How to Deal with Structural Breaks in Practical Cointegration Analysis

Roselyne Joyeux*

School of Economic and Financial Studies

Macquarie University

June 2002

ABSTRACT

Johansen, Mosconi and Nielsen (2000) generalize the likelihood-based cointegration analysis developed by Johansen (1988, 1996) to the case where structural breaks exist at known points in time. In this paper we provide a simple explanation of the specification of intervention dummies in VAR models used to test for cointegration, which is not present in the later paper. We also give practical guidelines for the inclusion and the specification of intervention dummies in VAR models and present simulation results, which illustrate the importance of specifying the dummy variables correctly.

JEL Classification Code: C32, C52, E43.

Keywords: structural break, dummy variable, cointegration,VAR models.

* Correspondence should be sent to: Dr Roselyne Joyeux
School of Economic and Financial Studies
Macquarie University
Sydney 2109
Australia
e-mail: rjoyeux@efs.mq.edu.au

1. Introduction

The empirical literature making use of unit root and cointegration tests has been growing over the last two decades. The application of those tests is challenging for many reasons including the treatment of deterministic terms (constant and trend) and structural breaks. Franses (2001) addresses the problem of how to deal with intercept and trend in practical cointegration analysis. In this note we use Franses (2001) approach to consider the treatment of structural breaks in VAR models used to tests for unit roots and cointegration. In what follows we assume that structural breaks occur at known break points.

There is a vast literature on structural breaks and unit root tests. If a series is stationary around a deterministic trend with a structural break we are likely to accept the null of a unit root even if we include a trend in the ADF regression. There is a similar loss of power in the unit root tests if the series present a shift in intercept. If the breaks are known the ADF test can be adjusted by including dummy variables in the ADF regression (Perron (1989, 1990), Zivot and Andrews (1992) among others).

In this note we show how intervention dummies should be specified and included in VAR models to test for unit roots and cointegration. Note that there is nothing new in this note, the material is basically covered in the Johansen, Mosconi and Nielsen (2000) paper (JMN (2000) thereafter). This note, however, provides a simple explanation of the specification of intervention dummies which is not present in the later paper. A survey of the applied literature using Johansen's test for cointegration in a VAR setting would reveal that intervention dummies are usually inappropriately specified. Indeed in empirical work it is often the case that structural breaks have to be accounted for. The inclusion of intervention dummies should improve the normality properties of the estimated residuals. This is, however, often not the case. The reason for this is that the dummy variables are incorrectly specified. It is the aim of this note to show how to specify and include intervention dummies and to make accessible to applied economists the latest development in the use of intervention dummies when testing for cointegration.

In Section 2 we present the results in the univariate case and in Section 3 we generalize to the multivariate case. Simulation results, which illustrate the importance of specifying the dummy variables correctly, are included in Section 4. Section 5 concludes.

2. Univariate Case

In this section we look at processes which can be modelled as autoregressive processes with possibly a trend or an intercept shift at some point in time.

Shift in Intercept Model

In this section we consider a univariate time series y_t, $t = 1, 2, \ldots, T$ which has a shift in mean at time T_1, $1 < T_1 < T$, and can be described by:

$y_t - \mu_1 = \phi_1(y_{t-1} - \mu_1) + \varepsilon_t$ when $t \leq T_1$

and

$y_t - (\mu_1 + \mu_2) = \phi_1(y_{t-1} - (\mu_1 + \mu_2)) + \varepsilon_t$ when $t > T_1$

where ε_t is a white noise process. The parameter ϕ_1 is assumed to be the same in all sub-samples. The model above is formulated conditionally on the first observations of each sub-sample: y_1 and y_{T_1+1}.

When $|\phi_1| < 1$, one can say that y_t is attracted by μ_1 for $t \leq T_1$ and by $(\mu_1 + \mu_2)$ for $t > T_1$.

This model can be rewritten as:

$$y_t - (\mu_1 + \mu_2 D_t) = \phi_1(y_{t-1} - (\mu_1 + \mu_2 D_{t-1})) + \varepsilon_t \qquad (1)$$

where

$D_t = 0$ if $t \leq T_1$

and

$D_t = 1$ if $t > T_1$.

If we let $\phi_1 = 1$ in equation (1) we get

$$y_t = y_{t-1} + \mu_2(D_t - D_{t-1}) + \varepsilon_t \qquad (2)$$

μ_1 is not identified when $\phi_1 = 1$ but the shift in mean μ_2 is.

We can rewrite (1) as:

$$\Delta y_t = (\phi_1 - 1)y_{t-1} + (1 - \phi_1)\mu_1 + \mu_2(D_t - \phi_1 D_{t-1}) + \varepsilon_t \qquad (5)$$

or

$$\Delta y_t = (\phi_1 - 1)y_{t-1} + (1 - \phi_1)(\mu_1 + \mu_2 D_{t-1}) + \mu_2 \Delta D_t + \varepsilon_t \qquad (6)$$

If we let $\rho_1 = \phi_1 - 1$, (6) can be rewritten as:

$$\Delta y_t = \rho_1 y_{t-1} - \rho_1(\mu_1 + \mu_2 D_{t-1}) + \mu_2 \Delta D_t + \varepsilon_t \qquad (7)$$

Since $\Delta D_t = 0$ if $t \leq T_1$ or if $t > T_1+1$, and $\Delta D_t = 1$ if $t = T_1+1$, the effect of ΔD_t corresponding to the observation y_{T_1+1} is to render the associated residual zero given the initial value in the second sub-sample. The inclusion of ΔD_t does not affect the asymptotic distribution of the t statistic of the estimated coefficient of y_{t-1}, $\hat{\rho}_1$, under the null of a unit root.

This representation also illustrates that when testing for a unit root the test regression should include both the lagged intervention dummy and the first difference of the intervention dummy, even though under the null of $\rho_1 = 0$ the lagged dummy disappears. Perron (1990) and Perron and Vogelsang (1992) tabulate the asymptotic distribution of the t statistic of the estimated coefficient of y_{t-1}, $\hat{\rho}_1$, under the null of a unit root. A better test would be to test for the joint significance of the coefficient of y_{t-1}, the intercept and the lagged intervention dummy in (7). In the multivariate case the test considered in this note is indeed a joint test of the above hypotheses.

Shift in Trend Model

In this section we consider a univariate time series y_t, $t = 1, 2, \ldots, T$ which has a shift in mean and a shift in trend at time T_1, $1 < T_1 < T$, and can be described by:

$$y_t - \mu_1 - \delta_1 t = \phi_1(y_{t-1} - \mu_1 - \delta_1 (t-1)) + \varepsilon_t \qquad \text{when } t \leq T_1$$

and

$$y_t - (\mu_1 + \mu_2) - (\delta_1 + \delta_2)t = \phi_1(y_{t-1} - (\mu_1 + \mu_2) - (\delta_1 + \delta_2)(t-1)) + \varepsilon_t \qquad \text{when } t > T_1$$

where ε_t is a white noise process. As before the model above is formulated conditionally on the first observations of each sub-sample: y_1 and y_{T_1+1}.

This model can be rewritten as:

$$y_t - (\mu_1 + \mu_2 D_t) - (\delta_1 + \delta_2 D_t)\, t = \phi_1 [y_{t-1} - (\mu_1 + \mu_2 D_{t-1}) - (\delta_1 + \delta_2 D_{t-1})(t-1)] + \varepsilon_t \qquad (8)$$

Alternatively (8) can be written as:

$$
\begin{aligned}
\boldsymbol{D}y_t = \boldsymbol{r}_1 y_{t-1} &+ \left[-\,\boldsymbol{r}_1(\,\boldsymbol{m}_1 + \boldsymbol{m}_2 D_{t-1}\,) + \boldsymbol{m}_2 \boldsymbol{D}D_t + \boldsymbol{f}_1(\,\boldsymbol{d}_1 + \boldsymbol{d}_2 D_{t-1}\,)\right] \\
&+ \left[\boldsymbol{d}_2 \boldsymbol{D}D_t - \boldsymbol{r}_1(\,\boldsymbol{d}_1 + \boldsymbol{d}_2 D_{t-1}\,)\right] t + \boldsymbol{e}_t
\end{aligned}
\tag{9}
$$

The effect of $\boldsymbol{D}D_t$ and $t\boldsymbol{D}D_t$, corresponding to the observation y_{T_1+1}, is to render the associated residual zero given the initial value in the second sub-sample. There is, however, no point in including both $\boldsymbol{D}D_t$ and $t\boldsymbol{D}D_t$ in (9) since $\boldsymbol{m}_2 \boldsymbol{D}D_t = \boldsymbol{m}_2$ if $t = T_1+1$ and 0 otherwise, and $\boldsymbol{d}_2\, t\boldsymbol{D}D_t = \boldsymbol{d}_2(T_1+1)$ for $t = T_1+1$ and 0 otherwise. We can thus rewrite (9) as:

$$
\boldsymbol{D}y_t = \boldsymbol{r}_1 y_{t-1} - \boldsymbol{r}_1\boldsymbol{d}_1 t - \boldsymbol{r}_1\boldsymbol{d}_2 D_{t-1} t + \boldsymbol{h}_1 + \boldsymbol{h}_2 D_{t-1} + \boldsymbol{k}_0 \boldsymbol{D}D_t + \boldsymbol{e}_t \tag{10}
$$

where $\boldsymbol{h}_1 = -\boldsymbol{r}_1\boldsymbol{m}_1 + \boldsymbol{f}_1\boldsymbol{d}_1$, $\boldsymbol{h}_2 = -\boldsymbol{r}_1\boldsymbol{m}_2 + \boldsymbol{f}_1\boldsymbol{d}_2$ and $\boldsymbol{k}_0 = \boldsymbol{m}_2 + \boldsymbol{d}_2(T_1+1)$.

As for the shift in intercept only case this representation shows that the test regression should include both the lagged intervention dummy and the first difference of the intervention dummy. It also shows that the lagged intervention dummy should be included both in the intercept and the deterministic trend variable, even though under the null of $\boldsymbol{r}_1 = 0$ the lagged dummy disappears in the trend component (but not in the intercept part). So the practical rule would be to include in the test regression the intercept, a lagged dummy intercept, a first difference dummy intercept, the trend, and the lagged dummy times the trend. Perron (1989) tabulates the asymptotic distribution of the t statistic of the estimated coefficient of y_{t-1}, $\hat{\boldsymbol{r}}_1$, under the null of a unit root. Note that the trend and the lagged dummy times the trend disappear under the null. A better test of the null of a unit root test is a joint test of the joint significance of the coefficients of y_{t-1}, the trend and the lagged dummy times the trend in (10).

Generalization to an AR(k) process

In the case where the process follows an AR(k) model with AR coefficients $\boldsymbol{f}_1,\ldots,\boldsymbol{f}_k$ equation (10) becomes:

$$
\boldsymbol{D}y_t = \boldsymbol{r}_k y_{t-1} - \boldsymbol{r}_k\boldsymbol{d}_1 t - \boldsymbol{r}_k\boldsymbol{d}_2 D_{t-k} t + \boldsymbol{h}_1 + \boldsymbol{h}_2 D_{t-k} + \sum_{i=1}^{k-1} \boldsymbol{G}_i \boldsymbol{D}y_{t-i} + \sum_{i=0}^{k-1} \boldsymbol{k}_i \boldsymbol{D}D_{t-i} + \boldsymbol{e}_t \tag{11}
$$

where

$\boldsymbol{r}_k = \boldsymbol{f}_1 + \boldsymbol{f}_2 + \ldots. + \boldsymbol{f}_k - 1$

and the model is formulated conditionally on the first k observations of each sub-sample. This representation shows that the test regression should include both the intervention dummy lagged k periods, the first difference of the intervention dummy and up to k-1 lags of the first difference of the intervention dummy. It also shows that the intervention dummy lagged k periods should be included both in the intercept and the deterministic trend variable, even though under the null of $\boldsymbol{r}_1 = 0$ the lagged dummy disappears in the trend component (but not in the intercept part). A unit root test should be a joint test of the joint significance of the coefficients of y_{t-1}, the trend and the dummy lagged k periods times the trend in (11).

Generalization to the case of more than one shift

We allow for q samples periods, $1 = T_0 < T_1 < T_2 < \ldots < T_q = T$. The last observation of the jth sample is T_j and the first period of the $(j+1)$th sample period is $T_j+1, j= 1,\ldots,q$. The model is formulated conditionally on the first k observations of each sub-sample, for example for the jth sub-sample, $y_{T_{j-1}+1},\ldots,y_{T_{j-1}+k}$. We also define q-1 intervention dummy variables[1]:

$$D_{j,t} = \begin{cases} 1 & \text{for } T_{j-1}+1 \le t \le T_j, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for } j = 2,\ldots,q$$

and

$$D_{j,t-k} = \begin{cases} 1 & \text{for } T_{j-1}+k+1 \le t \le T_j + k, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for } j = 2,\ldots,q.$$

Correspondingly we define:

$$I_{j,t} = \begin{cases} 1 & \text{for } t = T_{j-1}+1, \\ 0 & \text{otherwise,} \end{cases} \qquad .$$

When $q = 2$, $I_{j,t}$ is just $DD_{j,t}$. $I_{j,t-i+1}$ is an indicator variable for the ith observation in the jth period.

Equation (11), in the case of q periods becomes:

[1] Our notation for the intervention dummies differ from JMN (2000). In this later paper $D_{j,t}$ denotes an indicator function for the last observation in the j-1th sample.

$$
\begin{aligned}
Dy_t = {} & \boldsymbol{r}_k y_{t-1} - \boldsymbol{r}_k \boldsymbol{d}_1 t - \boldsymbol{r}_k t \sum_{j=2}^{q} \boldsymbol{d}_j D_{j,t-k} + \boldsymbol{h}_1 + \sum_{j=2}^{q} \boldsymbol{h}_j D_{j,t-k} + \sum_{i=1}^{k-1} G_i D y_{t-i} \\
& + \sum_{i=0}^{k-1} \sum_{j=2}^{q} \boldsymbol{k}_{j,i} I_{j,t-i} + \boldsymbol{e}_t
\end{aligned} \tag{12}
$$

As before the effect of $I_{j,t}, \ldots, I_{j,t-k+1}$, corresponding to the observations $y_{T_{j-1}+1}$, ..., $y_{T_{j-1}+k}$, is to render the respective residuals zero given the initial values in each period. In practice we need to include the intervention dummies for each sub-sample with the appropriate lags as well as the dummies times the trend and the indicator variables for the break points, again with the appropriate lags.

3. Multivariate Case

The most common method to test for the cointegration rank is the maximum likelihood cointegration test method developed by Johansen (1988, 1996). It is, however, the case that the inclusion of intervention dummies affects the distribution of cointegration tests. JMN (2000) generalize the likelihood-based cointegration analysis developed by Johansen (1988, 1996) to the case where structural breaks exist at known points in time. They show that new asymptotic tables are required. In this section we show how to obtain equation (2.6) of JMN (2000) by expanding the results from Section 2.

In what follows we assume that we have a p-vector process Y_t and that without structural breaks the model can be formulated conditionally on the first k observations by:

$$
DY_t = \boldsymbol{P} Y_{t-1} + \boldsymbol{P}_1 t + m + \sum_{i=1}^{k-1} G_i D Y_{t-i} + \boldsymbol{e}_t \tag{13}
$$

where $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_T$ are normal, independent and identically distributed p ´1 vectors with mean 0 and variance $\boldsymbol{W}$. We also assume that although some or all of the p time series in Y_t may have a time trend, none have a quadratic trend.

The hypothesis of cointegration can be reformulated as a reduced rank problem of the $\boldsymbol{P}$ matrix, in which case $\boldsymbol{P} = \boldsymbol{ab}'$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are $(p$ ´$r)$ full rank matrices, and Y_t has a quadratic trend. If none of the p time series displays a quadratic trend we need to assume that $\boldsymbol{P}_1 = \boldsymbol{ag}'$, where $\boldsymbol{g}$ is a $(1$ ´$r)$ full rank matrix.

If we now assume that we have *q-1* breaks (and *q* sub-samples), conditionally on the first *k* observations of each sub-sample the model can be rewritten as *q* equations:

$$DY_t = \left(\boldsymbol{P}, \boldsymbol{P}_j\right)\begin{pmatrix} Y_{t-1} \\ t \end{pmatrix} + m_j + \sum_{i=1}^{k-1} G_i DY_{t-i} + \boldsymbol{e}_t \qquad (14)$$

$j = 1,...,q$, where $\boldsymbol{P}_j$ and m_j are $(p\ ´1)$ vectors.

Under the null of cointegration, we restrict the trend to the cointegrating relationships to exclude the possibility of quadratic trends in any time series. This means that $\boldsymbol{P}_j = ag_j$'.

Instead of writing *q* equations we can define the following matrices:

$D_t = (1,...,D_{q,t})'$, $\quad m = (m_1,...,m_q)$, $\quad g = (g_1',...,g_q')'$

of dimensions $(q\ ´1)$, $(p\ ´q)$, $(q\ ´r)$ respectively, and rewrite (14) in a form similar to (12):

$$DY_t = a\begin{pmatrix} b \\ g \end{pmatrix}'\begin{pmatrix} Y_{t-1} \\ tD_{t-k} \end{pmatrix} + mD_{t-k} + \sum_{i=1}^{k-1} G_i DY_{t-i} + \sum_{i=0}^{k-1}\sum_{j=2}^{q} \boldsymbol{k}_{j,i} I_{j,t-i} + \boldsymbol{e}_t \qquad (15)$$

where the dummy variables $D_{j,t}$, $D_{j,t-k}$ and $I_{j,t}$ are defined as in the previous section, and the $\boldsymbol{k}_{j,i}$ are $(p\ ´1)$.vectors.

JMN (2000) develop a maximum likelihood cointegration test method based on the squared sample canonical correlations, $\hat{l}_i$, of DY_t and $\left(Y'_{t-1}, tD'_t\right)$ corrected for the regressors:

D_{t-k}, $\quad DY_{t-i}$, $i = 1,...,k-1$, $\qquad I_{j,t-i}$, $\quad i = 0,..., k-1$; $\quad j = 2,...,q$.

The likelihood ratio test statistic for the hypothesis of at most *r* cointegrating relations is given by:

$$\text{LR} = -T\sum_{i=r+1}^{p} log(1-\hat{l}_i) \qquad (16)$$

We consider next three cases:

1. none of the *p* time series displays a trending pattern, but the cointegrating relations have an intercept which can differ between the sub-samples;
2. some or all of the time series follow a trending pattern in each sub-sample and the cointegrating relations are trend stationary in each sub-sample; trend breaks are allowed both in the cointegrating relations and in the non-stationary series;

3. some or all of the time series follow a trending pattern in each sub-sample and the cointegrating relations are stationary in each sub-sample (with possibly a broken constant level); trend breaks are allowed only in the non-stationary series;

Shift in Intercept Model: None of the p time series have a deterministic trend

The only deterministic components in the model are the intercepts in the cointegrating relations which can differ between sub-samples. In that case we have:

$\Pi_1 = \Pi_2 = \ldots \Pi_q = 0$, moreover μ is restricted to the cointegrating relations.

The interpretation is that the cointegrating relations have an attractor μ_j which varies between sub-samples. This model is denoted by $H_c(r)$ in JMN (2000).

$$\Delta Y_t = \alpha \begin{pmatrix} \beta \\ \nu \end{pmatrix}' \begin{pmatrix} Y_{t-1} \\ D_{t-k} \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \sum_{i=0}^{k-1} \sum_{j=2}^{q} \kappa_{j,i} I_{j,t-i} + \varepsilon_t \qquad (17)$$

where $\alpha\nu' = \mu$

JMN (2000) show that the asymptotic distribution of the likelihood ratio test is well approximated by a Γ-distribution. The reader is referred to section 3.4 of JMN (2000) for the computation of the critical values depending on the number of non-stationary relations and the location of the break points.

Some or all of the time series follow a trending pattern

This model allows the individual time series to have broken trends, while the cointegrating relations may also broken trends. This model is denoted by $H_l(r)$ in JMN (2000). It is the most general case and is represented by equation (15). The derivation of the critical values for this model is also given in section 3.4 of JMN (2000).

Some or all of the time series follow a trending pattern in each sub-sample and the cointegrating relations are stationary in each sub-sample (with possibly a broken constant level); trend breaks are allowed only in the non-stationary series

This model is denoted $H_{lc}(r)$ in JMN (2000). The asymptotic distribution of the likelihood ratio test depends on nuisance parameters and cannot easily be obtained.

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \mu D_{t-k} + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \sum_{i=0}^{k-1}\sum_{j=2}^{q} \kappa_{j,i} I_{j,t-i} + \varepsilon_t \qquad (18)$$

Unit Root Tests

In the first two cases, models (15) and (17), JMN (2000) also show that test for linear restrictions on β, γ and ν are asymptotically χ^2-distributed. Such tests are particularly useful because they make it possible to test whether the individual time series are trend stationary on each sub-sample.

4. Simulation Results

Under construction…

5. Conclusion

In the last decade applied econometricians have usually treated structural breaks in VAR models in an ad hoc fashion. Intervention dummies have been included with little care given to their specification. In this note we have considered three models of interest in applications and have given a detailed account of the specification and inclusion of intervention dummies in those cases. Statistical theory for those cases has been developed in JMN (2000). Although there is no new statistical theory in this note, the discussion of the inclusion and specification of intervention dummies should be useful to applied economists. It is indeed often the case that including dummies does not solve the non-normality problems of the residuals encountered in the estimation of VAR and VECM models. The reason for this should now be clear.

References

Franses, P.H. (2001), How to Deal with Intercept and Trend in Practical Cointegration Analysis, *Applied Economics*, 33, 577-579.

Johansen, S., Moscow, R. and B. Nielsen (2000), Cointegration Analysis in the Presence of Structural Breaks in the Deterministic Trend, *Econometrics Journal*, 3, 216-249.

Perron, P. (1989), The Great Crash, the Oil Price Shock and the Unit Root Hypothesis, *Econometric*, 57, 1361-1401.

Perron, P. (1990), Testing for a Unit Root in a Time Series with a Changing Mean, *Journal of Business and Economic Statistics,* 8, 153-162.

Perron, P. and T.J. Vogelsang (1992), Nonstationarity and Level Shifts with a Application to Purchasing Power Parity, *Journal of Business and Economic Statistics*, 10, 301-320.

Zivot, E. and D.W.K. Andrews (1992), Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit Root Hypothesis, *Journal of Business and Economic Statistics*, 10, 251-270.