Journal of Business & Economics
Vol. 4 No.2 (July-December 2012) pp.177-199

A PLS Based Approach to Cointegration Analysis

Nelson Muriel*
Graciela González-Farías **
Rogelio Ramos***

Abstract

This paper addresses the testing for cointegrating vectors and the estimation of cointegrating relations using Partial Least Squares. Together with Harris (1997) and Bossaerts (1988), the PLS approach relies on a method of multivariate statistics and thus does not require identifying restrictions on the cointegrating vectors or of a full specification of the short-run dynamics of the process. The PLS estimator for the cointegrating vectors is found to be super consistent and robust to heavy-tailed innovations. A test is provided for the rank of cointegration which is assessed by means of Monte Carlo simulation. A brief application to Mexican inflation data is also provided.

Key Words: Partial Least Squares, non-stationarity, Cointegration analysis, Asymptotics.

1. Introduction

The procedures for estimation and testing of cointegrating vectors, that have been widely studied since the concept was first laid out in Granger (1983), are usually thought of as falling into one of the two categories. The first category requires the cointegrating vectors to be fully identified and so

* Nelson Muriel, Post doc Assistant Professor, Centro de Investigación en Matemáticas (CIMAT), Department of Probability and Statistics, Guanajuato, México. Corresponding author E-mail: nelson.muriel@cimat.mx
** Graciela González-Farías, Full Professor Centro de Investigación en Matemáticas (CIMAT), Department of Probability and Statistics, Guanajuato, México.
*** Rogelio Ramos, Associate Professor Centro de Investigación en Matemáticas (CIMAT), Department of Probability and Statistics, Guanajuato, México.

Acknowledgement: This research has been partially supported by CONACYT Ciencia Básica No. 105657 and ECO 2010-19357, Spain

is founded upon parametric restrictions. The other category does not impose identifying restrictions and deals instead with estimating a basis for the cointegrating space leaving the test for any particular restriction to be a subsequent step in the analysis. Ordinary Least Squares (OLS) regression and Nonlinear Least Squares is representative of the first type of procedure, whereas Johansen's approach to maximum likelihood in a fully specified error correction model is the most prominent in the second category.

The main advantage of the first type of tests is that the short–run dynamics of the process do not need to be fully parameterized, so that a great deal of data generating processes fall within their reach. On the other hand, the use of identifying restrictions from the start may be disadvantageous in some scenarios. In an attempt to test for cointegration without either imposing identifying restrictions or fully specifying the short–run dynamics of the process, Harris (1997) proposes the use of Principal Component Analysis (PCA) and the properties of these estimators are further studied by Snell (1999).

PCA may be thought as procedure alongside that of Bossaerts (1988), based on Canonical Correlation Analysis (CCA), as yet a third category in testing of cointegration. This category is based on methods of multivariate analysis and as such relies either on a decomposition of the column space of the data or on some sort of variance decomposition for it. More explicitly, express X_t for the cointegrated process with a Vector Error Correction Model (VECM) representation

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{j=1}^{k} \Gamma_j X_{t-j} + \varepsilon_t,$$

where $X_t \varepsilon R^p$, the matrices $\alpha \varepsilon M(p\times r)$and $\beta \varepsilon M(p\times r)$are of full rank and $\beta\perp$ is the orthogonal complement of β. As is well known, there are two dominant directions to the process, namely $\beta' X_t$, the stationary direction and $\beta\perp, Xt$, the nonstationary direction. PCA of Cointegrating relations as explained in Harris (1997) or Snell (1999) depend directly on these two directions and works because in the space generated by the stochastic trends, the process has a "higher" variance (roughly of the same order as the sample size) whereas in the space generated by the stationary process $ß'X_t$ variances are significantly smaller. On the other hand, CCA of Cointegrating relations as explained in Bossaerts (1988) rests on the fact that the components $ß'X_t$ and $ß_{\perp},Xt$ are orthogonal and correlation is maximized in the direction of $ß_{\perp},Xt$.

This paper examines the performance of another method of multivariate statistics to perform cointegration analysis, namely Partial Least Squares (PLS) which has become a very useful tool in several research fields such as psychology, chemistry, medicine or economics to mention only a few. Rather than standing for a specific method, PLS comprises a whole class of techniques whose aim is to model the association between two given blocks of observed variables by recursively constructing a set of latent variables (orthogonal score vectors). PLS is a nonparametric approach to regression and serves very well the tasks of classification and reduction of dimensionality. A survey of diverse PLS methods can be found in Wegelin (2000), where some interesting comparisons are given among the different techniques. The reader is also referred to Rosipal and Krämer (2006) for an overview of PLS.

The motivation to use PLS for cointegration analysis is twofold. On the one hand, we know that as a method in multivariate statistics, PLS is akin to CCA. Actually, Wegelin (2000) argues that CCA is but a way of performing PLS, so that we can read PLS as a generalization of CCA. On the other hand, decomposition of the space induced by the structure of cointegrating system as described earlier and the proven success of PCA and CCA suggest that in looking for latent variables (score vectors), PLS will be forced into estimating the stochastic trends first (since the latent variables $ß_{\perp},Xt$ drive the process in the directions of higher variance) and then the cointegrating relations (the relevant latent variables orthogonal to the ones already estimated).

PLS begins with a data matrix X, of dimensions $T\times p$, where T stands for the number of observations and p for the dimensionality of each observation which will be used to either explain or predict the variables contained in another matrix Y of dimensions $T\times p$. Almost needless to say, the aim is to relate the p variables contained in X with the k variables stored inY. Writing

X and *Y* for such variables, the CCA approach is to find μ a *p*–dimensional vector and v a *k*–dimensional one, such as to maximize the correlation between the transformed variables *u'X* and *v'Y*. We state the convention that *u*' stands for transposition in the remaining sections of the paper. In compact terms, we can say that the objective function of CCA is

$$\max(Corr(u'X, v'Y))$$

Alternatively, PLS will attempt to maximize the covariance, rather than the correlation, with the additional restriction of using only unit vectors to transform the data, that is, find *u* and *v* such that $|u||=||v||=1$ and for which

$$\max(Cov(u'X, v'Y))$$

is attainted in some sense[1]. The sense in which this happens is precisely the division line among the many algorithms included within PLS. We refer to Braak and De Jong (1998) for a presentation of the objective functions related to PLS. [2]

One interesting aspect of this attempt is that PLS is more of a structural than a parametric approach, i.e., it rests much more heavily on there being an implicit decomposition of the column space of data than on the precise parametric structure of the VECM or the VAR process itself.

By not imposing identifying restrictions and by not being based on the explicit and fully specified VECM representation, the PLS approach to cointegration analysis may be useful for a wide variety of data generating processes and may not need extra–hypothesis of normally distributed disturbances or even square–integrable ones. Thus, PLS is likely to provide a rather flexible tool for cointegration analysis.

The paper is structured as follows. In Section 2 we introduce the PLS algorithm and state some of its main properties. Section 3 is devoted to relate

[1] We now intuitively see that working with centered and standardized data, PLS is just CCA.
[2] PLS has also been related to regression analysis and Garthwaite (1994) is an excellent introduction to this subject matter.

the workings of PLS with the structure of cointegrating systems. We prove that PLS produces super consistent estimators of (an orthonormal basis for) the stochastic trends and the cointegrating space. The decomposition provided by PLS is used in section 4 to test for cointegration taking advantage of the fact that PLS will always estimate the *I(1)* processes $\beta_{\perp}X_t$ before estimating the basis for the cointegrating space. A procedure based on PLS and KPSS is studied with Monte Carlo simulations and its particular advantages are shown. Finally, in section 5 we apply the PLS-KPSS test to four variables relevant for forecasting inflation, namely the consumer price index, the monetary base, the equilibrium interest rate at 28 days and the industrial production index. We use Mexican data in a period containing T = 144 observations and find one cointegrating relation.

2. The PLS Algorithm

As mentioned earlier, PLS stands for a handful of methods or algorithms. In the present paper, we will be dealing with what has come to be known as PLS2, see Wegelin (2000). We present the algorithm as is given by Höskuldsson (1998): Before the algorithm starts the matrices X and Y may be centered or scaled which suggests subtracting mean values or working with correlations. The algorithm is as follows

Step 1: Set μ to the first column of Y,
Step 2: $w = X'u/(u'u)$
Step 3: Scale w to be of length one
Step 4: $t = X'w$
Step 5: $c = Y't/(t't)$
Step 6: Scale c to be of length one
Step 7: $u=Y\, c/(c'c)$
If convergence then 8, else 2
Step 8: X–loadings: $p = X't/(t't)$; Y-loadings $q = Y'u/(u'u)$
Step 9: Regression (u upon t): $b = u't/(u'u)$
Step 10: Residual matrices: $X \rightarrow X - tp'$ *and* $Y \rightarrow Y\text{-}btc'$

We then start over iteratively with these new X and Y matrices which result from the previous iteration. Iterations may continue until a stopping criterion is met or X becomes the zero matrix. As shown in Höskuldsson (1998), in convergence, the vectors w and c satisfy

$$X'YY'Xw = \lambda_1 w$$
$$X'YY'Xc = \lambda_1 c \qquad (1)$$

where λ_1 is the greatest eigenvalue of the matrix Y'XX'Y. Furthermore, the score vectors t_i are mutually orthogonal and the pair (w_i, c_i) satisfies the maximization

$$(\mathrm{Cov}(t, u))^2 = \max(\mathrm{Cov}(Xd, Ye))^2,\ \|d\| = \|e\| = 1 \qquad (2)$$

The components t and μ may be interpreted as orthogonal components in the X and Y space respectively that have maximal covariance among all components in those spaces at each iteration.

3. Estimating the Cointegrating Relations with PLS

We now consider a cointegrated system in R^p given in VECM representation as

$$\Delta X_t = \alpha \beta' X_{t-1} + \sum_{i=1}^{k} \Gamma_i \Delta X_{t-i} + \varepsilon_t \qquad (3)$$

and assume that ε_t is a white noise sequence with mean 0 and covariance matrix Ω. The matrices α and β are $p \times r$ of full rank and the cointegrating relations are given by β so that $\beta' X_t$ is an $I(0)$ process. Together with these matrices, we consider $\alpha_\perp$ and $\beta_\perp$ orthogonal complements so that $\beta' \beta_\perp = \alpha' \alpha_\perp = 0$ and assume that

$$\left| -\alpha_{\perp'} \left(-I_p - \alpha\beta' + \sum_{i=1}^{k} \Gamma_i \right) \beta_\perp \right| \neq 0$$

As is proven in Johansen (1995) among others, this implies that X_t admits the Granger representation

$$X_t = C\sum_{i=1}^{t}\varepsilon_i + C_1(L)\varepsilon_t. \tag{4}$$

The process $C_1(L)\varepsilon_t$ is $I(0)$ and the the matrix C is given by

$$C = \beta_{\perp}\left(\alpha_{\perp'}\left(-I_p - \alpha\beta' + \sum_{i=1}^{k}\Gamma_i\right)\beta_{\perp}\right)^{-1}\alpha_{\perp'}$$

and is of rank $p\text{-}r$. Furthermore, it is also known that the matrix $(\beta_{\perp}, \beta)$ is a $p \times p$ with full rank, meaning that the space of the process Xt is effectively divided into two orthogonal components, $sp(\beta_{\perp})$ and $sp(\beta)$. This allows us to write

$$X_t = P_{\beta_{\perp}} X_t + P_{\beta} X_t \tag{5}$$

where $P_{\beta} = \beta(\beta'\beta)^{-1}\beta'$ is the projection matrix on the column space of β and similarly for $P_{\beta_{\perp}}$. Equation (5) hints at us that using multivariate methods for space decomposition may be a very good way to estimate the cointegrating relations, since the very structure of the process is already decomposed. Two such approaches are the use of PCA and CCA in Stock and Watson (1988), Harris (1997), Snell (1999) and Bossaerts (1988).

A natural, but yet unexplored, extension of such results is the use of PLS to estimate the cointegrating relations. One of the advantages of doing this is that the dimensionality of the process X_t need not be smaller than the sample size for the estimation to produce meaningful results. Another advantage which we read in Höskuldsson (1998) is the stability of the predictors derived from the method.

We begin our exposition intuitively, thinking of the VECM model (3). Granger's representation (4) suggests that, in accordance with the underlying assumption of all PLS methods, the process is driven by a set of latent variables, namely the stochastic trends $\alpha_{\perp}, Xt$, and the stationary component $C_1(L)\varepsilon_t$. Representation (5) also suggests that there are two sets of latent variables: The first set drives the process in the direction of the columns of $\beta_\perp$ while the second does so in the column space of β. As a matter of fact, Johansen (1995) shows that the process $\beta_\perp' X_t$ is a possible representation for the stochastic trends in that it is an $I(1)$ process wich does not cointegrate. Observe that this partition of the space in two complementary subspaces is not only orthogonal, but also puts apart the random walk component from the stationary component which is, itself, a variance decomposition.

In order to use PLS, we define the natural data matrices as:

$$Y = (X_T', X_{T-1}', \ldots X_2') \text{ and } X = (X_{T-1}', X_{T-2}', \ldots X_1')$$

The main point to be made from the preceeding explanation here is that at any iteration of the PLS method, the vectors w and c of steps 1 through 7 of the algorithm are bound to be either in $sp(\beta_\perp)$ or in $sp(\beta)$. To work out our intuition on how PLS works with $I(1)$ variables, assume that a cointegration relation exists and that we pick w_1 and c_1 in the first iteration of PLS. If $w_1 \in sp(\beta_\perp)$ and $c_1 \in sp(\beta)$, then the covariance between $c_1' X_t$ and $w_1' X_{t-1}$ will equal zero, while if both lie on the same space the squared covariance is strictly possitive. Thus, at each step of the algorithm, PLS will pick w and c both lying either on the cointegrating space $sp(\beta)$ or on its orthogonal complement in the light of (2).

Now, the question is: Which space will PLS follow in its first step? To answer this question, observe that in the direction of $\beta_\perp$, the process X_t is a random walk. Writing η_t for the noise sequence generating it and Y_t for the projection over the span of $\beta_\perp$ of X_t, it is observed that:

$$Cov\,(u'Y_t, v'Y_{t-1}) = u'\mathrm{E}\left[\sum_{i=1}^{t}\eta_i \sum_{j=1}^{t-1}\eta_j\right]v = u'\big((t-1)\mathrm{E}\,[\eta\eta']\big)v.$$

Maximizing with the constraint that $|u||=||v||=1$ we have got u=v as the eigenvector corresponding to the maximum eigenvalue λ_1 of the matrix

E[$\eta\eta'$] and a covariance of (t-1)λ_1, which shows that this covariance increases in the same order as the sample size does. Now, in the direction of β, the process X_t is stationary, so that the variance and covariances are bounded. This means that by choosing w_1 in $sp(\beta\perp)$, and thus also c_1, the covariance is most likely to be maximized. Actually, the way to go is taking $w_1 = c_1$ which provides two perfectly correlated processes $w_1'X_{t-1}$ and $c_1'X_t$. Maximum correlation implies maximum squared covariance, so that this has to be the first PLS component. The situation is alike that of using CCA and a similar explanation was issued by Bossaerts (1988) as to how CCA selects its components.

Observe the relation established also with PCA which chooses the components with maximum covariance in decreasing order. Evidently, given there exists a stochastic trend and since its variance increases as the sample size does, PCA will select this as a first component. As a consequence, under the hypothesis of r cointegrating relations, the first p-r principal components will be random walks and the last r will provide a good estimation of the cointegrating space which is what Harris (1997) and Snell (1999) both prove by different means. Back to PLS, observe that the foregoing argument depends only on the existence of a stochastic trend. Since the PLS components are orthogonal, as is shown in Höskuldsson (1998), the space spanned by $\beta\perp$ will be estimated first with $p\text{-}r=dim(sp(\beta\perp))$ components and then the last r components will be forced to choose the vectors u and c from the cointegrating space. This parallelism with PCA is not just a coincidence as the following Lemma shows.

Lemma 3.1 *Let X_t be a cointegrated system with exactly $0<r<p$ cointegrating relations and VECM representation*

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k} \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

Then, the first PLS component coincides with the first Principal Component.

Proof. As was noted earlier, the first PLS component will correspond to u and v solving

$$\max(\mathrm{Cov}(u'X_{t-1}, v'X_t)), \text{ subject to } \|u\|=\|v\|=1$$

Let

$$\tilde{S}_{01} = \sum_{t=1}^{T} X_t X_{t-1}', \quad \tilde{S}_{10} = \sum_{t=1}^{T} X_{t-1} X_t'$$

A standard argument shows that u and v are the eigenvectors associated with the maximum eigenvalue of $\tilde{S}_{10}\tilde{S}_{01}$ and $\tilde{S}_{01}\tilde{S}_{10}$ respectively. Since the eigenvalues of $\tilde{S}_{01}$ are unchanged upon multiplication by a scalar, we may as well consider the normalized matrices $S_{01} = T^{-3/2}\tilde{S}_{01}$ and $S_{10} = T^{-3/2}\tilde{S}_{10}$. Using Granger's Representation (4), we know that X_t can be expressed as a sum of an $I(1)$ component, I_t and a stationary component S_t which implies that

$$\begin{aligned} S_{10} &= T^{-3/2}\sum_{t=1}^{T} X_{t-1}X_t' = T^{-3/2}\sum_{t=1}^{T}(I_{t-1}+S_{t-1})(I_t'+S_t') \\ &= T^{-3/2}\sum_{t=1}^{T} I_{t-1}I_t' + T^{-3/2}\left(\sum_{t=1}^{T}\left(I_{t-1}S_{t'} + S_{t-1}S_t' + S_{t-1}I_t'\right)\right) \end{aligned}$$

As shown in Appendix B.7 of Johansen (1995), the last term in the RHS converges to 0 in probability but the first term does not tend to zero and is asymptotically equivalent to $T^{-3/2}\sum X_{t-1}X_{t-1}'$. Therefore, we have shown that

$$S_{10} = S_{11} + o_p(1)$$

where $S_{11} = T^{-3/2}\sum X_{t-1}X_{t-1}'$. This immediately gives

$$S_{10}S_{01} = S_{11}S_{11}' = o_p(1)$$

Observe that S_{11} is the covariance matrix of X_{t-1} multiplied by $T^{1/2}$, which implies that the eigenvectors of S_{10} are, asymptotically, the Principal Components of X_{t-1}. To complete the proof, simply observe that the matrix S_{11} is symmetric so that its Singluar Value Decomposition can be written as $S_{11} = U\Lambda U'$ which tells us that the columns in U are eigenvectors of S_{11} S_{11}' corresponding to the eigenvalues in Λ^2.

Remark 3.1 *The former Lemma only applies to the first component of the PLS algorithm we are focusing on, since the next iterations depend on deflated data matrices. Nonetheless, the method known as Orthogonal PLS relies on the first eigenvectors of* $S_{01}S_{10}$ *which makes it equivalent to PCA. See Worsley et al. (1997) for a brief description of this method. Also, the reader is referred to Noes and Martens (1985), Helland (1988) and Helland (1990) where diverse comprisons between PLS and PCA regression are performed.*

Corollary 3.1 *Let* $\beta_{\perp,i}$ *be the columns of the matrix* $\beta_\perp$ *so that the stochastic trends are* $\beta_{\perp,i}'X_t$ *for* $i=1, p-r$. *The first PLS vector* w_1 *is a superconsistent estimator of the direction of a stochastic trend* $\beta_{\perp,i}$ *in the cointegrating system* X_t.

Proof. By Lemma 1 in Harris (1997) the first $p-r$ Principal Components are jointly superconsistent estimates of $\beta_\perp$ which together with Lemma 3.1 immediately yields the desired result.

In order to examine what comes next in PLS, we need to describe the deflated process $\tilde{X}_t$ after having undergone one PLS iteration. Lemma 3.1 shows that $w_1 = c1$ which in turn implies that

$$t_1 = w_1' X_{t-1}, \quad u_1 = w_1' X_t$$

Thus, the loadings p and q can be computed as

$$p_1 = \frac{X'Xw_1}{w'X'Xw_1} = \frac{\lambda_1 w_1}{\lambda_1} = w_1$$

and similarly $q_1 = w_1$. Thus, the new data matrices are given by

$$X^{(1)} = X - Xw_1w_1' \text{ and } Y^{(1)} = Y - Xw_1w_1'$$

In terms of the process X_t, we have the deflated processes

$$X_{t-1}^{(1)} = X_{t-1} - w_1 w_1' X_{t-1}$$

Since w_1:$\beta_{\perp,1}$ and taking into account the decomposition (5), an equivalent expression is

$$\begin{aligned} X_{t-1}^{(1)} &= X_{t-1} - w_1 w_1' \beta_\perp (\beta_\perp' \beta_\perp)^{-1} \beta_\perp' X_t \\ &= X_{t-1} - \beta_{\perp,1} \beta_{\perp,1}' \beta_\perp (\beta_\perp' \beta_\perp)^{-1} \beta_\perp' X_t \\ &= X_{t-1} - \beta_{\perp,1} \beta_{\perp,1}' X_{t-1} \\ &= X_{t-1} - P_{\beta_{\perp,1}} P_{\beta_\perp} X_{t-1} \\ &= X_{t-1} - P_{\beta_{\perp,1}} X_{t-1} \end{aligned}$$

where the normalization $\beta_{\perp,1}'\beta_{\perp,1}$ used among others by Snell (1999) has been used which entails that $\beta_{\perp,1}'\beta_{\perp,1}$ is just the projective matrix upon the space generated by $\beta_{\perp,1}$. This almost proves the following

Lemma 3.2 *Let X_t be a cointegrating system with exactly $0<r<p$ cointegrating relations and p-r stochastic trends. The deflated process $X_t^{(1)}$ obtained after one iteration of the PLS algorithm contains exactly p-r-1 stochastic trends and the same r stationary components.*

Proof. By orthogonality of the stochastic components $\beta_{\perp,i}$, i=1,2,…,p-r, we may write (5) as

$$X_t = \sum_{i=1}^{p-r} P_{\beta_{\perp,i}} X_t + P_\beta X_t$$

which, together with the former paragraph shows that

$$X_t^{(1)} = \sum_{i=2}^{p-r} P_{\beta_{\perp,i}} X_t + P_\beta X_t .$$

The first term of the expression on the RHS adds the $p-r-1$ stochastic trends and the last one includes r stationary components.

Remark 3.2 *As a consequence of the former and Lemma 3.1, the stochastic trends present in* $X_{t-1}^{(1)}$ *are the same as the ones present in* X_{t-1} *except for the one with maximum variance; and the cointegrating relations* β *are kept unchanged.*

We close this section adding up all our results in the following

Theorem 3.1 *Let* X_t *be a cointegrating system with exactly* $0<r<p$ *cointegrating relations and p-r stochastic trends. The first p-r PLS components are superconsistent estimators of an orthonormal basis for the space spanned by* $\beta_\perp$ *and the last r are superconsistent estimators for the cointegrating relations* β*.*

Proof. Assume that *p-r=1*. Then, by Lemma 3.1, the first PLS component is a superconsistent estimate of $\beta_\perp$. Suppose instead that p-r>1. Then, by Lemma 3.1, the first PLS component is a superconsistent estimator of $\beta_{\perp,1}$ and the deflated process $X_{t-1}^{(1)}$ contains *p-r-1* stochastic trends given by the direction vectors $\beta_{\perp,2}, \ldots \beta_{\perp,p-r}$ by Lemma 3.2. After another iteration of PLS, we may use again Lemma 3.1 each time obtaining a superconsistent estimate for $\beta_{\perp,i}$, *i=2,…,p-r*. Having depleted the span of $\beta_\perp$ we know from the orthogonality of the PLS components that all the following *r* components are a basis for the span of *ß*.

Remark 3.3 *In the framework of forecasting, PLS is usually thought of as a method of selecting components of* X *allowing for an accurate prediction of the variables stored in* Y. *Usually, some cross–validation or similar procedure is performed to guarantee that the chosen PLS components will account for most of the variance in* Y *and the last components will be dismissed as are only barely useful for forecasting. All the former is applicable to* $I(0)$ *variables and as has been repeatedly proven, it works quite well. Nonetheless when dealing with* $I(1)$ *variables, the first PLS components will estimate the random walk components, i.e., the non–predictable ones, which means that we should stick to the last latent variables obtainted. This seemingly paradoxical conclusion is a consequence of the structural fact that the variables to be predicted have themselves a non–predictable component, namely* $\beta_{\perp}{}'X_t$ *in the direction of which the PLS covariances are maximized.*

4. Testing for Cointegration

From the results in the previous section, it follows that given a cointegrating system X_t with r cointegrating relations, the PLS components T_i will be divided into two subgroups: $\{T_1,\ldots,T_p\}$which are all $I(1)$ and $\{T_{p-r+1},\ldots,T_p\}$ which will all be $I(0)$. Observe the ordering of the components, given that T_i is $I(1)$ it necessarily follows that T_{i-1} is also $I(1)$, that is, the estimators for the stochastic trends are not scattered, but ordered together.

Observe that the null that T_p, the last PLS component is $I(0)$ is equivalent to there being at least one cointegrating relation, whereas the alternative that T_p is $I(1)$ reduce into no cointegrating relations at all. An immediate consequence is that in order to test

$$H_0: r > 0 \text{ vs. } H_1: r = 0$$

it is enough to perform a stationarity test onT_p, the last PLS component. The chosen test for the procedure is the KPSS test of Kwiatkowski et al. (1992) and the results of its implementation will be detailed. Extending this idea, the null hypothesis that T_{p-r_0} is stationary for a given and fixed value of r_0 is equivalent to the null that there are at least r_0+1 cointegrating relations while the alternative that T_{p-r_0} is $I(1)$ means that the number of cointegrating relations do not exceed r_0. Thus, the component T_{p-r_0} is instrumental in deciding

$$H_0\text{: } r > r_0 \text{ vs. } H_1\text{: } r \leq r_0$$

by means of the KPSS test. Finally, an estimator of the cointegrating rank can be given as

$$\hat{r} = \#\{T_i : T_i \in I(0)\}$$

which can be decided also on the basis of the KPSS test. We call this testing strategy PLS–KPSS here onwards.

4.1 Simulation Results

We implement the PLS-KPSS testing procedure on simulated data and analyze its performance. Beginning with the estimators, we simulate 10000 paths of length $T = 1000$ of the cointegrated process

$$\Delta X_t = \alpha\beta' X_{t-1} + \varepsilon_t$$

with ε_t having a standard multinormal distribution on R^p. Our first example is given by choosing $\beta = (-1/4,1)'$. We compare the results with other well–known methods for estimating β such as full maximum likelihood in Johansen (1995), principal component analysis in Harris (1997), Snell (1999) and Stock and Watson (1988) and canonical correlations in Bossaerts (1988). We normalize the estimated vector by dividing it by its second coordinate in order to make normalizations homogeneous among the different methods. Table 1 shows the accuracy of PLS for estimating the cointegrating relation β.

Table 1
Estimates of β with Different Methods in a Two–Dimensional System

Method	β	$\hat{\beta} - P_\beta \hat{\beta}$
Johansen	$\begin{pmatrix} -0.2507 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.00044349 \\ 0.00011087 \end{pmatrix}$
PCA	$\begin{pmatrix} -0.2475 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.0022952 \\ 0.0005738 \end{pmatrix}$
CCA	$\begin{pmatrix} -0.2517 \\ 1 \end{pmatrix}$	$\begin{pmatrix} -0.0014104 \\ -0.0003526 \end{pmatrix}$
PLS	$\begin{pmatrix} -0.2506 \\ 1 \end{pmatrix}$	$\begin{pmatrix} -0.00053657 \\ -0.00013414 \end{pmatrix}$

We next consider a three dimensional system with $\beta = (1, 0, -2)$ and perform the same estimations obtaining the results summarized in Table 2.

Table 2
Estimates of β with Different Methods in a Three–Dimensional System

Method	β	$\beta - P_\beta \beta$
Johansen	$\begin{pmatrix} 1 \\ 0.00008 \\ -2.001032 \end{pmatrix}$	$\begin{pmatrix} -0.00018557 \\ .00004 \\ -0.00009 \end{pmatrix}$
PCA	$\begin{pmatrix} 1 \\ 0.00722 \\ -1.98577 \end{pmatrix}$	$\begin{pmatrix} -0.0025593 \\ -0.0032505 \\ -0.0012797 \end{pmatrix}$
CCA	$\begin{pmatrix} 1 \\ -0.00215 \\ -2.01117 \end{pmatrix}$	$\begin{pmatrix} -0.0020192 \\ -0.00097364 \\ -0.0010096 \end{pmatrix}$
PLS	$\begin{pmatrix} 1 \\ -0.00242 \\ -2.0045471 \end{pmatrix}$	$\begin{pmatrix} -0.00081645 \\ -0.0010837 \\ -0.00040822 \end{pmatrix}$

In order to assess how well the PLS–KPSS works, we first simulate 10000 trajectories of lengths $T = 150, 250, 500$ of the random walk

$$X_t = I_p X_{t-1} + \varepsilon_t$$

for the different dimensionalities p=2,3,5. Since the PLS–KPSS works in the opposite direction than the traditional ML test of Johansen (1995), we can only compare these two by testing

$$H_0: r = 0 \text{ vs. } H_1: r > 0$$

with Johansen's procedure and testing

$$H'_0: r > 0 \text{ vs. } CH_1: r = 0$$

with the PLS–KPSS test. The comparison can be made by means of the proportion of times that the corresponding test provides the "correct" decision, that is, leads us into recognizing X_t as a random walk. Observe that in Johansen's case we are just computing the empirical size of the test, whereas for the PLS–KPSS test, we are actually assessing the power. The comparison is nonetheless relevant in pragmatic terms and is presented in Table 3.

Table 3
Comparison between Johansen's and PLS–KPSS Tests

Dimensions	Path Size	Johansen's Test	PLS test
	150	0.9479	0.9756
2	250	0.9474	1
	500	0.9486	1
	150	0.9438	0.9975
3	250	0.9445	0.9998
	500	0.9459	1
	150	0.9411	0.9197
5	250	0.9381	0.9931
	500	0.9454	1

As can be seen directly, the PLS–KPSS procedure is always a better way to go if we intend to detect a random walk. In simple and practical terms, with the traditional Johansen's test, we will be working something slightly above 5 percent of the time with *spurious* cointegrating relations, while this proportion is reduced by the use of PLS–KPSS test.

Also relevant in pragmatic terms is the estimation of the cointegrating rank. Given a p–dimensional time series, Lütkepohl's sequential procedure presented in Chapter 8 of Lütkepohl (2005) consists in testing a sequence of null hypotheses,

$$H_0\!: r = 0, H_0\!: r = 1, \ldots, H_0\!: r = p\text{-}1$$

until the null cannot be rejected for the first time and choosing the cointegrating rank accordingly. This is a very widely used technique to estimate r, the cointegrating rank, which is why comparing it to the PLS–KPSS estimator discussed in the previous sections is relevant.

Table 4
Comparison between Johansen's And PLS–KPSS Tests for Estimating the Cointegrating Rank

Dimensions,	Cointegrating Rank	Path Size	Johansen's Test	PLS test
		150	0.9420	0.9860
2	1	250	0.9510	0.9890
		500	0.9420	0.9830
		150	0.9510	0.9210
3	1	250	0.9480	0.9160
		500	0.9459	1

The PLS–KPSS procedure for testing for cointegration and estimating the cointegrating rank is flexible in admitting heavy tailed noise. As studied in Caner (1998), the trace and maximum eigenvalue statistics of Johansen (1995) suffer from size distortions when the noise sequence is not square–integrable. However, the PLS–KPSS procedure rests mainly on the asymptotic symmetry of the covariance matrix.

Table 5 shows how well PLS–KPSS works with different heavy–tailed noise sequences.

Table 5
Testing for *r=0* with Heavy Tailed Noise

Dimensions	Path Size	Noise distribution	PLS–KPSS detection rate
	150		0.9987
2	250	Lognormal(0,1)	0.9991
	500		1
	150		0.9905
2	250	Power Law α =1.5	0.9956
	500		0.9985
	150		0.9985
3	250	Lognormal (0.1)	1
	500		1
	150		0.9786
3	250	Power Law α =1.5	0.9938
	500		0.9977

Another situation to which the PLS–KPSS procedure is robust is that of near cointegration. The size distortions for Johansen's statistics were studied in Hjalmarsson and Österholm (2010) and some partial remedies were suggested by the authors which do not, however, eliminate the problem. Their setup is simulating paths of the process

$$Y_t = I_p\left(1+\frac{c}{T}\right)+\varepsilon_t$$

where the dimensionality *p* is either 2 or 3; *εt* is a noise sequence and *c* ranges from 0 to-60. We replicate this procedure for our previous sample sizes of *T*=150, 250, 500 and tabulate the proportion of times that the PLS-KPSS procedure concludes no cointegration relations are present in the data.

Table 6
Testing for *r=0* With Neatly Integrated Variables

Dimension 2	$c=-40$		Dimension 2	$c=-20$	
T=150	T=250	T=500	T=150	T=250	T=500
0.9810	1	1	0.9740	1	1
Dimension 3	$c=-40$		Dimension 3	$c=-20$	
T=150	T=250	T=500	T=150	T=250	T=500
0.9400	0.9930	1	0.9390	0.9980	1

To conclude this section we point out that all our results depend on the sample size being large enough and that for smaller sample sizes Johansen's method offers a much greater precision. In fact, for sample sizes roughly bellow 90 observations, Johansen's test performs much better than the PLS–KPSS test, which makes this later method not advisable for small samples.

5. Example with Real Data

We use four indexes built with monthly Mexican data in the time span from January 2000 to December 2011, thus a sample size of T=144. The data consists of p, the consumer price index, m_0, the monetary base, r, and the equilibrium interest rate at 28 days and y, the industrial production index. The data is depicted in Figure 1. The left–hand axis corresponds to the measurements of (log(p), log(r), log(y)) while the right–hand axis measures log(m_0) .

The existence of an equilibrium relation between these variables is usually thought of as evidence that inflation is, in the long run, a monetary phenomenon.

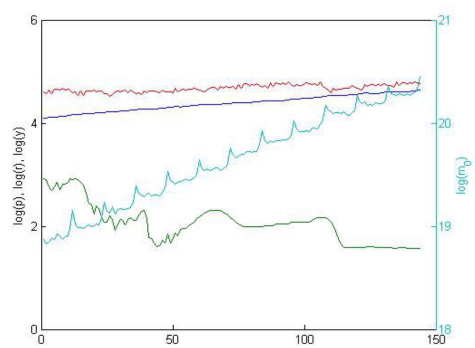

Fig1 Time Series for the Mexican Inflation Model

Applying the PLS–KPSS test to the data, we find that there is but one cointegrating relation. Normalizing the coefficients of β to the first component, the cointegrating relation is

$$\log(p)-0.3748\log(m_0) + 0.091\log(r)-0.6475\log(y)=0 \qquad (6)$$

which is congruent with the economic theory. For example, a reduction in the cost of money is thought to be accompanied by an increasing inflation, whereas a decrease of inflation is usually paired with a decrease in the monetary base. Both these relations are captured by (6) quite neatly.

References

Bossaerts, P. (1998). Common Nonstationary Components of Asset Prices. *Journal of Economic Dynamics and Control, 12,* 347–364.

Braak, C. J. F. T., and De Jong, S. (1988). The objective functions of Partial Least Squares Regression. *Journal of Chemometrics, 12,* 41–54, 1998.

Caner, M. (1998). Test for Cointegration with Infinite Variance Errors. *Journal of Econometrics*, 86, 155–175.

Garthwaite, P.H. (1994). An Interpretation of Partial Least Squares. *Journal of the American Statistical Association, 89*(425),122–127.

Granger, C. (1983). Co–Integrated Variables and Error–Correcting Models (Discussion paper 83–13). San Diego, CA: University of California.

Harris, D. (1997). Principal Component Analysis of Cointegrated Time Series. *Econometric Theory*, *13,* 529–557.

Helland, I.S. (1988). On the Structure of Partial Least Squares Regression. *Communications in Statistics - Simulation and Computation*, 17(2), 581–607.

Helland, I.S. (1990). Partial Least Squares Regression and Statistical

Models. *Scandinavian Journal of Statistics*, *17*(2), 97–114.

Hjalmarsson, E., and Österholm, P. (2010). Testing for Cointegration using the Johansen Mehodology when Variables are Near Integrated: Size Distortions and Partial Remedies. *Empirical Econometrics, 39,* 51–76.

Höskuldsson, A. (1998). PLS Regression Methods. *Journal of Chemometrics*, 2, 211–228.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models* (Advanced Texts in Econometrics). NY, USA: Oxford University Press.

Kwiatkowski, D., Phillips, P.C.B., Peter, S., and Yongcheol, S. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure are we that Economic Time Series have a Unit Root? *Journal of Econometrics*, *54,* 159–178.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin, Germany: Springer.

Næs, T., and Martens, H. (1985). Comparison of Prediction Methods for Multicollinear Data. *Communications in Statistics-Simulation and Computation, 14, 545–576.*

Rosipal, R., and Krämer, N. (2006). Overview and Recent Advances in Partial Least Squares. *In Subspace, Latent Structure and Feature Selection Techniques, Lecture Notes in Computer Science*, (pp.34–51). Berlin, Germany: Springer.

Snell, A. (1999). Testing for *r* versus *r-1* cointegrating vectors. *Journal of Econometrics, 88,* 151–191.

Stock, J.H., and Watson, M.W. (1988). Testing for Common Trends.

Journal of the American Statistical Association, *83*(404), 1097–1107.

Wegelin, J. A. (2000). A Survey of Partial Least Squares (PLS) Methods, with Emphasis on the Two–Block Case. Seattle, USA: Department of Statistics, University of Washington.

Worsley, K.J., Poline, J.B., Friston, K.J., and Evans, A.C. (1997). Characterizing the Response of PET and fMRI Data Using Multivariate Linear Models. *Neuroimage*, 6, 305–319.