Trade and Regional Growth in Spain: Panel Cointegration in a Small Sample

Frédéric Laurin
Université catholique de Louvain
June 2007

Abstract

In this paper, we test for the existence of a long-run relationship between per capita GDP and trade, for 15 Spanish Autonomous Communities between 1988 and 2004, using the cointegration methodology with panel data. In particular, we implement several panel unit root tests (Maddala and Wu, 1999; Levin, Lin and Chu, 2002; Im, Pesaran and Shin, 2003) and panel cointegration tests (Pedroni, 1999, 2004), with a special attention to their behaviour in a small sample. Based on Taylor and Sarno (1998), we also develop a SURE residual-based test, in order to explicitly take into account the cross-regional correlation pattern. Appropriate confidence intervals are estimated with a sieve bootstrap implemented for the small sample T=17, and generated in such a way that the dependence structure among cross-sectional units is preserved. We use alternatively exports, imports and total trade in the cointegration system to see the specific impact of each trade variable. Our cointegration tests reject the existence of a significant relationship between GDP per capita and exports. However, we do find some evidence of a significant relationship between GDP per capita and imports or with total trade (exports plus imports).

JEL classification: C22 - Time-Series Models; C23 - Models with Panel Data; F14 - Country and Industry Studies of Trade; F15 - Economic Integration; O18 - Regional, Urban, and Rural Analyses.

Address for correspondence: Frédéric LAURIN, Centre for Operations research and Econometrics (CORE), 34 Voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium, Ph: +32 (0)10 47 43 59,
e-mail: frederic.laurin@uclouvain.be

The relationship between trade and growth has generated an extensive empirical literature,[1] most of which have been written within the debate opposing the "export-led" and "import-substitution" policies for newly industrialized countries. For Spain, a member for the European Union, this question is not really relevant. However, it is often believed that Spain's recent economic successes are largely due to a strong expansion in trade since it has joined European Community in 1986. If this is the case, one should be able to observe a strong relationship between trade and GDP starting from Spain's entry into the European Community.

On theoretical grounds, the channels linking trade and economic growth are various and well documented: optimal reallocation of factors, economies of scale, foreign direct investments (FDI), productivity improvements stimulated by foreign competition, access to cheaper (or higher-quality) inputs, international transmission of ideas, "learning-on-imports", etc. At the same time, an increase in own productivity by local firms renders these firms more competitive on the world market, encouraging more exports. Hence, the causality linking trade and growth goes in both directions. Whatever the channels involved in this reciprocal causation process, we should observe a positive relationship between economic development and trade.

The objective of this study is to test the existence of such a positive relationship between GDP per capita and trade (exports and imports) for the Spanish Autonomous Communities (ACs) between 1988 and 2004. We rely on the cointegration technique, which allow one to assess whether GDP per capital and trade are linked through a long-run relationship. The period 1988-2004 is very interesting in Spain's case because it covers a phase of important industrial structural changes, following the country's accession to the European Community, especially after the set-up of the European Single Market program in 1992 (Laurin, 2007a). The basic assumption of this paper is that, in such a period of deep transformation, *the relationship between trade and growth will be sufficiently intense so as to confer to the cointegration tests a meaningful and significant interpretation*, even within a reduced time frame.

We will implement several panel cointegration tests, with a special attention to their behaviour in a small sample. The panel approach taken in this paper should contribute in strengthening the acknowledged low power of individual cointegration tests in small sample (we have T=17 annual

[1] See Lewer and Van den Berg (2003) and Ahmad (2001) for an empirical review of literature.

observations, from 1988 to 2004). We will employ some newly-developed panel unit root tests (Maddala and Wu, 1999; Levin *et al.*, 2002; Im *et al.*, 2003) and panel cointegration tests (Pedroni, 1999, 2004; Taylor and Sarno, 1998), which are likely to be more robust to the individual idiosyncratic variability within cross-units.

This paper contributes to the existing literature in several respects. First, we exploit a very extensive trade database, constructed by the *Agencia Estatal de Administración Tributaria*, that compiles all monthly international transactions in good and services by Spanish Autonomous Communities. Hence, the analysis will be undertaken at the regional level, in a panel setting. Existing studies have undertaken this kind of investigation at the national level, although the issue is likely to be more interesting at the regional level; the process of international specialization of production is mainly a dynamic regional process. Indeed, according to New Economic Geography (Fujita *et al.*, 1999), the agglomeration forces shaping the location of industry are by nature local. Furthermore, the external scale economies (such as technological externalities) are by definition limited in space. Thus, the impact of trade on growth is likely to unfold at the regional level.

Second, we will fully discuss and take into account the small sample bias inherent to any cointegration tests applied on a short period. In particular, the critical values for all panel cointegration tests are computed using a sieve bootstrap, following Li and Maddala (1997). This bootstrap is implemented for T=17, in order to get critical values adjusted for the small sample. But this is not the only usefulness of the bootstrap simulation. The Maddala and Wu (1999), Levin *et al.* (2002) and Im *et al.* (2003) tests are unit-root tests on a single series. These authors do not provide critical values of a residual-based cointegration test. Hence, the sieve bootstrap will also be used to compute the relevant critical values for theses tests. Moreover, most panel unit root or cointegration tests assume cross-section independence. This is hardly a realistic hypothesis for regions of a same country. For that matter, following Maddala and Wu (1999), the bootstrap samples are generated in a way such that the dependence structure among cross-sectional units is preserved. In addition, based on Taylor and Sarno (1998), we also develop a SURE (Seemingly Unrelated Regression Estimation) residual-based methodology to test for cointegration, in order to explicitly take into account the cross-regional correlation pattern.

Third, we also show that using exports as the sole variable expressing trade – as it has been done in numerous studies – can be misleading. We expect exports to have a positive effect on growth, but the impact of imports may be ambiguous. Imports can displace local production, leading to a negative effect on growth. On the contrary, imports could stimulate productivity through a better access to inputs and intermediates, through competitive pressure from trade liberalization, through learning-on-imports, etc. In this paper, we use exports, imports and total trade separately in the cointegration system to see the specific impact of each component of trade.

Our cointegration results reject the existence of a significant relationship between GDP per capita and exports. This negative result is coherent with the already mixed results obtained in past literature (Ahmad, 2001). However, we do find some evidence of a significant relationship between GDP per capita and imports or with total trade. One important caveat here is that all these results are conditional on the sample period 1988-2003, and cannot be extended to the longer run.

The paper is organized as follows. In the first section, we propose a short theoretical discussion reviewing the various economic channels relating trade and growth. In section 2, we introduce and define the various individual and panel cointegration tests computed in this paper, comparing the different assumptions underlying each test in their design or their distribution. We also fully discuss of the particular conditions implied by the small sample problem. In the third section, the cointegration results are presented and discussed, starting first with the individual tests, followed by the panel and SURE tests. We also test for the existence of a break in the cointegration relationship. In Annex 2, the bootstrap algorithm and properties are described in detail and we analyse the properties of the bootstrap simulations with some summary statistics.

1. Review of literature

What are the effects of international trade on economic growth? This issue has been analyzed abundantly in the literature. Here, we propose a schematic review on the numerous channels by which trade might affect growth.

In standard trade models, regions specialize according to comparative advantages. The efficiency gains from this specialization process lead to a decrease in prices and to an increase in real income. This is a "static" level effect on welfare. However, from the empirical point of view, this static gain of trade may unfold over many years, offering some short-term dynamics. Extending the standard models, some scholars have tried to introduce genuine growth effects by focusing on the standard neoclassical process of capital accumulation. Johnson (1967) mentions that if a region specializes in a capital-intensive sector, then, by the Rybczynski theorem, specialization will favour faster capital accumulation in the region, raising regional growth. In Corden (1971), some part of the increased income level gain from trade will be saved, leading to an increase in investments and growth. Similarly in Baldwin (1992a), by the Stolper-Samuelson theorem, trade liberalization raises the return to capital, inducing more capital accumulation, thereby increasing growth.

The introduction of formal "growth" effects in trade theory came with the development of the endogenous growth model (Arrow 1962; Romer 1986, 1990; Lucas 1988). In these models, the overall stock of knowledge in a region can contribute to the productivity of a single firm- through **technological externalities** for example - giving rise to regional scale economies. Trade can enlarge the regional stock of knowledge in numerous ways. Suppose that the region happens to specialize in a technological-intensive industry following trade liberalization. Since R&D efforts are undertaken mainly by the technological-intensive industry, this reallocation of resources increases the rate of innovation in the region, which generates even more technological externalities, enlarging the stock of knowledge. This is what Rivera-Batiz and Romer (1991) call the **allocation effect** of trade. **Learning-by-doing** can have a similar impact on economic growth (Young, 1991). If the region initially specializes in high and sophisticated technological goods, this will spur more learning-by-doing than for low-tech goods or for goods in which learning-by-doing has mostly exhausted. Moreover, the reduction of trade impediments favours the transmission of ideas across international borders, e.g. **international technological spillovers** (Grossman and Helpman, 1991b; Rivera-Batiz and Romer, 1991). Furthermore, assuming imperfect competition, the expansion of export markets increases monopolistic rents that can be used for R&D and innovation activities (Grossman and Helpman 1995; Baldwin 1992b). This is the **integration effect** of trade.

One important point to notice is the importance of the growth effect of imports. With trade liberalization, local firms can **access new and cheaper goods and services**, thereby increasing their competitiveness. In addition, consumers can enjoy a gain in utility through higher product diversity. Second, if these imports are sufficiently specialized or innovative, they join to the pool of intermediate goods that can directly increase the productivity of the end-users (Ethier, 1982; Grossman and Helpman, 1991a). Third, new knowledge can also be extracted by studying the nature and composition of imported products (**learning-on-imports** or reverse engineering). We can add a further trade effect that represents a very important component of the firms' own productivity growth: the **competition effect**. Trade liberalization amplifies the competitive pressure on local firms. The overall effect of competition can be ambiguous. On the positive side, it is well know that competition promotes efficiency, innovation, product quality, while favouring a better absorption of modern technology (Balassa, 1978; Krueger, 1980; Baldwin, 1992b). Second, trade liberalization favours the expansion of profitable firms, and inversely forces the closures of the less competitive firms (Bernard *et al*., 2000; Melitz and Ottaviano, 2005), contributing in aggregate to an increase in regional productivity. On the negative side, import competition can displace local production and lead to closure of uncompetitive firms and, in some circumstances, to unemployment. Finally, since competition reduces monopoly rents, firms are less capable of investing in R&D (Schumpeter, 1942; Romer 1986, 1990).

Note that the relationship between growth and trade is not one-sided. Own productivity growth should make firms more competitive on world market: they will export more. At the same time, economic development tends to go hand in hand with import expansion. On the consumption side, economic growth should increase average income, stimulating the demand for foreign consumption goods. On the production side, the expansion of regional production should increase the imports of differentiated inputs or intermediates.

Laurin (2007b) reveals that, at the industry level, the ACs having a relatively higher level of exports per capita tend to have also a higher level of imports per capita. As a result, testing the effects of exports on growth means implicitly that we are also testing part of the effects of imports. Both variables express the same economic phenomenon: the extent of the regions' integration into international trade. That could explain part of the mixed results of past literature on the effects of exports on growth[2]. Therefore, considering only exports for exploring the effects of trade on growth can be misleading for the interpretation of the cointegration results, especially for regions showing a reasonable level of trade openness. How can this common evolution between exports and imports be explained at the regional level[3]? Regions that export the most may have a greater "aptitude" for international trade, and consequently import more (and vice-versa). Trade openness favours all kinds of interaction with foreign firms and economic agents, unfolding all kinds of opportunities for trade. Another reason is that, to be successful on world markets, a firm needs to be competitive. That means using the lower-price or best-quality inputs possible. Hence, a firm must import some inputs and intermediate goods abroad. Thus, the capacity to export depends on access to imports. Finally, from the monopolistic competition trade models (Helpman, 1981; Ethier, 1979 and 1982), it is well-known that trade is mostly intra-industrial between developed countries, exchanging inputs or different product varieties within the same industry.

Hence, will test three main cointegration specifications:

1. Relationship between GDP per capita and exports;
2. Relationship between GDP per capita and imports;
3. Relationship between GDP per capita and total trade (exports plus imports).

2. Econometric Methodology

One difficulty with the estimation of an equation in the time dimension is the well-known problem of spurious regressions: if the variables follow a non-stationary process, the regression results might be spurious (Granger and Newbold, 1974). The earliest studies investigating the relationship between trade and economic growth have relied on simple bivariate correlation coefficients to test the causal relationship between export and growth (Emery, 1967; Michaely, 1977; Tyler, 1981) or on the estimation of a production function (Balassa, 1978; Sheehey, 1990, 1992). All these studies find unsurprisingly a positive effect of trade on growth, but this relation might just be spurious. Moreover, statistical association does not identify causality. However, there might exist a stationary linear

[2] Indeed, Shan and Gang Tian (1998) have shown the importance of controlling for imports to avoid a spurious causality result between exports and growth.

[3] At the macroeconomic level, a country having a flexible exchange rate should not experience a long period of current account imbalance. But regions sharing the same currency might be less affected by this mechanism constraining the current balance.

combination of the series, in which case the series are said to co-integrate. Cointegration indicates that the series share a common stochastic trend, e.g. they co-evolve together along a long-term path. In the absence of cointegration, the estimated relationship will have absolutely no economic meaning. The regression will be totally spurious.

A second type of literature has instead used the concept of Granger causality (Granger, 1969; Sims, 1972): a variable x_t is said to "Granger cause" another variable y_t when the past and present values of x_t provide some significant information to forecast $y_{t+1.}$ For example, Pereira and Xu (2000) devise a multivariate VAR testing the causality between four variables: growth, export, employment and investment. They find that export growth affects positively GDP growth in sample of 30 countries and that the effect of export on investment is the most substantial channel stimulating growth. However, if the variables are shown to be cointegrated, then an error-correction model (ECM) or Vector error-correction model (VECM) should be used in order to properly test for Granger-causality (Engle and Granger, 1987). Hence, the most advanced literature investigating the effects of trade on growth has proceed in two necessary steps: first, verify the cointegrating properties of the variables, then estimate a ECM or a VECM to test for Granger-causality (for example, see Marin, 1992; Xu, 1996; Shan and Gang Tian, 1998; Dritsakis 2004). In the case of Spain, Balaguer and Cantavella-Jorda (2004) find the existence of a long-run relationship among output, exports and structural change for the sample period 1961-2000. In the short-term, they also identify a bi-directional causality between exports and output. Also focusing on Spain, Alguacil and Orts (2002), for the sample period 1970-1992, observe a causal relationship running from FDI to exports, but not in the inverse direction. They also confirm a long-term equilibrium relationship between growth, exports, FDI, real exchange rate and OECD output (representing world demand). In these two papers on Spain, the analysis is undertaken at the national aggregate level, but not at the regional level.

For the present study, it is not obvious that trade should necessarily have a contemporaneous and immediate effect on economic growth: it might take some time for the trade effects to spread through the different channels linking trade and growth. One solution would be to add an appropriate number of lagged values in a VAR or VECM, but this will only restrain the degrees of freedom in an already small sample[4]. In addition, with annual data and a small sample, the lag structure might not be very efficient in detecting the short-term dynamics. Moreover, there may exist a cross-correlation not only between regions, but also between the equations of the multivariate VAR. This implies the estimation of a complex cross-correlation matrix, which may be difficult in a small sample. In view of all these econometrics issues, we believe that a causality analysis or the estimation of a structural model in a panel context should be developed in a separate study.

Besides, the existence of short-term effects is neither necessary nor sufficient for the existence of a long-term relationship. In fact, cointegration is a stronger proposition than causality. Cointegration between two or more variables is sufficient for the presence of causality in at least one direction (Engle and Granger, 1987). Cointegration tests are a very flexible and general statistical method that reveals the possible existence of a significant long-term relationship between a set of variables, whatever the actual mechanisms linking these variables together[5]. And our empirical hypothesis about the long-term relationship between GDP and trade in Spain is firmly rooted in the previous theoretical literature. In sum, the cointegration methodology has a convenient economic and econometric meaning. Hence, this paper deals exclusively with the long-term dynamics between GDP and trade, by implementing several - individual and panel - cointegration tests.

2.1 The residual-based methodology

All the cointegration tests implemented in this study are based on the two-step Engel-Granger residual-based methodology (Engle and Granger, 1987). The first step is to estimate by OLS the relationship between GDP per capita and a trade variable:

- **First step**: estimate by OLS

(1) $Y_{it} = \alpha_i + \beta_i X_{it} + e_{it} \qquad i = 1,\ldots, N, \quad t = 1,\ldots,T$

[4] As shown in Xu (1996), the failure to support causality between export and growth may be attributable to the choice of lags in the VAR system.
[5] But accordingly, this technique does not identify the particular channels governing this long-term relationship nor the direction of causality.

where Y_{it} is the log of GDP per capita, X_{it} is alternatively the log of exports, imports or total trade (exports + imports) and α_i is an individual effect. The index *i* denotes the individual cross-section unit (N=15 Autonomous Communities) and index *t* denotes the sample period (T=17). Note that β_i is known to be super consistent[6]. If the series are cointegrated, the residuals from this regression should be stationary. Therefore, the second step involves a unit root test on the predicted residuals $\hat{e}_{it}$:

- **Second step**: unit root test on $\hat{e}_{it}$.

If $\hat{e}_{it}$ has a unit root, then $Y_{it} - \alpha_i - \beta_i X_{it}$ is not a cointegrating relationship. Any unit root test can be used on $\hat{e}_{it}$, provided that we have the appropriate critical values. For example, one could apply the well-known **Augmented Dickey-Fuller** (ADF) (Dickey and Fuller 1979) and **Phillips-Perron** (PP) (Phillips and Perron, 1988) unit root test. MacKinnon (1991) has computed the appropriate critical values for the residual-based Engel-Granger ADF and PP cointegration test for any sample size T.

2.2 The panel cointegration tests

However, these last two unit root tests, applied individually to each AC, are known to have a low power in a small sample. Fortunately, we can exploit the panel dimension of our data set to strengthen the power of the residual-based test. The panel estimation might compensate for temporary and idiosyncratic noises disturbing the cointegration relationship within a particular cross-unit. We will first implement on the residual $\hat{e}_{it}$ the following panel unit root tests:

- The Levin-Lin-Chu (LLC) unit root test;
- Im-Pesaran-Shin (IPS) unit root test;
- Maddala-Wu Fisher (MWF) unit root test.

These tests are based on the augmented Dickey-Fuller test, using the following first-order autoregressive regression:

[6] The single equation regression allows to test that at least one vector of cointegrating coefficients exist, bearing in mind that doing so imposes a normalization on the variable that we arbitrarily choose on the left hand side of the regression equation.

(2) $\Delta \hat{e}_{it} = \alpha_i + \rho_i \hat{e}_{it-1} + lags(\Delta \hat{e}_{it-j})_i + u_{it}$

where *j* represents the appropriate number of lags necessary to correct for serial correlation. The null hypothesis of no cointegration implies that $\rho_i = 0$, e.g. there is a unit root in the residual series:

H0: $\rho_i = 0$: unit root in the residuals (no cointegration)

H1: $\rho_i \neq 0$: no unit root in the residuals (cointegration)

The null hypothesis is tested using the usual *t*-statistic associated with the autoregressive ρ_i coefficient. But since $\hat{e}_{it}$ is itself a predicted value, the usual Dickey-Fuller distribution is not valid for cointegration tests. Appropriate critical values must be used that take into account the error variance of the first step.

In the **Levin-Lin-Chu (LLC)** procedure (Levin, Lin and Chu, 2002) equation (2) is estimated by OLS as a pooled regression, imposing the ρ_i coefficients to be equal across units, but each unit has its own individual effect (α_i). The assumptions of the LLC test are rather restrictive. First, the test imposes homogeneity of the ρ coefficient, e.g. all units converge at same rate. The hypothesis should be formulated as:

H0: unit root for ALL units;

H1: no unit root for ALL units.

Second, all units are restricted to have an identical lag structure to correct for serial correlation.

The **Im-Pesaran-Shin (IPS)** unit root test (Im, Pesaran and Shin, 2003) and the **Maddala-Wu** Fisher unit root test **(MFW)** (Maddala and Wu, 1999) allow for more flexibility. It consists of estimating by OLS equation (2) unit by unit. Hence, each unit has its own coefficient (ρ_i), lag structure and individual effect (α_i). Then, the IPS statistic is computed using the average of the individual ADF *t*-statistics, while he MFW statistic combines the significance levels (*p*-values) of the individual ADF *t*-test (or the PP test[7]) in the following way:

$$\text{MWF statistic}_i = -2\sum_{i=1}^{N} \log_e \pi_i$$

[7] In fact, the MWF test can combine the *p*-value of any unit root test (Maddala and Wu, 1999).

where π_i is the *p*-value of the individual unit root test. The IPS and MFW tests are a way of combining evidence on N independent unit root tests. The hypothesis then becomes:

H0: unit root for all units;

H1: not all units contain a unit root.

Maddala and Wu (1999) show that their test has a higher power than the IPS and LLC tests.

Note that the asymptotic properties of the LLC, IPS and MWF tests are not valid for a residual-based cointegration test. For all three panel unit root tests, we need to compute the appropriate critical values. This is done with the sieve bootstrap described below.

Unlike the previous tests, the **Pedroni tests** (Pedroni, 1999, 2004) are specifically devised for the Engel-Granger residual-based methodology. Pedroni (1999, 2004) proposes seven different tests that can be categorized as "combining" tests or as "pooled" tests. The "combining" procedure calls for pooling along the within dimension of the panel: based on equation (2), each test is calculated individually unit by unit and then combined into an asymptotically converging statistics (Group Mean Statistics). This is analogous to the IPS and MWF tests. The three Pedroni "combining" tests are:

1. Group *rho*-stat (ρ-statistic): combining the Phillips-Perron *rho*-statistic;
2. Group PP *t*-stat (non-parametric): combining the Phillips-Perron *t*-statistic;
3. Group *t*-stat (parametric): combining the ADF *t*-statistic.

We can briefly summarize the "pooling" methodology by noting that the calculation of a unit root test usually involves a numerator and a denominator. In the case of the ADF unit root *t*-statistic:

$$t_i = \frac{(\hat{e}_{it-1}'\Delta\hat{e}_{it})}{\hat{\sigma}(\hat{e}_{it-1}'\hat{e}_{it-1})^{1/2}}$$

where $\hat{\sigma}$ is the estimated variance of the residual u_{it} in equation (2). Hence, the "pooled" Pedroni tests are constructed by summing both the numerator and the denominator terms of a given unit root statistic over the N dimension, and then computing an asymptotically converging statistic – which corresponds, in Pedroni's words, to pooling along the between dimension of the panel (Panel Statistics). The "pooled" Pedroni tests are based on four unit root statistics:

4. Panel variance ratio (*v*-statistic): based on Phillips and Ouliaris (1990) long-run variance ratio statistics for time-series;
5. Panel *rho*-stat (ρ-statistic): based on Phillips-Perron *rho*-statistic;
6. Panel PP *t*-stat (non-parametric): based on Phillips-Perron *t*-statistic;
7. Panel *t*-stat (parametric): based on the ADF *t*-statistic.

We refer to Pedroni (1999, 2004) for a detailed description of each test. All seven Pedroni statistics are transformed and normalized (see Pedroni 1999, 2004) so as to follow the standard normal distribution, taking into account some nuisance parameters, and given cross-section independence.

So far, all the above tests assume cross-section independence. This is hardly a realistic hypothesis for regions of a same country. If this assumption is not verified, Maddala and Wu (1999) suggest using a sieve bootstrap to find the appropriate critical values. Appendix 3 fully describes the construction of the sieve bootstrap. The bootstrap samples are generated in such a way that the dependence structure among cross-sectional units is preserved. The bootstrap is implemented for T=17, in order to get critical values adjusted for the small sample[8]. Hence, the sieve bootstrap will be useful for three reasons: 1. computing valid critical values for the LLC, IPS and MWF for a residual-based cointegration test; 2. relaxing the cross-section independence; 3. adjusting for the small sample. Chang (2004) examines the asymptotic validity of the bootstrap methodology applied to unit root tests for cross-sectionally dependent panels. Li and Maddala (1997) investigate the usefulness of the bootstrap methods for small sample inference in cointegrating regression models.

[8] Banerjee and Carrion-i-Silvestre (2004) also analyse the properties of the Pedroni tests in the presence of various types of structural breaks in the cointegration equation. They also devise a sieve bootstrap taking into account these breaks. See also Fachin (2005) for an application of the sieve bootstrap for panel cointegration at the regional level, taking into account the cross-correlation across Italian regions.

However, the cross-correlation pattern between regions conveys useful information that can be used to strengthen the estimates. Even if we can efficiently compute the appropriate critical values by bootstrapping, we are still not explicitly making use of this cross-sectional correlation. Following Taylor and Sarno (1998), we also estimate the residual-based equation (2) in a system of N equations: one equation for each AC. By estimating this system with Zellner's (1962) Seemingly Unrelated Estimator (SURE), we take into account of the contemporaneous correlation among the disturbances (u_{it} in equation 2). Using this approach, we can devise three SURE cointegration tests:

The unrestricted SURE Dickey-Fuller test (USURE): we estimate the N equations (2) in a SURE system and use the individual *t*-statistic of each ρ_i coefficient to test the cointegration hypothesis:

$$H0_i: \rho_i = 0$$

The USURE test is not a pooled procedure since homogeneity of the ρ_i coefficients is not imposed. The cointegrating properties are evaluated unit by unit. Nevertheless, the individual equations are not estimated in isolation, but in a system, taking into account the pattern of cross-correlation between ACs.

Taylo and Sarno's (1998) multivariate augmented Dickey-Fuller (MADF) test: the setting is the same as for the USURE, but the panel cointegration statistic involves testing the joint significance of all the ρ_i coefficients with a Fisher *F*-test (Wald test):

$$H0_i : \rho_i = 0 \qquad \text{for all } i = 1, \ldots, N$$

This test must be interpreted with caution since the null hypothesis implies a stationary process for all units. Taylo and Sarno (1998) show that the rejection of the null hypothesis may be due to only a few units having a stationary process.

The Restricted SURE Dickey-Fuller test (RSURE): we now constrain the ρ_i coefficients to be equal across ACs, keeping heterogeneity in the lag structure and in the individual effect. The cointegration hypothesis is tested using the *t*-statistic of the homogeneous ρ coefficient:

H0: $\rho = 0$

In all three SURE tests, we keep the cointegrating coefficients (β_i) in the first-step equation, the individual effects (α_i) and the lag structure completely heterogeneous. Again, critical values can be computed with the sieve bootstrap, implemented with the small sample size of T=17. Chang, Park and Song (2005) signal that the sieve bootstrap can be consistent for panel SURE cointegration tests.

There is a trade-off between complete individual flexibility and imposing some homogeneity by pooling (Banerjee, 1999). Allowing for total flexibility in each individual equation is a convenient feature. But in a small sample, there might be considerable gains in pooling the data and imposing homogeneity, in order to strengthen the estimates of the tests[9].

To sum up, Table 1 lists all the cointegration tests presented in this section. The panel statistics can be classified into two types, using the typology of Pedroni (1999, 2004): 1. the "pooled" tests: those that effectively pools the data; 2. the "combining" tests: those that combine each individual test statistic. Note that the USURE test is neither a pooled nor combining test, since it estimates an individual test value for each unit.

INSERT TABLE 1 ABOUT HERE.

2.3 A small sample exercise

Using the cointegration residual based-test procedure on a small sample raises two important econometric issues. First, the individual residual-based test is known to have low power in small samples. But, as already explained, this problem may be alleviated by taking advantage of the panel dimension of the data. Second, there is the problem of the small sample bias. Typically, for the individual residual-based tests, the critical values are higher (larger negative value) with a low sample

[9] Usually, the empirical literature imposes homogeneity of the cointegrating coefficients (the β coefficients in equation 1). This literature assumes that long-term coefficients are more likely to be homogeneous across units than short-term coefficients. Considering the theoretical literature on trade and regional growth, we rather take the opposite view. The long-term equilibrium between growth and trade can be heterogeneous across regions, because of divergent industrial structures, productivity gaps, capacity to adapt to trade liberalization, differences in geography, the cumulative process of agglomeration, etc. In the "new" trade models, economic integration can trigger a process of economic divergence across regions. Hence, we do not whish to impose homogeneity of the β coefficients.

size. This outcome occurs because the test procedure is trying to infer the long-term "generic" cointegrating properties of the series from a small sample. Any two independent random walks might temporarily and coincidentally show some spurious co-movements for a given short period, leading to what looks like a "cointegrating" phase. Hence, even if the unit root hypothesis is imposed, any Monte Carlo simulation will obtain a higher percentage of rejection of the null hypothesis (no cointegration) in a small sample than in a large sample. For all of our tests, the small sample bias will be fully taken into account in the sieve bootstrap simulation by imposing T=17. We also compute the bootstrap critical values for a larger sample (T=60) for comparison[10]. Indeed, the bootstrap summary statistics presented in Appendix 3 show that the small sample (T=17) critical values are higher than the T=60 critical values. Moreover, the panel tests may also alleviate the problem of the small sample bias. One might observe a coincidental "cointegrating" phase for some units in a given period, but as N grows large, the probability of having all units affected by this coincidence is significantly reduced.

Hence, the small sample distribution is more restrictive in that it corrects for "coincidental" cointegration that might be observed over a small period of time. However, in this paper, we are testing the significance of the relationship conditional on a particular set of years, whatever the behaviour of the relationship outside this time span. Suppose that H0 is rejected. Even if the critical values have been corrected for the small sample bias, there is still a probability (the significance level) that the series are not truly cointegrated in the long-term. Is the estimated relationship spurious? We cannot say. It may be that a particular and specific short-term mechanism truly relates the series together over the sample. Indeed, our objective is to investigate the existence of a relationship between trade and GDP specifically for a given set of years: 1988 to 2004. The implied assumption is that, in such a period of deep economic transformations in Spain, the relation between trade and growth will be sufficiently intense so as to confer to the cointegration tests a significant interpretation. Naturally, the reader must bear in mind that our intent is not to infer from a small sample the general cointegrating properties of the two series in a larger sample - what we call true "generic" cointegration - and that the results will be conditional on the 1988-2004 period. In these conditions, the cointegration test procedure is not less meaningful in a small sample, but only more stringent.

[10] T=60 is chosen to replicate the number of observations generally available in annual cross-country macroeconomic databases used by empirical papers (the Penn World Tables for example, starting in 1950).

2.4 Data

We have data on annual GDP, population, exports and imports for Spain's 17 Autonomous Communities. For the trade variables, we use a special database constructed by the *Agencia Estatal de Administración Tributaria,* the Spanish custom authorities, from which we can obtain the total annual value of exports and imports by Autonomous Communities, e.g. all trade going to or coming from foreign countries outside Spain[11]. Data on GDP and on population are taken from INE (*Instituto Nacional de Estadística*). To obtain real values, we use GDP, exports and imports deflators available from INE. GDP per capita is simply defined as real GDP divided by population. The trade series are available from 1988 to 2004, which defines our time sample. We exclude from the sample two Autonomous Communities, the small Islands of Canarias and Baleares, in order to discard influential observations having a negligible economic weight. In total, we have N=15 regional units (see Appendix 1 for the list and a map of ACs, and the abbreviations used) and T=17 years, for a total of 255 observations. The panel is balanced.

All series are characterized by a unit root process. Appendix 2 shows the detailed unit root tests for each variable.

3. Cointegration results

Before tackling the more robust panel cointegration tests, we present first the individual ADF and PP residual-based cointegration tests as a benchmark. For the ADF tests, we are constrained to a maximum of 2 lags because of the shortness of the time period. The lag length is selected by the general-to-specific approach (Hall, 1994; Campbell and Perron, 1991), subsequently deleting insignificant lags (at 10%) from two lags to none.

3.1 Individual results

We show in Table 2 the ADF and PP cointegration tests for the relationship GDP per capita-Exports and GDP per capita-Imports. For the ADF test, the Ljung-Box *Q*-statistic for serial correlation is also shown,

[11] Data on intra-Spanish trade between ACs themselves are unfortunately not available.

with its associate *p*-value in parenthesis. Fortunately, a maximum of two lags seems enough to correct for autocorrelation, since the Ljung-Box tests are all rejected.

INSERT TABLE 2 ABOUT HERE.

From the ADF tests, only three ACs (Catalonia, Murcia, Rioja) appear to have a significant (at 10%) stationary relationship between real GDP per capita and exports. For imports, only C.Mancha, Extremadura, Rioja and Valenciana have a significant relationship. Interestingly, the value of the ADF *t*-stat is higher with imports than with exports in a majority of ACs. Table 3 gives the results for the relationship between GDP per capita and total trade. There is now 6 ACs showing a significant relationship: Asturias (from the PP test), Catalonia, C.Mancha, Murcia, Rioja and Valenciana.

INSERT TABLE 3 ABOUT HERE.

3.2 Breaks in the cointegration relation

The presence of unknown structural breaks in the long-term relationship can greatly affect the cointegration results (Perron, 1989; Gregory and Hansen, 1996) In Laurin (2007c), we have found the existence of a break in both the trade and GDP series around 1993. This is explained mostly by the fact that exports start expanding just before the end of a recession. This period corresponds also to two important events. First, when Spain joined the EC in 1986, the membership treaty included a period of adaptation involving all kinds of temporary exemptions and protection measures that where due to gradually terminate up to 1992. Second, 1992 coincides with the consolidation of the European Single Market initiative by the European Commission.

Thus, we choose to verify the presence of a single downward shift around 1993 in the cointegration relation. To test that hypothesis, and following Gregory and Hansen (1996), we simply estimate the cointegration equation (1) including a level shift or a break in the cointegration relationship around 1993.

(7) $$Y_{it} = \alpha_i + \beta_{1i} X_{it} + \beta_{2i} DU_{Ti} + \beta_{3i} (DU_{Ti} * X_{it}) + e_{it}$$

where, T_i = year of break;

$DU_{Ti} = 1$ if $t > T_i$, 0 otherwise.

DU_t is a dummy variable representing a change in the level of the trend (change in intercept) and $DU_{Ti} * X_{it}$ is an interaction term representing a change in the coefficient of the cointegration relationship. The optimal year of break is selected by taking the highest cointegration ADF *t*-stat (largess negative value) across all break points over the sample range 1990-1996. Results are presented in Table 4.

INSERT TABLE 4 ABOUT HERE.

Indeed, most ACs have a significant downward shift or break in the cointegration relationship around 1993. Hence, part of the cointegration disequilibrium may be caused by this single-period event. As a result, in the next panel cointegration tests, we will also investigate two further relationships including a level break in 1993:

1. Relationship between GDP per capita and exports with a time dummy for 1993;
2. Relationship between GDP per capita and total trade with a time dummy for 1993.

3.3 Panel results

We now present the results for the panel cointegration tests. A bootstrap simulation must be replicated for each specification and each panel test. In the panel result tables, we indicate the *p*-value for the small sample (T=17) bootstrap distribution, as well as for the larger benchmark sample (T=60). A test statistic that appears to be significant at the T=60 benchmark, but non-significant for the T=17 distribution means that we observe to some extent a co-movement behaviour between the variables, but that this relationship might be "coincidental" over the sample. This conjecture would still fit our hypothesis of the existence of some relationship between GDP and trade, conditional on the 1988-2004 sample, but to be econometrically rigorous, the only admissible distribution for hypothesis testing is still the T=17 distribution.

The LLC, IPS and MWF unit root tests

Table 5 presents the panel LLC test. This test imposes homogeneity of the ρ coefficients and the same dynamic lag structure that we set at one lag. From the small sample bootstrap *p*-value, we cannot reject the hypothesis of a unit root in the cointegration residuals for any of the specifications. Note that the strongest results are still observed for the specifications using imports. Even by the large sample *p*-value standard, only the specifications GDP per capita-Exports with the 1993 time dummy would have been significant.

INSERT TABLE 5 ABOUT HERE.

The absence of any significant relationship for the LLC test may be due to the constraints on the lag structure. Adequately correcting for autocorrelation is an important issue for the residual-based test. Serial correlation can be more effectively corrected by allowing a different lag structure for each unit. Hence, in Table 6, we have estimated each N equation individually by OLS, where each unit has its own equation with a constant (fixed-effect) and its own lag structure, but constraining the ρ_i coefficients to be equal across units (homogeneity constraint). As before, the lag length of each individual equation is selected by the general-to-specific approach, with a maximum of 2 lags. Table 6 indicates the value of the restricted homogeneous coefficient and its associate *t*-statistic.

INSERT TABLE 6 ABOUT HERE.

Note that the results of Table 6 are not directly comparable to the LLC test since the latter is a transformed and normalized statistic. For the former, there is no need to transform the statistic value in any way since the unknown distribution can be estimated by bootstrap. Now, after relaxing the constraints on the lag structure, we get that the specification GDP per capita-Imports is significant at the 10% level, while the specification GDP per capita-Trade barely misses this significance level. This reveals the importance of appropriately correcting for serial correlation in the panel tests, as it is already known for the individual unit root tests. This issue will be again discussed latter with the SURE estimations.

The IPS (Im, Pesaran and Shin, 2003) and MWF (Maddala and Wu, 1999) tests allow for more flexibility, each unit having its own ρ coefficient and lag structure. Recall that the IPS test combines each individual ADF *t*-stat while the MWF test combines each individual *p*-value associated with the ADF or the PP *t*-statistic. The lag lengths are selected by the general-to-specific approach. Results are given in Table 7 for the IPS test and in Table 8 for the MWF test.

INSERT TABLE 7 ABOUT HERE.

Both the IPS test and MWF test (from the ADF *p*-value) appear to be significant (at 10%) for the relationship between GDP per capita with imports or with total trade, but not the other specifications. The MWF tests using the *p*-value from the PP statistics are systematically lower than with the *p*-value from the ADF statistics. Even by the standard of the large sample bootstrap critical values, none of the MWF-PP tests are significant. The non-parametric correction for serial correlation proposed by Phillips-Perron might be more sensitive in a small sample; in particular, the estimation of the long-term variance could be more imprecise with a reduced number of time observations[12].

INSERT TABLE 8 ABOUT HERE.

The Pedroni cointegration tests

The seven Pedroni cointegration tests are presented in Table 9. Essentially, the results are similar to the previous panel tests. The only specifications that appear to be significant are once more the relationship GDP per capita-Imports or GDP per capita-Trade. Again, the Group and Panel tests based on the Phillips-Perron *t*-stat are not significant. The peculiar result in Table 9 is the high significance of the variance ratio test relatively to the other tests. The variance ratio test takes the ratio of the short-term variance of the residual series over the long-term variance (as estimated by the Newey-West estimator). Again, it might be difficult to estimate correctly the long-term variance in a small sample. Hence, the variance ratio test might not be very reliable in a small sample.

---

[12] We have used the Newey-West estimator to compute the long-term variance. The truncation parameter for the Phillips-Perron test is 2, based on the Newey-West suggestion $q=4(T/100)^{2/9}$.

INSERT TABLE 9 ABOUT HERE.

In all the previous panel tests (LLC, IPS, MWF and Pedroni), the specifications including a time dummy for 1993 are not significant (except the Pedroni *v*-stat in the GDP per capita-Exports specification). In fact, including a level break in 1993 naturally increases the value of the test, but the bootstrap critical values are also higher with such a break (see bootstrap statistics in Appendix 3). By the OLS properties, adding a regressor in the cointegration equation (1) tends to make the residual series more stationary. Hence, it is always the case that the critical values must be adjusted upward for the number of variables in the cointegration relationship.

The SURE cointegration tests

The lag selection for each unit equation in the SURE system is a tricky issue. One method is to estimate each equation individually by OLS, determine the lag length using the usual general-to-specific approach, and then define the appropriate lag structure of each equation to be estimated in the SURE system. The other method is to apply the general-to-specific approach directly to the SURE regressions. But then, removing or adding a lag in one equation might change the significance of the lags in the other equations, and the general-to-specific approach becomes then a cumbersome exercise. We choose to work with the first method. The objective is to whiten the errors from any own autocorrelation within the residuals of a unit, and let the SURE procedure handle any type of cross-correlation patterns; hence, the lag selection from the simple OLS regressions should be sufficient to correct for own autocorrelation.

We exclude Extremadura and Rioja from the SURE system (two ACs having a small economic weight), in order to insure that, with T=17, the time dimension is sufficiently large relatively to the number of cross-sections. Even with this restriction, in the bootstrap simulation (see Appendix 3), the SURE estimator does not converge, on average, for about 10-12 bootstrap replications on 5000 (0,0002%). As for the distribution of the SURE tests in small sample, the standard deviations are much higher with T=17 than with T=60, revealing the instability of the SURE estimator in a small sample.

INSERT TABLE 10 ABOUT HERE.

Keeping in mind this caveat, we present in Table 10 the individual USURE tests. Each unit equation is estimated in the SURE system, keeping the ρ_i coefficients, the fixed-effect and the lag structure heterogeneous. None of the USURE individual *t*-stats are significant, except Catalonia and Murcia for the specification GDP per capita-Trade (akin to their individual ADF test), and Aragon, Catalonia, Navarra and Valenciana for the specification GDP per capita-Imports. For the latter specification, remember though that Extremadura and Rioja, excluded from the USURE system, had also a significant individual ADF *t*-statistic (Table 2).

Table 11 gives the Breusch-Pagan test of independence checking whether the residuals u_{it} (equation 2) are independent across the N equations, e.g. no significant cross-correlations between the regions[13]. As expected, the Breusch-Pagan test can be rejected for all specifications. Thus, cross-correlation between regions should indeed be taken into account in the testing procedure.

INSERT TABLE 11 ABOUT HERE.

Based on the individual USURE estimations, the Taylor and Sarno (1998) MADF test statistic simply corresponds to the Fisher *F*-test of joint significance of all ρ_i coefficients. The MADF tests are printed in Table 12. Given the lack of significance in the individual USURE results, we unsurprisingly find that none of the specifications are jointly significant either. The highest *p*-values are still associated with the specifications using imports or total trade.

INSERT TABLE 12 ABOUT HERE.

In the last panel of Table 12, we finally give the results for the RSURE test, constraining the ρ coefficient to be identical across units. With this homogeneity restriction imposed, we obtain RSURE results that are similar to the other panel tests: the only two significant specifications are the

---

[13] This is a sort of spatial autocorrelation test in time-series since it measures the correlation between regions in the residuals of their ADF equation. Usual spatial autocorrelation tests, such as Moran's statistic, are usually applied in cross-section, not in time series, and they are conditoned on a particular weighting matrix that has to be defined. Here, the SURE variance-covariance matrix allows for any type of correlation between regions. Moreover, it measures the cross-correlation in time between regions in their relationship between GDP per capita and trade. Hence, the time-dimension is taken into account. In Laurin (2007a), we investigate further the pattern of this cross-correlation between regions, using as determinants variables expressing geography, distance, gravity and industrial structure.

relationships GDP per capita-Imports and GDP per capita-Trade, but with a higher significance level (5%) than the previous panel tests (10%).

3.4 Discussion on the cointegration results

Our empirical hypothesis about a strong long-term relationship between GDP and exports seems to be rejected. How can this outcome be explained? First, recall that these results are conditional on the sample period 1988 to 2004. Our cointegration results cannot be generalized to other or longer time periods. In addition, even if they are implemented in panel, the cointegration tests may still have a low power in a small sample, particularly in the presence of breaks (identified in section 3.2) in the relationship which might poorly be taken into account by a time dummy. On top of this low power, we have shown that the cointegration tests in a small sample are very restrictive, since the critical values are adjusted upward. The low power and the more restrictive critical values make it harder to reject the null hypothesis in a small sample.

Similarly, these results are specific to Spain and cannot be generalized to other countries. Past cross-country studies typically show very different outcomes across countries. In particular, the existence (or not) of a cointegration relationship between trade and growth may depend on particular country characteristics (the level of economic development[14] or the level of "openness" to trade[15] for example). Indeed, in Laurin (2007a), we try to examine the extent to which the significance of the cointegration relationship between trade and GDP per capita may be conditional on structural change.

However, it may well be that the relationship between trade and growth is not as strong as hypothesized. The absence of significant results for exports is in line with the already mixed results already obtained in the empirical literature about the effects of exports on economic growth. Trade liberalization may affect economic development through other channels than exports directly (see for example Wacziarg, 2001): foreign direct investment, human or physical capital accumulation, etc. The general concept of openness to trade may be poorly approximated by trade variables alone. Moreover, import competition and trade liberalization can displace local production and lead to the closure of uncompetitive firms. This process

[14] See for example Balassa (1985), Moschos (1989), Sheehey (1992).
[15] See for example Calderón *et al*. (2001), Lee and Huang (2002).

could offset the expected positive effects of trade on growth, at least in the medium term, especially if these effects are dissolved in regional aggregates.

However, we find some evidence in favor of a long-term relationship between import (or total trade) and GDP per capita, contrary to exports. The mechanisms relating GDP to imports might be very different than for exports. For example, one could consider the following stylized story: suppose most productivity gains are to be caused not from exporting, but from internal efficiency improvement at the firm level and from the capital and technologies acquired via FDI. In particular, there are reasons to believe that FDI are a key variable explaining Spain's development[16]. Then, as own firm productivity increases, wages also increase. This higher income stimulates the demand for foreign products. Hence, we would see a close relationship between growth and imports. However, these productivity gains contribute in making Spain more competitive on world markets, and exports start increasing, but this effect only operates for the most productive firms (Melitz, 2003; Melitz and Ottaviano, 2005). Other firms maintain only a domestic market or close down. In total, we would thus see a stronger cointegration relationship for imports than for exports. Yet, the cointegration results for imports could be partly informative of the general relationship between economic integration and Spain's development, since both exports and imports seem to share common characteristics in their evolution.

3.5 Discussion on the cointegration methodology in a small sample

The empirical developments in this paper constitute an interesting exercise for implementing the panel cointegration methodology in a small sample. The gains from exploiting the panel dimension of the data - either by combining the individual statistics or by pooling the data - might compensate the low power of the cointegration tests in a small sample. However, as experienced in this paper, the researcher must confront four main difficulties when dealing with a small sample:

- First, the cointegration hypothesis must be interpreted with cautious. Mainly, in a small sample, the results should be conditional on the period sample. Even if the critical values are adjusted for the small sample bias, one should not try to infer the true long-term "generic" cointegration properties of the

[16] Regional Spanish data on foreign direct investments are available only since 1993. This is really unfortunate since FDI seem to be a key factor for Spain's development. See Alguacil and Orts (2002) on the causal relationship between FDI and exports in Spain.

series from a reduced number of time observations. On the other hand, with the proper caveats, the cointegration methodology should not necessarily be discarded all together with a small sample. The researcher could have some good theoretical or empirical grounds to believe that the variables will behave as cointegrated series for the given period sample, as least sufficiently intensively so as to confer to the cointegration tests a meaningful interpretation. There may exist a particular mechanism relating the variables over the given sample. In our case, we have conjectured that Spain's entry in the EC and the intense process of structural change that followed might represent such a mechanism. Another issue is that, in the panel tests, rejection of H0 is interpreted differently depending on the methodology used. For example, in the LLC test, rejecting H0 implies that there is no unit root for ALL units. In contrast, in the IPS and MWF tests, rejecting H0 implies that not all units contain a unit root, but some may. Hence, extra care should be given to the interpretation of the results.

The three other issues are already well-know facts in the econometric literature:

- The tests may be very sensitive to the lag selection chosen to correct for serial correlation, particularly in a small sample. The problem is even more acute for the SURE estimation method: adding or retrieving a single lag in one of the equation can substantially change the coefficients of the whole system.

- The SURE estimates may be very instable in a small sample, since the variance-covariance matrix may be difficult to estimate with a small number of observations per unit. Fortunately, our panel SURE results are not too far apart from the other panel tests, and we do not get extreme values such as those observed (albeit very infrequently) in the bootstrap simulations.

- The existence of a break in the cointegration relationship can greatly affect the cointegration tests in a small sample. But in a small sample, it is difficult to verify if the break is a true structural break or merely a temporary deviation from the equilibrium relationship.

Conclusion

In this paper, we have used various individual and panel cointegration tests to investigate the relationship between GDP per capita and trade in Spain's Autonomous Communities between 1988 and 2004. The main results are: i) there is some evidence of a significant relationship between GDP per capita and imports, and to a lesser extent, between GDP per capita and total trade (exports plus imports); ii) but we are unable to find a significant relationship between GDP per capita and exports. All these results are conditional on the sample period 1988 to 2004, and specific to Spain's case.

The empirical developments in this paper constitute an interesting exercise for implementing the panel cointegration methodology in a small sample. The gains from exploiting the panel dimension of the data might compensate the low power of the cointegration tests in a small sample. Moreover, the estimating constraints implied by a small sample have been fully taken into account by imposing T=17 in the bootstrap simulations. On the other hand, we have also discussed several important issues when using the cointegration methodology on a reduced time frame, notably the formulation of the hypothesis conditional on the sample period, the appropriate selection of the lag length, the presence of breaks in the cointegration relationship and the instability of the SURE estimates. In any way, using the proper caveats, the cointegration methodology should not necessarily be discarded all together in the presence of a small sample.

We have also shown the usefulness of the bootstrap methodology for panel cointegration tests in the presence of cross-section dependence and of a small sample. The theoretical literature on unit roots and cointegration has carried out great amounts of efforts in the derivation of the asymptotic distribution of the tests and in designing appropriate transformations or normalizations - corrected for nuisance parameters - so that the tests can follow a well-known distribution. The bootstrap procedure allows the researcher to be agnostic about the theoretical distribution of the tests (provided that the bootstrap distribution is well-behaved). It is very flexible and can be implemented for any kind of specifications, in particular when independence of cross-units must be rejected.

Appendix 1: List and Map of Autonomous Communities

INSERT FIGURE A1 ABOUT HERE.

INSERT TABLE A1 ABOUT HERE.

Appendix 2: unit root test on individual series

Tables A2 gives the augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) unit root test. A time trend and a constant are included in the specification of the tests. Except for some isolated cases, we can reject the hypothesis of stationarity: all series have a unit root.

INSERT TABLE A2 ABOUT HERE.

Appendix 3: the bootstrap procedure and properties

The bootstrap algorithm

Following Li and Maddala (1997), Maddala and Wu (1999), Chang, Park and Song (2005) and Banerjee and Carrion-i-Silvestre (2004), the sieve bootstrap is undertaken with the following steps:

Step 1: Fit $Y_{it} = \alpha_i + \beta_i X_{it} + e_{it}$ by OLS and obtain β_i, α_i and the fitted residual $\hat{e}_{it}$. Define $w_{it} = (\Delta x_{it}', \Delta\hat{e}_{it}')'$.

Step 2: Apply the sieve estimation method to the following VAR(p):

(A2) $$w_{it} = \theta_1 w_{it-1} + \ldots + \theta_p w_{it-p} + u_{it}$$

where the order of p is determined by the general-to-specific method, with a maximum of two lags. Calculate the centered fitted residuals $\tilde{u}_{it}$:

(A3) $\tilde{u}_{it} = \hat{u}_{it} - \frac{1}{T}\sum_1^t \hat{u}_{it}$ for $t = 1,\ldots,T$.

To preserve the cross-correlation between units, obtain $\tilde{u}_{it}^*$ by resampling $\tilde{u}_{it}$ with the time index fixed for all units, e.g. resample $\tilde{u}_t = (\tilde{u}_{1t}, \tilde{u}_{2t},\ldots,\tilde{u}_{NT})$. Construct the bootstrap samples w_{it}^* recursively using:

(A4) $w_{it}^* = \theta_1 w_{it-1}^* + \ldots + \theta_p w_{it-p}^* + u_{it}^*$ for $t = 1,\ldots, T+45$.

given the initial values $w_{it}^* = w_{it}$ for $t = 0,\ldots,1 - p$. The first 45 observations are dropped in order to eliminate the dependency on the choice of initial values. Hence, we must generate a bootstrap sample of length T+45. The bootstrap is implemented for T=17. For comparison, we also run the procedure over a bootstrap sample of T=60.

Step 3: Obtain bootstrap sample e_{it}^* and x_{it}^* by integrating w_{it}^*. Then, obtain y_{it}^* using:

(A5) $y_{it}^* = \alpha_i + \beta_i x_{it}^* + e_{it}^*$

Step 4: calculate any of the cointegration tests from this bootstrap sample.

We have considered 5000 bootstrap replications. The bootstrap must be replicated for each specification estimated, each specification having its own cointegrating coefficients (step 1) and cross-correlation structure. Notice in equation (A2) that we impose the unit root on the bootstrap residual e_{it}: the hypothesis of no cointegration (H0) is imposed by design. Therefore, the distribution of the test computed on the bootstrap sample should represent the lower-bound values under which H0 cannot be rejected. We can then find the appropriate critical values for the tests. For the tests involving the SURE estimation method, Extremadura and Rioja are excluded from the system.

Bootstrap statistics and properties

To illustrate the behaviour of the bootstrap simulations, Tables A4 shows the bootstrap statistics and critical values for each cointegration test, concentrating on the relationship GDP per capita–Export[17], and with a time dummy for 1993 (dum93). For the individual and USURE ADF *t*-tests, we have 75 000 bootstrap observations, since 15 individual statistics are computed in each bootstrap replication, except for the USURE test with T=17, for which there is 65 000 observations since we exclude two ACs.

Typically, the critical values are higher (larger negative values) for the small sample T=17 than the large sample T=60. There are two exceptions. The first exception is when a time dummy for 1993 is added to the cointegration equation, for some of the tests (MWF-PP test, the IPS test, the Pedroni non-parametric Group *t*-stat). Introducing a level break in the cointegration equation tends to have a larger effect in a small sample than a large sample. The second exception concerns two Pedroni tests: the variance ratio test and the Panel *rho*-stat. As already mentioned, in a small sample, it might be difficult to estimate correctly the long-term variance.

The IPS statistics have the exact same mean as the individual ADF *t*-tests, which is expected since the IPS test is an average of each individual *t*-stat. However, notice that the standard deviation – and therefore, the critical values - of the IPS statistics are systematically lower than for the individual *t*-stats. This is the panel effect. Hence, the information brought by other cross-units enhances the robustness of the test.

Contrary to the other tests, we know that the Pedroni statistics, specifically designed for the residual-based cointegration test, should follow a standard normal distribution *N*(0,1). Taking the large sample T=60, we can then compare the bootstrap distribution with the standard normal distribution to investigate the particular properties and behaviour of our bootstrap simulation. We observe that the mean of the bootstrap Pedroni statistics are different from zero and the standard deviation is higher than 1. This bias may be explained by two factors. First, the bias may represent the effect of the cross-correlation between units, which is taken into account in the bootstrap algorithm. Second, the bootstrap distribution depends of course on the particular model from which it is constructed. For example, in the

[17] Bootstrap statistics for the other specifications are available upon request.

case of the individual ADF *t*-stat, the MacKinnon (1991) estimated critical values (for two variables, a constant and T=17) are shown in Table A3 next to the bootstrap values. We can see that the bootstrap critical values are not quite equal to the ones estimated by MacKinnon (1991). The Dickey-Fuller and MacKinnon distributions are estimated by Monte Carlo simulations for very general cases, generating random regression coefficients, and involving a random number of lags. In our case, the bootstrap involves given coefficients and number of lags, as estimated in step 1 and 2.

The behaviour of the SURE tests in small sample is less satisfactory than for the other tests. For T=17, the SURE tests do not converge, on average, for about 10-12 bootstrap replications on 5000 bootstraps (0,0002%). As for the distribution of the SURE statistics, the standard deviation is much higher with T=17 than with T=60, revealing the instability of the SURE estimator in a small sample. We also find some extreme SURE test values (values laying far beyond the average distribution of the sample) for about 2,04% of bootstrap replications on average. This skewness and the large standard deviation in the small sample case can be explained by two factors. First, with N=13 and T=17, the SURE results may be instable, even if we have discarded two units. Second, the SURE system seems to be very sensitive to the lag structure. In fact, most extreme values are the results of the chosen lag structure. Adding or retrieving a single lag of the equation in one of the unit equation may be sufficient to change substantially the test values of the whole system, bringing any extreme value within the tails of the distribution. Otherwise, the large sample (T=60) SURE bootstrap distributions are well behaved.

INSERT TABLE A3 ABOUT HERE.

Bibliography

Ahmad, J. (2001) Causality between Exports and Economic Growth: What do the Econometric Studies Tell Us? *Pacific Economic Review*, **6**, 147-167.

Alguacil, M. T. and Orts, V. (2002) A Multivariate Cointegrated Model Testing for Temporal Causality between Exports and Outward Foreign Investment: The Spanish Case, *Applied Economics*, **34**, 119-132.

Arrow, K. J. (1962) The Economic Implications of Learning by Doing, *Review of Economics Studies*, **29**, 155-73.

Balaguer, J. and Cantavella-Jorda, M. (2004) Structural Change in Exports and Economic Growth: Cointegration and Causality Analysis for Spain, *Applied Economics*, **36**, 473-477.

Balassa, B. (1978) Export and Economic Growth: Further Evidence, *Journal of Development Economics*, **5**, 181-189.

_________ (1985) Exports, Policy choices and Economic Growth in Developing Countries after the 1973 Oil Crisis, *Journal of Development Economics*, **18**, 23-35.

Baldwin, R. (1992a) Measurable Dynamic Gains from Trade, *Journal of Political Economy*, **100**, 162-174.

_________ (1992b) On the Growth Effects of Import Competition, *NBER Working Paper* No. 4045.

Banerjee, A. (1999) Panel Data Units and Cointegration: an Overview, *Oxford Bulletin of Economics and Statistics*, **61**, 607-629.

Banerjee, A. and Carrion-I-Silvestre, J.L. (2004) Breaking Panel Data Cointegration, *Mimeo*.

Bernard, A. B. and Eaton, J. and Jensen, J.B., and Kortum, S.S. (2000) Plants and Productivity in International Trade, *NBER Working Paper* No. 7688.

Calderón, C. and Chong, A. and Zanforlin, L. (2001) On Non-linearities between Exports of Manufactures and Economic Growth, *Journal of Applied Economics*, **4**, 279-311.

Campbell, J.Y. and Perron, P. (1991) Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots, in *Macroeconomics Annual vol. 6*, (Eds) O.J. Blanchard and S. Fischer, The MIT Press, Cambridge.

Chang, Y. (2004) Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency, *Journal of Econometrics*, **120**, 263-293.

Chang, Y. and Park, J.Y. and Song, K. (2005) Bootstrapping Cointegrating Regressions, *Journal of Econometrics*, in press.

Corden, W. M. (1971) The Effects of Trade on the Rate of Growth, in *Trade, Balance of Payments and Growth*, (Eds) J. Bhagwati *et al.*, North-Holland, 117-143.

Dickey, D. and Fuller, W. (1979) Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, **74**, 427-431.

Dritsakis, N. (2004) Exports, Investments and Economic Development of pre-accession Countries of the European Union: an Empirical Investigation of Bulgaria and Romania, *Applied Economics*, **36**, 1831-1838.

Emery, R. F. (1967) The Relation of Exports and Economic Growth. *Kyklos*, **20**, 470-86.

Engle, R. F. and Granger, C. W. J. (1987) Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica*, **55**, 251-276.

Ethier, W. (1979) Internationally Decreasing Costs and World Trade, *Journal of International Economics*, **9**, 1-24.

_______________ (1982) National and International Returns to Scale in the Modern Theory of International Trade, *The American Economic Review*, **72**, 950-959.

Fachin, S. (2005) Long-Run Trends in Internal Migrations in Italy: a Study in Panel Cointegration with Dependent Units, *Econometrics EconWPA*, University of Rome La Sapienza.

Fujita, M. and Krugman, P. and Venables, A. J. (1999) *The Spatial Economy: Cities, Regions, and International Trade*, Cambridge University Press, New York.

Gregory, A. and Hansen, B. (1996) Residual-based Tests for Cointegration in Models with Regime Shifts, *Journal of Econometrics*, **70**, 99-126.

Granger, C. (1969) Investigating Causal Relationship by Econometric Models and Cross-Spectral Methods, *Econometrica*, **37**, 424-438.

Granger, C.W.J. and Newbold, P. (1974) Spurious Regression in Econometrics, *Journal of Econometrics*, **2**, 111-120.

Grossman, G. and Helpman, E. (1991a) Growth and Welfare in a Small Open Economy, in *International Trade and Trade Policy*, (Eds) E. Helpman and A. Razin, The MIT Press, Cambridge.

_______________________ (1991b) Endogeneous Growth: Trade, Knowledge Spillovers, and Growth, *European Economic Review*, **35**, 517-526.

_______________________ (1995) Technology and Trade, in *Handbook of International Economics vol. 3*, (Eds) G. Grossman and K. Rogoff, Elsevier Science, 1279-1335.

Hall, A.D. (1994) Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection, *Journal of Business and Economic Statistics*, **12**, 461-470.

Helpman, E. (1981) International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition: A Chamberlin-Heckscher-Ohlin Approach, *Journal of International Economics*, **11**, 305-340.

Im, K.S. and Pesaran, M.H. and Shin, Y. (2003) Testing for Unit Roots in Heterogeneous Panels, *Journal of Econometrics,* **115**, 53-74.

Johansen, S. (1988) Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, **12**, 231-254.

Johnson, H. G. (1967) The Possibility of Income Loss from Increased Efficiency of Factor Accumulation in the Presence of Tariffs, *Economic Journal*, **77**, 151-154.

Laurin, F. (2007a) Trade, Growth and Structural Change in Spain's Autonomous Communities, *Université catholique de Louvain*, *Centre for Operations research and Econometrics* (Louvain-la-Neuve, Belgium), June 2007.

_________ (2007b) Mapping Specialization into Exports in Spain's Regions, *Université catholique de Louvain*, *Centre for Operations research and Econometrics* (Louvain-la-Neuve, Belgium), June 2007.

_________ (2007c) *Trade, Specialization and Economic Growth in Spain's Autonomous Communities*, Doctoral dissertation, *Université catholique de Louvain,* Faculté des sciences économiques, sociales et politiques, Département des sciences économique.

Lee, C.-H and Huang, B.-N. (2002) The Relationship between Exports and Economic Growth in East Asian Countries: A Multivariate Threshold Autoregressive Approach, *Journal of Economic Development*, **27**, 45-67.

Levin, A. and Lin, C.F. (1992) Unit Root Test in Panel Data: Asymptotic and Finite Sample Properties, University of California at San Diego, *Discussion Paper* No. 92-93.

Levin, A. and Lin, C.F and James Chu, C.-S. (2002) Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties, *Journal of Econometrics*, **108**, 1-24.

Lewer, J.J. and Van den Berg, H. (2003) How Large is International Trade's Effect on Economic Growth? *Journal of Economic Surveys*, **17**, 363-396.

Li, H. and Maddala, G.S. (1997) Bootstrapping Cointegrating Regressions, *Journal of Econometrics*, **80**, 297-318.

Lucas, R. E. (1988) On the Mecanics of Economic Development, *Journal of Monetary Economics*, **22**, 3-42.

MacKinnon, J.G. (1991) Critical Values for Co-integration Tests, in *Long-Run Economic Relationships*, (Eds) R.F. Engle and C.W.J. Granger, Oxford University Press.

Maddala, G.S. and Wu, S. (1999) A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics*, **61**, number supplement 1, 631-652.

Marin, D. (1992) Is the Export-led Growth Hypothesis Valid for Industrialized Countries? *The Review of Economics and Statistics*, **74**, 678-88.

Melitz, M.J. and Ottaviano, G.I.P. (2003) Market Size, Trade, and Productivity, *NBER Working Paper Series* No. 11393, June 2005.

Michaely, M. (1977) Export and Growth: An Empirical Investigation, *Journal of Development Economics*, **4**, 49-53.

Moschos, G. (1989) Export Expansion, Growth and the Level of Economic Development, *Journal of Development Economics*, **30**, 93-102.

Pedroni, P. (1999) Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors, O*xford Bulletin of Economics and Statistics,* **61**, 653-670.

________ (2001) Purchasing Power Parity Tests in Cointegrated Panels, *The Review of Economics and Statistics*, **83**, 727-731.

________ (2004) Panel Cointegration; Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the Purchasing Power Parity Hypothesis, *Econometric Theory*, **20**, 597-325.

Pereira, A. M. and Xu, Z. (2000) Export Growth and Domestic Performance, *Review of International Economics*, **8**, 63-73.

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

Pesaran, M.H., and Smith, R. (1995) Estimating Long-run Relationships from Dynamic Heterogeneous Panels, *Journal of Econometrics*, 68, 79-113.

Pesaran, M.H. and Smith, R. and Shin, Y. (1999) Pooled Mean Group Estimation of Dynamic Heterogenous Panels, *Journal of the American Statistical Society,* **94**, 621-634.

Phillips, P.C.B. and Ouliaris, P. (1990) Asymptotic Properties of Residual Based Tests for Cointegration, *Econometrica*, **58**, 165-193.

Phillips, P.C.B. and Perron, P. (1988) Testing for a Unit Root in Time Series Regression, *Biometrika*, **75**, 335-346.

Rivera-Batiz, L.A. and Romer, P. (1991) International Trade with Endogenous Technological Change, *European Economic Review*, **35**, 971-1004.

Romer, P.M. (1986) Increasing Returns and Long-run Growth, *Journal of Political Economy*, **94**, 1002-1037.

___________ (1990) Endogenous Technological Change, *Journal of Political Economy*, **98**, 72-102.

Shan, J. and Gang Tian, G. (1998) Causality between Exports and Economic Growth: the Empirical Evidence from Shanghai, *Australian Economic Papers*, **37**, 195-202.

Sheehey, E.J. (1990) Exports and Growth: a Flawed Framework, *Journal of Development Studies*, **27**, 111-116.

___________ (1992) Exports and Growth: Additional Evidence, *Journal of Development Studies*, **28**, 730-34.

Sims, C. A. (1972) Money, Income, and Causality, *American Economic Review*, **62**, 540-552.

Taylor, M.P. and Sarno, L. (1998) The Behavior of Real Exchange Rates during the Post-Bretton Woods Period, *Journal of International Economics*, **46**, 281-312.

Tyler, W. G. (1981) Growth and Export Expansion in Developing Countries, *Journal of Development Economics*, **9**, 121-130.

Xu, Z. (1996) On the Causality between Export Growth and GDP Growth: An Empirical Reinvestigation, *Review of International Economics*, **4**, 172-184.

Young, A. (1991) Learning-by-doing and the Dynamic Effects of International Trade, *Quarterly Journal of Economics*, **106**, 369-405.

Zellner, A. (1962) An Efficient Method of Estimating Seemingly Unrelated Regressions and Test for Aggregation Bias, *Journal of the American Statistical Association*, **57**, 348-368.

Aknowledgement: The research presented in this paper was done at the Center for Operation Research and Econometrics (CORE). I would like to thank Jacques Thisse, Dominique Peeters, Hylke Vandenbussche and Luc Bauwens for helpful comments and suggestions. This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State,

Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the author.

Table 1: The cointegration tests

Individual tests	
• Engel-Granger Augmented Dickey-Fuller test (ADF)	
• Engel-Granger Phillips-Perron test (PP)	
• Unrestricted SURE ADF (USURE)	
Panel tests	
The "pooled" test	**The "combining" test**
• Levin-Lin-Chu (LLC)	• Im-Pesaran-Shin (IPS)
• Pedroni Panel variance ratio	• Maddala-Wu Fisher (MFW)
• Pedroni Panel *rho*-stat	• Pedroni Group *rho*-stat
• Pedroni Panel PP *t*-stat (non-parametric)	• Pedroni Group PP *t*-stat (non-parametric)
• Pedroni Panel ADF *t*-stat (parametric)	• Pedroni Group ADF *t*-stat (parametric)
• Restricted SURE ADF (RSURE)	• Sarno & Taylor's Multivariate ADF (MADF)

Table 2: Individual cointegration test (in log)

Cointegration between GDP per capita and exports						**Cointegration between GDP per capita and imports**					
	ADF Tests				**PP test**		**ADF Tests**				**PP test**
	***t*-stat**	**lags**	***Q*-statistic**		***t*-stat**		***t*-stat**	**lags**	***Q*-statistic**		***t*-stat**
Andalusia	-2.127	(1)	7,24	(0,20)	-1,582	Andalusia	-2,367	(0)	3,70	(0,72)	-2,470
Aragon	-2.333	(1)	1,78	(0,88)	-1,751	Aragon	-2,404	(1)	1,51	(0,91)	-1,677
Asturias	-3.145	(0)	4,74	(0,58)	-3,141	Asturias	-2,922	(0)	5,80	(0,45)	-2,973
Cantabria	-2.187	(1)	3,37	(0,64)	-1,363	Cantabria	-1,694	(1)	2,08	(0,84)	-1,347
Catalonia	-3.889**	(1)	2,95	(0,71)	-2,641	Catalonia	-2,537	(1)	2,78	(0,73)	-1,893
C.Leon	-1.415	(1)	2,19	(0,82)	-0,961	C.Leon	-1,397	(1)	1,51	(0,91)	-1,001
C.Mancha	-2.111	(0)	7,46	(0,28)	-2,335	C.Mancha	-3,782**	(2)	3,93	(0,56)	-3,430*
Extremadura	-2.440	(1)	1,02	(0,96)	-1,789	Extremadura	-4,436**	(1)	2,65	(0,75)	-2,509
Galicia	-2.549	(0)	2,64	(0,85)	-2,548	Galicia	-2,958	(0)	1,30	(0,97)	-2,936
Madrid	-2.196	(1)	0,44	(0,99)	-1,676	Madrid	-2,378	(0)	5,98	(0,43)	-2,443
Murcia	-5.052***	(2)	3,08	(0,69)	-2,231	Murcia	-2,860	(2)	1,17	(0,95)	-2,331
Navarra	-0.987	(0)	6,05	(0,42)	-1,232	Navarra	-2,440	(0)	2,34	(0,89)	-2,363
P.Vasco	-1.066	(0)	13,23	(0,04)	-1,371	P.Vasco	-2,448	(0)	2,70	(0,85)	-2,482
Rioja	-3.702*	(1)	4,60	(0,47)	-2,695	Rioja	-4,326**	(1)	0,63	(0,99)	-2,981
Valenciana	-3.044	(1)	3,08	(0,69)	-1,651	Valenciana	-3,620*	(1)	1,33	(0,93)	-1,747

The figure in parentheses besides the ADF *t*-stat is the lag selection. Lag selection: 10% significance of last lag, max lag = 2. The *Q*-statistic refers to the Ljung-Box test for serial correlation. The *p*-values are given in parentheses after the *Q*-statistic. The truncation parameter for the Phillips-Perron test is 2, based on the Newey-West suggestion $q=4(T/100)^{2/9}$. Significance based on the critical values estimated by MacKinnon (1991). *= significant at 10%; **=significant at 5%; ***=significant at 1%.

Table 3: Individual cointegration test between GDP per capita and total trade (in log)

	ADF Tests				**PPtest**
Region	***t*-stat**	**lags**	***Q*-statistic**		**t-stat**
Andalusia	-1.963	(1)	2.76	(0.74)	-1.767
Aragon	-2.595	(1)	3.51	(0.62)	-1.677
Asturias	-3.290	(0)	5.89	(0.44)	-3.320*
Cantabria	-2.136	(1)	2.50	(0.78)	-1.312
Catalonia	-3.773**	(1)	6.41	(0.27)	-2.529
C.Leon	-1.474	(1)	1.09	(0.95)	-0.963
C.Mancha	-3.559*	(2)	3.54	(0.62)	-2.961
Extremadura	-3.200	(1)	2.77	(0.73)	-2.028
Galicia	-2.651	(0)	1.82	(0.94)	-2.631
Madrid	-1.868	(0)	7.61	(0.27)	-2.004
Murcia	-5.786***	(2)	1.32	(0.93)	-2.370
Navarra	-1.426	(0)	3.08	(0.80)	-1.464
P.Vasco	-1.338	(0)	6.16	(0.41)	-1.538
Rioja	-3.481*	(2)	4.54	(0.47)	-2.835
Valenciana	-3.584*	(1)	1.31	(0.93)	-1.714

The figure in parentheses besides the ADF *t*-stat is the lag selection. Lag selection: 10% significance of last lag, max lag = 2. The *Q*-statistic refers to the Ljung-Box test for serial correlation. The *p*-values are given in parentheses after the *Q*-statistic. The truncation parameter for the Phillips-Perron test is 2, based on the Newey-West suggestion $q=4(T/100)^{2/9}$. Significance based on the critical values estimated by MacKinnon (1991). *= significant at 10%; **=significant at 5%; ***=significant at 1%.

Table 4: Gregory and Hansen (1996) methodology

	Optimal level break around 1993			Optimal trend break around 1993		
	coefficient	ADF *t*-stat	year	coefficient	ADF *t*-stat	year
Andalusia	-0.137***	-2.216	1993	-0.094***	-2.185	1993
Aragon	-0.117**	-2.022	1993	-0.046	-2.538	1991
Asturias	-0.008	-3.299	1994	0.092	-3.344	1994
Cantabria	-0.156***	-2.223	1993	0.359***	-3.078	1995
Catalonia	-0.066***	-3.116	1994	-0.027***	-3.061	1994
C.Leon	-0.111***	-1.780	1992	-0.125**	-1.684	1992
C.Mancha	-0.121**	-2.384	1994	0.279***	-3.123	1995
Extremadura	-0.114***	-3.678	1993	0.074***	-4.025	1994
Galicia	-0.063**	-3.941	1994	-0.154***	-3.303	1993
Madrid	-0.121***	-3.104	1992	-0.101***	-3.124	1992
Murcia	-0.082***	-3.507	1993	-0.807*	-3.412	1991
Navarra	-0.142***	-2.135	1992	0.044	-1.688	1994
P.Vasco	-0.160***	-2.508	1994	-0.101***	-2.387	1994
Rioja	-0.176**	-4.005	1994	0.091*	-4.033	1994
Valenciana	-0.081**	-2.636	1992	-0.044**	-2.576	1992

All variables in log. *= significant at 10%; **=significant at 5%; ***=significant at 1%.

Table 5: Panel LLC cointegration test with a fixed-effect and on one lag by unit.

			Bootstrap *p*-value	
Specification (in log)	**ρ coefficient**	**LLC *t*-stat**	**small**	**long**
GDP per capita–Exports	-0.539	-2.809	0.4170	0.1360
GDP per capita–Trade	-0.472	-2.388	0.6048	0.3222
GDP per capita–Imports	-0.527	-2.999	0.3922	0.1538
GDP per capita–Exports–dum93	-0.839	-5.130	0.1346	0.0444
GDP per capita–Trade–dum93	-0.814	-4.550	0.2208	0.1190

Note: the LLC t-stat is a normalized *t*-statistic. See Levin *et al*. (2002). Total number of observations: 230, N,T = (15,17). *= significant at 10%; **=significant at 5%; ***=significant at 1%.

Table 6: Unit root test with constrained ρ coefficient, but a flexible lag structure

			Bootstrap *p*-value	
Specification (in log)	**ρ coefficient**	***t*-stat**	**small**	**long**
GDP per capita–Export	-0.3618	-7.976	0.2250	0.2060
GDP per capita–Trade	-0.4102	-8.646	0.1062	0.0812
GDP per capita–Import	-0.4656	-8.942*	0.0784	0.0432
GDP per capita–Export–dum93	-0.6791	-9.338	0.3792	0.5872
GDP per capita–Trade–dum93	-0.6705	-9.297	0.3898	0.5992

Total number of observations: 230, N,T = (15,17). Lag selection: 10% significance of last lag, max lag = 2. *= significant at 10%; **=significant at 5%; ***=significant at 1%.

Table 7: Panel IPS cointegration test with a constant for each unit

		Bootstrap *p*-value	
Specification (in log)	**IPS *t*-stat**	**small**	**long**
GDP per capita–Exports	-2.470	0.1986	0.1082
GDP per capita–Trade	-2.721*	0.0676	0.0154
GDP per capita–Imports	-2.739*	0.0556	0.0086
GDP per capita–Exports–dum93	-3.090	0.3018	0.2542
GDP per capita–Trade–dum93	-3.159	0.2384	0.1920

Note: total number of observations: 230. N,T = (15,17). Lag selection: 10% significance of last lag, max lag = 2. *= significant at 10%; **=significant at 5%; ***=significant at 1%.

Table 8: Panel MWF cointegration test

	With ADF test			With PP test		
		Bootstrap *p*-value			Bootstrap *p*-value	
Specification (in log)	**MWF F-stat**	**small**	**long**	**MWF F-stat**	**small**	**long**
GDP per capita–Export	80.92	0.2094	0.0716	38.11	0.7108	0.7182
GDP per capita–Trade	101.94*	0.0734	0.0082	45.43	0.5198	0.5282
GDP per capita–Import	94.08*	0.0962	0.0102	56.54	0.3444	0.2724
GDP per capita–Export–dum93	121.88	0.3924	0.3464	98.47	0.5210	0.7752
GDP per capita–Trade–dum93	126.95	0.3276	0.2768	106.03	0.4064	0.6682

Note: total number of observations: 230. N,T = (15,17). Lag selection: 10% significance of last lag, max lag = 2.
*= significant at 10%; **=significant at 5%; ***=significant at 1%.

Table 9: Pedroni cointegration tests

	Panel Statistics ("pooled")											
	Variance ratio			*rho*-stat (ρ-statistic)			PP *t*-stat (non-parametric)			ADF *t*-stat (parametric)		
		Boot. *p*-value			Boot. *p*-value			Boot. *p*-value			Boot. *p*-value	
Specification (in log)	*v*-stat	small	long	ρ-stat	small	long	*t*-stat	small	long	*t*-stat	small	long
GDP per capita–Export	1.747*	0.0774	0.1636	-0.210	0.3660	0.4794	-0.547	0.5194	0.4380	-2.002	0.2686	0.1348
GDP per capita–Trade	1.948**	0.0462	0.1222	-0.527	0.2266	0.3642	-1.054	0.3374	0.2692	-3.238*	0.0644	0.0162
GDP per capita–Import	2.373**	0.0158	0.0616	-1.255*	0.0860	0.1974	-2.131	0.1202	0.0814	-4.238**	0.0118	0.0010
GDP per capita–Export–dum93	2.147	0.1072	0.1338	-0.483	0.3960	0.7556	-2.446	0.3300	0.3692	-2.702	0.3674	0.2294
GDP per capita–Trade–dum93	2.253*	0.0822	0.1160	-0.630	0.3174	0.7098	-2.880	0.2226	0.2366	-3.685	0.1546	0.0614

	Group Statistics ("combining")								
	rho-stat (ρ-statistic)			PP *t*-stat (non-parametric)			ADF *t*-stat (parametric)		
		Boot. *p*-value			Boot. *p*-value			Boot. *p*-value	
Specification (in log)	ρ-stat	small	long	*t*-stat	small	long	*t*-stat	small	long
GDP per capita–Export	1.401	0.5452	0.7730	0.465	0.6608	0.6258	-2.477	0.2662	0.0960
GDP per capita–Trade	1.020	0.3428	0.6790	-0.223	0.4644	0.4290	-3.863*	0.0798	0.0100
GDP per capita–Import	0.400	0.1534	0.5246	-1.326	0.2396	0.1826	-3.708*	0.0822	0.0112
GDP per capita–Export–dum93	0.863	0.3744	0.9802	-2.497	0.3942	0.5478	-3.578	0.3420	0.2016
GDP per capita–Trade–dum93	0.718	0.3020	0.9746	-2.824	0.3160	0.4388	-3.829	0.2864	0.1590

Note: total number of observations: 230. N,T = (15,17). Lag selection: 10% significance of last lag, max lag = 2. *= significant at 10%; **=significant at 5%; ***=significant at 1%.

Table 10: Individual unrestricted SURE Dickey-Fuller test (USURE)

	GDP p.c.-Exports			GDP p.c.-Trade			GDP p.c.-Imports		
		Boot *p*-value			Boot *p*-value			Boot *p*-value	
	t-test	small	large	t-test	small	large	t-test	small	large
Andalusia	-2.688	0.881	0.634	-4.618	0.576	0.110	-5.027	0.486	0.030
Aragon	-4.266	0.713	0.166	-5.245	0.487	0.109	-12.47**	0.045	0.000
Asturias	-7.311	0.328	0.000	-6.518	0.327	0.006	-3.433	0.780	0.375
Cantabria	-4.818	0.683	0.122	-5.980	0.414	0.021	-6.985	0.257	0.001
Catalonia	-7.579	0.344	0.001	-11.195*	0.054	0.000	-9.987*	0.076	0.000
C.Leon	-2.953	0.879	0.630	-3.936	0.719	0.370	-7.635	0.245	0.001
C.Mancha	-5.292	0.575	0.069	-5.823	0.406	0.030	-8.249	0.156	0.000
Extremadura	-	-	-	-	-	-	-	-	-
Galicia	-6.171	0.464	0.011	-7.436	0.199	0.001	-9.247	0.117	0.000
Madrid	-4.151	0.778	0.362	-4.946	0.593	0.137	-7.092	0.260	0.001
Murcia	-7.868	0.329	0.001	-11.126*	0.052	0.000	-9.597	0.110	0.000
Navarra	-1.921	0.929	0.828	-3.857	0.673	0.232	-10.364*	0.072	0.000
P.Vasco	-2.416	0.921	0.760	-5.270	0.478	0.045	-8.921	0.136	0.000
Rioja	-	-	-	-	-	-	-	-	-
Valenciana	-4.208	0.741	0.236	-8.154	0.175	0.000	-13.713**	0.030	0.000

	GDP p.c.-Exports-dum93			GDP p.c.-Trade-dum93		
		Boot *p*-value			Boot *p*-value	
	t-test	small	large	t-test	small	large
Andalusia	-4.702	0.919	0.643	-7.001	0.725	0.074
Aragon	-8.887	0.580	0.017	-9.104	0.556	0.012
Asturias	-12.185	0.247	0.000	-8.940	0.521	0.006
Cantabria	-7.163	0.719	0.101	-8.353	0.586	0.021
Catalonia	-3.957	0.955	0.879	-6.744	0.761	0.148
C.Leon	-5.181	0.873	0.435	-6.839	0.708	0.071
C.Mancha	-4.888	0.885	0.430	-3.948	0.941	0.743
Extremadura	-	-	-	-	-	-
Galicia	-9.006	0.493	0.003	-10.522	0.352	0.000
Madrid	-12.987	0.178	0.000	-10.576	0.331	0.000
Murcia	-8.942	0.552	0.012	-6.438	0.811	0.259
Navarra	-5.675	0.809	0.143	-4.757	0.887	0.375
P.Vasco	-5.311	0.867	0.491	-8.736	0.510	0.004
Rioja	-	-	-	-	-	-
Valenciana	-7.572	0.695	0.082	-8.923	0.550	0.010

Note: All variables in logs. Extremadura and Rioja excluded from SURE system. Total number of observations: 221. N,T = (13,17). Bootstrap implemented for T=17 and N=13. Each individual equation has is own constant and lag structure. Lag selection: 10% significance of last lag, max lag = 2, from the OLS individual regressions. *= significant at 10%; **=significant at 5%; ***=significant at 1%.

Table 11: Breusch-Pagan test of independence

Specification	Breusch-Pagan	*p*-value
GDP per capita–Exports	227.682	(0.000)
GDP per capita–Trade	157.261	(0.000)
GDP per capita–Imports	119.938	(0.001)
GDP per capita–Exports–dum93	316.502	(0.000)
GDP per capita–Trade–dum93	266.919	(0.000)

Independence of the residuals from the USURE system of N equations.

Table 12: Multivariate augmented Dickey-Fuller (MADF) test and Restricted SURE test (RSURE)

	MADF test			Constrained RSURE			
		Boot *p*-value				Boot *p*-value	
Specification (in log)	F-test	small	large	Coefficient	*t*-stat	small	large
GDP per capita–Exports	371.10	0.668	0.000	-0.3810	-11.235	0.507	0.000
GDP per capita–Trade	747.83	0.115	0.000	-0.3532**	-18.470	0.035	0.000
GDP per capita– Imports	665.85	0.135	0.000	-0.4098**	-33.328	0.003	0.000
GDP per capita–Exports–dum93	616.09	0.438	0.000	-0.7501	-10.673	0.865	0.259
GDP per capita–Trade–dum93	298.97	0.890	0.000	-0.6696	-13.371	0.512	0.000

Note: All variables in logs. Extremadura and Rioja excluded from SURE system. Total number of observations: 221. N,T = (13,17). Bootstrap implemented for T=17 and N=13. Lag selection: 10% significance of last lag, max lag = 2, from the OLS individual regressions. Taylor-Sarno MADF test with a flexible lag structure and a constant for each unit. RSURE test with homogeneous coefficients, but flexible lag structure and a constant for each unit. *= significant at 10%; **=significant at 5%; ***=significant at 1%.

Table A1: List of Autonomous Communities (and abbreviations)

Andalusia	Galicia
Aragon	Madrid
Asturias	Region of Murcia (Murcia)
Cantabria	Comunidad Floral de Navarra (Navarra)
Catalonia	Basque country (P.Vasco)
Castilla y Leon (C.Leon)	La Rioja (Rioja)
Castilla y Mancha (C.Mancha)	Comunidad Valenciana (Valenciana)
Extremadura	

Table A2: Unit root test: variables in level (trend and constant included)

Real GDP (in log)	ADF				PP
	t-stat	lags	Ljung-Box		*t*-stat
Andalusia	-1.501	(1)	4.74	(0.8560)	-1.704
Aragon	-2.503	(1)	6.77	(0.6610)	-2.392
Asturias	-1.850	(1)	19.76	(0.0195)	-2.080
Cantabria	-2.220	(1)	5.89	(0.7507)	-1.565
Catalonia	-3.624*	(1)	5.11	(0.8246)	-3.080
C.Leon	-1.393	(1)	11.38	(0.2500)	-2.849
C.Mancha	-3.716**	(2)	6.61	(0.6768)	-2.475
Extremadura	-2.138	(1)	11.65	(0.2335)	-2.179
Galicia	-1.335	(1)	7.35	(0.6913)	-1.088
Madrid	-1.987	(1)	8.89	(0.4474)	-1.601
Murcia	-1.467	(1)	6.28	(0.7113)	-1.264
Navarra	-2.506	(2)	3.12	(0.9592)	-1.479
P.Vasco	-1.723	(1)	10.21	(0.3333)	-1.019
Rioja	-2.707	(1)	10.91	(0.2815)	-2.576
Valenciana	-2.334	(1)	8.97	(0.4398)	-1.536

Real Export (in log)	ADF				PP
	t-stat	lags	Ljung-Box		*t*-stat
Andalusia	-2.839	(1)	3.58	(0.6109)	-2.499
Aragon	-2.172	(1)	3.37	(0.6418)	-1.940
Asturias	-1.406	(0)	2.32	(0.8876)	-1.139
Cantabria	-3.051	(1)	4.18	(0.5233)	-2.481
Catalonia	-2.101	(1)	2.18	(0.8233)	-2.190
C.Leon	-2.954	(1)	2.66	(0.7515)	-2.006
C.Mancha	-3.093	(0)	5.53	(0.4768)	-3.136
Extremadura	-2.727	(1)	1.39	(0.9244)	-2.346
Galicia	-2.074	(0)	4.14	(0.6564)	-2.076
Madrid	-1.873	(1)	5.95	(0.3104)	-1.978
Murcia	-1.995	(1)	1.94	(0.8571)	-2.297
Navarra	-2.159	(0)	6.54	(0.3649)	-2.121
P.Vasco	-2.124	(0)	6.85	(0.3348)	-2.163
Rioja	-1.703	(1)	2.65	(0.7524)	-1.604
Valenciana	-1.813	(0)	5.11	(0.5296)	-1.900

Real Import (in log)	ADF				PP
	t-stat	lags	Ljung-Box		*t*-stat
Andalusia	-3.042	(1)	3.03	(0.6953)	-2.389
Aragon	-3.020	(1)	1.56	(0.9051)	-2.128
Asturias	-1.438	(0)	2.32	(0.8878)	-1.729
Cantabria	-2.279	(0)	3.44	(0.7507)	-2.136
Catalonia	-1.269	(0)	3.71	(0.7153)	-1.405
C.Leon	-3.380**	(1)	5.47	(0.3603)	-1.717
C.Mancha	-1.662	(2)	2.52	(0.7720)	-2.237
Extremadura	-1.952	(1)	5.29	(0.3807)	-1.174
Galicia	-1.832	(0)	4.31	(0.6339)	-1.367
Madrid	-2.374	(1)	4.22	(0.5176)	-1.561

Real GDP per capita (in log)	ADF				PP
	t-stat	lags	Ljung-Box		*t*-stat
Andalusia	-2.356	(1)	2.40	(0.8790)	-1.646
Aragon	-3.524*	(1)	2.48	(0.8703)	-2.133
Asturias	-1.866	(0)	3.53	(0.8317)	-1.877
Cantabria	-2.280	(1)	3.48	(0.7453)	-1.816
Catalonia	-2.803	(1)	6.31	(0.3885)	-1.506
C.Leon	-1.214	(1)	3.80	(0.7037)	-1.253
C.Mancha	-3.761**	(1)	5.31	(0.5037)	-2.429
Extremadura	-1.295	(0)	11.84	(0.1057)	-1.581
Galicia	-2.043	(0)	4.06	(0.7717)	-2.073
Madrid	-2.573*	(1)	4.83	(0.5654)	-1.795

Murcia	-0.857	(0)	4.16	(0.6541)	-1.112	Murcia	-3.399*	(2)	1.07	(0.9824)	-1.851
Navarra	-2.220	(0)	6.03	(0.4188)	-2.138	Navarra	-2.048	(1)	4.00	(0.6761)	-2.307
P.Vasco	-1.437	(0)	2.22	(0.8981)	-1.787	P.Vasco	-1.576	(1)	2.47	(0.8714)	-1.312
Rioja	-2.584	(0)	2.91	(0.8196)	-1.476	Rioja	-1.436	(0)	6.20	(0.5162)	-1.511
Valenciana	-1.216	(0)	2.84	(0.8285)	-1.960	Valenciana	-2.658	(1)	3.75	(0.7095)	-1.859

All variables are taken in logs. The figure in parentheses besides the ADF *t*-stat is the lag selection. Lag selection: 10% significance of last lag, max lag = 2. The *Q*-statistic refers to the Ljung-Box test for serial correlation. The *p*-values are given in parentheses after the *Q*-statistic. The truncation parameter for the Phillips-Perron test is 2, based on the Newey-West suggestion $q=4(T/100)^{2/9}$. *= significant at 10%; **=significant at 5%; ***=significant at 1%.

Table A3: Bootstrap statistics for the specification GDP per capita and exports.

		Nb of obs.	Mean	Std. Dev.	Min	Max	Critical values 10%=	5%=	1%=
Individual ADF *t*-stat									
Two variables	T=60	750000	-2,060	0,914	-6,630	2,902	-3,184	-3,552	-4,267
						MacKinnon (1991) T=60	**-3,305**	**-3,720**	**-4,624**
	T=17	750000	-2,159	1,137	-12,455	5,124	-3,537	-4,018	-5,106
						MacKinnon (1991) T=17	**-3,116**	**-3,440**	**-4,084**
With dum93	T=60	750000	-2,894	1,139	-10,844	-0,534	-4,065	-5,043	-7,411
	T=17	750000	-2,910	1,198	-11,776	1,843	-4,440	-4,954	-6,077
IPS test									
Two variables	T=60	5000	-2,060	0,346	-3,475	-0,383	-2,486	-2,624	-2,830
	T=17	5000	-2,159	0,393	-3,993	0,571	-2,636	-2,789	-3,094
with dum93	T=60	5000	-2,894	0,296	-3,984	-1,874	-3,281	-3,374	-3,583
	T=17	5000	-2,910	0,357	-4,314	-1,534	-3,367	-3,504	-3,729
MWF test (ADF)									
Two variables	T=60	5000	53,724	17,279	13,066	152,393	76,720	84,642	100,599
	T=17	5000	65,272	23,293	10,244	216,752	93,770	105,806	138,580
With dum93	T=60	5000	111,479	27,716	37,760	214,211	148,188	158,631	178,290
	T=17	5000	116,238	29,045	37,799	250,918	154,330	167,851	193,301
MWF test (PP)									
Two variables	T=60	5000	47,408	14,930	10,755	176,568	66,676	73,962	89,443
	T=17	5000	49,171	18,381	6,083	184,699	72,055	81,909	109,401
With dum93	T=60	5000	117,529	24,260	46,624	219,716	149,016	159,368	178,413
	T=17	5000	103,649	29,819	33,279	269,318	142,543	158,995	191,578
Pedroni Panel Variance ratio test									
Two variables	T=60	5000	0,406	1,391	-3,067	6,801	2,222	2,840	4,082
	T=17	5000	0,307	0,986	-2,375	4,859	1,581	2,017	2,709
With dum93	T=60	5000	0,738	1,247	-2,676	5,751	2,373	2,847	3,816
	T=17	5000	1,219	0,753	-1,305	3,816	2,177	2,425	2,868
Pedroni Panel *rho*-stat									
Two variables	T=60	5000	-0,230	1,227	-7,164	3,681	-1,819	-2,379	-3,366
	T=17	5000	0,062	0,882	-3,589	3,355	-1,099	-1,462	-2,133
With dum93	T=60	5000	-1,359	1,200	-6,445	1,935	-2,955	-3,429	-4,378
	T=17	5000	-0,278	0,718	-2,880	2,287	-1,186	-1,451	-1,925
Pedroni Panel *t*-stat (non-parametric)									
Two variables	T=60	5000	-0,318	1,306	-7,058	5,595	-1,914	-2,412	-3,290
	T=17	5000	-0,609	1,214	-5,160	5,331	-2,137	-2,616	-3,562
With dum93	T=60	5000	-2,081	1,115	-6,926	1,487	-3,550	-3,921	-4,740
	T=17	5000	-1,894	1,301	-6,758	2,742	-3,552	-4,077	-5,024
Pedroni Panel *t*-stat (parametric)									
Two variables	T=60	5000	-0,537	1,354	-6,026	4,185	-2,204	-2,717	-3,656
	T=17	5000	-1,211	1,330	-7,278	5,853	-2,888	-3,361	-4,348
With dum93	T=60	5000	-1,807	1,186	-6,907	1,807	-3,369	-3,826	-4,634
	T=17	5000	-2,232	1,413	-7,534	3,114	-4,052	-4,560	-5,458
Pedroni Group rho-stat									
Two variables	T=60	5000	0,459	1,207	-5,946	4,375	-1,118	-1,637	-2,506
	T=17	5000	1,302	0,788	-1,686	4,811	0,294	-0,024	-0,637
With dum93	T=60	5000	-1,598	1,226	-6,062	2,196	-3,171	-3,639	-4,475
	T=17	5000	1,071	0,673	-1,421	3,559	0,223	-0,046	-0,536
Pedroni Group t-stat (non-parametric)									
Two variables	T=60	5000	0,061	1,505	-8,000	7,581	-1,792	-2,306	-3,361
	T=17	5000	-0,128	1,467	-6,401	7,700	-1,934	-2,545	-3,779
With dum93	T=60	5000	-2,644	1,304	-7,714	1,441	-4,335	-4,805	-5,673
	T=17	5000	-2,093	1,674	-8,878	3,780	-4,228	-4,947	-6,275
Pedroni Group t-stat (parametric)									
Two variables	T=60	5000	-0,464	1,610	-6,945	5,786	-2,441	-3,085	-4,131
	T=17	5000	-1,489	1,670	-8,807	7,476	-3,577	-4,196	-5,534
With dum93	T=60	5000	-2,333	1,463	-7,726	2,473	-4,236	-4,736	-5,685
	T=17	5000	-2,877	1,744	-9,586	2,937	-5,091	-5,770	-6,869
Individual USURE test									
Two variables	T=60	75000	-3,316	1,261	-9,483	2,428	-4,914	-5,377	-6,296
	T=17	65000	-5,628	3,281	-86,693	5,876	-9,407	-11,283	-16,313

With dum93	T=60	75000	-5,153	1,316	-12,835	-0,493	-6,872	-7,423	-8,506
	T=17	65000	-9,955	5,176	-158,404	1,794	-15,764	-18,714	-27,251
MADF test									
Two variables	T=60	5000	77,2	18,5	25,1	180,311	101,7	109,1	127,6
	T=17	5600	744,3	3861,2	60,2	267977,6	1253,5	2097,0	5846,0
With dum93	T=60	5000	154,6	24,7	85,9	279,1	186,9	197,9	224,3
	T=17	5000	1030,5	5206,7	128,3	293126,9	1536,2	2420,6	6894,5
RSURE test									
Two variables	T=60	5000	-6,7	0,94	-10,2	-3,106	-7,9	-8,3	-9,0
	T=17	5600	-10,2	7,00	-439,0	-0,588	-14,3	-16,7	-26,7
With dum93	T=60	5000	-10,1	0,87	-13,7	-7,456	-11,2	-11,6	-12,2
	T=17	5000	-16,1	11,48	-360,8	-6,410	-23,2	-29,2	-55,7

Figure A1: Maps of Spain's Autonomous Communities