h_t \end{pmatrix}' \sim I(1)$ but not cointegrated.
- This implies that $\Delta y_t \sim I(0)$.
- Testable hypotheses:
    - In the VAR model with Δy_t we should find full rank, rank = 2. Trace test should reject that $r = 1$.
    - In the VAR model with y_t, rank should be zero such that $y_t \sim I(1)$ and $\Delta y_t \sim I(0)$.
- Testing these two hypotheses we find that rank=2 in the first difference model, but we also find rank=2 in the levels model.
- How are impulse responses affected if h_t is stationary?
- In this situation, we have $z_t = \begin{pmatrix} \Delta(gdp_t - h_t) & h_t \end{pmatrix}' \sim I(0)$. Identification identical to Blanchard-Quah.

WHAT HAPPENS IF $y_t \sim I(1)$ AND β UNRESTRICTED?

- Use the closed form solution.
- Example: PPP analysis in Bergman, Cheung & Lai (2010).

BERGMAN, CHEUNG & LAI (2010)

- Background:
  - PPP Puzzle: “How can we reconcile the enormous short–term volatility of real exchange rates with the extremely slow rate at which shocks appear to damp out?” Rogoff (1996)
  - Empirical evidence suggest large and persistent deviations from PPP.
  - What is generating these deviations is an unresolved empirical issue.
- Extensive literature analyzing the sources of PPP deviations.
- In the literature they either assume that PPP holds (demand shocks dominate) or that PPP does not hold (real shocks dominate).
- No paper looking at the case when PPP does not hold, but exchange rates and relative prices cointegrate.

Bergman, Cheung & Lai (2010)

- We argue that both real and monetary factors determine the real exchange rate even when the real exchange rate is stationary.
- Both real and monetary factors can have permanent effects on the nominal exchange rate and the relative price but these effects cancel out.
- We decompose the joint dynamics of the nominal exchange rate and the relative price into a common–trend and a transitory component.
- We investigate whether our decomposition can yield insights that may help resolve the PPP puzzle.
- Finally, we test whether these two components (estimated structural shocks) reflect real and nominal variables such as labor productivity, money supply, interest rates or oil price shocks.

Bergman, Cheung & Lai (2010)

Assume that the nominal exchange rate ($e_t = \ln \mathcal{E}_t$) and the relative price levels ($p_t = \ln P_t^d / P_t^f$) can be modeled as the bivariate VEC system:

$$\Delta X_t = \mu + \Sigma_{j=1}^{k} \Gamma_j \Delta X_{t-j} - \Pi X_{t-k} + \varepsilon_t \tag{2}$$

where $X_t = \begin{bmatrix} e_t & p_t \end{bmatrix}'$, $-\Pi = \alpha\beta'$, rank(Π) = 1, $\alpha = \begin{bmatrix} \alpha_1 & \alpha_2 \end{bmatrix}'$ contains adjustment coefficients, β is the cointegration vector, and ε_t are i.i.d. with mean zero and covariance matrix, Ω.

- If $\beta = \begin{pmatrix} 1 & -1 \end{pmatrix}$ then PPP holds.
- If $\beta = \begin{pmatrix} 1 & -\beta_{12} \end{pmatrix}$ then PPP does not hold but exchange rates and relative price cointegrate.
- How can we identify the structural model?

BERGMAN, CHEUNG & LAI (2010)

- Apply the Warne approach!

Identification

- Identification is based on Warne and the estimation of a restrictive VAR.
- Let the cointegration vector be given by $\beta = \begin{bmatrix} 1 & \beta_{12} \end{bmatrix}'$,

$$M = \begin{bmatrix} 1 & 0 \\ 1 & \beta_{12} \end{bmatrix}$$

  and

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

- Redefine the time series vector as $z_t = \begin{bmatrix} \Delta e_t & e_t - \beta_{12} p_t \end{bmatrix}'$.

IDENTIFICATION

- Estimate the restricted VAR
$$B(L)\, z_t = \theta + \eta_t$$
- Compute $B(1)$ and $C(1) = M^{-1} DB(1)^{-1} M$ and $\Sigma_u = M\Omega M^{-1}$.
- Given $\Upsilon_0 = \beta_\perp$ we can compute π using equation (16) (note this will be a scalar) which will give us Υ and the $F_k = (\Upsilon'\Upsilon)^{-1} \Upsilon' C(1)$.
- The adjustment coefficients can be computed using equation (26): $\alpha = M^{-1} B(1)\, P_r$ where $P_r = \begin{bmatrix} 0 & 1 \end{bmatrix}$. Let $U = \begin{bmatrix} 0 & 1 \end{bmatrix}$ and compute $\xi = \alpha\,(U\alpha)^{-1}$, $Q = \xi'\Sigma_u\xi$ and finally $F_r = Q^{-1}\xi\Sigma_u^{-1}$.
- We then have that $B_0^{-1} = \begin{bmatrix} F_k \\ F_r \end{bmatrix}^{-1}$

DATA

Consumer price indices and dollar–based exchange rates, April 1973 through December 2005, three major currencies — the German mark, the Japanese yen, and the British pound.

Preliminary data analysis

- Lag order in VAR model. We use two lags for the case of Germany and three lags for the cases of Japan and the United Kingdom.
- Unit root and cointegration tests (Table 1).

COINTEGRATION TESTS

Table 1. Results of cointegration and unit-root tests

	Bartlett corrected trace test for cointegration				Restriction test of $\beta' = [1 \quad -1]$	
	$r=0$	p-value	$r=1$	p-value	Estimated β'	p-value
Panel A: Cointegration analysis						
Germany	26.91**	0.035	3.85	0.761	$[1 \quad -3.03]$	0.235
Japan	15.54**	0.048	0.00	0.962	$[1 \quad -1.02]$	0.898
UK	37.31**	0.001	5.60	0.522	$[1 \quad 40.43]$	0.000
	ADF	ADF–GLS				
Panel B: Unit-root analysis						
Germany	−2.244	−2.216**				
Japan	−2.101	−1.042				
UK	−2.616*	−1.837*				

Notes: The trace test is Johansen's (2000) Barlett-corrected Likelihood Ratio (LR) trace test (for which r indicates the number of cointegration vector being tested under the null) and the estimated normalized cointegration vector (i.e. the β estimate) is also reported in each case. The lag order used for the Johansen procedure is determined using the SIC with a maximum of 12 lags allowed. The restriction test for $\beta' = [1 \quad -1]$ examines the null hypothesis that the long-run PPP condition holds. For unit-root tests, ADF is the standard ADF test, whereas ADF–GLS represents the more efficient GLS-based test devised by Elliott *et al*. (1996). The lag parameters for individual unit-root tests are chosen using the SIC with the maximum lag order set to 16.

* and ** denote statistical significance at the 10 and 5% levels, respectively.

Sources of the Common–Trend and Transitory Innovations

- We have found that both CT and transitory innovations are important contributors to real exchange rate dynamics.
- We then explore whether the CT and transitory innovations are attributable to changes in both real and monetary macroeconomic variables — including changes in productivity differentials, relative money supply, interest rate differentials and oil price changes.

Sources of the Common–Trend and Transitory Innovations

- Both permanent and transitory shocks are important contributors to real exchange rate dynamics.
- What is generating estimated permanent and transitory shocks?
- Hypothesis: Both real and monetary macroeconomic variables — changes in productivity differentials, relative money supply, interest rate differentials and oil price changes.

- Let ΔZ_t denote the macroeconomic variable (productivity) and w_t be the estimate of the two structural shocks, then we can write down the regression model

$$w_t = S(L)\,\Delta Z_t + u_t. \tag{10}$$

To handle problems of measurement errors, we estimate the bivariate VAR model and the regression above as a system.

- Combine this with the structural VAR

$$B(L)\,y_t = \theta + MF^{-1}w_t \tag{11}$$

where

$$M = \begin{bmatrix} 1 & 0 \\ 1 & -\beta_{12} \end{bmatrix}$$

- Such that

$$B(L)\,y_t = \theta + MF^{-1}S\Delta Z_t + MF^{-1}u_t = \theta + S^{*}\Delta Z_t + \nu_t \tag{12}$$

Note: Same lag structure for both endogenous and macro factors.

Sources of the Common–Trend and Transitory Innovations

Table 3. Causality tests for the sources of the CT and transitory innovations of RERs

	US macro variables			Domestic macro variables			
	ΔPR	ΔMS	ΔIR	ΔPR	ΔMS	ΔIR	OIL
Panel A: With the PPP condition imposed							
CT innovations							
Germany	1.322	1.121	12.778**	3.195	4.076	5.755*	7.555**
	(0.516)	(0.571)	(0.002)	(0.202)	(0.130)	(0.056)	(0.023)
Japan	2.736	7.385*	13.720**	6.169	1.179	2.271	1.972
	(0.434)	(0.061)	(0.003)	(0.104)	(0.758)	(0.518)	(0.578)
UK	2.393	7.967**	0.474	1.184	2.358	0.136	3.557
	(0.495)	(0.047)	(0.924)	(0.757)	(0.502)	(0.987)	(0.313)
Transitory innovations							
Germany	1.881	7.341**	14.101**	3.460	0.553	21.381**	2.993
	(0.391)	(0.025)	(0.001)	(0.177)	(0.759)	(0.000)	(0.224)
Japan	3.265	6.089	4.289	0.270	0.146	2.566	0.972
	(0.353)	(0.107)	(0.232)	(0.966)	(0.986)	(0.464)	(0.808)
UK	5.345	8.330**	11.553**	2.371	0.041	22.549**	5.488
	(0.148)	(0.040)	(0.009)	(0.499)	(0.998)	(0.000)	(0.139)
Panel B: Without the PPP condition imposed							
CT innovations							
Germany	2.172	3.012	19.724**	5.497*	1.926	9.982**	5.689*
	(0.337)	(0.222)	(0.000)	(0.064)	(0.382)	(0.007)	(0.058)
Japan	2.734	7.378*	13.714**	6.176*	1.181	2.267	1.976
	(0.434)	(0.061)	(0.003)	(0.100)	(0.758)	(0.519)	(0.577)
UK	5.252	5.193	12.613**	2.271	0.045	21.204**	5.382
	(0.154)	(0.158)	(0.006)	(0.518)	(0.998)	(0.000)	(0.146)
Transitory innovations							
Germany	2.644	5.272*	21.993**	6.564**	0.319	16.940**	3.456
	(0.267)	(0.072)	(0.000)	(0.038)	(0.853)	(0.000)	(0.178)
Japan	3.024	6.175	3.593	0.387	0.200	2.467	0.953
	(0.388)	(0.103)	(0.309)	(0.943)	(0.978)	(0.481)	(0.813)
UK	5.764	6.515*	12.495**	1.712	0.122	20.551**	7.266*
	(0.124)	(0.089)	(0.006)	(0.634)	(0.989)	(0.000)	(0.064)

Notes: The different macro fundamental variables considered in the causality tests include US and domestic labour productivity (ΔPR), US and domestic money supply (ΔMS), and US and domestic interest rates (ΔIR), and oil prices (ΔOIL). The Wald test statistics presented above examine the null hypothesis of no causal effects from the corresponding variables on the structural shocks to the RER. The numbers in parentheses are p-values.
* and ** denote statistical significance at the 10 and 5% levels, respectively.

Sources of the Common–Trend and Transitory Innovations

- The overall results suggest that CT innovations cannot be associated with productivity shocks.
- Productivity shocks are not related to transitory shocks.
- Both CT and transitory innovations are driven by monetary variables.
- Oil price shocks may play only a minor role.
- Results are not very sensitive to the assumption that PPP holds.

Bergman, Cheung & Lai (2010): Conclusions

- CT dynamics are important for the real mark and the real yen rate, not for the real pound rate.
- Estimated contribution of common–trend dynamics sensitive to the assumption of PPP.
- Overall results suggest a relevant but limited role of macro fundamentals in estimated shocks. Results not robust to currency but to PPP assumption.
- Some interesting patterns still emerge.
  - Money supply and interest rate changes are more important than productivity and oil price changes in explaining both CT and transitory dynamics.
  - U.S. money supply changes are more important than domestic money supply changes.
- Our findings are consistent with recent open–economy models discussed by Chari, Kehoe and McGrattan (2002), Benigno (2004), Hairault and Sopraseuth (2005).
- Half–lives of the real exchange rate in response to transitory shocks ranging from 4.1 to 4.6 years and in response to CT shocks ranging from 2.2 to 4.4 years. Our results cannot explain or solve the PPP puzzle.
- Replication: BCLreplication.m and BCLreplication.py

CAN STRUCTURAL VAR MODELS RECOVER RESPONSES IN DSGE MODELS?

- Galí used structural VAR models to evaluate responses in DSGE models, in particular RBC models.
- Conclusion was: RBC models imply impulse responses rejected by data, therefore, we should reject RBC models.
- Galí paper started a debate!
- Christiano, Eichenbaum and Vigfusson (2004) argued that hours should be modeled as a stationary variable (but they use hours worked per capita, not total hours worked as in Galí). If so, then structural VAR analysis suggest that technology drives up hours and output (the opposite to what Galí found).
- Francis and Ramey (2005) showed that RBC models with richer preference and technology structures produce impulse responses consistent with the Galí result.

Structural VARs and DSGE models

- Chari, Kehoe and McGrattan (2008) challenged the Galí results on methodological grounds, the estimates have severe bias. Based on simulations of an RBC where the response of hours is positive, the data suggests the opposite. Therefore, severe bias. They conclude: VAR models with long-run restrictions should not be used.
- This conclusion has been challenged by Christiano, Eichenbaum and Vigfusson (2006). Three arguments: criticize Chari et.al. model specification, hours should be stationary, and impulse responses should take sampling uncertainty into account.
- Kehoe (2006) criticized the econometric approach suggested by Christiano et.al.
- Debate still going on.
- Conclusion from this debate: Better to allow for more than two shocks, important to use enough lags (VARMA structure in DSGE models translate into VAR(∞) models), identification and specification similar across VAR and DSGE models, use confidence bands (different approaches suggested), consider alternatives to long-run restrictions.
- An excellent idea for an MSc thesis!

Lessons for empirical work

- Well specified model is essential!
- Better to have more than 2 structural shocks in the model, but a trade-off.
- Always check whether minor changes in specification/identification matter.
- Important to base identification on theory.
- Choice of approach (KL, Warne, CT based, stationary representation based and so on) is a matter of taste.
- Next topic: Inference on IRFs and FEVDs, KL ch. 12

Advanced Macroeconomics: Structural Vector Autoregressive Analysis

Lectures 20 to 21

U. Michael Bergman
University of Copenhagen

Fall 2024

THESE SLIDES COVER

- Further perspectives on the identification of permanent and transitory shocks.
- Historical decompositions as counterfactuals.
- Forecast scenarios.
- Overidentifying restrictions.
- Practical issues when estimating VAR models.
- VAR, DESM and DSGE models. A comparison.

Identification of permanent and transitory shocks

Curriculum: Pagan and Pesaran (2008)

- Paper focuses on the generalization of the Blanchard-Quah paper. In that paper we had one permanent and one transitory shock. The question is how to handle models with more than one shock of each type.
- First part of the paper (sections 2-3) covers the econometric approach to estimate such models. The same topic is covered by KL also.
- The second part discusses a couple of well-known models in the literature and provides further perspectives and also points out some mistakes.
- First, they revisit the Blanchard-Quah paper. Then, they discuss the King, Plosser, Stock and Watson (1991) and the Galí (1992) papers, also covered previously and in the homework assignment.
- Finally, they also discuss some other models. We will only cover the Shapiro & Watson (1988) model.

Notes from sections 2 and 3

- Let the structural VAR(2) model with n variables be given by (using Pagan & Pesaran notation)

$$A_0 z_t = A_1 z_{t-1} + A_2 z_{t-2} + \varepsilon_t$$

where we assume that $\Sigma = I_n$.

- Rewrite the VAR model by subtracting z_{t-1} from both sides such that

$$A_0 \Delta z_t = -A(1)\, z_{t-1} - A_2 \Delta z_{t-1} + \varepsilon_t$$

where $A(1) = A_0 - A_1 - A_2$.

- The associated reduced form model is then given by

$$\Delta z_t = -A_0^{-1} A(1)\, z_{t-1} - A_0^{-1} A_2 \Delta z_{t-1} + A_0^{-1} \varepsilon_t$$

or as

$$\Delta z_t = -\Pi z_{t-1} - \Psi \Delta z_{t-1} + e_t$$

- Assume now that there are $r < n$ cointegration relations in the system implying that Π has reduced rank and that $\Pi = \alpha\beta'$.
- The structural form VECM can then be written as

$$A_0 \Delta z_t = -\alpha^* \beta' z_{t-1} - A_2 \Delta z_{t-1} + \varepsilon_t$$

where $\alpha^* = A_0 \alpha$.

Notes from sections 2 and 3

- To identify the structural model we need to estimate the parameters in A_0. There are n^2 coefficients in A_0 but only n equations implying that we can only identify $n(n-1)$ coefficients.
- To identify these parameters we need to apply *a priori* assumptions derived or inspired by economic models/reasoning!
- The number of cointegration vectors r uniquely determines the number of transitory shocks present in the underlying VEC/VAR model. If the cointegraiton rank is r then we can decompose the structural shocks into r transitory shocks and $n-r$ permanent shocks.
- Identification can be achieved by imposing restrictions on the long-run effects of permanent shocks (restrictions on Υ using our standard notation) and transitory shocks can be identified by imposing restrictions on the contemporaneous effects of transitory shocks (restrictions on B_0^{-1}).
- How many restrictions do we need? To identify $n-r$ permanent shocks we need $(n-r)(n-r-1)/2$ restrictions and to identify the transitory shocks we need to impose $r(r-1)/2$ restrictions.
- Pagan & Pesaran then continue discussing identification in terms of instrumental variables. This is not covered in the curriculum!

- Consider the six variable model in KPSW comprised of output y_t consumption c_t, investment inv_t, real money balance $(m_t - p_t)$, nominal interest rate i_t and inflation Δp_t.
- They assume that there are three permanent shocks in this system.
- Therefore, there are three cointegration vectors. Let $z_t = (y_t, i_t - \Delta p_t, m_t - p_t, c_t, inv_t, i_t)'$ and assume the following cointegration vectors (they are now treated as known)

$$\beta' = \begin{pmatrix} -1 & -\phi_1 & 0 & 1 & 0 & 0 \\ -1 & -\phi_2 & 0 & 0 & 1 & 0 \\ -\beta_y & 0 & 1 & 0 & 0 & \beta_r \end{pmatrix}$$

- The long-run effects of the three permanent shocks are given by

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta_y & -\beta_r & -\beta_r \\ 1 & \phi_1 & 0 \\ 1 & \phi_2 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

(this is $\tilde{A}$ in the KPSW paper or Υ_0 in the KL notation).

- The long-run impact of the permanent shocks are given by

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta_y & -\beta_r & -\beta_r \\ 1 & \phi_1 & 0 \\ 1 & \phi_2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \pi_{11} & 0 & 0 \\ \pi_{21} & \pi_{22} & 0 \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix}$$

- And we then find that the long-run effects are given by

$$\begin{bmatrix} \pi_{11} & 0 & 0 \\ \pi_{21} & \pi_{22} & 0 \\ \beta_y\pi_{11} - \beta_r\pi_{21} - \beta_r\pi_{31} & -\beta_r\pi_{22} - \beta_r\pi_{32} & -\beta_r\pi_{33} \\ \pi_{11} + \phi_1\pi_{21} & \phi_1\pi_{22} & 0 \\ \pi_{11} + \phi_2\pi_{21} & \phi_2\pi_{22} & 0 \\ \pi_{21} + \pi_{31} & \pi_{22} + \pi_{32} & \pi_{33} \end{bmatrix}$$

- Compare these long-run effects to the ones in the original KPSW model.

- Equation (8) in KPSW implies that

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & \phi_1 \\ 1 & 0 & \phi_2 \\ \beta_y & -\beta_r & -\beta_r \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \pi_{11} & 0 & 0 \\ \pi_{21} & \pi_{22} & 0 \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix}$$

$$= \begin{bmatrix} \pi_{11} & 0 & 0 \\ \pi_{11} + \phi_1 \pi_{31} & \phi_1 \pi_{32} & \phi_1 \pi_{33} \\ \pi_{11} + \phi_2 \pi_{31} & \phi_2 \pi_{32} & \phi_2 \pi_{33} \\ \beta_y \pi_{11} - \beta_r \pi_{21} - \beta_r \pi_{31} & -\beta_r \pi_{22} - \beta_r \pi_{32} & -\beta_r \pi_{33} \\ \pi_{21} + \pi_{31} & \pi_{22} + \pi_{32} & \pi_{33} \\ \pi_{21} & \pi_{22} & 0 \end{bmatrix}$$

- How many restrictions have we imposed here?
- How many restrictions are imposed in PP?

Pagan & Pesaran on Galí

- Galí estimates a four variable VAR inspired by an IS-LM model.

Pagan & Pesaran on Galí

- Galí estimates a four variable VAR inspired by an IS-LM model.
- All four variables are assumed to be $I(1)$ and there are two cointegration vectors present in the system. Let $z_t = (y_t, \Delta m_t, \pi_t, i_t)$ and that we have the following two cointegration vectors

$$\beta' = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

  where the first vector is the change in the real money balance and the second is the real interest rate.

Pagan & Pesaran on Galí

- Galí estimates a four variable VAR inspired by an IS-LM model.
- All four variables are assumed to be $I(1)$ and there are two cointegration vectors present in the system. Let $z_t = (y_t, \Delta m_t, \pi_t, i_t)$ and that we have the following two cointegration vectors

$$\beta' = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

  where the first vector is the change in the real money balance and the second is the real interest rate.
- In this system we have two permanent and two transitory shocks. We need one restriction to identify the permanent shocks and one to identify the transitory shocks.

Pagan & Pesaran on Galí

- Galí is estimating a VAR using the time series vector $z_t = (\Delta y_t, \Delta i_t, i_t - \Delta p_t, \Delta m_t - \Delta p_t)$ and then suggest that 6 restrictions are needed to identify the system. This is correct if we are not taking cointegration into account!

Pagan & Pesaran on Galí

- Galí is estimating a VAR using the time series vector $z_t = (\Delta y_t, \Delta i_t, i_t - \Delta p_t, \Delta m_t - \Delta p_t)$ and then suggest that 6 restrictions are needed to identify the system. This is correct if we are not taking cointegration into account!
- For example, Galí is using three instead of one restriction to identify the permanent shocks and three instead of one restriction to identify the transitory shock. This is not necessary if we estimate the VECM instead!

PAGAN & PESARAN ON GALÍ

- Galí is estimating a VAR using the time series vector $z_t = (\Delta y_t, \Delta i_t, i_t - \Delta p_t, \Delta m_t - \Delta p_t)$ and then suggest that 6 restrictions are needed to identify the system. This is correct if we are not taking cointegration into account!
- For example, Galí is using three instead of one restriction to identify the permanent shocks and three instead of one restriction to identify the transitory shock. This is not necessary if we estimate the VECM instead!
- We recognize that the VAR estimated by Galí is in fact a restricted VAR suggested by Warne. And then it is straightforward to use the closed form solution to estimate the IRF's.

Historical decompositions as counterfactuals

- We have previously discussed how to use historical decompositions.
- The historical decomposition illustrates the cumulative effect of the structural shocks on the variables allowing us to evaluate what type of shock is relatively more important than other shocks are.
- One example is illustrated in KL Figure 4.2. The global oil market model focusing on the real oil price from the late 1970's.

HISTORICAL DECOMPOSITIONS AS COUNTERFACTUALS

- Replicate Figure 4.2: See Figure4_2.zip

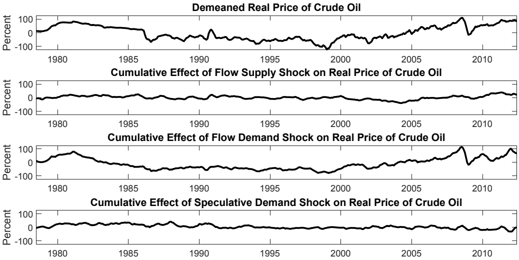


Figure 4.2: Historical decomposition of the real price of crude oil in percent deviations from the mean.
Source: Kilian and Lee (2014).

Historical decompositions as counterfactuals

- An alternative to the standard historical decomposition is to construct counterfactuals.
- A simple approach is to ask how the variable would evolve if one structural shock is absent. Let y_{kt} denote the k^{th} actual time series and let $\bar{y}^j_{kt}$ be the cumulative contribution of shock j to the variable k up to date t. Then the counterfactual is given by

$$y_{kt} - \bar{y}^j_{kt}$$

- The counterfactual shows how the variable k would have evolved if shock j is zero for all time periods.
- Interpretation: If $y_{kt} - \bar{y}^j_{kt} > y_{kt}$, then shock j lowered the actual value. Opposite if counterfactual is below the actual value.

Historical decompositions as counterfactuals

- KL Figure 4.4 illustrates. Replication file: Figure4_4.zip

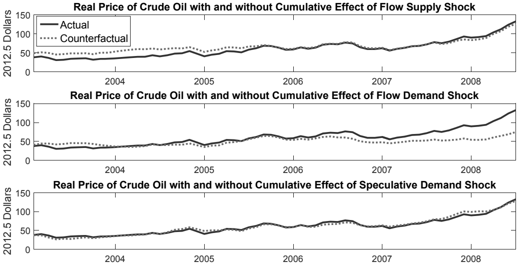


Figure 4.4: Historical counterfactuals for the real price of crude oil from January 2003 to June 2008
Source: Kilian and Lee (2014).

Forecast scenarios

- We can use the estimates of a VAR to asses the consequences of different forecast scenarios.
- This is different from standard forecasts (something we have not covered during the course) since these are based on estimated structural VARs. Note: The objective is not to conduct forecasting exercises! It is a hypothetical “what-if” question.
- Given our estimated model, we may be interested in how a change either in the variables themselves or in the structural shocks would affect a certain variable in our model.
- For example, we might be interested in how the effects of a certain change in the future interest rate will affect house prices, how shocks during historical episodes affect variables for example how monetary policy shocks during the Bernanke era would affect the economy. Both these approaches are based on the assumption that parameters in the structural VAR do not change over time.
- The latter approach is easier to implement if it is possible to quantify shocks and deviations from estimated shocks.

Forecast scenarios

- Consider the estimated structural VAR model

$$y_{t+h} = \sum_{i=0}^{\infty} \Theta_i w_{t+h+i} = \overbrace{\sum_{i=0}^{h-1} \Theta_i w_{t+h-i}}^{\text{cumulative effect of previous shocks}} + \underbrace{\sum_{i=h}^{\infty} \Theta_i w_{t+h-i}}_{\text{unconditional expectation}}$$

  Note that the second term is the unconditional forecast where all future shocks are zero and can be computed iterating the autoregressive reduced form VAR model.
- Forecast scenarios can be constructed by adding a sequence of nonzero future shocks. The forecasts we compute then represent conditional point forecasts.
- It is necessary to normalize all forecasts to a baseline case (the unconditional forecast) to construct the conditional point forecast.

FORECAST SCENARIOS

- To illustrate: Assume that the vector of future structural shocks is

$$\mathbb{E}\left(\sum_{i=0}^{h-1}\Theta_i w_{t+h-i} + \sum_{i=h}^{\infty}\Theta_i w_{t+h-i} \mid \left\{w_{t+h-i} = w_{t+h-i}^{\text{scenario}}\right\}_{i=0}^{h-1}, \Omega_t\right)$$

$$-\mathbb{E}\left(\sum_{i=0}^{h-1}\Theta_i w_{t+h-i} + \sum_{i=h}^{\infty}\Theta_i w_{t+h-i} \mid \left\{w_{t+h-i} = w_{t+h-i}^{\text{baseline}}\right\}_{i=0}^{h-1}, \Omega_t\right)$$

$$= \sum_{i=h}^{\infty}\Theta_i w_{t+h-i} \mid \left\{w_{t+h-i} = w_{t+h-i}^{\text{scenario}}\right\}_{i=0}^{h-1}, \Omega_t = \sum_{i=0}^{h-1}\Theta_i w_{t+h-i}^{\text{scenario}}$$

  where Ω_t denotes the information set available at time t. To arrive at the result, note that $w_{t+h-i}^{\text{baseline}} = 0$ for $i = 0, \ldots, h-1$.
- Note 2: This is similar to the definition of a an impulse response function.
- In practice we then only have to compute the deviation of the sequence of scenario shocks from 0 and then multiply with the baseline forecast.
- The difference between these scenarios and IRF's is that we use a sequence of shocks to compute the forecasts based on scenarios whereas IRF's are computed assuming a unit shock in period 1 where all other shocks are equal to zero.

FORECAST SCENARIOS

- Note: We need to use the full sample to estimate the structural VAR and we don't need to re-estimate the model for each point forecast.
- Example: Baumeister & KIlian (2014) global oil model.
- They construct a number of forecast scenarios for the real crude oil price; (i) a return of Iraqui oil production to full capacity, (ii) a supply disruption in Libya, (iii) a strong recovery of the global economy, (iv) a financial meltdown similar to Lehman Brothers collapse, and two contagion scenarios reflecting political events in Middle East.
- They use historical precedent or construct purely hypothetical scenarios (and shocks).
- Figure 4.5 illustrates. Replication: Figure4_5.zip

FORECAST SCENARIOS

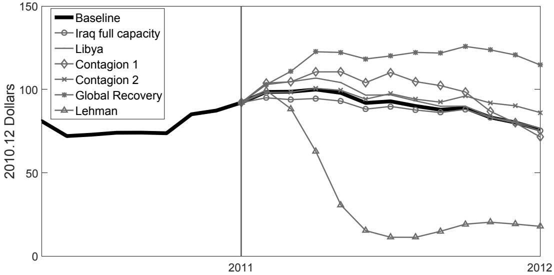


Figure 4.5: Selected real-time forecast scenarios for the real price of crude oil as of December 2010.
Source: Adapted from Baumeister and Kilian (2014).

HOW TO HANDLE OVERIDENTIFYING RESTRICTIONS

Curriculum: KL section 9.2.3 and 12.13.3.

- We have discussed just identified structural VARs where we apply
  - Cholesky decompisitions in stable VARs (recursive identification).
  - Alternative approaches to identify non-recursive systems.
  - Alternative approaches to identify long-run and short-run restrictions.
- In some cases depending on the underlying economic model (reasoning) we may want to impose additional restrictions on the B_0^{-1} matrix (or on the B_0 matrix).
- One example discussed in KL is the Kilian (2009) global market for oil model. This model was used to illustrate recursive identification (KL section 8.4.2) and to illustrate different bootstrap methods (we replicated Figure 12.6).
- Recall the following:
- Trivariate model comprised of world crude oil production, global real economic activity and the price of oil.
- $y_t = (\Delta prod_t, rea_t, rpoil_t)'$.
- Structural shocks are identified as follows

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 & 0 \\ b_0^{21} & b_0^{22} & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} \end{bmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix}$$

How to handle overidentifying restrictions

- From previous slide

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 & 0 \\ b_0^{21} & b_0^{22} & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} \end{bmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix}$$

- This identification is based on the idea of a vertical short-run oil supply curve and a downward-sloping short-run oil demand curve.
- There are two demand shocks, aggregate and oil-specific where there is a delay such that oil-specific demand shocks raising the oil price cannot immediately affect global economic activity or the production of oil.
- Aggregate demand affects the price of oil immediately but oil production with a delay.
- And finally we have oil production shocks having effects on all variables immediately.
- Kilian discusses whether we should add the restriction that $b_0^{21} = 0$. The argument is: Higher oil prices caused by oil-supply shocks should not affect global economic activity immediately.

How to handle overidentifying restrictions

- The Cholesky estimate of B_0^{-1} is found to be

$$\hat{P} = \mathrm{chol}\left(\hat{\Sigma}_u\right) = \begin{bmatrix} 1.5617 & 0 & 0 \\ 0.0735 & 4.0588 & 0 \\ -0.0044 & 0.0047 & 0.0594 \end{bmatrix}$$

  suggesting that $b_0^{21} \approx 0$.
- How should we estimate B_0^{-1} if we impose the over-identifying restriction that $b_0^{21} = 0$?
- We may use GMM! Note that there are more moment conditions than unknown parameters and therefore the method-of-moments estimator (using the Matlab/Python solver) cannot satisfy all conditions at the same time.
- But, what we can do is to minimize a weighted average of the moment conditions. Weights are given by the inverse of the variance-covariance matrix of the sample moment conditions.

How to handle overidentifying restrictions

- We will use the same approach as when using the method-of-moments approach for just identified recursive or non-recursive models.
  1. Estimate the reduced form VAR.
  2. Equate the sample second moment of the reduced form VAR innovations with the same second moment expressed as function of the structural B_0^{-1} matrix.
- This implies that the relevant moment condition is

$$\mathbb{E}\left(\text{vech}\left(u_t u_t'\right) - \text{vech}\left(B_0^{-1} B_0^{-1\prime}\right)\right) = 0$$

  where we as usual normalize the structural shocks such that $\sum_w = I_K$.
- Rewrite this expression in terms of estimated residuals from the reduced form VAR

$$\text{vech}\left(\frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t'\right) - \text{vech}\left(B_0^{-1} B_0^{-1\prime}\right) = \text{vech}\left(\tilde{\sum}_u\right) - B_0^{-1} B_0^{-1\prime}$$

How to handle overidentifying restrictions

- The idea is now to search for B_0^{-1} that minimizes this expression. Find the unknown elements in B_0^{-1} that minimizes

$$J = T\left(\text{vech}\left(\widetilde{\Sigma}_u\right) - \left(\tilde{\sum}_u\right) - B_0^{-1}B_0^{-1\prime}\right)' \widehat{W}\left(\text{vech}\left(\widetilde{\Sigma}_u\right) - \left(\tilde{\sum}_u\right) - B_0^{-1}B_0^{-1\prime}\right)$$

  where $\widehat{W}$ is the inverse of the estimated variance-covariance matrix of the sample moments, i.e.,

$$\widehat{W} = \left[\frac{1}{T}\sum_{t=1}^{T}\left(\text{vech}\left(\hat{u}\hat{u}'\right) - \overline{\text{vech}\left(\hat{u}\hat{u}'\right)}\right)\left(\text{vech}\left(\hat{u}\hat{u}'\right) - \overline{\text{vech}\left(\hat{u}\hat{u}'\right)}\right)'\right]^{-1}$$

  where

$$\overline{\text{vech}\left(\hat{u}\hat{u}'\right)} = \frac{1}{T}\sum_{t=1}^{T}\text{vech}\left(\hat{u}\hat{u}'\right)$$

- In order to test the overidentifying restriction we know that

$$\widehat{J} \xrightarrow{d} \chi_n^2$$

  where n is the number of overidentifying restrictions. Note: Given that the remaining restrictions are correct!

How to handle overidentifying restrictions

- Estimation of B_0^{-1} under overidentifying restrictions. We want to impose the restriction that the element $b_0^{21} = 0$.
- Consider the trivariate VAR describing the global oil market. Assume that the VAR is stable and that the lag length is equal to 24.
- Estimating this VAR and if we apply the Cholesky decomposition we find that

$$\hat{P} = \mathrm{chol}\left(\hat{\Sigma}_u\right) = \begin{bmatrix} 1.5617 & 0 & 0 \\ 0.0735 & 4.0588 & 0 \\ -0.0044 & 0.0047 & 0.0594 \end{bmatrix}$$

  as mentioned above.
- Then we need to impose the restriction and minimize

$$J = T\left(\mathrm{vech}\left(\widetilde{\Sigma}_u\right) - \left(\tilde{\sum}_u\right) - B_0^{-1}B_0^{-1\prime}\right)' \widehat{W}\left(\mathrm{vech}\left(\widetilde{\Sigma}_u\right) - \left(\tilde{\sum}_u\right) - B_0^{-1}B_0^{-1\prime}\right)$$

- We can use the Matlab function fminsearch or Python function fmin (scipy.optimize.fmin) to find a vector x that minimizes a function $f(x)$ in the same manner as we used the solver to find the solution of B_0^{-1} in just identified systems.
- We need to set up this function and then supply starting values.

HOW TO HANDLE OVERIDENTIFYING RESTRICTIONS

- A good guess of B_0^{-1} could be

$$\begin{bmatrix} 1.5617 & 0 & 0 \\ 0 & 4.0588 & 0 \\ -0.0044 & 0.0047 & 0.0594 \end{bmatrix}$$

- Then we need to specify the function J to be minimized.
- Then we need apply the functions in Matlab or in Python to obtain an estimate of the 5 parameters in B_0^{-1}.

How to handle overidentifying restrictions

- How do we handle bootstrap if we want to compute confidence bands around the IRFs?
- In the same way as we used bootstrap in models identified with the help of the solver!
- One suggestion in KL is to recenter the bootstrap moment condition in each bootstrap replication. They use the “bias” correction $\mathrm{vech}\Sigma_u - \mathrm{vech}\left(B_0^{-1} B_0^{-1\prime}\right)$ to correct each replication of B_0^{-1}.
- Replication: Figure12_8.zip

Practical Issues: KL 19

These slides cover:

- Deterministic components and trends.
- Seasonality
- Structural change

Deterministic components and trends

- Straightforward to include, constant and linear trends in VARs.
- Linear trend in first differences? If so, then quadratic trend in levels.
- Allow for linear trend in cointegration vector? Five alternatives in general! No deterministic components (unrealistic), constant in cointegration vector but no trend in levels data (less likely specification), constant in cointegration vector and linear trend in levels data (a common case), constant and linear trend in cointegration vector and linear trend in levels data (often the case), constant and linear trend in cointegration vector and linear trends in first differences and quadratic trend in levels (less likely in standard applications).
- Test the null that the linear trend is absent in the cointegration vector.
- Should we de-trend data prior to estimating a VAR?
- Depends on the focus of the analysis. But, all univariate filters (HP-filter and band-pass filters) are less reliable. They find cyclical variation even when it is absent in the series.
- Could bias VAR analysis.

Seasonality

- Use seasonally adjusted data.
- Straightforward to include seasonal dummies in the VAR model. Only if seasonality is deterministic!
- Centered seasonal dummies should be used! No effect on distribution of Johansen trace test.
- Stochastic seasonality: Distinguish between stationary and integrated seasonality.
  - Stationary seasonality: seasonal effects vary over time. Estimate VAR for different seasonal periods.
  - Integrated seasonality: Use seasonal-differencing (month-to-month growth rates) and then use the transformed series in the VAR analysis.
- Apply seasonal filters: X-12-ARIMA or TRAMO-SEATS methods used by statistical agencies. If data is not seasonally adjusted, use any of these methods to filter the date and then use the seasonally adjusted data in the VAR analysis.
- Note: All filters used may introduce unwanted dynamics in the filtered series that could affect inference and conclusions.

Structural change

- Straightforward to allow for structural breaks if the timing is known.
- Examples: Monetary regimes, chairman of the central bank and so on.
- Outliers can be handled by adding dummies.
- Main problem: Are parameters time-invariant?
- If parameters are not time-invariant we need to model this variation.
- How to detect time variation?
    - Split sample if timing is known.
    - Estimate the VAR using rolling windows.
    - Estimate a time varying VAR model (TVC-VAR). Difficult to estimate, small samples, need to specify regime switches (immediate or smooth changes). Some models work well for certain data and samples but not when extending the sample or when adding variables. How many regimes?
- Advice: Check for outliers in the data (and in the estimated residuals). Add dummies to account for these outliers. Check the data for apparent changes. Use exogenous information to split the sample (monetary regime, for example). For cointegration tests, use rolling windows to check that the rank is time-invariant.
- Next topic: DEMS, DSGE and SVAR. Note: We have covered some aspects of DSGE and SVAR previously!

VAR in a historical perspective

- Pioneered by Sims (1980) Econometrica paper.
- Dominant approach in empirical macro since then.
- Cowles Foundation macro models dominant prior to 1980 (Klein, 1950 US model). Dynamic Simultaneous Equations Model (DSEM) built on Keynes. Could contain a large number of equations (even hundreds of equations).

VAR AND DSGE: KL CH. 6

- DSGE models: based on microstructure but assume that model structure is invariant to policy interventions.
- Utility and profit maximization with exogenous shocks (technology shocks, for example). Log-linearizing around a nonstochastic steady-state path dependent on state variables known to economic agents at each point in time.
- Simulate the dynamic behavior of the endogenous variables subject to shocks.
- The state-space representation of a linearized DSGE model can be written as a reduced-form VARMA models which under certain conditions can be approximated by a finite-order reduced form VAR model.
- The link between DSGE and structural VARs is less straightforward! It depends on how the VAR is identified.

VAR AND DSGE

- Main problem: Dimensionality! Number of shocks in DSGE models is smaller than the dimensionality of the model. (One technology shock in textbook RBC model.)
- KPSW model: Three variables but one shock in the DSGE model but three shocks when estimating the structural VAR. Structural VAR will not be consistent with DSGE.
- Need to:
  - Use reduced-rank estimation methods when estimating the structural VAR.
  - Add shocks to DSGE model (in an ad-hoc way, measurement errors).
  - Reduce the number of variables in the VAR.
  - Add structural shocks to the DSGE model (no ad-hoc shocks, economic shocks such as preference shocks).
- May need to consider a partial DSGE model, scale down the number of variables. Exclude variables where we don't have good data, for example the capital stock.
- Problem: Even if the time series vector is a VAR(p) process, a subset of this time series vector may not, in general, be a finite order VAR process. It will be a VARMA process.
- Potential solution: Approximate a VAR (∞) process by a finite order VAR process. But, how many lags to include? Open area of research.
- Why not use VARMA models instead? Difficult to estimate large scale VARMA processes.

VAR and DSGE

- DSGE models are evaluated using calibration and by comparing IRFs with IRFs from structural VARs or we could just estimate the DSGE model.
- General problem: How to interpret failure of matching model data with actual data?
- Typical response is to: refine the model, not reject the model!
- Problem how to compare IRFs (apart from problems using calibration and identification of structural VAR).
- Recent advances in impulse-response matching. Different approaches suggested, but no consensus.
- Open area of research: How good approximation is a VAR for DSGE models used by policy makers?
- What is the null hypothesis? Is the DSGE model the correct model? Or, is the structural VAR the correct model?
- Problem with convergence when estimating DSGE models, attempts to use Bayesian econometrics.

CONCLUSION

- Mechanical application of any of the three models should be avoided!
- DSGE and VAR models are complementary! Not, necessarily substitutes.
- DSGE and structural VARs have their own weaknesses and strengths.
- KL: “no basis for claims that one approach dominates the other”.

VAR and DESM

- Comparing VAR approach and Dynamic Simultaneous Equations Models (DESM) (and DSGE models).
- What is a DESM model?
- Underlying idea: to build empirical macro model to gain policy insights.
- Small number of endogenous variables and large number of exogenous (no contemporaneous or lagged feedback) variables.
- For example, we may illustrate DSEM models as

$$\begin{bmatrix} H_{11}(L) & H_{12}(L) \\ 0 & H_{22}(L) \end{bmatrix} \begin{pmatrix} z_t \\ x_t \end{pmatrix} = \begin{bmatrix} F_{11}(L) & 0 \\ 0 & F_{22}(L) \end{bmatrix} \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix}$$

  where $H_{ij}(L)$ and $F_{ij}(L)$ are matrix polynomials in the lag operator and $w_t = \begin{pmatrix} w'_{1t} & w'_{2t} \end{pmatrix}'$ are mutually uncorrelated white noise. Note: x_t is exogenous.
- In principle, this is a structural Vector Autoregressive Moving Average model with exogenous variables, a VARMAX model.
- Structural equations are typically estimated using single-equation methods where, in the reduced form, z_t is expressed as a function of endogenous and exogenous variables and contemporaneous exogenous variables.

VAR and DSEM

- Final form found by inverting the autoregressive lags for z_t and expressing z_t as a function of current and lagged values of the shocks and the exogenous variables.
- Advantages: allow for policy analysis using different scenarios (changes in exogenous variables) and an economic structure.
- Still used but not in academic research mainly since these models could not predict developments in the 1970s. (Lucas critique).
- Response: Developing new macro models based on rational expectations, DSGE models.
- KL section 6.1.1 illustrate that reduced form VAR models encompass many structural DSEM models.
- Structural VARs are estimated adding cross-equation restrictions and exclusion restrictions on the reduced form VAR parameters. These restrictions affect the dynamics.
- DSEMs built by independent blocks where all variables not determined within the block are given, there is no feedback between blocks.
- Structural VAR models can be seen as small-scale DSEM models where we impose restrictions on the impact matrix but leave lagged values unrestricted.

Comparing VAR, DSEM and DSGE

Features	DSGE	DSEM	SVAR
Exogeneity restrictions	Few	Many	None
Dynamic exclusion restrictions	Few	Many	Few
Number of variables	Large	Very large	Small
Number of shocks	Few	Many	Few
Trend treatment	Explicit	Implicit	Explicit
Micro structure required	Yes	No	No

NEXT LECTURE

- Summary of the course
- Few words on the exam.

Advanced Macroeconomics: Structural Vector Autoregressive Analysis

Lecture 19

U. Michael Bergman
University of Copenhagen

Fall 2024

These slides cover

- Sign restrictions in the Blanchard-Quah model.
- Sign restrictions in the general case.
- Application: Uhlig (2005)
- Identification of fiscal policy shocks: Blanchard-Perotti (2002).

CURRICULUM: KL 13.1-13.2

- Reconsider the Blanchard-Quah bivariate VAR model with supply and demand shocks.
- The relationship between reduced form residuals and structural shocks is

$$\begin{pmatrix} u_t^q \\ u_t^p \end{pmatrix} = B_0^{-1} \begin{pmatrix} w_t^{\text{supply}} \\ w_t^{\text{demand}} \end{pmatrix}$$

  where u_t^q is quantity (output) and u_t^p is the price level.
- If the short-run supply curve is vertical, then we expect the following relationship between structural shocks and output and prices.

$$\begin{pmatrix} u_t^q \\ u_t^p \end{pmatrix} = \begin{bmatrix} * & 0 \\ * & * \end{bmatrix} \begin{pmatrix} w_t^{\text{supply}} \\ w_t^{\text{demand}} \end{pmatrix}$$

SIGN RESTRICTIONS

- The underlying argument for this restriction was an AD-AS type model stating that a *positive supply shock* should has a *positive effect on output* in both the short- and long-run. In the short-run, a *positive supply shock* should have a *negative effect on prices* (if short-run supply curve is not vertical). A *positive demand shock* should tend to *increase both output and prices* in the short-run. Therefore:

$$\begin{pmatrix} u_t^q \\ u_t^p \end{pmatrix} = \begin{bmatrix} + & + \\ - & + \end{bmatrix} \begin{pmatrix} w_t^{\text{supply}} \\ w_t^{\text{demand}} \end{pmatrix}$$

- Comparing to standard identification we find that the structural shocks are not identified point-wise, but set identified.

Sign restrictions

- Another example is the effect of monetary policy. A monetary contraction should (i) raise federal funds rate, (ii) lower prices, (iii) decrease non-borrowed reserves and (iv) reduce output.
- But, empirical studies often find:
    - Liquidity puzzle: Monetary policy shocks (positive shocks to money stock) increase money stock and reduces the interest rate.
    - Price puzzle: Positive monetary policy shocks reduce inflation.
- Identification should be compatible with economic model predictions.
- How to implement restrictions on the sign of the responses?
- Static sign restrictions:
    - Sign restrictions on IRFs in impact period.
    - Sign restrictions on the contemporaneous effects, sign restrictions on B_0^{-1}.
- Dynamic sign restrictions: Restrictions in IRFs beyond the impact period.

SIGN RESTRICTIONS IN BIVARIATE MODELS

- Sign restrictions first suggested by: Faust (1998), Canova and De Nicolo (2002) and Uhlig (2005)
- Using our standard notation, we assume that the variance-covariance matrix of the structural shocks is normalized such that

$$\mathbb{E}\left(w_t w_t'\right) \equiv \Sigma_w = I_K$$

- Let $u_t = P\eta_t$ where $PP' = \Sigma_u$ and the shocks in η_t are uncorrelated and have variance equal to unity.
- We are searching for a candidate solution for structural shocks w_t^* that satisfies the sign restrictions on B_0^{-1}.
- Let $u_t = P\eta_t$ and we are searching for $w_t^* = Q'\eta_t$ where Q' is a square orthogonal matrix such as $Q'Q = QQ' = I_K$. Note: $u_t = PQQ'\eta_t = PQw_t^*$ and $B_0^{-1} = PQ$.
- Given that B_0^{-1} satisfies the sign restrictions, we can compute IRFs.
- Two approaches to compute candidate solutions for B_0^{-1}: Use Givens rotation matrices or Householder transformation.

- The structural VMA model is given by

$$y_t = C(L)B_0^{-1}B_0 u_t = C(L)B_0^{-1}w_t$$

- Blanchard and Quah impose the restriction that demand shocks have no long-run restrictions on output. This restriction implies that

$$\Upsilon = C(1)B_0^{-1} = \Xi B_0^{-1}$$

- Sign restrictions on B_0^{-1}?
    - A positive supply shock has a positive impact on both output and unemployment on impact.
    - A positive demand shock has a positive effect on output but a negative effect on unemployment on impact.

  This implies the following impact matrix

$$B_0^{-1} = \begin{bmatrix} + & + \\ + & - \end{bmatrix}$$

- How to implement this sign restriction?

Givens rotation matrix: Blanchard-Quah model

- Above we have that $\Sigma_u = PP'$. Let Q be an orthonormal matrix such that $QQ' = I_2$. Then we have that $B_0^{-1} = PQ$ and $\Sigma_u = PQ(PQ)' = PP'$.
- Assume that Q is given by the Givens matrix

$$Q = \begin{bmatrix} -\sin\phi & \cos\phi \\ \cos\phi & \sin\phi \end{bmatrix}$$

where ϕ is an unknown parameter.
- We know that

$$C(1)B_0^{-1} = C(1)PQ$$

- Using these relationships

$$C(1)\, B_0^{-1} = \begin{bmatrix} C_{11}(1) & C_{12}(1) \\ C_{21}(1) & C_{22}(1) \end{bmatrix} \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}.$$

Givens rotation matrix: Blanchard-Quah model

- The upper RHS element in $C(1)B_0^{-1}$ is equal to zero

$$\left(C_{11}\left(1\right)P_{11}+C_{12}\left(1\right)P_{21}\right)Q_{12}+C_{12}\left(1\right)P_{22}Q_{22}=0$$

  implying that

$$\frac{Q_{22}}{Q_{12}}=-\frac{C_{11}\left(1\right)P_{11}+C_{12}\left(1\right)P_{21}}{C_{12}\left(1\right)P_{22}}=\frac{\sin\phi}{\cos\phi}=\tan\phi$$

  and finally we obtain

$$\phi=\arctan\left[-\frac{C_{11}\left(1\right)P_{11}+C_{12}\left(1\right)P_{21}}{C_{12}\left(1\right)P_{22}}\right].$$

- Given the estimate of ϕ and $\Sigma_u=PP'$ we can now compute the matrix B_0^{-1} and use this to obtain IRFs.

Givens rotation matrix: Blanchard-Quah model

- Check that we find identical B_0^{-1} matrix using the Givens matrix.
- Add the code above to BQ_decomp.m file.
- Using the Givens matrix, we have imposed a sign restriction!

Householder transformation

- Instead of using the Givens rotation matrix, we can generate solutions for B_0^{-1} by using QR factorizations.
- Idea is based on the fact that any real square matrix W can be decomposed as $W = QR$ where Q is an orthogonal matrix (columns are orthogonal unit vectors) such that $QQ' = I$ and R is an upper triangular matrix. W is unique if it is invertible and if the diagonal of R is positive.
- Rubio-Ramírez, Waggoner and Zha (2010) algorithm: draw each column of the $K \times K$ matrix W at random from a $\mathcal{N}(0, I_K)$ distribution and apply the QR factorization to each draw. This provides us with Q.
- Then we compute $B_0^{-1} = PQ$ where P is the lower-triangular Cholesky decomposition of Σ_u.
- Potential pitfalls: MATLAB has a function computing the QR factorization. But, must ensure that the diagonal elements of R is positive. If a diagonal element in R is negative, then we need to switch sign of all elements in the row associated with a negative diagonal element and we need to adjust Q such that $W = QR$ holds.
- Compute the IRFs and check that the sign restrictions are satisfied, if not discard the IRFs and generate a new draw.
- To increase computational gain one could flip columns of the Q matrix or multiply by -1.

EXAMPLE: UHLIG (2005)

- Estimate the reduced form VAR/VEC model to obtain parameters (A/Γ) and residuals (u_t and Σ_u).
- Draw a random orthonormal matrix S', compute $P' = \mathrm{chol}(\Sigma_u)$ and finally compute $B_0^{-1} = PS$.
- Compute the IRFs.
- Are the sign restrictions satisfied? If yes, save the IRFs, if no discard the IRFs.
- Replicate N times and report the median IRFs together with confidence bands.

EXAMPLE: UHLIG (2005)

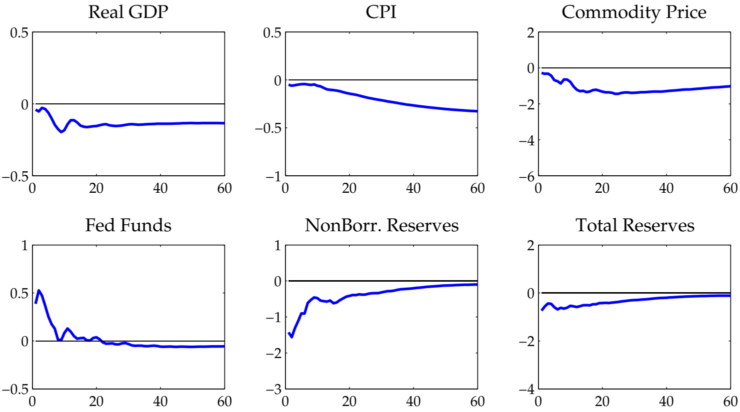
- US monthly data: GDP, CPI, Commodity prices, Federal Funds rate, non-borrowed reserves and total reserves. Sample: 1965:1-2003:12.
- Monetary contractions should
  - raise the Federal Funds rate
  - lower prices
  - decrease non-borrowed reserves
  - reduce output
- Empirical studies often find:
  - Liquidity puzzle: Monetary policy shocks (positive shocks to money stock) increase money stock and increase the interest rate.
  - Price puzzle: Contractionary monetary policy shocks (interest rate increases) increases inflation.

EXAMPLE: UHLIG (2005)

- According to Uhlig, a contractionary monetary policy shock does not lead to:
  - Increases in prices.
  - Increases in non-borrowed reserves.
  - Decreases in the Federal Funds rate.
- No restrictions on the response of output.
- Apply the method outlined above.

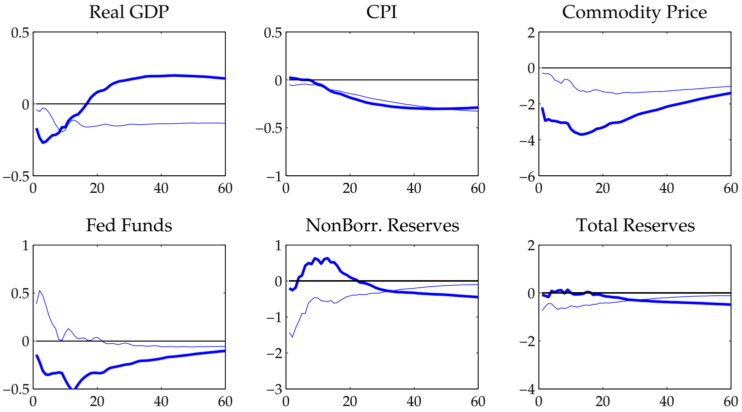
EXAMPLE: UHLIG (2005)

First draw: all signs are correct so save this



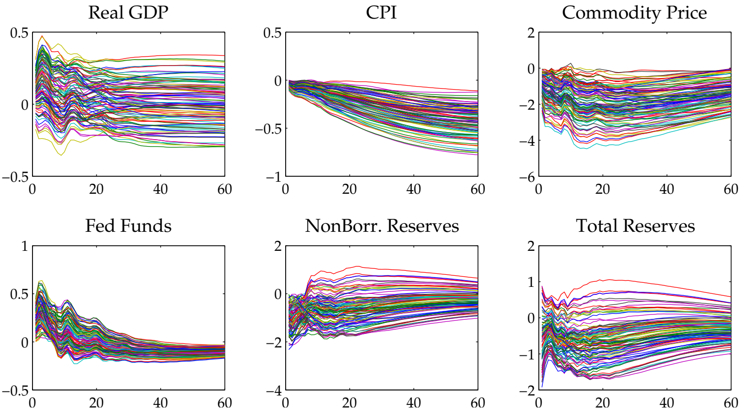
Example: Uhlig (2005)

Second draw: signs are incorrect, discard this



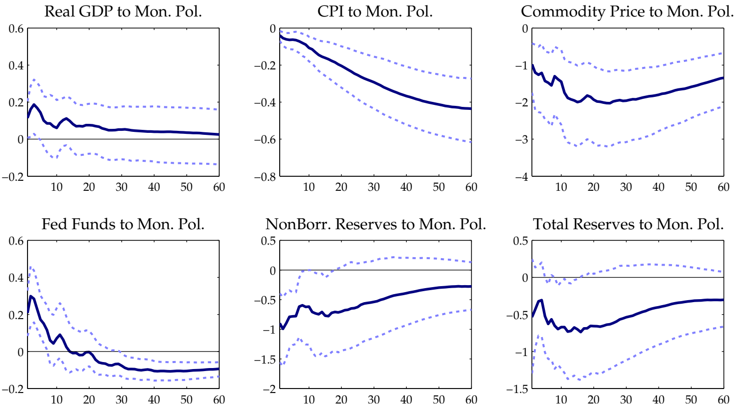
EXAMPLE: UHLIG (2005)

After a while:



EXAMPLE: UHLIG (2005)

IRFs and confidence bands: correct signs



EXAMPLE: UHLIG (2005)

Replication package for matlab: UhligReplication.m

Concluding remarks

- Sign restrictions is of interest as an alternative.
- Requires some additional coding and could be time consuming.
- Sign restrictions are used but not very often.
- Next topics: Other non-recursive Blanchard-Perotti identification of fiscal policy shocks

CURRICULUM:

Blanchard and Perotti (2002) and KL section 8.5.1

- Short-run restrictions but not recursive!
- The effects of fiscal policy (tax shocks and government spending shocks).
- Idea: Condition on institutional information: tax system, transfer system and timing of tax collection.
- Specifically, they focus on the 1975 temporary tax cut.

CURRICULUM:

Blanchard and Perotti (2002) and KL section 8.5.1

- Model: Quarterly tri-variate VAR model

$$Y_t = A(L, q)\, Y_{t-1} + U_t$$

  where $Y_t = \begin{bmatrix} T_t & G_t & X_t \end{bmatrix}'$ and U_t is the reduced form vector of residuals. All variables measured in real terms and per capita.

- Why quarterly model? Taxes that are paid in the last quarter of each year depends on GDP in the current and past three quarters. In the past three quarters, there is no dependence on GDP.

- Identification: Structural shocks $e_t = \begin{bmatrix} e_t^t & e_t^g & e_t^x \end{bmatrix}'$.

$$\begin{bmatrix} 1 & 0 & -a_1 \\ 0 & 1 & -b_1 \\ -c_1 & -c_2 & 1 \end{bmatrix} \begin{bmatrix} t_t \\ g_t \\ x_t \end{bmatrix} = \begin{bmatrix} 1 & a_2 & 0 \\ b_2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e_t^t \\ e_t^g \\ e_t^x \end{bmatrix}$$

- Is the model identified? No! We need to impose 3 restrictions.

CURRICULUM:

Blanchard and Perotti (2002) and KL section 8.5.1

- First: Need to construct the elasticity of government spending to output (b_1) and of taxes (a_1) to output. They cannot find any feedback from output on government spending, therefore $b_1 = 0$. They collect data to compute the elasticity of taxes to output (at a quarterly frequency). The parameter a_1 is set equal to the sample average.
- Second: Using the assumed values of a_1 and b_1 we can compute cyclically adjusted reduced form tax and spending residuals:

  $$t_t' = t_t - a_1 x_t \quad \text{and} \quad g_t' = g_t.$$

  Since t_1' and g_t' are not correlated to e_t^x we can use them as instruments to estimate c_1 and c_2.
- Third: No convincing argument to identify a_2 and b_2! Therefore, they propose two alternatives: Either $a_2 = 0$ and $b_2 \neq 0$ or $a_2 \neq 0$ and $b_2 = 0$. Estimate these using standard regression.

- Imposing these restrictions, they find that output is falling following tax shocks and that output is increasing following spending shocks.
- Replication: BPreplication.m
- Next topic: Further issues on the identification of long-run and short-run shocks. KL section 11.3.1. Then we turn to forecast scenarios and counterfactuals. KL sections 4.3 and 4.4.

Advanced Macroeconomics: Structural Vector Autoregressive Analysis

Lecture 18

U. Michael Bergman
University of Copenhagen

Fall 2024

CURRICULUM: KL 12.8

What are Local Projections?

- Suggested by Jordà (2005) as an alternative to estimate IRFs using a VAR model.
- Attempts to solve the bias problem, that VAR slope estimates of the VAR are biased and the bias stemming from the non-linear transformations of VAR slope estimates (the IRFs).
- Has attracted interest recently and is used in empirical applications.
- Idea is to estimate the impulse response function without any non-linear transformation of VAR estimates.
- Works if DGP is a stationary and linear VAR.

IRFs from VAR

- The K-dimensional reduced form VAR(p) model can be written as

$$y_t = \nu + A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + u_t \tag{1}$$

- Under the assumption that the VAR model is stable, we can rewrite our model as a moving average model (VMA)

$$\begin{aligned} A(L) y_t &= \nu + u_t \\ &= A(L)^{-1} (\nu + u_t) \\ &= A(1)^{-1} \nu + A(L)^{-1} u_t \\ &= \delta + \Phi(L) u_t \end{aligned} \tag{2}$$

  where $\Phi(L) = I_K + \sum_{i=1}^{\infty} \Phi_i L^i$ and $\Phi(L)A(L) = I_K$.
- To compute the impulse response function we define

$$J = \begin{bmatrix} I_K & 0 & \cdots & 0 \end{bmatrix}$$

  which has the dimension $K \times Kp$ and compute the impulse response function using

$$\Phi_i = J A^i J' \tag{3}$$

  where the matrices Φ_i have the dimension $K \times K$.

IRFs from VAR

- The structural form VAR(p) model can be written as

$$B_0 y_t = B_1 y_{t-1} + \ldots + B_p y_{t-p} + w_t$$

  where the matrix B_0^{-1} can be computed using a Cholesky decomposition of Σ_u

- The structural IRFs are given by $\Theta_j = \Phi_j B_0^{-1}$

IRFs from VAR

- We know that any VAR(p) model can be written as a VAR(1) model (disregarding deterministic components)

$$Y_t = \mathbf{A} Y_{t-1} + U_t$$

where

$$Y_t \equiv \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad \mathbf{A} \equiv \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_K & 0 & \dots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_K & 0 \end{bmatrix}, \text{ and } \quad U_t \equiv \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- Successive substitution of Y_{t-i} yields

$$Y_{t+i} = \mathbf{A}^{i+1} Y_{t-1} + \sum_{j=0}^{i} \mathbf{A}^j U_{t+i-j}$$

Local Projections

- Consider the VAR representation from the previous slide

$$Y_{t+i} = \mathbf{A}^{i+1} Y_{t-1} + \sum_{j=0}^{i} \mathbf{A}^{j} U_{t+i-j}$$

- Left multiplying this by J yields

$$y_{t+i} = J\mathbf{A}^{i+1} Y_{t-1} + \sum_{j=0}^{i} J\mathbf{A}^{j} U_{t+i-j}$$

$$= J\mathbf{A}^{i+1} Y_{t-1} + \sum_{j=0}^{i} J\mathbf{A}^{j} J' u_{t+i-j}$$

$$= J\mathbf{A}^{i+1} Y_{t-1} + \upsilon_{t+i}$$

where we remember that $\Phi_i = J\mathbf{A}^i J'$.

Local Projections

- Let the $K \times Kp$ matrix $A^j = J\mathbf{A}^j$.
- Note that the first $K \times K$ block of A^j is equal to the reduced-form impulse response matrix Φ_j.
- We can then compute the structural impulse response as for the VAR model

$$\Theta_j = \Phi_j B_0^{-1}$$

- Jordà notes that each equation of the VAR system can be estimated equation-by-equation using least squares to obtain A^i and therefore also Φ_i.

Local Projections

- Stack the system of equations

$$\begin{pmatrix} y_t \\ \vdots \\ y_{t+H-1} \end{pmatrix} = \begin{bmatrix} A^1 \\ \vdots \\ A^H \end{bmatrix} Y_{t-1} + \begin{pmatrix} \upsilon_t \\ \vdots \\ \upsilon_{t+H-1} \end{pmatrix}$$

  containing impulse responses up to horizon H.
- Estimate these equations as a system using GLS.
- It is then straight forward to estimate the implied structural Local Projections.
- Asymptotic distribution is derived in Jordà (2005) or by using a block bootstrap methods as considered by Kilian and Kim (2009).

Comparing IRFs from VAR and Local Projections

- An alternative way is to compute the linear projection directly using a sequence of OLS regressions for each horizon h

$$y_{t+h} = \mu + F_1 y_t + F_2 y_{t-1} + \cdots + F_q y_{t-q+1} + u_{t+h} \quad \text{for} \quad h = 1, \ldots, H \tag{4}$$

- Note: u_{t+h} may be serially correlated or heteroscedastic (use robust standard errors). Lag length q must be common across all horizons.
- The response of y_{t+h} to a reduced form shock in t is F_1. Therefore,

$$\Phi_h^{LP(q)} = F_1$$

  with $\Phi_0^{LP(q)} = I_K$.
- The structural Local Projection is:

$$\Theta_h^{LP(q)} = \Phi_h^{LP(q)} B_0^{-1}$$

- Note: There is no algorithm to compute B_0^{-1}, it is computed using the reduced form VAR(p) model.

Why Local Projections?

- Replicating the analysis in Jordà.
- Two different experiments are use to evaluate local projections vs. structural VAR impulse responses.
- The first example is a standard VAR used to study the effects of monetary policy (Christiano et.al., 1996). We will replicate this example.
- The second example is based on simulations of a SVAR-GARCH model.
- Monetary policy example: six variable VAR(12) model comprised of non-agricultural payroll employment, personal consumption expenditures deflator, annual growth rate of the index of sensitive materials prices, federal funds rate, the ratio of nonborrowed reserves plus credit to total reserves, and the annual growth rate of M2.
- Model is identified using a Cholesky decomposition.
- There is also an empirical application, output-inflation trade-off in a New Keynesian model.
- Matlab replication: JordaReplication.m Python replication: JordaReplication.py

WHY LOCAL PROJECTIONS?

- Jordà suggests that LPs are:
  - Easy to compute, can be estimated directly using GLS (or LS) regressions.
  - Pointwise and joint inference of LPs is straightforward.
  - Structural LPs are more robust to model misspecification than standard VAR estimates.
- KL argues that none of these arguments hold!

Simplicity

- LS is simple and no need to use non-linear transformation.
- But, increase efficiency using GLS.
- However, estimating a large number of parameters and many of these are not of interest. In LP we estimate HqK^2 but are only interested in HK^2 of these, HqK^2 could be a very large number.
- In a VAR model we estimate pK^2 parameters independent on the horizon H.
- Kilian and Kim (2009) show that if the DGP is a VAR model, then structural LPs have larger bias as well as higher variance compared to IRFs from the VAR. ("Simplicity comes at a considerable cost")

INFERENCE

- Asymptotic distribution of LPs derived using the same assumptions as used for deriving the distribution of IRFs.
- Asymptotic distribution of LPs found to be less valid. Coverage accuracy of LPs is lower than the coverage accuracy of IRFs.
- Bootstrap might help, but very limited evidence suggesting that bootstrap methods are useful.

ROBUSTNESS

- KL: LPs are not more robust to misspecification than IRFs.
- IRFs are non-linear transformations of linear VAR estimates. LPs are linear approximations to non-linear IRFs.
- LPs suffer from the same misspecification problem as IRFs.
- LPs may not even be better approximations of VARMA processes. Simulations suggest the opposite!
- LPs are not superior to IRFs if the DGP is non-linear as long as the VAR has enough number of lags. Also, if DGP is non-linear, then B_0^{-1} estimates using linear VAR are invalid.

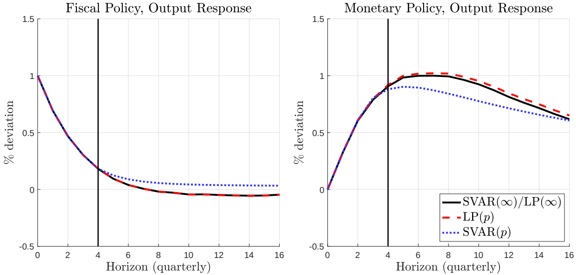
Robustness

- Important to reduce bias in IRFs, but LPs (in its current form) may not be the solution. More research warranted. Non-linear LPs?
- But, all this discussion about the pros and cons of local projection vs. VAR may be irrelevant!
- Plagborg-Møller and Wolf (2022) prove that local projections and VAR estimates the same impulse responses, any local projection IRF can be obtained from a recursive VAR and vice versa! Using the same lag length, the two methods provide approximately the same results but not for long horizons.
- But, in principle this is also implied in KL ch. 12.8!
- Then replicating figure 2 in Plagborg-Møller & Wolf (2022).
- The main point in this paper is that asymptotically there is no difference between local projections and structural VAR analysis.
- Another result (proposition 2) states that local projection and VAR impulse responses will be very similar at short horizons. In general, for p lags in the local projection and the VAR, then impulse responses are approximately equivalent. For horizons $> p$, they do not.

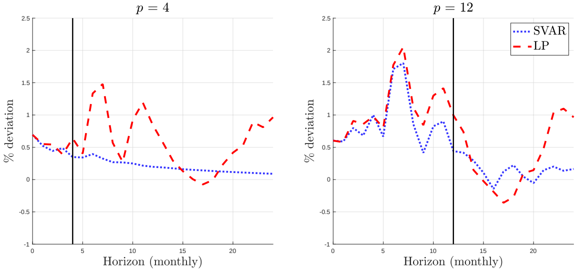
Plagborg-Møller and Wolf (2022) illustration

- They present two illustration, one where they use the Smets & Wouters (2007) DESGE model to generate the data and then an empirical example based on Gertler & Karadi (2015) stud of the effects of monetary policy on th excess bond premium.
- Replication files can be found in PMW.zip
- Replicate figures 1 and 2.

Plagborg-Møller and Wolf (2022) illustration



Plagborg-Møller and Wolf (2022) illustration



Plagborg-Møller and Wolf (2022) illustration

- Next topic: Identification by sign restrictions KL ch. 13.

Advanced Macroeconomics: Structural Vector Autoregressive Analysis

Introduction

Lecturer: U. Michael Bergman
TA: Peter Ravn
University of Copenhagen

Fall 2024

Plan for today

1. Curriculum and overview of the course
2. Properties of univariate AR models: A recap
3. What is a VAR model?
4. Why study VAR models?
5. Survey
6. Python and Matlab

Overview of the course

- Time series analysis with an emphasis on applications in macroeconomics and international finance.
- The aim of the course is to provide students with:
  - A working knowledge of structural vector autoregressive (VAR) models.
  - Emphasis on the development of programming skills either in Python or in Matlab.

Course description

TEACHING METHOD: A combination of lectures and practical sessions

HOMEWORK ASSIGNMENTS: Two compulsory assignments that must be submitted and approved before the final exam. The assignments focus on the estimation of a simple exchange rate model and must be prepared individually (but allowed to work together). Individual submission in Absalon.

Note that it is not possible to “save” the approval from an earlier semester. All students must hand in the required homework assignments.

COURSE DESCRIPTION

EXAM: 48 hours take home exam

PREREQUISITES: Basic knowledge of time series econometrics, autoregressive processes, theory for likelihood estimation and hypothesis testing and unit root testing.

It is recommended, but not a prerequisite, that students follow the Matlab basics tutorial available on the course homepage in Absalon prior to the start of the course or the Python tutorial.

REQUIREMENTS: Practical sessions will be held in a lecture room. Participants must bring a laptop in order to follow these sessions. Matlab must be installed even though you plan to use Python. We need to run Matlab code from Python.

CURRICULUM

- Kilian, L., and H. Lütkepohl (2017), Structural Vector Autoregressive Analysis, Cambridge University Press.
- Journal articles available electronically from the course homepage where you also can find lecture notes, slides, cases, due dates for cases and some other relevant information.

Other useful books
Walter Enders (2014), Applied Econometric Time Series, Wiley, Helmut Lütkephol (2005), New Introduction to Multiple Time Series Analysis, Springer, Katarina Juselius (2009), The Cointegrated VAR Model: Methodology and Applications, Oxford University Press.

Lecture plan, reading list, Matlab/Python example codes and course description

- Available in Absalon

LEARNING OBJECTIVES:

The aim of this course is to provide the students with a theoretical and practical knowledge of structural vector autoregressive (VAR) models within stationary and non-stationary frameworks as well as important econometric methods widely used in macroeconomics, financial economics and international finance.
After completion of the course, students should be able to carry out the analysis of economic data using structural VAR models, assessing the empirical results, use the approach to identify the model given the data generating process and be able to program the chosen method in either Python or Matlab.

Knowledge:

- Identify and distinguish between stationary and nonstationary VAR models.
- Estimate, interpretate and identificate the structural VAR models.
- Distinguish and assess alternative approaches to identify structural VARs.
- Inference in structural VARs.
- Evaluate and compare empirical results from other approaches (DSGE models) with structural VARs.
- Formulate economic hypotheses used as restrictions when identifying structural VARs including cointegration restrictions.

LEARNING OBJECTIVES:

Skills

- Specify and estimate structural VAR models.
- Estimate structural VAR models applying different types of identification and assess whether the model is exact-, under- or overidentified.
- Apply structural VARs to the analysis of macroeconomic
- Analyze the VAR model for variables integrated of order two and perform a nominal-to-real transformation.
- Analyze economic data using structural VAR models and assess the empirical results.
- Use Python/Matlab to analyze new data sets using pre-programmed modules and code new functions.
- Identify the model given the data generating process and to program the chosen method in Python/Matlab.

LEARNING OBJECTIVES:

Competences

- Independently formulate and analyze structural VARs for new economic problems.
- Formulate hypotheses used to identify structural VARs derived from economic theory.
- Apply economic theory to obtain an understanding of the mechanisms governing the dynamics of a certain data set.
- Use and design new programs in Python or Matlab.

Course outline

The course will be divided into four parts:

1. The first part will provide an introduction to Python/Matlab including data handling, running programs and the basics of programming.
2. The second part introduces the basic VAR model as well as the vector error correction (VEC) model. We discuss the fundamentals of VARs, including the Wold theorem, specification issues, prediction, Granger causality tests and non-stationarity.
3. In the third part we focus on structural VARs, that is the transformation of reduced form information into structural relationships. Topics include structural impulse response analysis, forecast error variance decompositions, historical decompositions, forecasts and counterfactual analysis. Four different approaches to identification will be discussed, identification using short-run restrictions, long-run restrictions, combinations of short- and long-run restrictions, the narrative approach and sign restrictions. These approaches will be illustrated with applications in macroeconomics and international finance. Inference in these models will also be discussed.
4. The fourth part focuses on the relationship between structural VARs and other macroeconomic models such as, for example, the DSGE model. We will assess structural VARs and compare to other approaches and discuss, among other things, policy evaluations using structural VARs and DSGE models and how these approaches can be combined. These issues will also be illustrated using empirical examples from the literature.

Properties of univariate AR models: A recap

- Univariate autoregressive model (AR): A recap
- From univariate AR to multivariate vector autoregressive model (VAR)

Properties of univariate AR models: A recap

Consider the AR(1), model

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t \tag{1}$$

where we assume that $|\alpha_1| < 1$.

Definition

A stochastic process y_t with finite mean and variance is covariance stationary if for all t and $k \neq 0$,

$$\begin{aligned} E[y_t] &= \mu \\ E\left[(y_t - \mu)^2\right] &= \sigma_y^2 \\ E[(y_t - \mu)(y_{t-k} - \mu)] &= \rho(k) \end{aligned}$$

where μ, σ_y^2 and $\rho(k)$ are all constants.

Add a linear trend and remove the autoregressive component

$$y_t = \beta t + \varepsilon_t.$$

where t is a linear trend.

Definition

A stochastic process y_t is trend stationary if for all t and $k \neq 0$, $\tilde{y}_t \equiv y_t - E[y_t]$ is covariance stationary.

Properties of univariate AR models: A recap

What happens if $\alpha_1 \geq 1$?

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t \tag{2}$$

Definition

If the stochastic process y is not trend stationary, but Δy_t is covariance stationary, then y_t is integrated of order 1 denoted $I(1)$.

Definition

If the stochastic process y is not trend stationary, but $\Delta^d y_t$ is covariance stationary, then y_t is integrated of order d denoted $I(d)$.

Matlab example: Example1.m Python example: Example1.py

GRAPHS FROM LECTURE NOTE ON AR PROCESSES

Figure 1: Simulated unit root and trend stationary processes.

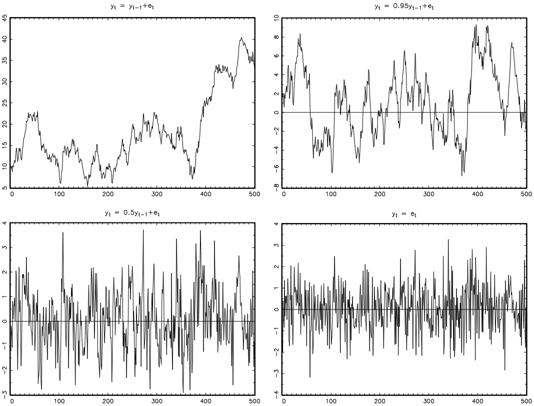
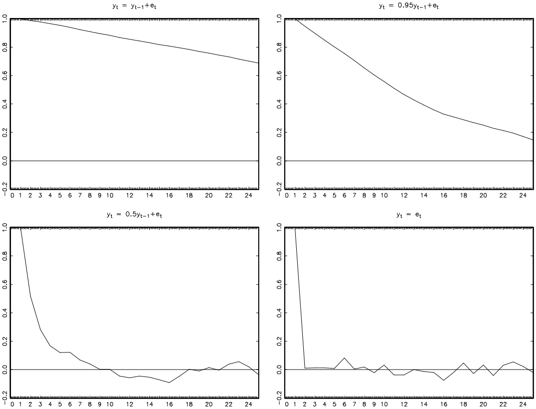
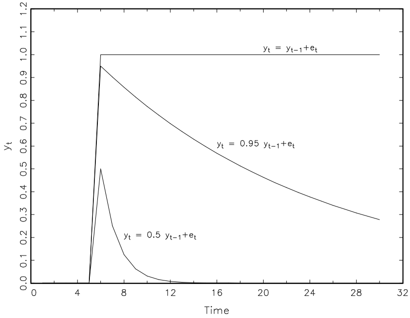


Figure 2: Autocorrelations for simulated unit root and trend stationary processes.



Graphs from lecture note on AR processes

Figure 3: Mean reversion of unit root and covariance stationary time series.



From AR to Vector Autoregressive (VAR) model

Economic applications

- Estimating AD-AS model.
- RBC model, effects of technology shocks on GDP and hours worked.
- New-Keynesian Phillips curve.
- Estimating the effects of monetary policy.
- Exchange rate determination.
- Effects of supply and demand shocks in standard AD-AS model.

What is an Impulse Response function?

- Solving AR models, from AR to moving average model (MA). See Appendix B.1 in lecture note on AR processes.
- Examples:Example1.m and Example1.py
- Interpretation of moving average representation.
- How does it work in VAR models?

Why study VAR models?

- Used extensively in macroeconomics, international finance and financial economics (also agricultural and energy economics).
- Used to evaluate theoretical models
- Allow us to conduct counter-factual and policy analysis.
- A standard tool used by central banks and governments.
- Used in other courses such as Advanced Macroeconomics: Business Cycles and Advanced Macroeconomics: Monetary Aspects. And numerous courses at other highly regarded universities.

WHY PYTHON/MATLAB?

- Flexible software used by practitioners and researchers.
- Matlab is based on vectors and matrices whereas Python is based on arrays.
- Matlab is a commercial software while Python is free.
- Many built-in functions. Most time series analysis covered by Matlab less by Python.
- Simple introductions available online and extensive help functions.
- Functions and procedures often published online by individual researchers. (But, always make sure that these functions are cross-validated!)

NEXT LECTURE/EXERCISE

- Next lecture: Stable VAR models. Curriculum: see Lecture Plan
- Next exercise: Introduction to Python/Matlab, Homework assignment 1.