A Note on Cointegration and Weak Exogeneity

Jiachuan Tian

09/21/2014

Update: 10/27/2014

Revised R.D. Weaver

11/03/2014

Update: 11/07/2014

Introduction
Econometric modeling of time series faces a variety of unique specification issues that challenge the interpretability of results and credibility of the underlying research. Among these issues a central issue is the classification of variables in terms of causal order or exogeneity (Johansen, 1992). However, within the context of time series, a related issue is the specification of the comovement or cointegration of variables.

Such specification involves structural breaks and

Linear Regression vs Cointegration
Many economic variables shows trend (or low frequencies) behavior over time (Engle and Granger, 1987; Johansen, 1992). In the most general case, such trends are stochastic trend, and in any case, such behavior implies the underlying time series are nonstationary. If a k-dimension vector time series process $\{X_t\}$ consists of I(1) processes, then there could exist a vector of real numbers v such that $v'X_t$ is stationary. Such vector v is called a cointegration vector vector for process $\{X_t\}$. Denote $Z_t = v'X_t$, then Z_t is stationary. Following Engle and Granger (1987), if we define $\{X_t\}$ as describing economic processes, we can define an economic equilibrium as occurring when $v'X_t = 0$. As they note, most often $v'X_t = Z_t$ and we interpret Z_t as equilibrium error. It is of interest whether the error exhibits systematic structure or random properties.
Thus, in general, we can consider observations of Z_t as including deviations from such an equilibrium. Further, the cointegration vector v can be interpreted as characterizing the long-run (low frequency) relation between variables in the system. The relationships are distinguished as long-run given we implicitly assume the process $\{X_t\}$ takes time to adjust to equilibrium (Engle and Granger, 1987; Johansen, 1992).
<elaborate….intuition referring to notation….>

To see the difference between traditional linear regression and cointegration, without loss of generality, consider a bivariate system $\{X_t, Y_t\}$ as an example. If $\{X_t, Y_t\}$ has no time properties, then we can simply write it as $\{X,Y\}$. Thus the linear regression problem can be written as $Y=b'X+e$. Note by tradition asymptotic results, we have:

$$\sqrt{N}(\hat{b}-b) \xrightarrow{d} N(0, E(XX')^{-1}E(e^2XX')\,E(XX')^{-1})$$

If we have homoskedasticity, then we have the standard OLS result of consistency and asymptotic efficiency. Importantly, this result requires the expectation of XX' exists as a characterizing statistic of an underlying distribution $f(XX')$ that is time invariant and generates each of the observations XX'.

In constrast, if $\{X_t, Y_t\}$ is a time variant I (1) system and if X_t and Y_t are cointegrated, then we have $Y_t=b'X_t+e_t$ with e_t stationary. Following Johansen (1990), we can then write (X_t, Y_t) representation:

and it follows:

$$N(\hat{b}-b) \xrightarrow{d} (I-bc')(\gamma'\gamma)^{-1}\gamma(\int G(u)G(u)'du)^{-1} \times \int G(dV)'(c'b)^{-1}$$

where c is the normalized vector ….., τ is the coefficient of t in the (X_t, Y_t) representation and γ is orthogonal to (b,τ) so that $(b,\gamma,\tau) spans\ \mathbb{R}^n\ and\ \gamma,\tau\ are\ orthogonal.$ Note in this case, $\{\Delta X_t, \Delta Y_t\}$ is by definition stationary, but it is not meaningful to construct a VAR on $\{\Delta X_t, \Delta Y_t\}$.

While such a VAR would describe relationships within the differenced system, it will in general fail to characterize the long-run relationship between $\{X_t, Y_t\}$ (Engle and Granger, 1987).

While the above argument motivates the need to focus on what can be viewed as a "full system" representation of the cointegrating relationship when $\{X_t, Y_t\}$ are nonstationary, it is most often the case that the dimensions and content of the full system is uncertain for the econometrician. In that case, the question arises of whether it is useful and how estimates can be interpreted to estimate the cointegrating relationships of a partial representation of the full system. Weak exogeneity resolves this issue by clarifying that only if the partial system is in fact a weakly exogenous subsystem can we we interpret estimated parameters of such a partial system as related to those in the full system.

Weak Exogeneity: Definition and General Notation

Economic systems often include substantial numbers of variables and motivate interest in specification and estimation of parts of such systems. Indeed, where specification of the overall system is uncertain, or where interest focuses only on a subsystem such partial systems are both of interest and may allow explicit consideration of the scope of an uncertain system. This opens the question of under what conditions estimation of a subsystem provides useful information (consistency and efficiency) concerning parts of a larger system. Johansen (1992) introduced the concept of weak exogeneity to respond to this question. Consider a full linear system in the form of a vector error correction model (VECM):

$$\Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \alpha\beta' X_{t-k} + \phi D_t + \mu + \varepsilon_t, t = 1,..,T$$

where $X_{-k+1},..,X_0$ are fixed and $\varepsilon_1, \ldots, \varepsilon_T$ are independent p-dimensional Gaussian variable with mean zero and variance matrix Λ. The vector D_t denotes seasonal dummies centered at zero. The parameters in the model are short-run effects $\Gamma_1, \ldots, \Gamma_{k-1}$, the seasonal coefficients ϕ, the constant term μ, the covariance matrix Λ, the $p \times r$ matrices α (the adjustment coefficients) and β (the cointegrating relations).

It is often the case that the econometrician is unable to specify or measure all elements of the full system represented by the VECM. Such model uncertainty leads to estimation of what we could call a "partial system" that implicitly defines a "residual system" that is left out of consideration. Granger's definition of weak exogeneity allows us to consider implications of estimation of the partial system. Granger's approach asks under what conditions will the expectation of the partial system conditioned on the residual system be

Let *a* be a $p \times m$ matrix of *m* the selection matrix we use to define the partial system. Let $b = a_\perp$ be a $p \times (p-m)$ full rank matrix of vectors orthogonal to *a* that we use to define the residual system.

To investigate the weak exogeneity of the residual system $b'\Delta X_t$, relative to the partial system, we need to consider the conditional expectation of the complement $a'\Delta X_t$ given the residual system $b'\Delta X_t$. To derive this, by pre-multiplication of the full VECM, we have $a'\Delta X_t$
<show derivation of

$$E(a'\Delta X_t | b'\Delta X_t, \Delta X_{t-1}, \dots, \Delta X_{t-k+1}, \Delta X_{t-k})$$
$$= (a - b\Lambda_{bb}^{-1}\Lambda_{ba})'(\sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \alpha'\beta X_{t-k} + \phi D_t + \mu) + \Lambda_{ab}\,\Lambda_{bb}^{-1} b'\Delta X_t$$

and

$$var(a'\Delta X_t | b'\Delta X_t, \Delta X_{t-1}, \dots, \Delta X_{t-k+1}, \Delta X_{t-k}) = \Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba} = \Lambda_{aa.b}$$

where $\Lambda_{aa} = a'\Lambda a, \Lambda_{ab} = a'\Lambda b, etc.$

Then we have

$$a'\Delta X_t = (a - b\Lambda_{bb}^{-1}\Lambda_{ba})'(\sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \alpha'\beta X_{t-k} + \phi D_t + \mu) + \Lambda_{ab}\,\Lambda_{bb}^{-1} b'\Delta X_t + u_t, t = 1, \dots T$$

where $u_t = (a - b\Lambda_{bb}^{-1}\Lambda_{ba})'\varepsilon_t$ are independent Gaussian variables with zero mean and variance matrix $\Lambda_{aa.b}$. Such model is called a partial VAR model or a conditional model for $a'\Delta X_t$ given $b'\Delta X_t$ and past information. Using this notation, Johansen defines weak exogeneity of $b'\Delta X_t$ for α and β as the parameters of the distribution of $a'\Delta X_t$ given $b'\Delta X_t$ and the past information $\Delta X_{t-1}, \dots, \Delta X_{t-k+1}, \Delta X_{t-k}$ are variation independent of parameters of the distribution of $\Delta X_{t-1}, \dots, \Delta X_{t-k+1}, \Delta X_{t-k}$(Johansen, 1992). In particular, note we can write $b'\Delta X_t$ as

$$b'\Delta X_t = \sum_{i=1}^{k-1} b'\Gamma_i \Delta X_{t-i} + b'\alpha\beta' X_{t-k} + b'\phi D_t + b'\mu + b'\varepsilon_t, t = 1, .., T$$

If $b'\alpha = 0$, then $b\,'\Delta X_t = \sum_{i=1}^{k-1} b'\Gamma_i \Delta X_{t-i} + b'\phi D_t + b'\mu + b'\varepsilon_t, t = 1, .., T$. This means $b'\Delta X_t$ does not contain information on the cointegration relations (vector β).
Following the definition by Engle et.al (1983) and Johansen (1992), this indicates $b'\Delta X_t$ is weakly exogenous for the parameter (α, β).

Hence the maximum likelihood estimator for (α, β) in the full model is the same as that in partial model. Equivalently, Johansen (1992) shows that the hypothesis of weak exogeneity of $Z_t = b'X_t$ for α and β can be formlateed as

$$H: \alpha_z = 0$$

From above, it is clear that Z may be of smaller dimension than X. We define α_z as composed of the rows of α corresponding to Z_t.Then under the null hypothesis H the maximum likelihood estimation of the parameters could be performed by reduced rank regression and that the rest of H in H_r: ($\Pi = \alpha'\beta$ where α and β are $p \times r$ matrices) consists in comparing the eigenvalues $\widehat{\lambda_i}$ and $\widetilde{\lambda_i}$, where $\widehat{\lambda_i}$ is eigenvalue without the restriction and $\widetilde{\lambda_i}$ is eigenvalue with the restriction respectively. The test statistic is

$$T \sum_{i=1}^{r} \ln \left\{ \frac{(1 - \widetilde{\lambda_i})}{(1 - \widehat{\lambda_i})} \right\} \xrightarrow{d} \chi^2(rp_z)$$

where p_z is the dimension of Z_t.

Weak Exogeneity: Software Implementation

This paper employs the R-package 'urca' by Pfaff to test cointegration and weak exogeneity. To test exogeneity using such package, we need to first identify the restriction matrix $\boldsymbol{A}$ such that the corresponding rows of $\boldsymbol{A}$ to Z_t have zero entries, where other rows are unit vectors. Suppose we have a vector X_t of dimension 5, and we want to test weak exogeneity of its second component X_{2t}, then the restriction matrix is defined as

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

Then we can apply function 'alrtest' to test weak exogeneity of X_{2t}, with argument 'A' (restriction matrix) equaling to $\boldsymbol{A}$ as defined above and the argument 'r' equaling to the cointegration rank of the full system. The value of the test statistic is stored in the slot **teststat** and its marginal level of significance, the p-value, in the slot **pval**. (Pfaff, 2008).

Weak Exogeneity: Application and Results

We suspect macroeconomic factors may play a role in each bivariate VECM. To motivate the choice of the macroeconomic factors, we restrict our interest on those macroeconomic factors that can be directly adjusted by policies (e.g. interest rate, money supply, etc.) rather than 'natural' factors (GDP growth, etc.). We choose money supply, exchange rate and interest rate to be our macroeconomic factors to consider. Such choice made by such criteria also agree with past literature. (********can insert previous note on the lit. review of macro-factors as well as a better motivation*******)

TABLE-1 Bivariate Pair and Macro Factors (M2, 10 year maturity rate and US/EU exchange rate) Cointegration Weak Exogeneity Test Table

As table 1 shows, money supply (M2) is not exogenous in every pair of commodity and energy pair. In addition, the US/Euro exchange rate is showed to be not exogenous in most pairs. The results imply that the money supply and exchange rate are endogenous within the system defined as commodity and energy prices and hence cannot be omitted models considering the cointegration of energy and commodity prices. This means that the presence of structural breaks cannot be considered within the context of bivariate pairs of commodity and energy prices. The significant role of M2 plays in the price systems is consistent with monetary theories (Bordo 1980, Chambers and Just 1982, Gilbert 2008). And the result of exchange rate being endogenous is also consistent with previous literature (Baek and Koo 2009, Gilbert 2008, Harri et al. 2009, Nazioglu and Soytas 2011).

Given the results in table 1, we want to take a closer look at the role of exchange rate and interest rate play in the price transmission. In addition, we conclude that long run interest rate (10 year maturity rate) could be exogenous to the system. We suspect short term interest rate may play a role in the price transmission. This motivates the refinement of new macro-factors chosen (3 month maturity rate and exchange rate).

TABLE-2 Cointegration Rank Table each bivariate pair with Macro-factors

TABLE-3 Cointegration Rank Table of each bivariate pair without Macro-factors

Table 3 reports cointegration ranks of bivariate VECM consisting of two price variables, while table 2 reports cointegration ranks of multivariate VECM consisting of the exact two price variables together with two Macro factor (3 month maturity rate and exchange rate). Comparing cointegration ranks reported in two tables, we find the cointegration rank is not increasing as we add in macro-factors. This impies that the maro-factors are not significant in the bivariate VECM system and hence we may exclude them to estimate a bivariate VECM consisting price variables only. To make sure of this, we conduct weak exogeneity test in the VECM.

TABLE-4 Bivariate Pair and Macro Factors (3-month Maturity Rate and US/EU exchange rate) Weak Exogeneity Test

For those VECMs with strict positive cointegration rank, we test weak exogeneity of each variable in the four-variable VECMs. The results are reported in table 4. As shown in table 4, short run interest rate is not (weakly) exogenous in four pairs, while interest rate is (weakly) exogenous in every pair. This result basically agrees with the previous conclusion we draw from the comparison of table 2 and table 3, which means we can exclude macroeconomic factors in bivariate VECM for further analysis.

Appendix A. Table-1 Bivariate Pair and Macro Factors (M2, 10-year Maturity Rate and US/EU exchange rate) Weak Exogeneity Test

	Stat	P		Stat	P		Stat	P		Stat	P		Stat	P		Stat	P
It_corn	15.1	0.001	It_corn	0.6	0.731	It_corn	0.9	0.340	It_corn	17.1	0.000	It_corn	0.9	0.628	It_corn	0.4	0.541
It_wheat	6.5	0.039	Itsyb	32.0	0.000	US_corn	0.0	0.992	USwheat	8.8	0.012	US_syb	36.8	0.000	Brent	1.2	0.267
M2	86.7	0.000	M2	92.6	0.000	M2	90.6	0.000	M2	91.1	0.000	M2	80.6	0.000	M2	74.2	0.000
X10.year	8.9	0.012	X10.year	14.4	0.001	X10.year	2.1	0.145	X10.year	4.4	0.108	X10.year	17.9	0.000	X10.year	0.2	0.621
US.Euro	5.2	0.076	US.Euro	5.2	0.075	US.Euro	5.8	0.016	US.Euro	8.6	0.013	US.Euro	5.0	0.081	US.Euro	5.0	0.025
It_wheat			It_wheat	2.5	0.280	It_wheat	3.5	0.062	It_wheat	18.2	0.000	It_wheat	3.4	0.184	It_wheat	1.8	0.175
It_wheat			It_syb	25.7	0.000	US_corn	0.2	0.698	USwheat	8.5	0.014	US_syb	31.2	0.000	Brent	0.8	0.382
M2			M2	94.0	0.000	M2	97.1	0.000	M2	92.7	0.000	M2	86.4	0.000	M2	83.7	0.000
X10.year			X10.year	7.3	0.026	X10.year	2.0	0.154	X10.year	0.6	0.727	X10.year	7.9	0.019	X10.year	0.0	0.884
US.Euro			US.Euro	5.3	0.071	US.Euro	5.7	0.017	US.Euro	8.5	0.014	US.Euro	5.2	0.074	US.Euro	5.1	0.024
It_syb			It_syb			It_syb	0.0	0.830	It_syb	12.0	0.002	It_syb	14.0	0.001	It_syb	1.1	0.292
It_wheat			It_syb			US_corn	0.0	0.947	USwheat	1.3	0.510	US_syb	5.1	0.077	Brent	0.9	0.333
M2			M2			M2	105	0.000	M2	99.4	0.000	M2	98.6	0.000	M2	93.0	0.000
X10.year			X10.year			X10.year	3.1	0.079	X10.year	3.2	0.198	X10.year	2.0	0.370	X10.year	1.3	0.248
US.Euro			US.Euro			US.Euro	5.2	0.023	US.Euro	4.3	0.117	US.Euro	4.5	0.106	US.Euro	4.2	0.040
US_corn			US_corn			US_corn			US_corn	0.2	0.625	US_corn	0.0	0.862	US_corn	0.1	0.716
It_wheat			It_syb			US_corn			USwheat	0.7	0.387	US_syb	0.1	0.729	Brent	0.4	0.519
M2			M2			M2			M2	94.6	0.000	M2	94.6	0.000	M2	91.9	0.000
X10.year			X10.year			X10.year			X10.year	1.7	0.195	X10.year	2.4	0.124	X10.year	1.6	0.206
US.Euro			US.Euro			US.Euro			US.Euro	4.2	0.041	US.Euro	5.7	0.017	US.Euro	6.0	0.014
USwheat			USwheat			USwheat			USwheat			USwheat	1.2	0.553	USwheat	0.2	0.651
It_wheat			It_syb			US_corn			USwheat			US_syb	7.8	0.020	Brent	0.9	0.350
M2			M2			M2			M2			M2	89.7	0.000	M2	85.7	0.000
X10.year			X10.year			X10.year			X10.year			X10.year	4.0	0.133	X10.year	0.0	0.914
US.Euro			US.Euro			US.Euro			US.Euro			US.Euro	3.0	0.221	US.Euro	2.4	0.124
US_syb			US_syb			US_syb			US_syb			US_syb			US_syb	1.0	0.318
It_wheat			It_syb			US_corn			USwheat			US_syb			Brent	1.0	0.307
M2			M2			M2			M2			M2			M2	75.7	0.000
X10.year			X10.year			X10.year			X10.year			X10.year			X10.year	0.9	0.342
US.Euro			US.Euro			US.Euro			US.Euro			US.Euro			US.Euro	4.0	0.045

Note: The weak exogeneity test is conducted in R using function 'alrtest'. The 5% significance test statistic is marked by red color, which means the variable is not (weakly) exogenous in the five variable (VECM) system.

Appendix B. Table-2 Cointegration Rank of each bivariate pair with Macro-factors (3 month Maturity Rate and Exchange Rate)

	EU_corn	EU_wheat	EU_soybean	US_corn	US_wheat	US_soybean	Brent Blend
EU_corn		1	1	0	0	1	0
EU_wheat	1		1	0	0	1	0
EU_soybean	1	1		0	0	1	0
US_corn	0	0	0		0	0	0
US_wheat	0	0	0	0		0	0
US_soybean	1	1	1	0	0		0
Brent Blend	0	0	0	0	0	0	

Note: The cointegration rank test is conducted in R using function 'ca.jo'. The test results is based on 10% significance level. The number in ij-th slot (i-th row and j-th column) reports the cointegration rank of the VECM system consisting of i-th row variable, j-th column variable, 3 month maturity rate and exchange rate.

Table-3 Cointegration Rank of each bivariate pair without Macro-factors

	EU_corn	EU_wheat	EU_soybean	US_corn	US_wheat	US_soybean	Brent Blend
EU_corn		1	1	0	1	1	0
EU_wheat	1		1	0	1	0	0
EU_soybean	1	1		0	1	1	0
US_corn	0	0	0		0	0	0
US_wheat	1	1	1	0		0	0
US_soybean	1	0	1	0	0		0
Brent Blend	0	0	0	0	0	0	

Note: The cointegration rank test is conducted in R using function 'ca.jo'. The test results is based on 10% significance level. The number in ij-th slot (i-th row and j-th column) reports the cointegration rank of the VECM system consisting of i-th row variable, j-th column variable (without Macro-factors).

Appendix C. Table-4 Bivariate Pair and Macro Factors (3-month Maturity Rate and US/EU exchange rate) Weak Exogeneity Test

	Stat	P		Stat	P		Stat	P		Stat	P		Stat	P		Stat	P
It_corn	6.938	0.008	It_corn	0.062	0.803	It_corn			It_corn			It_corn	0.191	0.662	It_corn		
It_wheat	9.062	0.003	Itsyb	29.616	0.000	US_corn			USwheat			US_syb	38.034	0.000	Brent		
3 month	4.066	0.044	3 month	8.469	0.004	3 month			3 month			3 month	9.191	0.002	3 month		
US.Euro	1.568	0.211	US.Euro	0.644	0.422	US.Euro			US.Euro			US.Euro	2.422	0.120	US.Euro		
It_wheat			It_wheat	0.139	0.709	It_wheat			It_wheat			It_wheat	1.924	0.165	It_wheat		
It_wheat			It_syb	26.797	0.000	US_corn			USwheat			US_syb	31.184	0.000	Brent		
3 month			3 month	5.522	0.019	3 month			3 month			3 month	3.777	0.052	3 month		
US.Euro			US.Euro	0.134	0.714	US.Euro			US.Euro			US.Euro	0.907	0.341	US.Euro		
It_syb			It_syb			It_syb			It_syb			It_syb	15.907	0.000	It_syb		
It_wheat			It_syb			US_corn			USwheat			US_syb	4.734	0.030	Brent		
3 month			3 month			3 month			3 month			3 month	0.019	0.890	3 month		
US.Euro			US.Euro			US.Euro			US.Euro			US.Euro	0.574	0.449	US.Euro		
US_corn			US_corn			US_corn			US_corn			US_corn			US_corn		
It_wheat			It_syb			US_corn			USwheat			US_syb			Brent		
3 month			3 month			3 month			3 month			3 month			3 month		
US.Euro			US.Euro			US.Euro			US.Euro			US.Euro			US.Euro		
USwheat			USwheat			USwheat			USwheat			USwheat			USwheat		
It_wheat			It_syb			US_corn			USwheat			US_syb			Brent		
3 month			3 month			3 month			3 month			3 month			3 month		
US.Euro			US.Euro			US.Euro			US.Euro			US.Euro			US.Euro		
US_syb			US_syb			US_syb			US_syb			US_syb			US_syb		
It_wheat			It_syb			US_corn			USwheat			US_syb			Brent		
3 month			3 month			3 month			3 month			3 month			3 month		
US.Euro			US.Euro			US.Euro			US.Euro			US.Euro			US.Euro		

Note: The weak exogeneity test is conducted in R using function 'alrtest'. The 5% significance test statistic is marked by red color, which means the variable is not (weakly) exogenous in the four variable (VECM) system.

Reference

Engle, R. F., Hendry, D. F., & Richard, J. F. (1983). Exogeneity. *Econometrica: Journal of the Econometric Society*, 277-304.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251-276.

Johansen, S., & Juselius, K. (1992). Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK. *Journal of econometrics*, *53*(1), 211-244.

Johansen, S. (1992). Testing weak exogeneity and the order of cointegration in UK money demand data. *Journal of Policy Modeling*, *14*(3), 313-334.

Johansen, S. (1992). Cointegration in partial systems and the efficiency of single-equation analysis. *Journal of Econometrics*, *52*(3), 389-402.

Pfaff, B. (2008). *Analysis of integrated and cointegrated time series with R*. Springer.