Economics 6003
Quantitative Economics

1. VAR Application
2. Nonstationarity and Cointegration
3. Identification and Invertibility
4. VAR Impulse Response Alternatives

John Bluedorn

Lecturer, Economics Division

Monday/Thursday, 16/19 March 2009

CEE (1999) MP App.

- CEE use a recursive VAR to investigate the effect of monetary policy upon the US economy. The data are quarterly, covering 1965q3 to 1995q2. The reduced form VAR consists of:

$$[Y, \quad P, \quad PCOM, \quad FF, \quad NBR, \quad TR, \quad M1]$$

where Y is log real output, P is log price level, $PCOM$ is the log of a world commodity price index, FF is the federal funds rate in percentage points (short-term interest rate), NBR is log non-borrowed reserves in the banking system, TR is log total reserves (borrowed plus non-borrowed) in the banking system, and $M1$ is log money supply (M1).

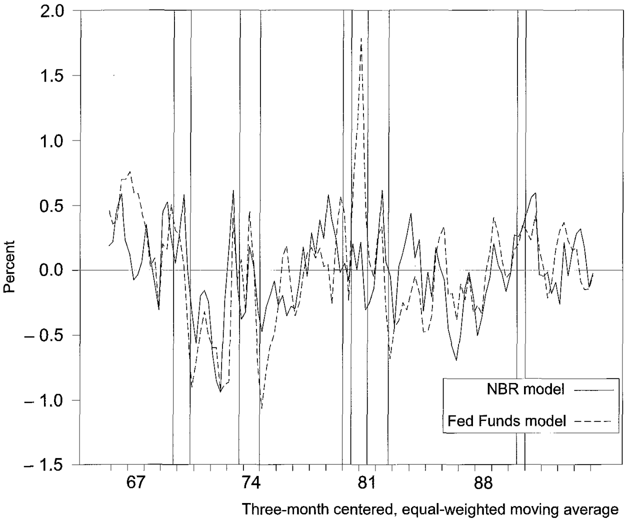
CEE (1999) MP App.

- CEE consider 3 identification schemes, each of which reflects a different set of assumptions about how monetary policy is implemented. They are represented by the recursive orderings:
  1. The short-term interest rate is the policy instrument:
  $$[Y, \quad P, \quad PCOM, \quad \underline{FF}, \quad NBR, \quad TR, \quad M1]$$
  2. Non-borrowed reserves are the policy instrument:
  $$[Y, \quad P, \quad PCOM, \quad \underline{NBR}, \quad FF, \quad TR, \quad M1]$$
  3. The ratio of non-borrowed reserves to total reserves is the policy instrument:
  $$[Y, \quad P, \quad PCOM, \quad TR, \quad \underline{NBR}, \quad FF, \quad M1]$$

CEE MP App.

- Notice that in the absence of shocks to demand in the reserves market that identification schemes 1 and 2 deliver similar results.
- How do the *ff* shock under identification 1 compare with the contemporaneous response of *ff* to contractionary *nbr* shocks under identification 2?
  - Why is this interesting? The argument would be that similar series tell us that inference should be robust to the two identification schemes.
  - There are limits to such logic. Sims (1998) in a response to a paper by Rudebusch (1998) notes that two very different instruments in 2SLS can give identical, consistent coefficient estimates. Thus, the two series need not be highly correlated for inference under *both* identification schemes to be valid.
  - Consider the 3-month moving average to reduce effect of pure noise. Vertical lines represent NBER recessions.

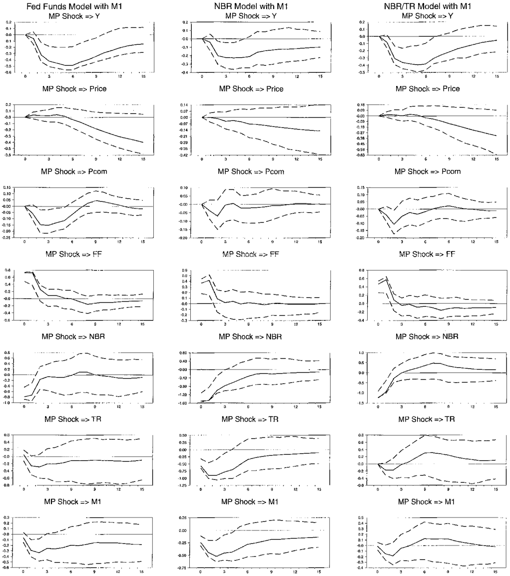
CEE (1999) MP App.



CEE (1999) MP App.

- The correlation between the two series is 0.51.
- Standard deviations of the "shock" series are 0.71 for identification 1 and 0.39 for identification 2.
- Comparing the timing of contractionary policy shocks and the timing of NBER recessions reveals:
  - Tendency for MP to be contractionary just before a recession begins.
  - Tendency for MP to be expansionary just after the end of a recession.
  - Does this indicate that MP leads to recessions and then lifts us out? Such an interpretation would accord well with a classic monetarist stance on the business cycle.
  - We should be cautious - such an interpretation requires that the MP identification is correct (with respect to both simultaneity and omitted variables' biases). The fact that the MP expansion occurs *after* the recession terminates suggests that the identifications are *not* fully valid.

CEE (1999) MP App. IRFs



CEE (1999) MP App. IRF findings

- A rise in *ff* leads to a fall in *nbr* – liquidity effect?
- *tr* is stickier than *nbr*. Do non-borrowed reserves play a buffering role for banks?
- *y* declines, but it takes 2 quarters for it to be significant. Output response is notably hump-shaped.
  - 6 quarters until a turning point in output response.
  - a 71 basis point rise in *ff* ultimately leads to only 0.5% fall in output.
  - If order *ff* above *y*, perverse effect on output - it goes up! The identification choice matters a lot.
- *p* response is flat for 6 quarters, and then falls. But, no significant decline and after 4 years, total price fall is only 0.1%.

CEE (1999) MP App. FEVDs

Table 3
Percent of *k*-period ahead forecast error variance due to policy shock: quarterly results

	FF policy shock				*NBR* policy shock				*NBR*/*TR* policy shock			
	2	4	8	12	2	4	8	12	2	4	8	12
Y	0.4 (0, 5)	21 (7, 41)	44 (18, 56)	38 (15, 50)	0 (0, 3)	7 (1, 21)	10 (1, 28)	8 (2, 12)	0 (0, 7)	17 (5, 36)	30 (9, 47)	22 (7, 42)
P	0.5 (0, 3)	0.3 (0, 6)	0.4 (0, 8)	2.5 (0, 17)	0 (0, 2)	0 (0, 5)	1 (0, 9)	1 (0, 12)	0 (0, 3)	0 (0, 6)	0 (0, 8)	1 (0, 14)
PCOM	1.0 (0, 5)	10 (1, 26)	13 (3, 26)	11 (2, 27)	0 (0, 3)	1 (0, 10)	1 (0, 11)	1 (1, 12)	1 (0, 4)	3 (0, 16)	3 (1, 16)	2 (1, 15)
FF	65 (43, 81)	35 (211, 51)	19 (12, 38)	17 (11, 36)	19 (7, 34)	10 (4, 23)	5 (3, 19)	5 (3, 19)	34 (17, 48)	18 (9, 21)	10 (5, 25)	9 (5, 25)
NBR	22 (8, 38)	9 (4, 21)	3 (2, 15)	2 (1, 17)	85 (64, 92)	46 (27, 63)	19 (10, 38)	13 (7, 33)	47 (29, 61)	19 (11, 31)	8 (5, 23)	6 (4, 21)
TR	2 (0, 11)	2 (0, 15)	1 (0, 15)	1 (0, 17)	37 (20, 52)	29 (11, 47)	12 (4, 33)	8 (2, 29)	1 (0, 5)	1 (0, 7)	1 (0, 10)	1 (0, 16)
M1	8 (2, 23)	9 (2, 27)	5 (1, 24)	4 (1, 25)	25 (10, 40)	25 (7, 44)	13 (2, 34)	9 (3, 29)	2 (0, 11)	2 (0, 12)	1 (0, 10)	1 (0, 13)
M2	36 (16, 54)	39 (17, 57)	35 (10, 56)	30 (5, 52)	34 (17, 51)	34 (14, 53)	28 (7, 50)	24 (4, 46)	35 (17, 51)	46 (22, 62)	45 (15, 60)	41 (10, 58)

CEE (1999) MP App. FEVD findings

- If *ff* is the monetary policy instrument (identification 1), then:
  - up to 44% of forecast errors in *y* can be explained by monetary policy!
  - only a *tiny* proportion of forecast errors in *p* can be explained by monetary policy.
  - a large proportion of the short-run forecast errors in *ff* are due to *ff* shocks – in other words, exogenous policy dominates endogenous policy in determining short-run variability of the policy rate. Policymakers would not be pleased by this finding.
- What do these FEVDs mean?
  - First, subject to all the issues discussed last time, so should be hesitant to conclude too much.
  - Second, the appropriate identification of monetary policy may have changed over the sample. If this matters, then we have not isolated exogenous policy changes. See Bernanke and Mihov (1998) for a hybrid strategy.

Nonstationarity and Cointegration

- Up to now, we have always implicitly been assuming that all the variables we are considering are weakly stationary or I(0) (we'll describe this terminology in a moment).
- If a variable is *non*stationary or *not* weakly stationary, then we have a problem in estimating dynamic models:
  - Test statistics will have non-standard limiting distributions (non-normal). So, we have to be careful with our statistical inference.
  - Nonstationarity implies that history can have a strong influence on the future $\Rightarrow$ an economic shock can have a *permanent* effect.
    - A random walk:

$$\begin{aligned} Y_t &= Y_{t-1} + \varepsilon_t \\ \Rightarrow Y_t &= \sum_{i=0}^{\infty} \varepsilon_{t-i} \end{aligned}$$

Nonstationarity and Cointegration

- As we can see, the problem here is that the AR representation is *non*invertible. Notice that a "solution" can still be found by recursive substitution, but there is no guarantee that it is convergent unless we put additional restrictions upon the ε sequence (e.g., initial conditions).
- Using the lag operator notation, the random walk can also be written as:

$$\begin{aligned} Y_t &= LY_t + \varepsilon_t \\ (1 - L) &= \varepsilon_t \\ \Delta Y_t &= \varepsilon_t \end{aligned}$$

Thus, the first difference in Y is weakly stationary if Y is a random walk or I(1). The I(1) means that we can difference the series once to retrieve a weakly stationary transformation (integrated of order one).

How to detect nonstationarity?

- A variety of tests have been developed to test for nonstationarity. They involve finding the nonstandard limiting distribution for some estimator (e.g., t-statistic) involving the random walk model. Thus, you use the nonstandard limiting distribution to test for a *unit root* under the null of a unit root (the alternative is stationarity).
- The most well-known tests are:
  - Dickey-Fuller $\rightarrow$ in Stata, `dfuller` - usual augmented test, `dfgls` - somewhat modified test, `madfuller` - a panel version.
  - Phillips-Perron $\rightarrow$ in Stata, `pperron`.
  - other variants.
- If fail to reject the null, then first difference (unless cointegrated - more later). This solution generalizes for I(2), etc.

Spurious Regression

- Granger and Newbold (1974) discovered that a regression of an I(1) variable on another, completely unrelated I(1) variable by OLS would lead a researcher to find a statistically significant relationship between the two variables!
- We can check this by looking at the behavior of the residuals from the regression. If they appear to show a unit root, then the relationship between the dependent and explanatory variables is spurious. There are different test statistics in this case; see Hamilton (1994) for details.
- If the two I(1) variables *are* related and the residuals are weakly stationary, then we have *cointegration*.

Cointegration

- If a linear combination of I(1) variables is weakly stationary, then we say that the set of variables is cointegrated.

$$Y_t = \beta X_t + \varepsilon_t$$

  Note that the time indices match up on Y and X. β is referred to as the *cointegrating vector* for Y and X.
- **Interpretation:** There is some long-run equilibrium linear relationship between Y and X. So, Y and X do not wander away from each other (at least not for long). Theoretical examples:
  - Consumption rates $\rightarrow \ln(C_t) - \ln(Y_t)$ is stationary (Davidson et al., 1978 prediction).
  - Real exchange rates $\rightarrow \ln(P_t^*) + \ln(s_t) - \ln(P_t)$ is stationary (purchasing power parity prediction).
  - Rational forecast of I(1) variable $\rightarrow \hat{Y}_t - Y_t$ is stationary, where hat denotes a forecast (rational expectations prediction).

How to detect cointegration?

- Undertake a levels regression for the I(1) variables (e.g., for two, regress the level of one on the other). If the residuals pass the unit root test (so reject the null), then there is cointegration. If not, the relationship is spurious and you should difference the I(1) variables.
  - If there is cointegration, then you should include an *error correction* term in the VAR model. Why? It accounts for possible convergence towards the long-run equilibrium relationship. The system is a generalization of the usual VAR to include the error correction term. It is known as a *vector error correction* model or VECM.
  - In Stata, VECM's are implemented via the `vec` routine. There are IRFs and dynamic forecasts and a subset of the usual VAR diagnostics that can then be called in Stata. See the help for more details.
  - Cointegration is also sometimes referred to as the existence of a common stochastic trend.

Why worry about cointegration?

- If there is cointegration, then excluding an error correction term from the VAR can lead to bias in the other estimators. Why? We need to account for the internal dynamics due to convergence towards the long-run equilibrium. Otherwise, we may have an omitted variable.
- In applied work, we can get an idea of how economically important the error correction mechanism is by comparing the IRFs from a VECM to the corresponding VAR. You will find that oftentimes a VAR in levels is still OK, even if there is an omitted. In a sense, things "balance-out". Toda and Yamamoto (1995) show that this can happen in a VAR, so that inference will still be OK despite the misspecification.
- We won't delve into the VECM. Things can get technical quite quickly; see Hamilton (1994) for more details.

Identification and Invertibility

- Recall that the invertible representation of an MA process Y allows us to write the process as an AR(∞) process, where it is solely a function of its past values; we solve the model backwards. As we noted, there is also a *non-invertible* representation, where we solve the model forwards.
- The two representations are statistically *indistinguishable*. From a practitioner's perspective, the invertible representation is preferred, since the usual AR method of modelling actually makes sense then.
- However, sometimes a theoretical economic model will lead to an MA representation that is *non*-invertible. Kasa (2003) makes this point with respect to Campbell-Shiller type tests of present-value models of the current account.

Identification and Invertibility

- The *identified* economic model (the structural model) is not invertible. If we use an AR representation, we will have a misspecification that can lead us towards bad inference. With respect to present-value models, Kasa (2003) describes how noninvertibility of the structural MA representation can lead you to either over-reject or over-accept tests of the intertemporal approach to the current account.
- What to do? Hansen and Sargent (1991) highlight the issue (based on earlier work). They recommend working *directly* with the *theoretical, identified* MA representation, rather than attempting to use an AR representation. This implies the use of MA estimation methods (e.g., maximum likelihood), which are more complicated. However, it will be accurate under the null that the identification is correct (unlike the AR representation).

Example non-invertible MA process

- Consider the MA(1) process:

$$
\begin{aligned}
Y_t &= \varepsilon_t + 2\varepsilon_{t-1} \\
\Rightarrow Y_t &= (1+2L)\,\varepsilon_t
\end{aligned}
$$

The backwards solution does not converge – AR representation is not well-defined with this MA coefficient. So, we have to solve for ε forwards.

$$
\begin{aligned}
Y_t &= \varepsilon_t + 2\varepsilon_{t-1} \\
\frac{0.5L^{-1}}{(1+0.5L^{-1})} Y_t &= \varepsilon_t
\end{aligned}
$$

where $(1+2L)^{-1} = \frac{0.5L^{-1}}{(1+0.5L^{-1})}$. ε is a function of the future Y!

VAR IRFs and Alternatives

- VARs are powerful tools to estimate dynamic responses. As with any tool, there are some shortcomings:
  - Asymptotic or delta-method standard errors for VAR-based impulse responses can be poorly behaved and difficult to calculate, since the VAR IRFs are highly non-linear functions of the estimated VAR coefficients.
  - VAR-based IRFs can be sensitive to the lag order chosen for the VAR. Lag order misspecification gets compounded as we consider further horizons of the IRF.
- Beyond such statistical reasons, there are also conceptual reasons to be hesitant about VAR-based IRFs.

VAR IRFs and Alternatives

- Since the IRFs are based upon powering-up an underlying dynamic model that is linear, the resulting IRFs display:
  - *symmetry*. The response to a positive shock is the mirror image of the response to a negative shock.
  - *shape invariance*. The response to a scaled-up shock is essentially the original IRF scaled-up.
  - *history independence*. The response does not depend on the local conditional history (e.g., responses might be different in recessions than in booms).
- We can accommodate some of these concerns in the VAR framework, but usually only by introducing non-linearities into the VAR model itself (e.g., Markov-switching regimes, etc.).

The Local Projections Alternative

- Jorda (2005) proposed a neat alternative based upon the literature on multi-step, direct forecasting – *local projections*.
  - There is *no* identification step, as in structural VARs. This is purely an estimation method for impulse responses. The identified impulses come from somewhere else.
  - It only uses simple linear regression and so is extremely easy to calculate (both responses and their standard error).
  - Each of limitations of VAR-based IRFs is relaxed, allowing for interesting comparisons across positive/negative, shock-scales, and local histories, in a straightforward manner.
  - He finds that the resulting IRFs are much less sensitive to lag order misspecification.

The Local Projections Alternative

- Let Y denote the response variable and let Z denote the impulse variable. Then, the impulse response at horizon h can be estimated by:

$$Y_{t+h} = B(L)\,Y_{t-1} + \delta_h Z_t + C(L)\,Z_{t-1} + \varepsilon_{t+h}$$

where $B(L)$ and $C(L)$ are lag polynomials. δ_h represents the impulse response of Z_t upon Y in h steps ahead.
- This is just a usual linear regression (in parameters), so there is no problem estimating H (max horizon for the IRF) regressions by OLS and harvesting the δs to construct an IRF.
- Standard errors should be calculated via a heteroskedasticity-autocorrelation (HAC) robust estimator (e.g., Newey-West, clustering, etc.), because the ε will follow an MA structure as a function of h (this is OK because the lagged dependent variables are $h+1$ steps away).

The Local Projections Alternative

- If we are concerned about symmetry, shape invariance, and/or history independence, we can readily introduce them into the regression through interactions with the impulse variable:
  - asymmetry – an impulse for positive changes and an impulse for negative changes
  - shape variance – a squared impulse term. The IRF is then the derivative of the regression with respect to the impulse; with a squared term, the marginal effect depends upon the scale of the impulse.
  - history independence – an interaction with the past Y. Thus, the marginal effect depends upon the past history.
- Thus, if these elements are important, then we can readily include them in our dynamic response estimates without complicated calculations. Even if they are not, the local projections alternative is a nice robustness check on VAR-based IRFs.

LPM/VAR IRF comparison

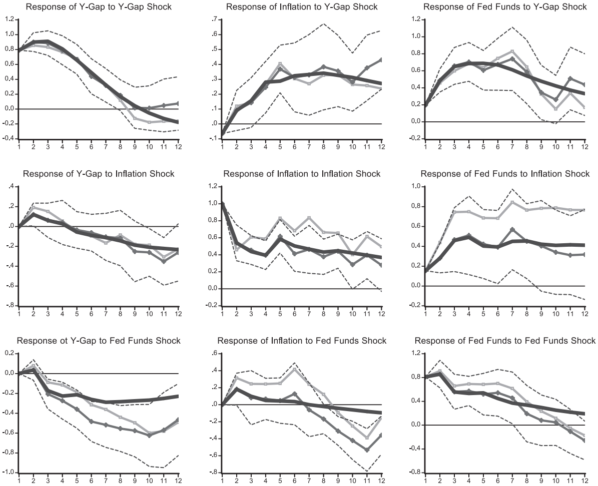


FIGURE 5. IMPULSE RESPONSES FOR THE NEW KEYNESIAN MODEL BASED ON A VAR, AND LINEAR AND CUBIC PROJECTIONS

Notes: The thick line is the response calculated from a VAR; the solid line with crosses is the response calculated by linear projection; the two dashed lines are 95-percent confidence level error bands for the individual coefficients of the linear projection response; and the solid line with circles is the response calculated by cubic projection evaluated at the sample mean. All responses calculated with four lags.

Summary

- Nonstationarity (unit roots) can lead to erroneous inference. Differencing is usually recommended, but it is often not essential to when using a VAR.
- A caveat with unit-root tests is that they are of low power. If the true eigenvalue is just under 1, then the process is weakly stationary. However, the unit-root test will most likely fail to reject the null of nonstationarity. Differencing does result in information loss, so we shouldn't go crazy with it. Of course, if a stationary process is close to a unit-root, things may be better behaved in small samples if we treat it as one.
- Error correction (EC) terms may be important if cointegration occurs. Then, a VECM model may be warranted. However, a simple VAR may still have robust IRFs, even if the EC term is omitted. It's an empirical issue.

Summary

- If we have a theoretically identified model, it need not result in an invertible MA representation. In that case, it is better to model the MA process (or ARMA, VARMA) directly rather than try to approximate it by an AR/VAR.
- Local projection methods offer a nice, simple, flexible alternative way to generate impulse responses without recourse to the usual powering-up of VAR-based IRFs. Recall however that issues of identification are treated completely separately when using LPM, unlike a VAR, where identification and estimation are linked.