607
I(0) vs. I(1) in the multivariate case

Jon Faust

http://e105.org/e607

November 30, 2015

▶ **Multivariate**

- We have discussed a bunch of result about I(0) and I(1) variables in the univariate case.
- The multivariate case adds some dimensions to this discussion.

  That's a pun.

▶ **Vector case**

- Our process is a VARMA:

$$\begin{aligned} \tilde{A}(L)y_t &= B(L)\varepsilon_t \\ A(L)x_t &= B(L)\varepsilon_t \end{aligned}$$

- This is the same notation as for the univariate case.
- The only change in notation we'll make is that $E\varepsilon\varepsilon' = \Sigma$ (full rank).
- We also assume that $A_0 = \tilde{A}_0 = I$.

  Natural generalization of $A_0 = 1$ in the univariate case. So long as Σ is full rank, this is only a normalization
- Once again, assume that $A(L)$ and $B(L)$ have no common factors and no unit roots, and that the roots of $A(L)$ are all outside the unit circle.
- Also assume that $\tilde{A}(L)$ is such that the marginal process for each ys is either I(0) or I(1).
- In the scalar case, we said $\tilde{A}(L) = (1 - L)A(L)$, but in the vector case we want to allow that some but not all elements are I(1).
- And in some sense, we'll need to know the number of unit root variables

Which will essentially come down to the rank of the I(0) stuff vs. the rank of the $I(1)$ stuff.

- To start, consider the case in which each element of y is I(1)
- In this case, do we want to difference both variables and then proceed?

The answer is maybe, maybe no.

▶ **Example:**

- The process:

$$\begin{aligned} y_{1t} &= y_{1,t-1} + \varepsilon_{1t} \\ y_{2t} &= \beta y_{1,t-1} + \varepsilon_{2t} \end{aligned}$$

with the εs mutually independent.

- Note that $y_{2,t} - \beta y_{1t}$ is I(0)

$$y_{2t} - \beta y_{1,t} = \varepsilon_{2t} - \beta \varepsilon_{1t}$$

- Actually, $y_{2,t} - \beta y_{1,s}$ is I(0) for any s: we just need to subtract out the level.
- But note, by looking at the second equation, that one observation on this sytem gives information about β.

In differencing we would lose 1 observation and in this case, that one observation would have information about β.

▶ **Aside:: Univariate case**

- In the univariate case, we had

$$(1 - L)y_t = C(L)\varepsilon_t$$

- One observation would tell us nothing about the parameters. Only the changes in y are informative.

▶ **Conclusion: In the vector case, if any (nontrivial) linear combination of the variables are I(0), differencing all variables involves a loss of information about the parameters.**

▶ **Thus,**

- We are going to need to think about not only whether each variable is I(0) or I(1), but also whether each linear combination of the variables is I(0) or I(1).

▶ **Definition: cointegration**

- If each element of a vector of variables, y_t, is I(1) and some linear combination, $\beta' y_t$ is I(0), we say the variables are cointegrated and we call β the cointegrating vector.
- Where we are headed: In a K-variables system,
  - there can be between $K-1$ and 0 cointegrating relations

    That is, between $K-1$ and 0 linearly independent combinations that are I(0)
  - And this is determined by the rank of $\tilde{A}(1)$

▶ **Proof-ish**

- We start with our system:

$$\tilde{A}(L)y_t = B(L)\varepsilon_t$$

  where y is $(K \times 1)$
- Re-write the system:

$$\begin{aligned} A(L)y_t &= B(L)\varepsilon_t \\ \delta y_t &= -A(1)y_{t-1} + G(L)\Delta y_t + B(L)\varepsilon_t \end{aligned}$$

  You did this on the problem set.
- Under our assumption that each variable is at most $I(1)$, the LHS is I(0), so the RHS must be as well.
- And the final two terms on the RHS are obviously I(0)
- Thus, $\tilde{A}(1)y_t$ must be I(0)

▶ $\tilde{A}(1)y_t$ **I(0) implies...**

- If $\tilde{A}(1)y_t$ is I(0), then every linear combination

$$\lambda' \tilde{A}(1)y_t$$

  is I(0) as well
- Let us suppose that rank$(A(1)) = J$
- In this case, we can always write

$$\tilde{A}(1) = RS'$$

  where R and S are each $K \times J$ and have full column rank, J
- Our logic says that,

$$\lambda \tilde{A}(1)y_t \text{ is } I(0) \text{ for all } \lambda$$

or

$$\lambda R z_t \text{ is } I(0) \text{ for all } \lambda$$

where

$$z_t = S' y_t$$

- Since R has rank J, we can choose λ to pick out any single element of z

  (That is, we pick λ such that

  $$\lambda R = \chi_j$$

  where χ_j is a vector of zeros except for a 1 in the j^{th} position.)

- Thus, each z must be $I(0)$.
- In other words, the columns of S are each cointegrating vectors.

▶ **A couple of cases:**

- If rank($A(1)$) = 0, then $A(1) = 0$ and there are no cointegrating vectors.

  equivalently: each element of y must be $I(1)$.

- If rank($A(1)$) = K then every element of y must be I(0)

  (we can select a λ that picks out any element of y.)

- Semantics: If all variables are I(0), we could say there is a K-dimensional space of cointegrating relations, but by convention we say there is no cointegration without at least some I(1) element.
- Otherwise, we have rank $0 < J < K$ and there are $K - J$ unit roots and J cointegrating relations.

▶ **Cointegrating rank**

- This all came from decomposing $A(1) = RS'$.
- But this decomposition is not unique
- We could use $\tilde{R}\tilde{S'} = [RQ][Q^{-1}S]$ for any full rank Q
- In short, we have identified a J-dimensional space of cointegrating vectors.

  Any nontrivial linear combination of cointegrating vectors is also cointegrating.

▶ **One more point: Error correction mechanism**

▶ **ECM**

- Our proof'ish about cointegration essentially derived what is called a the Vector Error Correction Model (VECM) or representation of the system.

- This is worth discussing.

▶ **Lets take the two-variable sense**

- If we have two I(1) variables that are cointegrated, then we can always write a univariate process for each variable of the form:

$$A(L)\Delta y_t \quad = \quad \gamma_1 z_{t-1} + B(L)\varepsilon_t$$

  where $z_t = \phi' y_t$ (scalar) is $I(0)$
- z_t is $I(0)$ and reflects the cointegrating relation.
- The z_t term must affect the changes in at least one of the two variables, and this is what keeps the levels of the ys in proper alignment.

▶ **Lessons**

- In a K variable system each element of which is at most $I(1)$ there can be between 1 and $K-1$ cointegrating relations
- If there is at least 1 cointegrating relation, then there is a VECM representation
- There must be predictable variation in at least 1 of the ys
- More precisely, lagged levels of the ys (in the form of an ECM) must predict changes in one of the ys.
- When cointegrated ys are far apart, we must be predicting that they will come back together if the error correction term, z_t is I(0).

▶ **Application**

- Campbell and Shiller provide a nice application that explores most of the issues we've just discussed

  Campbell, John Y., and Robert J. Shiller. 1987. Cointegration and tests of present value models. Journal of Political Economy 95(5): 1062-1088. go

  https://dash.harvard.edu/handle/1/3122490

- Understanding this paper deeply will give you many insights.

▶ **Present value**

- Standard theory says stock prices are a present value of expected dividends

$$P_t = \sum_{j=0}^{\infty} \frac{E_t d_j}{(1+r)^j}$$

- If dividends, d_j are I(1) then prices are as well and stock prices and dividends are cointegrated.
- This means that a model purely in differences of prices and dividends leaves out information
- Further, the error correction term between prices and dividends *must* predict either stock prices or dividends
- You should read this article to see a nice application of all of this machinery

▶ First para. of conclusion

- p.1086

  In this paper we have shown how a present value model may be tested whe the variables of the model, y_t and Y_t, follow linear stochastic processes that are stationary in first differences rather than in levels. If the present value model is true, a linear combination of the variables—which we call the spread—is stationary. Thus, y_t and Y_t are cointegrated. The model implies that the spread is linear in th optimal forecast of the one-period change in Y_t and also in the optimal forecast of teh prestent value of all future changes in y_t. We have shown how to conduct formal Wald tests of these implications.

▶ Teaser: they reject the model

- but you should have known that: Shiller is an author, not Fama.
- When is the last time an economist found a result at odds with his/her prior?

▶ Note

- This paper was written in the same era as the Campbell Mankiw consumption paper
- This is back when we thought this analysis of I(0) and I(1) systems might provide powerful insights.
- As we'll see, most folks reject this perspective at present.

▶ Hint about the problem

- Forecasting was one area where folks saw a big potential for unit roots to be helpful operating through the error correction term
- Thus, we got a lot of forecasting models in which the error correction term played a big role.

  That is, $z_t = y_t - \beta x_t$ where x and y look $I(1)$ was an important predictor
- Unfortunately, these models generally were a disaster in forecasting.
- Any misspecification or structural break in the error correction mechanism tended to put an I(1) element in the forecast error.

  and misspecification and breaks seem ubiquitous in macro.

- If you put weight on explosive variables hoping they net out somehow to be nonexplosive, you better get it right. Otherwise, you make big mistakes.
- Hendry was a major advocate of ECMs in modelling and forecasting and then led the analysis of why they ultimately failed

  The latter comes in several papers/books with Clements.

- E.g., Hendry, David F. and Michael P. Clements. 2001. Economic Forecasting: Some lessons from recent research. ECB Working paper No. 82.