

Nonlinear Error Correction Models With An Application To Commodity Prices*

Rafael Magri**

Marcelo C. Medeiros***

Abstract

Existing tests for nonlinearity in vector error correction models are highly intensive computationally and have nuisance parameters in the asymptotic distribution, what calls for cumbersome bootstrap calculations in order to assess the distribution. Our work proposes a consistent test which is implementable in any statistical package and has Chi-Squared asymptotics. Moreover, Monte Carlo experiments show that in small samples our test has nice size and power properties, often better than the preexisting tests. We also provide a condition under which a two step estimator for the model parameters is consistent and asymptotically normal. Application to international agricultural commodities prices show evidence of nonlinear adjustment to the long run equilibrium on the wheat prices.

Keywords: Cointegration, Nonlinear Models, Linearity Testing, Asymptotic Theory.

JEL Codes: C12, C13, C32.

*Submitted in May 2014. Revised in October 2014. The partial support of CNPq/Brazil is gratefully acknowledged. The authors wish to thank João Victor Issler and Adrian Pizzinga for insightful discussions.

**Department of Economics, Pontifical Catholic University of Rio de Janeiro, Brazil.
E-mail: rafaelmagri@gmail.com

***Department of Economics, Pontifical Catholic University of Rio de Janeiro, Brazil.
E-mail: marcelo.c.medeiros@gmail.com

1. Introduction

The works of Engle and Granger (1987) and Johansen (1991) laid the ground for cointegration analysis. In their framework, the number and shape of linear long-run relationships can be estimated, as well as the linear adjustment of the system towards these equilibria. Readily the possibility of nonlinear behavior was advanced; see, for example, Granger and Teräsvirta (2006).¹ However, not until Hansen and Seo (2002) did the matter received a fully formal treatment, latter expanded by Bec and Rahbek (2004), Saikkonen (2005, 2008), Seo (2006, 2007), and Kristensen and Rahbek (2009), among others. Virtually all the tests and estimators developed involve complex calculations and non-standard asymptotic distributions.

All possible advantages of taking nonlinear relations into account in a stationary environment are present in this framework; see, for instance, van Dijk et al. (2002). Furthermore, there is the possibility to model persistent behavior in a globally stationary setting, as noted in Gonzalo and Pitarakis (2006). Furthermore, contrary to a linear adjustment function where every deviation from the cointegration relation is corrected, a nonlinear adjustment function may accommodate a deviation for a long period, until another shock takes the deviation to a value where the adjustment function does not equal zero anymore.

In this paper, we propose a very simple linearity test and an estimation procedure for a nonlinear Error Correction Model (ECM).² Our test is based on the Taylor expansion of the adjustment function of the ECM, in the same spirit of Luukkonen et al. (1988). The test is calculable as a standard F test. We show that, for a broad class of nonlinear adjustment functions, the test is consistent and χ^2 asymptotically distributed. Monte Carlo experiments show that in finite samples the test has nice size and power properties, often better than the preexisting tests in the literature. In comparison to the existing literature, our tests are in the same spirit of Hansen and Seo (2002) and Seo (2007). Each of these tests has one specific form of nonlinearity as alternative. In the former, it is a two-regime threshold regression (TR) model and in the later, it is a smooth transition regression (STR) model. As in many cases of linearity testing, under the null hypothesis there are unidentified parameters, turning the usual maximum likelihood approach impossible. The solution used in both papers, which was proposed in Davies (1987), is to perform a supLM-type test, which has a nonstandard distribution under the null and is frequently computer intensive to implement. First, the likelihood functions are discontinuous or too flat, requiring grid searches over some parameters. Second, bootstrap simulations are needed in order to assess the

¹Nonlinearities in the cointegration relation itself, as in Saikkonen and Choi (2004), may not have an Error Correction representation, as noted in Gonzalo and Pitarakis (2006). This makes the two approaches rather separate from each other.

²We assume cointegration rank is known. For cointegration tests under nonlinearity, see Kapetanios et al. (2006) and Seo (2006).

asymptotic distributions, which are frequently not free from nuisance parameters. Moreover, although the tests are derived against a very specific alternative hypothesis, it is sensible to expect a high power against many classes of nonlinearities.

We also provide a mild condition on the derivatives of the nonlinear function under which a two step estimator for the parameters in the model is consistent and has a normal asymptotic distribution. The first stage in our approach is a Ordinary Least Squares (OLS) regression, while the second step is a Nonlinear Least Squares (NLS) estimation. In a recent paper, Kristensen and Rahbek (2009) show, in a maximum likelihood framework, that it is possible to have asymptotic normal estimators, although a nonstandard distribution is the general case. Whereas they point that the linear model is in the subset of models with normal estimators, they do not provide any general condition to determine which models pertain to each case. While there is no formal way to determine which function to estimate, we provide a heuristic approach based on a semi-parametric investigation of the data to help the choice process.

Finally, we apply the test and estimation procedure to international agricultural commodities prices. We find evidence of nonlinear behavior in the wheat prices. Our results point to a behavior consistent with the presence of transaction costs. The estimates of the cost are 17.4% for the pair Argentina and United States and 12.1% for the pair Argentina and Brazil. These figures make sense, putting higher costs between more distant markets.

The rest of the paper is organized as follows. Section 2 presents the model and discusses its properties and applications. Section 3 describes the test procedure and establishes its asymptotic distribution. Section 4 describes the estimation procedures and establishes asymptotic distribution of the estimated parameters. Section 5 compares the small sample properties of the proposed test with the preexisting ones. Section 6 applies the methodology developed throughout the paper to international agricultural commodities prices. Section 7 concludes.

2. The Model

We consider the following generalization of the linear ECM:

$$\Delta \mathbf{y}_t = \mathbf{f}(\boldsymbol{\beta}' \mathbf{y}_{t-1}) + \sum_{i=1}^p \boldsymbol{\Gamma}_i \Delta \mathbf{y}_{t-i} + \boldsymbol{\epsilon}_t, \quad (1)$$

where $\mathbf{y}_t \in \mathbb{R}^n$ is a $I(1)$ vector of cointegrated series, the scalar $z_t = \boldsymbol{\beta}' \mathbf{y}_{t-1} \sim I(0)$ is a unique linear cointegration relationship, and $\mathbf{f}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}^n$ is a possibly nonlinear function. The model has a linear cointegration relation but a nonlinear dynamics towards the long-run equilibrium. Although it is not explicitly considered in the paper, the long-run relation may include deterministic terms, such as a constant or a linear trend.

The first issue is to establish the existence of such a model. It must be shown

that a cointegrated vector \mathbf{y}_t may have an error correction representation as in (1). In the linear case we have the Granger representation theorem of Engle and Granger (1987). In the nonlinear framework a similar result have been established under three different set of assumptions. In Saikkonen (2005), ϵ_t may have a GARCH structure but $\mathbf{f}(x) - (\alpha_0 + \alpha_1 x) = 0$ when $|x| \rightarrow \infty$. Kristensen and Rahbek (2009) make the same hypothesis on the limits of \mathbf{f} , but ϵ_t must be an independent and identically distributed (IID) innovation. Saikkonen (2008) can be seen as a generalization of Bec and Rahbek (2004), where the function \mathbf{f} must be a linear combination of linear functions, but not necessarily the same in the extremes while ϵ_t may have a GARCH structure.

The most important assumption is the linearity of \mathbf{f} in the limit. Since it is present in every existent proof, we will use it throughout the paper.

Assumption 1 $\mathbf{f}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}^n$ is such that, for some $\mathbf{K}_1, \mathbf{K}_2 \in \mathbb{R}^n$ and $\alpha_1, \alpha_2 \in \mathbb{R}^n$, $\lim_{x \rightarrow \infty} \mathbf{f}(x) - (\mathbf{K}_1 + \alpha_1 x) = \mathbf{0}$ and $\lim_{x \rightarrow -\infty} \mathbf{f}(x) - (\mathbf{K}_2 + \alpha_2 x) = \mathbf{0}$.

This restriction leads to the wide usage of smooth transition models, where a weighting function (called transition function) is used to combine two or more linear functions. The most common transition functions are the logistic and the exponential. If

$$\mathbf{f}(\beta' \mathbf{y}_{t-1}) = \alpha \beta' \mathbf{y}_{t-1} + \{1 - \exp[-\lambda(\beta' \mathbf{y}_{t-1} - c)^2]\} \delta \beta' \mathbf{y}_{t-1} \quad (2)$$

we have the exponential smooth transition model, where λ is the smoothness (velocity) of the transition and c is the location parameter. If

$$\mathbf{f}(\beta' \mathbf{y}_{t-1}) = \alpha \beta' \mathbf{y}_{t-1} + \frac{1}{1 + \exp[-\lambda(\beta' \mathbf{y}_{t-1} - c)]} \delta \beta' \mathbf{y}_{t-1} \quad (3)$$

we have the logistic smooth transition model, again with λ as the velocity of transition and c as the location parameter. Another example, less widespread, is a combination of logistic functions, from Suárez-Fariñas et al. (2004)

$$\mathbf{f}(\beta' \mathbf{y}_{t-1}) = \alpha \beta' \mathbf{y}_{t-1} + \left\{ 1 + \frac{1}{1 + \exp[-\lambda(\beta' \mathbf{y}_{t-1} - c)]} - \frac{1}{1 + \exp[-\lambda(\beta' \mathbf{y}_{t-1} + c)]} \right\} \delta \beta' \mathbf{y}_{t-1}. \quad (4)$$

Figure 1 illustrates the shape of the transition functions discussed above. These transition functions generates adjustment functions \mathbf{f} as shown in Figure 2. The exponential and double logistic models may be very similar depending on the value of the parameters. The exponential model has been used for no-arbitrage

conditions in the presence of transaction costs, for example, addressing the PPP puzzle in Michael et al. (1997). However, the exponential model may account for little deviations from linearity, sometimes fitting its curve to better accommodate an outlier. The logistic model is appropriate for cointegration of variables with different behavior when above or under a certain value, possibly zero. A very similar model was used in Hansen and Seo (2002) for long and short bonds interest rates.

3. Testing Linearity

We want to test whether \mathbf{f} is a linear function or not. If \mathbf{f} is the exponential or logistic function, this could be done by testing $\mathcal{H}_0 : \lambda = 0$ or $\mathcal{H}_0 : \boldsymbol{\delta} = 0$ in Equations (2), (3), or (4). The issue is that in any case there are unidentified parameters under the null hypothesis. One possible solution is to follow Davies (1987) and perform a supLM test, as done in Hansen and Seo (2002) and Seo (2007). Another possibility is to follow the ideas in Luukkonen et al. (1988) and substitute \mathbf{f} by its Taylor expansion around the null hypothesis and test the polynomial coefficients. The latter will be our approach.

We will make two extra assumptions.

Assumption 2 *The function $\mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}^n$ is three times continuously differentiable.*

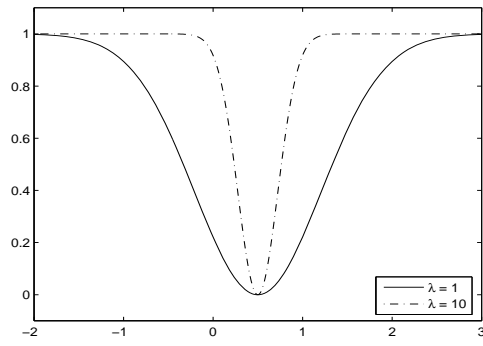
Assumption 3 *$\{\boldsymbol{\epsilon}\}_t$ is a sequence of independent and normally distributed random vectors with zero mean and positive definite covariance matrix $\boldsymbol{\Upsilon}$, $\boldsymbol{\epsilon}_t \sim \text{NID}(\mathbf{0}, \boldsymbol{\Upsilon})$.*

Assumption 4 *Set $z_t = \boldsymbol{\beta}'\mathbf{y}_{t-1} \sim I(0)$. $\mathbb{E}(z_t^\delta) < \infty$ and $\mathbb{E}(\Delta\mathbf{y}_{t-i}z_t^\delta) < \infty$, for $\delta > 6$ and $i = 1, \dots, p$.*

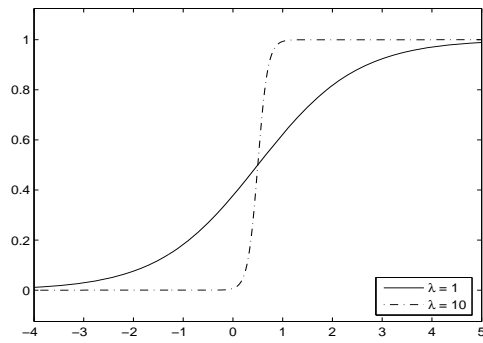
The differentiability is needed in order to guarantee the validity of the Taylor expansion, while Assumption 3 is common in error correction models. A important restriction is threshold models. Since the threshold function is not differentiable everywhere, it does not satisfy Assumption 2, which is essential to the validity of the test. However, in finite samples, a threshold model can always be approximated through a logistic smooth function by taking a large enough λ .

This approach has two advantages over the supLM one. First, it is extremely simple and much faster computationally. While supLM demands bootstrap calculations, grid searches and involves non-standard asymptotic distributions, the Taylor expansion approach demands only a simple F-test. The computing time difference is of the order of 10^4 . Second, since the test does not have a specific alternative hypothesis, it is consistent against a large set of nonlinearities. The supLM approach will be consistent only against the specific alternative being tested,

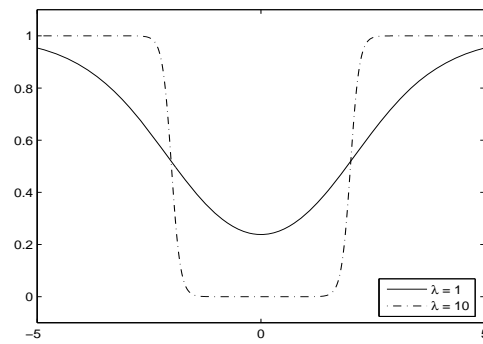
Figure 1
Transition functions versus the transition variable



(a) Exponential Function



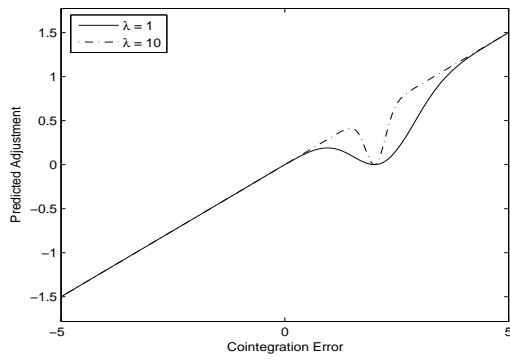
(b) Logistic Function



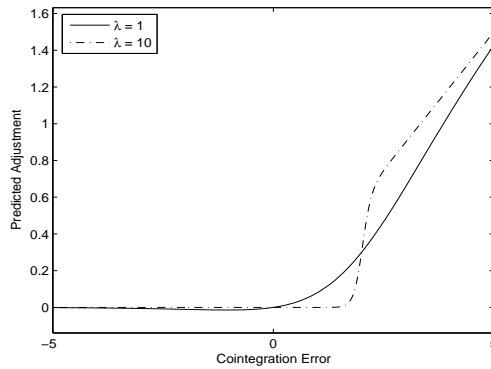
(c) Double Logistic Function

Each function is depicted with two different values for the parameter λ . Dashed lines ($\lambda = 10$) and continuous line ($\lambda = 1$).

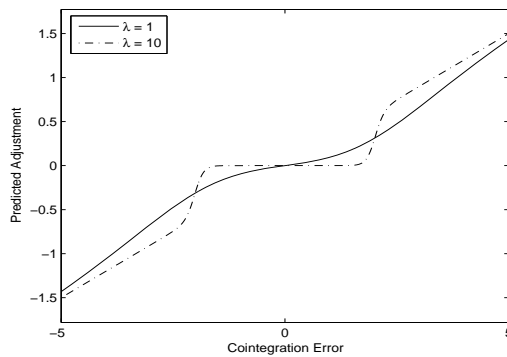
Figure 2
Adjustment Functions



(a) Exponential Model Adjustment



(b) Logistic Model Adjustment



(c) Double Logistic Model Adjustment

Each function is depicted with two different λ values, $c = 2$, $\alpha = 0$ and $\delta = 0.3$. Dashed lines ($\lambda = 10$) and continuous line ($\lambda = 1$).

but will also have power against other alternatives as well. Thus, since it is possible that the process is nonlinear but is not the alternative hypothesis, accepting \mathcal{H}_0 does not mean the process is linear and rejecting \mathcal{H}_0 does not guarantees the nonlinearity has the form being tested. This is true even for large samples.

Using a third-order expansion of \mathbf{f} around $\beta' \mathbf{y}_{t-1} = 0$, (1) becomes³

$$\Delta \mathbf{y}_t = \boldsymbol{\theta}_0 + \boldsymbol{\theta}_1(\beta' \mathbf{y}_{t-1}) + \boldsymbol{\theta}_2(\beta' \mathbf{y}_{t-1})^2 + \boldsymbol{\theta}_3(\beta' \mathbf{y}_{t-1})^3 + \sum_{i=1}^p \boldsymbol{\Gamma}_i \Delta \mathbf{y}_{t-i} + \boldsymbol{\epsilon}_t^*, \quad (5)$$

where $\boldsymbol{\epsilon}_t^* = \boldsymbol{\epsilon}_t + (\frac{1}{6}) \mathbf{f}^{(4)}(k_t)(\beta' \mathbf{y}_{t-1})^4$, for some $k_t \in \mathbb{R}$, $\boldsymbol{\theta}_0 = \mathbf{f}(0)$, $\boldsymbol{\theta}_1 = \mathbf{f}^{(1)}(0)$, $\boldsymbol{\theta}_2 = (\frac{1}{2}) \mathbf{f}^{(2)}(0)$, and $\boldsymbol{\theta}_3 = (\frac{1}{6}) \mathbf{f}^{(3)}(0)$. $\mathbf{f}^{(i)}(0)$ is the i th-order derivative of \mathbf{f} evaluated at 0. When \mathbf{f} is linear, we have $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_3 = \mathbf{0}$. When \mathbf{f} is nonlinear, $\mathbf{f}^{(2)}(x) \neq 0$ for almost every $x \in \mathbb{R}$. The inequality will not hold if x is a point of inflexion of the function. This will be true for $x = 0$ if, for example, \mathbf{f} is an odd function. We include the third term of the expansion to deal with this situations. To keep it concrete, take the exponential model where $\boldsymbol{\theta}_0 = \mathbf{0}$, $\boldsymbol{\theta}_1 = \boldsymbol{\alpha} + \boldsymbol{\delta} \exp(-\lambda c^2)$, $\boldsymbol{\theta}_2 = -4c\boldsymbol{\delta}\lambda \exp(-\lambda c^2)$, $\boldsymbol{\theta}_3 = 3\boldsymbol{\delta}[2\lambda \exp(-\lambda c^2) - 4c^2\lambda^2 \exp(-\lambda c^2)]$. If the location parameter, c , is zero, the function is odd, and $\boldsymbol{\theta}_2 = \mathbf{0}$, but $\boldsymbol{\theta}_3 \neq \mathbf{0}$. Therefore, the test will be able to detect the nonlinearity. To test $\mathcal{H}_0 : \mathbf{f}$ is linear against $\mathcal{H}_A : \mathbf{f}$ is nonlinear, we propose the following procedure:

- (a) Estimate $\widehat{\boldsymbol{\beta}}$ super-consistently. In our framework, it is enough to run an OLS in the equation $y_{1t} = \beta_1 + \beta_2 y_{2t} + \dots + \beta_n y_{nt} + u_t$;⁴
- (b) Estimate (5) by OLS using $\widehat{\boldsymbol{\beta}}$ in place of $\boldsymbol{\beta}$. Then, perform a F-test for the following null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_3 = \mathbf{0}$.

Proposition 3.1 *Under Assumptions 1–4 and $\mathcal{H}_0 : \boldsymbol{\theta}_2 = \boldsymbol{\theta}_3 = \mathbf{0}$, the F-statistic on the second stage of the proposed test has a $\chi^2(2n)$ asymptotic distribution. Moreover, under \mathcal{H}_A , the test is consistent.*

4. Estimation and Asymptotic Properties

Given a function \mathbf{f} parameterized by a vector $\boldsymbol{\psi} \in \mathbb{R}^m$, $\mathbf{f}(x, \boldsymbol{\psi})$, the parameters $\boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_p$ and $\boldsymbol{\psi}$ are estimated in two stages. Set $\boldsymbol{\eta}' = [\boldsymbol{\psi}', \text{vec}(\boldsymbol{\Gamma}_1)', \dots, \text{vec}(\boldsymbol{\Gamma}_p)']'$ and define the Nonlinear Least Squares (NLLS) estimator of $\boldsymbol{\eta}$ as

$$\widehat{\boldsymbol{\eta}} = \underset{\boldsymbol{\eta}}{\text{argmin}} \mathcal{Q}_T(\mathbf{Y}, \boldsymbol{\eta}) = \underset{\boldsymbol{\eta}}{\text{argmin}} \sum_{t=1}^T \boldsymbol{\epsilon}_t(\boldsymbol{\eta})' \boldsymbol{\epsilon}_t(\boldsymbol{\eta}), \quad (6)$$

³See Lemma C.1 in the Appendix.

⁴In case of endogenous regressors, Dynamic OLS (DOLS) may be used.

where $\epsilon_t(\boldsymbol{\eta}) = \Delta \mathbf{y}_t - \mathbf{f}(\boldsymbol{\beta}' \mathbf{y}_{t-1}, \boldsymbol{\psi}) - \sum_{i=1}^p \boldsymbol{\Gamma}_i \Delta \mathbf{y}_{t-i}$ and $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$ is a $(T \times n)$ matrix representing the dataset.

Consider the following estimation procedure:

- (a) Estimate $\widehat{\boldsymbol{\beta}}$ super-consistently as discussed in Section 3.
- (b) Estimate (1) by NLLS using $\widehat{\boldsymbol{\beta}}$ instead of $\boldsymbol{\beta}$.

To establish the asymptotic normality of the estimators, we will need two extra assumptions. First, define the following partial derivatives:

$$\mathbf{f}_{\psi_i}(x, \boldsymbol{\psi}) = \frac{\partial \mathbf{f}(x, \boldsymbol{\psi})}{\partial \psi_i}, \mathbf{f}_{\psi_i x}(x, \boldsymbol{\psi}) = \frac{\partial^2 \mathbf{f}(x, \boldsymbol{\psi})}{\partial \psi_i \partial x}, \mathbf{f}_{\psi_i \psi_j}(x, \boldsymbol{\psi}) = \frac{\partial^2 \mathbf{f}(x, \boldsymbol{\psi})}{\partial \psi_i \partial \psi_j},$$

and so on.

Assumption 5 *The derivatives $\mathbf{f}_{\psi_i x}(x, \boldsymbol{\psi})$ and $\mathbf{f}_{\psi_i \psi_j x}(x, \boldsymbol{\psi})$ are limited in $x \forall i, j \in \{1, \dots, m\}$.*

Note, for example, that the models discussed in the previous sections attend Assumption 5.

Proposition 4.1 *Under Assumptions 1–5, $\sqrt{T}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Furthermore, the matrix $\boldsymbol{\Sigma}$ is consistently estimated by*

$$\widehat{\boldsymbol{\Sigma}} = \left[\sum_{t=1}^T \mathbf{J}(\mathbf{y}_t, \widehat{\boldsymbol{\eta}})' \mathbf{J}(\mathbf{y}_t, \widehat{\boldsymbol{\eta}}) \right]^{-1} \left[\sum_{t=1}^T \mathbf{J}(\mathbf{y}_t, \widehat{\boldsymbol{\eta}})' \epsilon_t(\widehat{\boldsymbol{\eta}}) \epsilon_t(\widehat{\boldsymbol{\eta}})' \mathbf{J}(\mathbf{y}_t, \widehat{\boldsymbol{\eta}}) \right] \left[\sum_{t=1}^T \mathbf{J}(\mathbf{y}_t, \widehat{\boldsymbol{\eta}})' \mathbf{J}(\mathbf{y}_t, \widehat{\boldsymbol{\eta}}) \right]^{-1},$$

where

$$\mathbf{J}(\mathbf{y}_t, \boldsymbol{\eta}) = \frac{\partial \epsilon_t(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}.$$

Again, as a consequence of the faster convergence rate of the cointegrating vector $\widehat{\boldsymbol{\beta}}$, the nonlinear least squares of the second stage has standard asymptotics. Using a maximum likelihood approach, Kristensen and Rahbek (2009) showed that the general distribution of the parameters estimator is non-standard, drawing attention to the fact that it was possible that some models would yield normal distributions, for example, the linear model. However, having rejected the null hypothesis of linearity in the test presented in the previous section, what model should a researcher estimate? In other words, which function \mathbf{f} to choose? In some applications, it is possible that the researcher has a specific function in mind. For

example, in Kapetanios et al. (2006) it is shown that a incomplete information model may lead exactly to a model with logistic transition equation. Yet, this will not always be the case. When there is no theoretical function available, we propose a heuristic procedure based on a semi-parametric approach. Comparing the semi-parametric estimate with the existing functions in the nonlinear literature, it is possible to choose the most adequate.

5. Small Sample Properties

In this section we conduct Monte Carlo experiments to assess size and power of the test in small samples. We will compare the performance of the proposed test (henceforth refereed as Taylor) with the other two available tests for nonlinearity in Error Correction Models, Hansen and Seo (2002) and Seo (2007), which henceforth will be referred respectively as HS and Seo.⁵ These are supLM tests, designed against a specific alternative hypothesis. In HS, the alternative is a threshold model, whereas in Seo it is a logistic or an exponential smooth transition model.

The models used in each simulation will be the same used in Hansen and Seo (2002) and Seo (2007), so as to render the results directly comparable. The sample size is always 250 and the number of repetitions is fixed in 1000. The error is independent multivariate normal with unit variance. First, to asses the empirical size, the Data Generating Process (DGP) will be

$$\Delta \mathbf{y}_t = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (y_{1t-1} - \beta_2 y_{2t-1}) + \mathbf{\Gamma} \Delta \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t,$$

where $\mathbf{\Gamma}$ may assume three values:

$$\mathbf{\Gamma}_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \mathbf{\Gamma}_1 = \begin{pmatrix} -0.2 & 0 \\ -0.1 & -0.2 \end{pmatrix}, \text{ or } \mathbf{\Gamma}_2 = \begin{pmatrix} -0.2 & -0.1 \\ -0.1 & -0.2 \end{pmatrix}.$$

The parameters $\alpha_1 = -1$ and $\beta_2 = -1$ are fixed and α_2 varies among 0, -0.5 and 0.5. The results of the simulation for the Taylor test are in Table 1, as well as the results from Hansen and Seo (2002) and Seo (2007). The results are somewhat homogenous, showing good size in almost every case. The largest deviation being 1.4% for Seo L, followed by 0.9% for Seo E, 0.8% for HS and 0.8% for Taylor. The mean deviation is 0.38% for HS, 0.84% for Seo L, 0.24% for Seo E and 0.4% for Taylor.

To compare the power, the first DGP is

$$\Delta \mathbf{y}_t = \begin{pmatrix} \alpha_1 \\ 0 \end{pmatrix} (y_{1t-1} - \beta_2 y_{2t-1}) + \begin{pmatrix} -\delta_1 \\ 0 \end{pmatrix} (y_{1t-1} - \beta_2 y_{2t-1}) f(z_{t-1}, \lambda, c) + \boldsymbol{\epsilon}_t,$$

⁵Seo L will refer to the test against a logistic function and Seo E will refer to the test against the exponential function.

Table 1
EMPIRICAL SIZE.

This table presents for the featured tests the empirical size for a linear ECM Data Generating Process calculated by a Monte Carlo experiment with 1000 repetitions for a 250 observations sample size.

α_2	0	-0.5	0.5	0	0
Γ	Γ_0	Γ_0	Γ_0	Γ_1	Γ_2
HS	0.054	0.056	0.047	0.054	0.052
Seo L	0.059	0.058	0.036	0.051	0.060
Seo E	0.046	0.051	0.047	0.049	0.047
Taylor	0.054	0.053	0.055	0.042	0.050

where f is the exponential function, as in Equation (2) or the logistic function, as in Equation (3). The parameters δ_1 and λ take the values (0.4, 0.8) and (0.75, 3, 9) respectively, while α_1 is fixed in -0.2 and β_2 is fixed in 1. The simplicity of the model under the alternative is due to the computational requirements of the two supLM tests. In Hansen and Seo (2002) we read “To keep the calculations manageable, we generate the data from the simple process,” while Seo (2007) uses “25 grid points (...) to reduce the computational costs.”

It is expected that higher δ_1 values will yield higher powers, since the nonlinearity will be more pronounced. Also, for the logistic function, higher λ values should yield higher powers, since, as can be seen in Figure 2, a smaller λ makes the function more similar to a linear one. For the exponential function, higher powers should be associated with smaller λ values. As shown in Figure 2, high lambda values turn the nonlinear region very small.

In Table 2 we can see that, even though this model is the alternative hypothesis under which the Seo test was developed, the Taylor test has a better power for all but one of the parameters combinations for the logistic case. Even the HS test is better than the Seo test in this case. In the exponential case we have the inverse situation: the Seo test is better than the other two tests in all but one of the parameters combinations. However, the power is very small in most of the cases.

The second DGP is

$$\Delta \mathbf{y}_t = \begin{pmatrix} \alpha_1 \\ 0 \end{pmatrix} (y_{1t-1} - \beta_2 y_{2t-1}) + \begin{pmatrix} \delta_1 \\ 0 \end{pmatrix} (y_{1t-1} - \beta_2 y_{2t-1}) I(y_{1t-1} - \beta_2 y_{2t-1} \leq \lambda) + \epsilon_t,$$

where $I(\cdot)$ is the indicator function. Parameters α_1 and β_2 will be held fixed at -1 and 1 respectively, while δ will take values (0.2, 0.4, 0.6, 0.8). The parameter λ is set so that $\omega = P(y_{1t-1} - \beta_2 y_{2t-1} \leq \lambda)$ equals 0.5 or 0.25. This model is

Table 2
EMPIRICAL POWER.

This table presents for the featured tests the empirical power for a Smooth Transition ECM Data Generating Process calculated by a Monte Carlo experiment with 1000 repetitions for a 250 observations sample size.

δ_1	0.4	0.4	0.4	0.8	0.8	0.8
λ	0.75	3	9	0.75	3	9
DGP with Logistic Function						
HS	0.514	0.626	0.610	0.927	0.904	0.897
Seo L	0.212	0.596	0.554	0.520	0.966	0.953
Taylor	0.625	0.648	0.620	0.984	0.970	0.943
DGP with Exponential Function						
HS	0.066	0.064	0.058	0.096	0.069	0.043
Seo E	0.285	0.077	0.057	0.869	0.124	0.057
Taylor	0.119	0.065	0.043	0.379	0.064	0.058

Table 3
EMPIRICAL POWER.

This table presents for the featured tests the empirical power for a Threshold ECM Data Generating Process calculated by a Monte Carlo experiment with 1000 repetitions for a 250 observations sample size.

δ_1	0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8
ω	0.5	0.5	0.5	0.5	0.25	0.25	0.25	0.25
HS	0.156	0.450	0.878	0.997	0.187	0.527	0.844	0.933
Taylor	0.151	0.503	0.891	0.994	0.179	0.592	0.888	0.958

the alternative for which the HS test was developed. The empirical power is in Table 3.

It is worth remembering that we cannot establish the consistency of the Taylor test against the threshold alternative. Despite this, the power in small samples is the same or better than the HS test in all parameters used in this Monte Carlo.

6. Application: International Commodities Prices

A leading application of nonlinearity in ECMs is price adjustment of identical products traded in separate markets. It is reasonable to expect these prices to be cointegrated, otherwise there would be perennial arbitrage opportunities. Yet, for each arbitrageur there must be a minimum deviation to cover transaction costs. Therefore, the adjustment should be nonlinear. Examples are tradable goods in the American CPI in Lo and Zivot (2001), stocks and their respective Depositary Receipts in Chung et al. (2005), agricultural commodities prices in Balcombe et al.

Table 4
LINEARITY TESTS.

Commodity	Country Pair	F statistic	P-Value
Wheat	US ARG	35.60026	0.0000
	US BR	5.313077	0.2567
	BR ARG	38.48891	0.0000
Soybean	US ARG	7.716483	0.1025
	US BR	8.660560	0.0702
	BR ARG	2.307507	0.6794
Maize	US BR	3.175390	0.5289

(2007), among many others.

In Balcombe et al. (2007), Bayesian estimation is used to avoid the issues already mentioned in nonlinear ECMs. They find evidence of nonlinearity in adjustment of Brazilian and American prices of maize, soybeans and wheat. We will apply our methodology to a database similar to the one used there. It is a monthly prices series from United States Department of Agriculture (USDA) to USA and Argentina prices and from Institute for Pure and Applied Economics Research (IPEA) to Brazil prices of each one of the three products, consisting of 151 observations from October/1996 to April/2009.

All series are $I(1)$, either by Augmented Dickey-Fuller or Phillips-Perron tests. Both Engle-Granger and Johansen tests for linear cointegration indicate the presence of cointegration vector $(1, -1)$. For the sake of brevity we do not report these results here.

On Table 4 we see the results for our nonlinearity test. We find strong evidence of nonlinearity on the adjustments of wheat prices from Argentina (both to Brazil and United States) and weak evidence on the adjustment of soybeans prices from the United States (again, both to Brazil and Argentina).

We will estimate the nonlinear ECM for the two wheat prices pairs. But first, we estimate the linear ECMs for both pairs, in order to compare the results afterwards. The lag length selection is made by the Schwartz Information Criterion. The ECMs are estimated without constant both in the cointegration equation and in the VAR. The equation to be estimated is

$$\begin{pmatrix} \Delta p_{it} \\ \Delta p_{jt} \end{pmatrix} = \begin{pmatrix} \alpha_i \\ \alpha_j \end{pmatrix} (p_{it-1} - \beta_2 p_{jt-1}) + \Gamma_1 \Delta \mathbf{p}_{t-1} + \mathbf{u}_t, \quad (7)$$

where i and j are the countries. Our parameters of interest are α_i and α_j , which tells us how much prices adjust in response to deviations from the long run equilibrium and β_2 , which gives us the long run equilibrium of the prices. The results are in Table 5. Both pairs have a $(1, -1)$ cointegration vector, meaning the prices are equal in equilibrium. Furthermore, in both cases it is the price in Argentina that moves in response to disequilibria.

Table 5
Linear ECM Estimations for Wheat Prices

United States and Argentina			Brazil and Argentina		
	Estimate	Std. Deviation		Estimate	Std. Deviation
β_2	1.012794	0.004632	β_2	1.004575	0.004522
α_{US}	-0.009278	0.038997	α_{Br}	0.057601	0.060062
α_{Arg}	0.235855	0.099806	α_{Arg}	-0.218631	0.049419
R^2 US equation		0.0524	R^2 Br equation		0.2048
R^2 Arg equation		0.2683	R^2 Arg equation		0.1939

To decide which nonlinear model to estimate, we make a semi-parametric regression. The results, seen in Figure 3, panels (a), (b), (c) and (d), show that for small deviations there is no adjustment. After a threshold the adjustment is proportional to the deviation size. This features are consistent with the presence of transaction costs. Moreover, two graphics show no reaction to deviations, which are exactly the non-significant coefficients from the linear estimates. The function shape is very similar to the Smooth Transition Model from Suárez-Fariñas et al. (2004), of which the graphic is in Figure 3(e). We will estimate the model with this nonlinear function.

The expression for the function is

$$F(p_{it-1} - \beta_2 p_{jt-1}, \lambda, c) = 1 + \frac{1}{1 + \exp[-\lambda(p_{it-1} - \beta_2 p_{jt-1} - c)]} - \frac{1}{1 + \exp[-\lambda(p_{it-1} - \beta_2 p_{jt-1} + c)]} \tag{8}$$

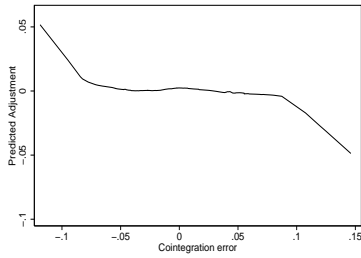
and the model to be estimated is

$$\begin{pmatrix} \Delta p_{it} \\ \Delta p_{jt} \end{pmatrix} = \begin{pmatrix} \alpha_i \\ \alpha_j \end{pmatrix} (p_{it-1} - \beta_2 p_{jt-1}) F(p_{it-1} - \beta_2 p_{jt-1}, \lambda, c) + \mathbf{\Gamma}_1 \Delta \mathbf{p}_{t-1} + \mathbf{u}_t.$$

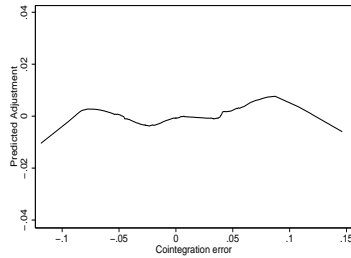
The estimation procedure is not straightforward. As noted in Teräsvirta (1994), the joint estimation of λ and the rest of the parameters is difficult. Numerical problems arise, making the convergence too slow and inflating the estimated parameter. We adopt the same solution proposed in that paper, which is to perform a grid search over (λ, c) and then estimate the rest of the parameters given these numbers. Afterwards we use these estimates as starting points for the minimization problem in Equation (6).

The results are in Table 6. For the prices from United States and Argentina, the estimated threshold is 0.1742. In our motivation, this means that if the difference between prices is less than 17.42%, there is not enough arbitrage pressure to drive

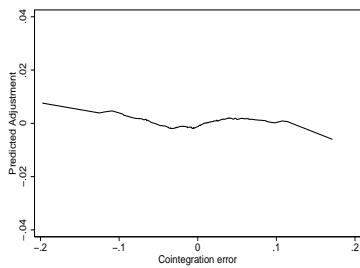
Figure 3
Adjustment functions



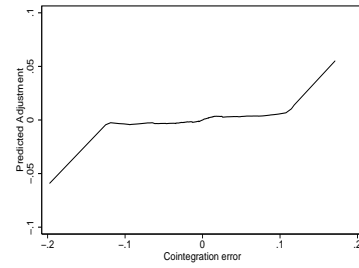
(a) Nonparametric Argentina's reaction to Brazil



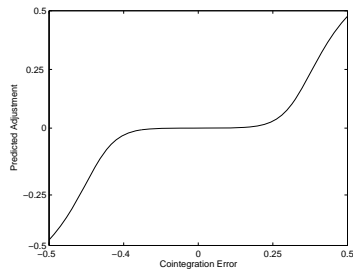
(b) Nonparametric Brazil's reaction to Argentina



(c) Nonparametric United States' reaction to Argentina



(d) Nonparametric Argentina's reaction to United States



(e) Proposed parametric function

Panels (a), (b), (c) and (d) display the nonparametric estimations. Panel (e) displays the function used in the estimation.

Table 6
Nonlinear ECM Estimations for Wheat Prices

United States and Argentina			Brazil and Argentina		
	Estimate	Std. Deviation		Estimate	Std. Deviation
λ	28.0915	10.8536	λ	65.2485	18.1752
c	0.1742	0.0284	c	0.1211	0.0112
β_2	1.0135	0.0069	β_2	1.0038	0.0068
α_{US}	-0.1516	0.1083	α_{Br}	0.1440	0.1010
α_{Arg}	0.8819	0.2073	α_{Arg}	-0.9164	0.1726
1 st Regime %		0.054	1 st Regime %		0.044
R^2 US equation		0.0716	R^2 Br equation		0.2204
R^2 Arg equation		0.4129	R^2 Arg equation		0.3871

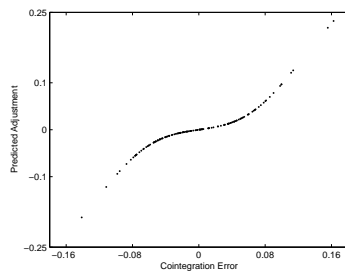
the prices back to the equilibrium. When the difference is bigger than 17.42%, arbitrageurs enter the market, forcing the prices back to the long run equilibrium. For the prices from Brazil and Argentina, this threshold is 12.11%, less than the previous one. This makes sense, since we expect the transaction costs to be smaller between markets closer to each other.

As expected, in comparison with the linear model, the estimated long run equilibrium, β , has almost not changed. The adjustment coefficients estimates, α , are not directly comparable. They have the same meaning only for sufficiently large cointegration errors, when the nonlinear adjustment function is already proximate to its linear limit. For small deviations, Figure 4 shows the shape of the estimated functions in each equation of the two nonlinear ECMs. Since in the linear ECM, this function would be linear with inclination equal to α_i , it is clear that our nonlinear estimates predict smaller adjustment for smaller cointegration errors and higher adjustment for higher cointegration errors.

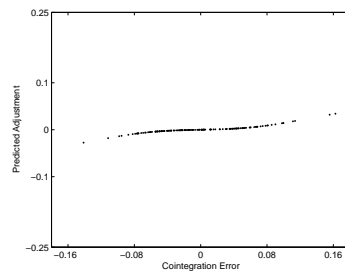
7. Conclusions

Exiting tests for nonlinearity in Error Correction Models (ECM) are of the supLM type, with non standard asymptotic distribution. Furthermore, even though they are designed against specific alternative hypothesis, they have considerable power against other nonlinearities, even though lacking consistency. We have proposed a simple F type test, with χ^2 asymptotic distribution, designed against a more general alternative. In small samples, the proposed test has similar power when testing for the specific alternatives for the existing tests and better power when testing for a diverse nonlinearity. We also provide a condition on the derivatives of the nonlinear function, which is attended by any Smooth Transition Model, under which a two step estimator for the parameters in the model has normal asymptotic distribution. The first stage is a Ordinary Least Squares, while the second is a Nonlinear Least Squares. While there is no formal way to determine

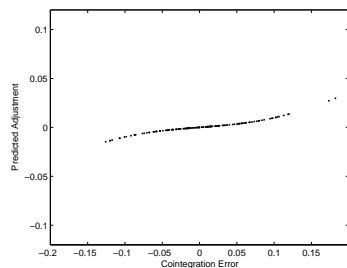
Figure 4
Estimated Adjustment Functions



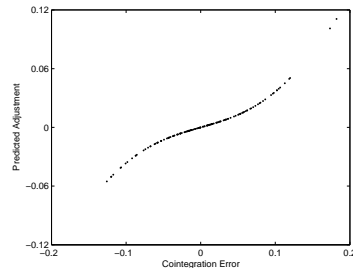
(a) Estimated Argentina's reaction to Brazil



(b) Estimated Brazil's reaction to Argentina



(c) Estimated United States' reaction to Argentina



(d) Estimated Argentina's reaction to United States

which function to estimate, we provide a heuristic approach based on a semi-parametric investigation of the data to help the choice process.

Both contributions follow the literature in assuming a known number of cointegration relations. A important aim for future research in the field is a test for the cointegration rank in the presence of this kind of nonlinear behavior.

Testing for non linearities in the adjustment of agricultural commodities prices in different countries shows strong evidence of nonlinear adjustment between wheat prices from Argentina, Brazil and United States. A semi-parametric regression points to a Smooth Transition ECM of the type used in Suárez-Fariñas et al. (2004), in which there is no price adjustment for disequilibria below a certain threshold. The estimated threshold is 17.4% for the equilibrium between prices in United States and Argentina and 12.1% between prices in Brazil and Argentina. This result is consistent with the presence of transaction costs, putting higher costs between more distant markets.

References

- Apostol, T. (1967). *Calculus*. Wiley, New York, 2 edition.
- Balcombe, N., Bailey, A., & Brooks, T. (2007). Threshold effects in price transmission: The case of Brazilian wheat, maize and soya prices. *American Journal of Agricultural Economics*, 89:308–323.
- Bec, F. & Rahbek, A. (2004). Vector equilibrium correction models with non-linear discontinuous adjustments. *Econometrics Journal*, 7:628–651.
- Chung, H., Ho, T.-W., & Wei, L.-J. (2005). The dynamic relationship between the prices of ADRs and their underlying stocks: Evidence from the threshold vector error correction model. *Applied Economics*, 37:2387–2394.
- Davies, R. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74:33–43.
- Engle, R. & Granger, C. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55:251–276.
- Gonzalo, J. & Pitarakis, J.-Y. (2006). *Threshold Effects in Multivariate Error Correction Models*. Palgrave Handbook of Econometrics.
- Granger, C. & Teräsvirta, T. (2006). *Modelling Nonlinear Dynamic Relationships*. Oxford University Press, Oxford.
- Hansen, B. & Seo, B. (2002). Testing for two-regime threshold cointegration in vector error-correction models. *Journal of Econometrics*, 110:293–318.
- Ibragimov, R. & Phillips, P. (2008). Regression asymptotics using Martingale convergence methods. *Econometric Theory*, 24:888–947.

- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59:1551–1580.
- Kapetanios, G., Shin, Y., & Snell, A. (2006). Testing for cointegration in nonlinear smooth transition error correction models. *Econometric Theory*, 22:279–303.
- Kristensen, D. & Rahbek, A. (2009). Likelihood-based inference in nonlinear error-correction models. *Journal of Econometrics*, 158:78–94.
- Lo, M. & Zivot, E. (2001). Threshold cointegration and nonlinear adjustment to the law of one price. *Macroeconomic Dynamics*, 5:533–576.
- Luukkonen, R., Saikkonen, P., & Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75:491–499.
- Michael, P., Nobay, A., & Peel, D. (1997). Transaction costs and nonlinear adjustment in real exchange rates: An empirical investigation. *Journal of Political Economy*, 105:862–879.
- Saikkonen, P. (2005). Stability results for nonlinear error correction models. *Journal of Econometrics*, 127:69–81.
- Saikkonen, P. (2008). Stability of regime switching error correction models under linear cointegration. *Econometric Theory*, 24:294–318.
- Saikkonen, P. & Choi, I. (2004). Cointegrating smooth transition regression. *Econometric Theory*, 20:301–340.
- Seo, B. (2007). Testing for smooth transition nonlinearity in partially nonstationary vector autoregressions. *Journal of the Korean Statistical Society*, 36:257–274.
- Seo, M. H. (2006). Bootstrap testing for the null of no cointegration in a threshold vector error correction model. *Journal of Econometrics*, 134:129–150.
- Suárez-Fariñas, M., Pedreira, C., & Medeiros, M. (2004). Local global neural networks: A new approach for nonlinear time series modeling. *Journal of the American Statistical Association*, 99:1092–1107.
- Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89:208–218.
- van Dijk, D., Teräsvirta, T., & Franses, P. (2002). Smooth transition autoregressive models – A survey of recent developments. *Econometric Reviews*, 21:1–47.
- Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge.

A. Appendix

The proofs to Propositions 3.1 and 4.1 are in the first section, while the proofs of all lemmas used are in the second section of this Appendix.

B. Propositions

Proof. (Proposition 3.1) Suppose \mathcal{H}_0 is true. Without loss of generality and or the sake of clarity, consider a system of two variables, only one lag of only one of the variables and no constant. Under \mathcal{H}_0 and the knowledge of β all the regressors are stationary. Therefore, the asymptotic distributions would be standard. We will show that using super-consistent $\hat{\beta}$ yields the same limiting distributions. Consider the first equation of the system:

$$\Delta y_{1t} = \theta_{11}(\hat{z}_{t-1}) + \theta_{21}(\hat{z}_{t-1})^2 + \theta_{31}(\hat{z}_{t-1})^3 + \gamma \Delta y_{1t-1} + \tilde{\epsilon}_{1t}, \quad (9)$$

where $z_t = \beta' \mathbf{y}_{t-1}$, $\hat{z}_t = \hat{\beta}' \mathbf{y}_{t-1}$ and $\tilde{\epsilon}_{1t} = \epsilon_{1t} + \theta_{11}(\hat{\beta}_2 - \beta_2)y_{2t-1} + \theta_{21}[2z_{t-1}(\hat{\beta}_2 - \beta_2)y_{2t-1} + 3(\hat{\beta}_2 - \beta_2)^2 y_{2t-1}^2] + \theta_{31}[3z_{t-1}^2(\hat{\beta}_2 - \beta_2)y_{2t-1} - 3z_{t-1}(\hat{\beta}_2 - \beta_2)^2 y_{2t-1}^2 + 5(\hat{\beta}_2 - \beta_2)^3 y_{2t-1}^3]$. We will prove the result for this equation, the extension to both equations being straightforward but involving much longer and tedious manipulations.

First, note that

$$\sqrt{T} \begin{pmatrix} \widehat{\gamma}_1 - \gamma_1 \\ \widehat{\theta}_{11} - \theta_{11} \\ \widehat{\theta}_{21} - \theta_{21} \\ \widehat{\theta}_{31} - \theta_{31} \end{pmatrix} = T \begin{pmatrix} \sum_{t=1}^T \Delta y_{1t-1}^2 & \sum_{t=1}^T \Delta y_{1t-1} \widehat{z}_{t-1} & \sum_{t=1}^T \Delta y_{1t-1} \widehat{z}_{t-1}^2 & \sum_{t=1}^T \Delta y_{1t-1} \widehat{z}_{t-1}^3 \\ \sum_{t=1}^T \Delta y_{1t-1} \widehat{z}_{t-1} & \sum_{t=1}^T \widehat{z}_{t-1}^2 & \sum_{t=1}^T \widehat{z}_{t-1}^3 & \sum_{t=1}^T \widehat{z}_{t-1}^4 \\ \sum_{t=1}^T \Delta y_{1t-1} \widehat{z}_{t-1}^2 & \sum_{t=1}^T \widehat{z}_{t-1}^3 & \sum_{t=1}^T \widehat{z}_{t-1}^4 & \sum_{t=1}^T \widehat{z}_{t-1}^5 \\ \sum_{t=1}^T \Delta y_{1t-1} \widehat{z}_{t-1}^3 & \sum_{t=1}^T \widehat{z}_{t-1}^4 & \sum_{t=1}^T \widehat{z}_{t-1}^5 & \sum_{t=1}^T \widehat{z}_{t-1}^6 \end{pmatrix}^{-1} \times \frac{1}{\sqrt{T}} \begin{pmatrix} \sum_{t=1}^T \Delta y_{1t-1} \widetilde{\epsilon}_t \\ \sum_{t=1}^T \widehat{z}_{t-1} \widetilde{\epsilon}_t \\ \sum_{t=1}^T \widehat{z}_{t-1}^2 \widetilde{\epsilon}_t \\ \sum_{t=1}^T \widehat{z}_{t-1}^3 \widetilde{\epsilon}_t \end{pmatrix}.$$

Hence, to establish our result, it suffices to show that:

- (a) $\text{plim } T^{-1} \sum_{t=1}^T \Delta y_{1t-1} \widehat{z}_{t-1}^l = \text{plim } T^{-1} \sum_{t=1}^T \Delta y_{1t-1} z_{t-1}^l, \forall l = 1, 2, 3.$
- (b) $\text{plim } T^{-1} \sum_{t=1}^T \widehat{z}_{t-1}^k = \text{plim } T^{-1} \sum_{t=1}^T z_{t-1}^k, \forall k = 2, 3, 4, 5, 6.$
- (c) $T^{-1/2} \sum_{t=1}^T \Delta y_{1t-1} \widetilde{\epsilon}_t$ has the same asymptotic distribution of $T^{-1/2} \sum_{t=1}^T \Delta y_{1t-1} \epsilon_t.$
- (d) $T^{-1/2} \sum_{t=1}^T \widehat{z}_{t-1}^h \widetilde{\epsilon}_t$ has the same asymptotic distribution of $T^{-1/2} \sum_{t=1}^T z_{t-1}^h \epsilon_t$ for $h=1,2,3.$

(a) and (b) follow directly from Lemma C.4.

To prove (d), note that the expression $\widehat{z}_{t-1}^2 \widetilde{\epsilon}_{1t}$ equals $\widehat{z}_{t-1}^2 \epsilon_{1t}$ plus a number of terms in the form

$$(\widehat{\beta}_2 - \beta_2)^i \sum_{t=1}^T \widehat{z}_{t-1}^k y_{2t-1}^i.$$

Following what was shown in Lemma C.4, as long as $i \geq 1$ and $k > 0$, the expression is $o_p(1)$, i.e., the limit when $T \rightarrow \infty$ is zero. If $k = 0$, we need $i \geq 2$. In this case, the inequality is respected in all expressions.

Hence, $T^{-1/2} \sum_{t=1}^T \widehat{z}_{t-1}^h \widetilde{\epsilon}_{1t} = T^{-1/2} \sum_{t=1}^T \widehat{z}_{t-1}^h \epsilon_{1t} + o_p(1)$. Again from Lemma C.4, $T^{-1/2} \sum_{t=1}^T \widehat{z}_{t-1}^h \epsilon_{1t} = T^{-1/2} \sum_{t=1}^T \widehat{z}_{t-1}^h \epsilon_{1t} + o_p(1)$, and from here the result follows.

Proof of claim (c) is analogous to the proof to claim (d).

Now, suppose \mathcal{H}_A is true. To prove the consistency we will show that, under \mathcal{H}_A , the F-statistic diverges to infinity. Under the alternative, Δy_{1t} follows Equation (9) except for the error, which becomes $\widetilde{\epsilon}_{1t}^* = \widetilde{\epsilon}_{1t} + \frac{1}{6} f^{(4)}(k_t, \psi)(z_{t-1})^4$ for some fixed $k_t \in \mathbb{R}$.

Let $\mathbf{Z}_t = (\widehat{z}_{t-1}, \widehat{z}_{t-1}^2, \widehat{z}_{t-1}^3, \Delta y_{1t-1})'$ and $\mathbf{Z}_T = (\mathbf{Z}_1, \dots, \mathbf{Z}_T)'$. Then, $\text{plim} (\frac{1}{T} \mathbf{Z}_T \mathbf{Z}_T')^{-1} = \mathbf{\Omega}$ is unchanged whether \mathcal{H}_0 is true or not. We will show that under \mathcal{H}_A , $T^{-1/2}(\widehat{\theta}_{21} - 0)$ diverges. Let $\widetilde{\mathbf{\Omega}}$ be the relevant partition of $\mathbf{\Omega}$.

Note that

$$\begin{aligned} T^{-1/2} \widehat{\theta}_{21} &= \widetilde{\mathbf{\Omega}} T^{-1/2} \sum_{t=1}^T \widehat{z}_{t-1}^2 \Delta y_{1t} \\ &= T^{1/2} \theta_{21} + \widetilde{\mathbf{\Omega}} T^{-1/2} \sum_{t=1}^T \widehat{z}_{t-1}^2 \widetilde{\epsilon}_{1t}^* \\ &= T^{1/2} \theta_{21} + \widetilde{\mathbf{\Omega}} T^{-1/2} \sum_{t=1}^T \left[\widehat{z}_{t-1}^2 \widetilde{\epsilon}_{1t} + \widehat{z}_{t-1}^2 \frac{1}{6} f^{(4)}(k_t, \psi)(z_{t-1})^4 \right]. \end{aligned}$$

We know, from Assumptions 1 and 5, that $\frac{1}{6} f^{(4)}(k_t, \psi)$ is bounded. So, we can write

$$-K \sum_{t=1}^T \widehat{z}_{t-1}^2 (z_{t-1})^4 < \sum_{t=1}^T \widehat{z}_{t-1}^2 \frac{1}{6} f^{(4)}(k_t, \psi)(z_{t-1})^4 < K \sum_{t=1}^T \widehat{z}_{t-1}^2 (z_{t-1})^4.$$

Pre-multiplying by $T^{-1/2}$, taking limits, and using the results in Lemma C.4, we get

$$O_p(1) < \sum_{t=1}^T \widehat{z}_{t-1}^2 \frac{1}{6} f^{(4)}(k_t, \psi)(z_{t-1})^4 < O_p(1).$$

From (d) we know that $T^{-1/2} \sum_{t=1}^T \widehat{z}_{t-1}^2 \widetilde{\epsilon}_{1t}$ is $O_p(1)$. Therefore, we have two limited terms plus $T^{1/2}\theta_{21}$, which will diverge to ∞ , giving us the result. The same argument applies to the F-test, only with lengthier calculations.

Finally, from (b) it is easy to see that $\text{plim}T^{-1}(\widehat{\epsilon}_{1t}^2)$ exists.

■

Proof. (Proposition 4.1) Again, for the sake of simplicity, let us consider only one lag of only one variable. In addition, without loss of generality, we will assume ψ is scalar. The NLLS problem is

$$\min \frac{1}{2} T^{-1} \sum_{t=1}^T \left[\Delta y_{1t} - f(\widehat{\beta}' \mathbf{y}_{t-1}, \psi) - \gamma \Delta y_{1t-1} \right]^2.$$

The first order conditions are:

$$T^{-1} \sum_{t=1}^T \widehat{\mathbf{s}}_t(\widehat{\psi}, \widehat{\gamma}) = 0,$$

where

$$\widehat{\mathbf{s}}_t(\widehat{\psi}, \widehat{\gamma}) = \left[\Delta y_{1t} - f(\widehat{\beta}' \mathbf{y}_{t-1}, \widehat{\psi}) - \widehat{\gamma} \Delta y_{1t-1} \right] \begin{bmatrix} f_{\psi}(\widehat{\beta}' \mathbf{y}_{t-1}, \widehat{\psi}) \\ \Delta y_{1t-1} \end{bmatrix}.$$

We will always use the hat to make clear whether the function is calculated with $\widehat{\beta}' \mathbf{y}_{t-1}$ or $\beta' \mathbf{y}_{t-1}$.

We can make a mean-value expansion around (ψ, γ) :

$$\sum_{t=1}^T \widehat{\mathbf{s}}_t(\psi, \gamma) + \sum_{t=1}^T \widehat{\mathbf{H}}_t(\widetilde{\psi}, \widetilde{\gamma}) \begin{pmatrix} \psi - \widehat{\psi} \\ \gamma - \widehat{\gamma} \end{pmatrix} = 0,$$

where

$$\widehat{\mathbf{H}}_t(\psi, \gamma) = \begin{bmatrix} \frac{\partial \widehat{\mathbf{s}}_t(\psi, \gamma)}{\partial \psi} \\ \frac{\partial \widehat{\mathbf{s}}_t(\psi, \gamma)}{\partial \gamma} \end{bmatrix}',$$

$(\widetilde{\psi}, \widetilde{\gamma}) = (\widehat{\psi}, \widehat{\gamma}) + t(\psi, \gamma)$, for some $t \in (0, 1)$.

From Lemma C.5

$$\text{plim}T^{-1} \sum_{t=1}^T \widehat{\mathbf{H}}_t(\widetilde{\psi}, \widetilde{\gamma}) = \text{plim}T^{-1} \sum_{t=1}^T \mathbf{H}_t(\psi, \gamma) = \mathbf{H}(\psi, \gamma),$$

for some fixed $\mathbf{H}(\psi, \gamma)$.

Therefore,

$$T^{1/2} \begin{pmatrix} \psi - \hat{\psi} \\ \gamma - \hat{\gamma} \end{pmatrix} = [\mathbf{H}_t(\psi, \gamma)]^{-1} \left(-T^{-1/2} \right) \sum_{t=1}^T \hat{\mathbf{s}}_t(\psi, \gamma) + o_p(1).$$

All we have to show now is that

$$(T^{-1/2}) \sum_{t=1}^T \hat{\mathbf{s}}_t(\psi, \gamma) = (T^{-1/2}) \sum_{t=1}^T \mathbf{s}_t(\psi, \gamma) + o_p(1).$$

It is sufficient to show that

- (a) $T^{-1/2} \sum_{t=1}^T \Delta y_{1t} f_{\psi}(\hat{\boldsymbol{\beta}}' \mathbf{y}_{t-1}, \hat{\psi}) = T^{-1/2} \sum_{t=1}^T \Delta y_{1t} f_{\psi}(\boldsymbol{\beta}' \mathbf{y}_{t-1}, \psi) + o_p(1)$
- (b) $T^{-1/2} \sum_{t=1}^T f(\hat{\boldsymbol{\beta}}' \mathbf{y}_{t-1}, \hat{\psi}) f_{\psi}(\hat{\boldsymbol{\beta}}' \mathbf{y}_{t-1}, \hat{\psi}) = T^{-1/2} \sum_{t=1}^T f(\boldsymbol{\beta}' \mathbf{y}_{t-1}, \psi) f_{\psi}(\boldsymbol{\beta}' \mathbf{y}_{t-1}, \psi) + o_p(1)$

Claims (a) and (b) follow directly from Lemma C.4 and Assumption 5.

As to the covariance matrix estimator, the proof is standard. Since we only need to use the Law of Large Numbers, the non-stationarity of the variables does not bring any extra complications Wooldridge (2001).

■

C. Lemmas

Lemma C.1 *Suppose f is a function which is n times continuously differentiable on the closed interval $[a - r, a + r]$ and $n + 1$ times differentiable on the open interval $(a-r, a+r)$. If there exists a positive real constant M_n such that $|f^{(n+1)}(x)| < M_n, \forall x \in (a - r, a + r)$, then*

$$f(x) = f(a) + f'(a) \frac{(x - a)}{1!} + f''(a) \frac{(x - a)^2}{2!} + \dots + f^{(n)}(b) \frac{(x - a)^n}{n!}$$

for some $b \in (a, x)$.

Proof. See Apostol (1967).

■

From Ibragimov and Phillips (2008):

Theorem C.2 *Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a twice continuously differentiable function such that f' satisfies the growth condition $|f'(x)| \leq K(1 + |x|^\alpha), \forall x \in \mathbf{R}$ for some constants $K > 0$ and $\alpha < 0$. Suppose that u_t and v_t are two linear processes $u_t = \sum_{j=1}^{\infty} \gamma_j \epsilon_{t-j}$ and $v_t = \sum_{j=1}^{\infty} \delta_j \epsilon_{t-j}$ where $\sum_{j=1}^{\infty} j |\gamma_j| < \infty, \sum_{j=1}^{\infty} j |\delta_j| < \infty$ and*

$(\epsilon_t)_{t \in \mathbf{Z}}$ are zero-mean i.i.d. random variables with $E[\epsilon_0^2] < \infty$ and $E[|\epsilon_0|^p] < \infty$ for $p \geq \max(6, 4\alpha)$. Then

$$\frac{1}{\sqrt{T}} \sum_{t=2}^{[Tr]} f \left(\frac{1}{\sqrt{T}} \sum_{i=1}^{t-1} u_i \right) v_t \xrightarrow{d} \lambda_{uv} \int_0^r f'(\omega_u W(v)) dv + \omega_v \int_0^r f(\omega_u W(v)) d(W(v)),$$

where $\omega_u = E[u_t^2]$, $\omega_v = E[v_t^2]$ and $\lambda_{uv} = \sum_{j=1}^{\infty} E[u_0 v_0]$.

The exact form of the limiting distribution is not relevant for our results. What we need is the following corollary.

Lemma C.3 Under the conditions of Theorem C.2,

$$T^{-1/2} \sum_{t=2}^T f \left(T^{-1/2} \sum_{i=1}^{t-1} u_i \right) v_t = O_p(1)$$

Note that the derivatives of any polynomial function satisfy the growth condition.

Lemma C.4 Let v_t be a stationary process, \mathbf{y}_t be an $I(1)$ cointegrated vector, with cointegration vector β and $\hat{\beta}$ a super-consistent estimate of β . Let also, for some $d < \infty$, $f : \mathbb{R} \rightarrow \mathbb{R}$ be d times continuously differentiable and the d -th derivative of f be limited. Then, $T^{-1/2} \sum_{t=1}^T f(\beta' \mathbf{y}_t) v_t = T^{-1/2} \sum_{t=1}^T f(\hat{\beta}' \mathbf{y}_t) v_t + o_p(1)$, and $\text{plim } T^{-1} \sum_{t=1}^T f(\beta' \mathbf{y}_t) = \text{plim } T^{-1} \sum_{t=1}^T f(\hat{\beta}' \mathbf{y}_t)$.

Proof. For the first result, consider first a two dimensional case $\beta' \mathbf{y}_t = y_{1t} + \beta_2 y_{2t}$. Using Lemma C.1 to expand f around $\beta_2 y_{2t}$,

$$f(\hat{\beta}' \mathbf{y}_t) = f(\beta' \mathbf{y}_t) + f'(\beta' \mathbf{y}_t) (\hat{\beta}_2 - \beta_2) y_{2t} + \dots + \frac{f^{(d-1)}(\beta' \mathbf{y}_t) (\hat{\beta}_2 - \beta_2)^{d-1} y_{2t}^{d-1}}{(d-1)!} + \frac{f^d(\widetilde{\beta}' \mathbf{y}_t) (\hat{\beta}_2 - \beta_2)^d y_{2t}^d}{d!}$$

for some $\widetilde{\beta}' \mathbf{y}_t \in (\hat{\beta}' \mathbf{y}_t, \beta' \mathbf{y}_t)$. Taking the k -th term, such that $3 \leq k \leq d-1$, we have, by Lemma C.3,

$$\begin{aligned} (\hat{\beta}_2 - \beta_2)^k \sum_{t=1}^T f^k(\beta' \mathbf{y}_t) v_t y_{2t}^k / k! &= T^{-(k-1)/2} \left[T(\hat{\beta}_2 - \beta_2) \right]^k \\ &\quad \times \left[T^{-(k+1)/2} \sum_{t=1}^T \frac{f^k(\beta' \mathbf{y}_t) v_t y_{2t}^k}{k!} \right] \\ &= T^{-(k-1)/2} O_p(1) O_p(1) = o_p(1). \end{aligned}$$

For $k = 2$ we get a $O_p(1)$, but it will be further divided by $T^{1/2}$, giving us an $o_p(1)$. Since the d -th derivative is limited, for some $M \in \mathbb{R}$, the sum of the d -th term is bounded by

$$\begin{aligned} \pm M \sum_{t=1}^T (\widehat{\beta}_2 - \beta_2)^d v_t y_{2t}^d &= \pm M T^{-(d-1)/2} \left[T(\widehat{\beta}_2 - \beta_2) \right]^d \left[T^{-(d+1)/2} \sum_{t=1}^T v_t y_{2t}^d \right] \\ &= T^{-(d-1)/2} O_p(1) O_p(1). \end{aligned}$$

Again, if $d \neq 1$ we have an $o_p(1)$ expression, if $d = 1$, we are back to the $k = 2$ case. Therefore, the only remaining term is the first, which gives us the result. For the multidimensional case, just repeat the reasoning for each dimension of β' .

The second result is proven by the same line of reasoning. The only difference is that in the end of each expression we will have $T^{-(i-2)/2} O_p(1) O_p(1)$, giving us an $o_p(1)$ except for $i = 1, 2$. But since the expression will be divided by T , we will have $o_p(1)$ for every i . ■

Lemma C.5 $\text{plim } T^{-1} \sum_{t=1}^T \widehat{\mathbf{H}}_t(\tilde{\psi}, \tilde{\gamma}) = \text{plim } T^{-1} \sum_{t=1}^T \mathbf{H}_t(\psi, \gamma)$.

Proof. $\widehat{\mathbf{H}}_t(\tilde{\psi}, \tilde{\gamma})$ equals

$$\begin{bmatrix} L & \Delta y_{1t-1} f_{\psi}(\widehat{\beta}' \mathbf{y}_{t-1}, \tilde{\psi}) \\ \Delta y_{1t-1} f_{\psi}(\widehat{\beta}' \mathbf{y}_{t-1}, \tilde{\psi}) & \Delta y_{1t-1}^2 \end{bmatrix}$$

where $L = f_{\psi\psi}(\widehat{\beta}' \mathbf{y}_{t-1}, \tilde{\psi})(-\Delta \mathbf{y}_{1t} + f(\widehat{\beta}' \mathbf{y}_{t-1}, \tilde{\psi}) + \tilde{\gamma} \Delta y_{1t-1}) + f_{\psi}(\widehat{\beta}' \mathbf{y}_{t-1}, \tilde{\psi})^2$. From Lemma C.4 and Assumption 5:

- (a) $\text{plim } T^{-1} \sum_{t=1}^T \Delta y_{1t-1} f_{\psi}(\widehat{\beta}' \mathbf{y}_{t-1}, \tilde{\psi}) = \text{plim } T^{-1} \sum_{t=1}^T \Delta y_{1t-1} f_{\psi}(\beta' \mathbf{y}_{t-1}, \tilde{\psi})$
- (b) $\text{plim } T^{-1} \sum_{t=1}^T f_{\psi\psi}(\widehat{\beta}' \mathbf{y}_{t-1}, \tilde{\psi}) \left[-\Delta y_{1t} + f(\widehat{\beta}' \mathbf{y}_{t-1}, \tilde{\psi}) + \tilde{\gamma} \Delta y_{1t-1} \right] =$
 $\text{plim } T^{-1} \sum_{t=1}^T f_{\psi\psi}(\beta' \mathbf{y}_{t-1}, \tilde{\psi}) \left[-\Delta y_{1t} + f(\beta' \mathbf{y}_{t-1}, \tilde{\psi}) + \tilde{\gamma} \Delta y_{1t-1} \right]$
- (c) $\text{plim } T^{-1} \sum_{t=1}^T f_{\psi}(\widehat{\beta}' \mathbf{y}_{t-1}, \tilde{\psi})^2 = \text{plim } T^{-1} \sum_{t=1}^T f_{\psi}(\beta' \mathbf{y}_{t-1}, \tilde{\psi})^2$

Therefore, we have established that $\text{plim } T^{-1} \sum_{t=1}^T \widehat{\mathbf{H}}_t(\tilde{\psi}, \tilde{\gamma}) = \text{plim } T^{-1} \sum_{t=1}^T \mathbf{H}_t(\tilde{\psi}, \tilde{\gamma})$. Usual nonlinear least squares approach, c.f. Wooldridge (2001), may easily be used to establish $\text{plim } T^{-1} \sum_{t=1}^T \mathbf{H}_t(\tilde{\psi}, \tilde{\gamma}) = \text{plim } T^{-1} \sum_{t=1}^T \mathbf{H}_t(\psi, \gamma)$. These two equalities give us the result. ■