2017

Panel regression with nonstationary variables

Seyedsoroosh Azizi

ABSTRACT

PANEL REGRESSION WITH NONSTATIONARY VARIABLES

SeyedSoroosh Azizi, M.S.
Division of Statistics
Northern Illinois University, 2017
Dr. Duchwan Ryu, Director

I examine the panel regression when variables are nonstationary. Using nonstationary variables might cause the spurious regression problem. A spurious relationship or spurious regression is a mathematical relationship in which two or more events or variables are not causally related to each other, yet it may be wrongly inferred that they are due to either coincidence or the presence of a certain third, unseen factor. In this research, I start by examining the definition of stationary and nonstationary variables and then I move on to cointegration relationship in time series. Then I examine different methods we should use in different scenarios. I investigate the same problems in panel data. Next, I examine motivations behind remittances and I provide an empirical model for how workers' remittances and host and home countries' GNI per capita are nonstationary variables and then I prove that there is a cointegration relation between them. Then I compare the results of the OLS regression with fully-modified OLS and dynamic OLS and conclude that the results are almost similar.

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

AUGUST 2017

PANEL REGRESSION WITH NONSTATIONARY VARIABLES

BY

SEYEDSOROOSH AZIZI

A THESIS SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

DIVISION OF STATISTICS

Thesis Director:
Dr. Duchwan Ryu

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table | Page

CHAPTER 1

REGRESSION WITH TIME SERIES DATA: NONSTATIONARY VARIABLES

1.1 Stationary and Nonstationary Variables in Time Series

This chapter provides a review of the theoretical literature on testing for unit roots and cointegration in time series. The analysis of time series data is of vital interest to many groups, such as macroeconomists studying the behavior of national and international economies, finance economists analyzing the stock market, and agricultural economists predicting supplies and demands for agricultural products. For example, if we are interested in forecasting the growth of gross domestic product or inflation, we look at various indicators of economic performance and consider their behavior over recent years. Alternatively, if we are interested in a particular business, we analyze the history of the industry in an attempt to predict potential sales. In each of these cases, we are analyzing time series data.

1.1.1 Stationary Variables

A time series y_t is stationary if its mean and variance are constant over time and if the covariance between two values from the series depends only on the length of time separating the two values, not on the actual times at which the variables are observed. That is, the time series y_t is stationary if for all values and every time period it is true that

$$E(y_t) = \mu \quad (constant\ mean) \tag{1.1}$$

$$var(y_t) = \sigma^2 \quad (constant\ variance) \tag{1.2}$$

$$cov(y_t, y_{t+s}) = cov(y_t, y_{t-s}) = \gamma_s \quad (covariance\ depends\ on\ s,\ not\ t) \tag{1.3}$$

The autoregressive model of order one, the AR(1), is a useful univariate time series model for explaining the difference between stationary and nonstationary series. It is given by

$$y_t = \rho y_{t-1} + v_t, \quad |\rho| < 1 \tag{1.4}$$

where the errors v_t are independent, with zero mean and constant variance σ_v^2, and may be normally distributed.

The AR(1) process shows that each realization of the random variable y_t contains a proportion ρ of last period's value y_{t-1} plus an error v_t drawn from a distribution with mean zero and variance σ_v^2. Since we are concerned with only one lag, the model is described as an autoregressive model of order one. In general an AR(p) model includes lags of the variable y_t to y_{t-p}. By recursive substitution we get

$y_1 = \rho y_0 + v_1$

$y_2 = \rho y_1 + v_2 = \rho(\rho y_0 + v_1) + v_2 = \rho^2 y_0 + \rho v_1 + v_2$

.

.

.

$y_t = v_t + \rho v_{t-1} + \rho^2 v_{t-2} + ... + \rho^t y_0$

The mean of y_t is $E(y_t) = E(v_t + \rho v_{t-1} + \rho^2 v_{t-2} + ... + \rho^t y_0) = 0$ since the error v_t has zero mean and the value of $\rho^t y_0$ is negligible for a large t. The variance can be shown to be constant $\sigma_v^2/(1-\rho^2)$ while the covariance between two errors s period apart ,γ_s, can be shown to be $\sigma_v^2 \rho^s/(1-\rho^2)$. Thus, the AR(1) model in equation (1.4) is a classic example of a stationary process with a zero mean.

Real-world data rarely have a zero mean. We can introduce a nonzero mean μ by replacing y_t in equation (1.4) with $(y_t - \mu)$ as follows:

$$(y_t - \mu) = \rho(y_{t-1} - \mu) + v_t$$

which can be rearranged as

$$y_t = \alpha + \rho y_{t-1} + v_t, \quad |\rho| < 1 \tag{1.5}$$

where $\alpha = \mu(1-\rho)$. That is, we can accommodate a nonzero mean in y_t by either working with "de-meaned" variable $(y_t - \mu)$ or introducing the intercept term α in the autoregressive process of y_t as in equation (1.5). Corresponding to these two ways, we describe the "de-meaned" variable $(y_t - \mu)$ as being stationary around zero, or the variable y_t as stationary around its mean value $\mu = \alpha/(1-\rho)$.

Another extension to equation (1.4) is to consider an AR(1) model fluctuating around a linear trend $(\mu + \delta t)$. Some real-world data appear to exhibit a trend. In this case we let the "de-treand" series $(y_t - \mu - \delta t)$ behave like an autoregressive model:

$$(y_t - \mu - \delta t) = \rho(y_{t-1} - \mu - \delta(t-1)) + v_t, \quad |\rho| < 1$$

which can be rearranged as

$$y_t = \alpha + \rho y_{t-1} + \lambda t + v_t \tag{1.6}$$

where $\alpha = (\mu(1-\rho) + \rho\delta)$ and $\lambda = \delta(1-\rho)$. The de-trended series $(y_t - \mu - \delta t)$ also has a constant variance and covariance that depend only on time separating observations, not the time at which they are observed. In other words, the "de-trended" series is stationary.

1.1.2 Nonstationary Variables: Random Walk Models

Consider the special case of $\rho = 1$ in equation (1.4):

$$y_t = y_{t-1} + v_t \tag{1.7}$$

This model is known as the random walk model. These time series are called random walks because they appear to wander slowly upward or downward with no real pattern; the values of sample means calculated from subsamples of observations will be dependent on the sample period. This is a characteristic of nonstationary series. For random walks represented by equation (1.7), we have

$y_1 = y_0 + v_1$

$y_2 = y_1 + v_2 = (y_0 + v_1) + v_2 = y_0 + \sum_{s=1}^{2} v_s$

.

.

.

$y_t = y_{t-1} + v_t = y_0 + \sum_{s=1}^{t} v_s$

Recognizing that the v_t are independent, taking the expectation and the variance of y_t yields, for a fixed initial value y_0,

$E(y_t) = y_0 + E(v_1 + v_2 + ... + v_t) = y_0$

$var(y_t) = var(v_1 + v_2 + ... + v_t) = t\sigma_v^2$

The random walk has a mean equal to its initial value and a variance that increases over time, eventually becoming infinite. Although the mean is constant, the increasing variance implies that the series may not return to its mean, and so sample means taken for different periods are not the same.

Another nonstationary model is obtained by adding a constant term to equation (1.7):

$$y_t = \alpha + y_{t-1} + v_t \tag{1.8}$$

This model is known as the random walk with drift. It should be noted that it is important to know whether a time series is stationary or nonstationary before one embarks on a regression analysis because there is a danger of obtaining apparently significant regression results from unrelated data when nonstationary series are used in regression analysis.

1.2 Spurious Regressions and Unit Root Tests in Time Series

Regressions that show significant result with unrelated data are said to be spurious. When nonstationary time series are used in a regression model, the results may spuriously indicate a significant relationship when there is none. In these cases the least squares estimator and least squares predictor do not have their usual properties, and t statistics are not reliable. Since many macroeconomic time series are nonstationary, it is particularly important to take care when estimating regressions with macroeconomic variables.

There are many tests for determining whether a series is stationary or nonstationary. The most popular one is the Dickey-Fuller test. There are three variations of the Dicky-Fuller test designed to take into account the role of the constant term and the trend.

1.2.1 Dickey-Fuller Test 1 (No Constant and No Trend)

The AR(1) process $y_t = \rho y_{t-1} + v_t$ is stationary when $|\rho| < 1$, but when $\rho = 1$ it becomes nonstationary. Hence, one way to test for stationarity is to examine the value of ρ. In other words, we test ρ is equal to one or significantly less than one. Tests for this purpose are known as unit root tests for stationarity. Consider AR(1) model

$$y_t = \rho y_{t-1} + v_t \tag{1.9}$$

where the v_t are independent random errors with zero mean and constant variance σ_v^2. We can test for nonstationarity by testing the null hypothesis that $\rho = 1$ against the alternative that $|\rho| < 1$, or simply $\rho < 1$. This one-sided (left tail) test is put into a more convenient form by subtracting y_{t-1} from both sides of (1.9) to obtain

$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + v_t$

$\Delta y_t = (\rho - 1)y_{t-1} + v_t = \gamma y_{t-1} + v_t$

where $\gamma = \rho - 1$ and $\Delta y_t = y_t - y_{t-1}$. Then, the hypotheses can be written in terms of either ρ or γ:

$H_0 : \rho = 1 \Longleftrightarrow H_0 : \gamma = 0$

$H_1 : \rho < 1 \Longleftrightarrow H_1 : \gamma < 0$

Note that we conclude the null hypothesis is that the series is nonstationary. In other words, if we do not reject the null, we conclude that it is a nonstationary process; if we reject the null hypothesis, then we conclude that the series is stationary.

1.2.2 Dickey-Fuller Test 2 (with Constant but No Trend)

The second Dickey-Fuller test includes a constant term in the test equation:

$$\Delta y_t = \alpha + \gamma y_{t-1} + v_t$$

The null and the alternative hypotheses are the same as before.

1.2.3 Dickey-Fuller Test 3 (with Constant and with Trend)

The third Dickey-Fuller test includes a constant and a trend in the test equation:

$$\Delta y_t = \alpha + \gamma y_{t-1} + \lambda t + v_t$$

The null and the alternative hypotheses are the same as before. Note that the critical values for the Dickey-Fuller test is different than standard critical values (Fuller, 1976).

Table 1.1: Critical Values

	Without trend		With trend	
Sample size	1%	5%	1%	5%
T = 25	-3.75	-3.00	-4.38	-3.60
T = 50	-3.58	-2.93	-4.15	-3.50
T = 100	-3.51	-2.89	-4.04	-3.45
T = 250	-3.46	-2.88	-3.99	-3.43
T = 500	-3.44	-2.87	-3.98	-3.42
T = ∞	-3.43	-2.86	-3.96	-3.41

An important extension of the Dickey-Fuller test allows for the possibility that the error term is autocorrelated. Using the model with an intercept and a trend as an example, the extended test equation is

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{s=1}^{m} a_s \Delta y_{t-s} + v_t \tag{1.10}$$

The number of lagged terms can be determined by examining the autocorrelation function (ACF) of the residuals v_t, or the significance of the estimated lag coefficients a_s. The unit root tests based on the equation (1.10) and its variants (intercept excluded or trend excluded) are referred to as augmented Dickey-Fuller tests. When $\gamma = 0$, in addition to saying that the series is nonstationary, we also say the series has a unit root. In practice, we always use the augmented Dickey-Fuller test (rather than the nonaugmented version) to ensure the errors are uncorrelated.

Note that we reject the null hypothesis of nonstationarity if $\tau < \tau_c$. If $\tau > \tau_c$ then we do not reject the null hypothesis that the series is nonstationary.

1.2.3.1 Order of Integration

Recall that if y_t follows a random walk, then $\rho = 1$ (or equivalently $\gamma = 0$) and the first difference of y_t becomes

$$\Delta y_t = y_t - y_{t-1} = v_t$$

An interesting feature of the series $\Delta y_t = y_t - y_{t-1}$ is that it is stationary since v_t, being an independent $(0, \sigma_v^2)$ random variable, is stationary. Series like y_t, which can be made stationary by taking the first difference, are said to be integrated of order one, and denoted as I(1). Stationary series are said to be integrated of order zero, I(0). In general, the order of integration of a series is the minimum number of times it must be differentiated to make it stationary.

1.3 Cointegration

As a general rule, nonstationary time series variables should not be used in regression models. However, there is an exception to this rule. If y_t and x_t are nonstationary I(1) variables, then we expect their difference, or any linear combination of them, such as $e_t = y_t - \beta_1 - \beta_2 x_t$, to be I(1) as well. However, there is an important case when $e_t = y_t - \beta_1 - \beta_2 x_t$ is a stationary I(0) process. In this case y_t and x_t are said to be cointegrated. Cointegration implies that y_t and x_t share similar stochastic trends, and, since the difference e_t is stationary, they never diverge too far from each other.

A natural way to test whether y_t and x_t are cointegrated is to test whether the errors $e_t = y_t - \beta_1 - \beta_2 x_t$ are stationary. Since we cannot observe e_t, we test the stationarity of the least square residuals, $\hat{e}_t = y_t - \beta_1 - \beta_2 x_t$, using a Dickey-Fuller test. The test for cointegration is effectively a test of the stationarity of the residuals. If the residuals are stationary, then y_t and x_t are said to be cointegrated; if the resuduals are nonstationary, then y_t and x_t are not cointegrated, and apparent regression relationship between them is said to be spurious.

The test for stationarity of the residuals is based on the test equation

$$\Delta \hat{e}_t = \gamma \hat{e}_{t-1} + v_t \tag{1.11}$$

where $\Delta \hat{e}_t = \hat{e}_t - \hat{e}_{t-1}$. Note that the regresion has no constant term because the mean of the regression residuals is zero. Also, since we are basing this test upon estimated values of the residuals, the critical values will be different from those in Table 2.1, and are provided in

Table 2.2 (Hamilton 1994, p.766). The test equation can also include extra terms like $\Delta\hat{e}_{t-1}$, $\Delta\hat{e}_{t-2}$,... on the right-hand side if they are needed to eliminate autocorrelation in v_t. The null and alternative hypotheses in the test for cointegration are

H_0: the series are not cointegrated $\Longleftrightarrow$ residuals are nonstationary.

H_1: the series are cointegrated $\Longleftrightarrow$ residuals are stationary.

Similar to the one-tail unit root test, we reject the null hypothesis of no cointegration if $\tau \leqslant \tau_c$, and we do not reject the null hypothesis that the series are not cointegrated if $\tau > \tau_c$.

Table 1.2: Critical Values for the Cointegration Test

Regression model	1%	5%	10%
$y_t = \beta x_t + e_t$	-3.39	-2.76	-2.45
$y_t = \beta_1 + \beta_2 x_t + e_t$	-3.96	-3.37	-3.07
$y_t = \beta_1 + \delta t + \beta_2 x_t + e_t$	-3.98	-3.42	-3.13

1.3.1 The Error Correction Model

A relationship between I(1) variables is also often referred to as a long-run relationship while a relationship between I(0) variables is often referred to as a short-run relationship.

If we start from the autoregressive distributed lag (ARDL) model

$$y_t = \delta + \theta_1 y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + v_t$$

With a bit of algebra we can get to the error correction model:

$$\Delta y_t = -\alpha(y_{t-1} - \beta_1 - \beta_2 x_{t-1}) + \delta_0 \Delta x_t + v_t \tag{1.12}$$

where $\alpha = (1 - \theta_1)$. The expression in parentheses is the cointegrating relationship. In other words, we have embedded the cointegration relationship between y_t and x_t in a general ARDL framework. Equation (1.12) is called an error correction equation for the following two reasons: (1) the expression $(y_{t-1} - \beta_1 - \beta_2 x_{t-1})$ shows the deviation of y_{t-1} from its long-run value, $\beta_1 + \beta_2 x_{t-1}$ (the "error" in the previous period); (2) the term $\alpha = (\theta_1 - 1)$ shows the "correction" of Δy_t to the "error." More specifically, if the error in the previous period is positive so that $y_{t-1} > (\beta_0 + \beta_1 x_{t-1})$, then y_t should fall and Δy_t should be negative. Conversely, if the error in the previous period is negative so that $y_{t-1} < (\beta_0 + \beta_1 x_{t-1})$, then y_t should rise and Δy_t should be positive. This means that if a cointegrating relationship between y_t and x_t exists, so that adjustments always work to "correct the error," then empirically we should also find that $\alpha = (1 - \theta_1) > 0$, which implies that $\theta_1 < 1$. If there is no evidence of cointegration between the variables, then the term θ_1 would be insignificant.

1.4 Regression When There Is No Cointegration

Thus far, we have shown that regression with I(1) variables is acceptable providing those variables are cointegrated and avoid the problem of spurious results. We also know that regression with stationary I(0) variables is acceptable. What happens when there is no cointegration between I(1) variables? In this case, the sensible thing to do is to convert the nonstationary series to stationary series.

The conversion of nonstationary series to stationary series and the kind of model we estimate depend on whether the variables are difference stationary or trend stationary. In the former case, we convert the nonstationary series to its stationary counterpart by taking first differences. In the latter case, we convert the nonstationary series to its stationary counterpart by de-trending.

1.4.1 First Difference Stationary

Consider a variable y_t that behaves like the random walk model:

$$y_t = y_{t-1} + v_t$$

This is a nonstationary series with a "stochastic" trend, but it can be rendered stationary by taking the first difference:

$$\Delta y_t = y_t - y_{t-1} = v_t$$

The variable y_t is said to be a first difference stationary series. Recall that this means that y_t is said to be integrated of order 1. Now suppose that Dickey-Fuller tests reveal that two variables, y_t and x_t, which you would like to relate in a regression, are first difference stationary, I(1), and not cointegrated. Then, a suitable regression involving only stationary variables relates changes in y_t to changes in x_t, with relevant lags included and no intercept. For example, using one lagged Δy_t and a current and lagged Δx_t, we have

$$\Delta y_t = \theta \Delta y_{t-1} + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + e_t \tag{1.13}$$

Now consider a series y_t that behaves like a random walk with drift:

$$y_t = \alpha + y_{t-1} + v_t$$

and note that y can be rendered stationary by taking the first difference:

$$\Delta y_t = \alpha + v_t$$

The variable y_t is also said to be a first difference stationary series even though it is stationary around a constant term. Now suppose again that y_t and x_t are I(1) and not cointegrated. Then an example of a suitable regression equation, again involving stationary variables, is obtained by adding a constant to (1.13). That is,

$$\Delta y_t = \alpha + \theta \Delta y_{t-1} + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + e_t \tag{1.14}$$

Although there is often doubt about the role of the constant term, the usual practice is to include an intercept term in the regression.

1.4.2 Trend Stationary

Consider a model with a constant term, a trend term, and a stationary error term:

$$y_t = \alpha + \delta t + v_t$$

The variable y_t is said to be trend stationary because it can be made stationary by removing the effect of the deterministic (constant and trend) components

$$y_t - \alpha - \delta t = v_t$$

This series is, strictly speaking, not an I(1) variable, but it is described as stationary around a deterministic trend. In this case we estimate the equation:

$$y_t = \alpha + \delta t + \theta y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + e_t \tag{1.15}$$

CHAPTER 2

UNIT ROOTS AND COINTEGRATION IN PANELS

2.1 Introduction

This chapter provides a review of hypotheses tests for unit roots and cointegration in panels when the time dimension (T) and the cross-section dimension (N) are relatively large. When N is small (say less than ten) and T is relatively large, standard time series techniques apply to systems of equations, such as the seemingly unrelated regression equations (SURE), and the panel aspect of the data should not pose new technical difficulties.

In panels, after establishing a cointegration relationship, the long-run parameters can be estimated efficiently using techniques similar to those proposed in the case of single time series models. Specifically, fully modified OLS procedures, the dynamic OLS estimator and estimators based on a vector error correction representation are adapted to panel data structures. Note that the panel test outcomes are often difficult to interpret if the null of the unit root or cointegration is rejected. The best conclusion is that a significant fraction of the cross-section units is stationary or cointegrated.

2.2 Model and Hypotheses to Test

Assume that time series $\{y_{i0}, ..., y_{iT}\}$ on the cross-section units i=1,2,...,N are collected for each i by a simple first-order autoregressive, AR(1), process

$$y_{it} = -(\alpha_i - 1)\mu_i + \alpha_i y_{i,t-1} + \epsilon_{it} \tag{2.1}$$

where the initial values, y_{i0}, are given, and the errors ϵ_{it} are identically, independently distributed (iid) across, i and t with $E(\epsilon_{it}) = 0$, $E(\epsilon_{it}^2) = \sigma_i^2 < \infty$ and $E(\epsilon_{it}^4) < \infty$. These processes can also be written equivalently as simple Dickey-Fuller (DF) regressions:

$$\Delta y_{it} = -\phi_i \mu_i + \phi_i y_{i,t-1} + \epsilon_{it} \tag{2.2}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\phi_i = \alpha_i - 1$. In further developments of the model it is also helpful to write equations 2.1 and 2.2 in mean deviations forms $\hat{y}_{it} = \alpha_i \hat{y}_{i,t-1} + \epsilon_{it}$ where $\hat{y}_{it} = y_{it} - \mu_i$. The corresponding DF regression in $\hat{y}_{it}$ is given by

$$\Delta \hat{y}_{it} = \phi_i \hat{y}_{i,t-1} + \epsilon_{it} \tag{2.3}$$

Most panel unit root tests are designed to test the null hypothesis of a unit root for each individual series in a panel. Accordingly, the null hypothesis is

$$H_0 : \phi_1 = \phi_2 = ... = \phi_N = 0 \tag{2.4}$$

i.e. all time series are independent random walks. The formulation of the alternative hypothesis is instead a controversial issue that critically depends on which assumptions one makes about the nature of the homogeneity/heterogeneity of the panel. First, under the assumption that the autoregressive parameter is identical for all cross-section units, we can consider

$$H_1^a : \phi_1 = \phi_2 = ... = \phi_N = \phi \textit{ and } \phi < 0$$

The panel unit root statistics motivated by H_1^a pools the observations across the different cross-section units before forming the "pooled" statistics. One drawback of tests based on such alternative hypotheses is that they tend to have power (reject the null) even if only a few of the units are stationary. Hence a rejection of the null hypothesis, H_0, is not convincing evidence that a significant proportion of the series is indeed stationary. At another extreme, there is the alternative hypothesis stating that at least one of the series in the panel is generated by a stationary process:

$$H_1^b : \phi_i < 0, \; for\ one\ or\ more\ i$$

In the case of of large N and T, panel unit root tests will lack power if the alternative, H_1^b, is adopted. A more appropriate alternative (Im, Pesaran, and Shin, 2003) is given by the heterogeneous alternative

$$H_1^c : \phi_i < 0, \; i = 1, 2, ..., N_1, \; \phi_i = 0, i = N_1 + 1, N_2 + 1, ..., N \tag{2.5}$$

such that

$$\lim_{N \to \infty} \frac{N_1}{N} = \delta, 0 < \delta \leq 1 \tag{2.6}$$

Using the above specification the null hypothesis is $H_0 : \delta = 0$, while $H_1^c : \delta > 0$.

The Im, Pesaran, and Shin test also allows for some (but not all) of the individual series to have unit roots under the alternative hypothesis. But the fraction of the individual processes that are stationary is positive. The t-bar statistic, denoted by $t - bar_{NT}$, is formed as a simple average of the individual t statistics for testing the null hypothesis of $\phi_i = 0$. If $t_{iT}(p_i, \rho_i)$ is the standard t statistic, then

$$t - bar_{NT} = \frac{1}{N}\sum_{i=1}^{N} t_{iT}(p_i, \rho_i) \tag{2.7}$$

If T $\to \infty$, then for each i the t statistic (without time trend) converges to the Dickey-Fuller distribution, η_i, defined by

$$\eta_i = \frac{\frac{1}{2}\{[W_i(1)]^2 - 1\} - W_i(1)\int_0^1 W_i(u)du}{\int_0^1 [W_i(u)]^2 du - [\int_0^1 W_i(u)du]^2} \tag{2.8}$$

where W_i is standard Brownian motion. The limiting distribution is different when a time trend is included in the regression (Hamilton 1994, p. 499).

2.3 Estimation of Cointegrating Relations in Panels

We consider a single-equation framework where it is assumed that y_{it} and the k $\times$ 1 vector of regressors $\mathbf{x}_{it}$ are I(1) with at most one cointegrating relation among them, namely that there exists a linear relationship such that the error u_{it} is stationary. It is assumed that $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}^{'})^{'}$ is independently and identically distributed across i, and the regressors, $\mathbf{x}_{it}$, are not cointegrated.

It is assumed that the vector of coefficients, **fi**, is the same for all cross-sectional units. That is, a homogeneous cointegration relationship is assumed. Alternatively, it may be assumed that the cointegration parameters are cross-section specific (heterogenous cointegration).

By applying a sequential limit theory it can be shown that the OLS estimator of **fi** is $T\sqrt{N}$ consistent and, therefore, the time series dimension is more informative on the long-run coefficients than the cross-section dimension. Furthermore, it is important to notice that (as in the time series framework) the OLS estimator is consistent but inefficient in the model with endogenous regressors.

Fully modified OLS (FM-OLS) approach is proposed to obtain an asymptotically efficient estimator for homogeneous cointegration vectors. This estimator adjusts for the effects of endogenous regressors and short-run dynamics of the errors. To correct for the effect of endogeneity of the regressors, the dependent variable is adjusted for the part of the error that is correlated with the regressor. An alternative approach is the dynamic OLS (DOLS) estimator.

CHAPTER 3

EMPIRICAL EXAMPLE: MOTIVATIONS BEHIND REMITTANCES

3.1 Data and Model

The regression model is as follows:

$$RPM_{it} = \beta_0 + \beta_1 Y_{it} + \beta_2 X_{it} + \beta_3 E_{it} + \beta_4 R_{it} + \mu_i + \delta_t + e_{it} \tag{3.1}$$

where RPM_{it} is remittances per migrant inflow to recipient country i at year t. It is the ratio of "remittances inflow to country i" to "emigrant stock from country i." Y_{it} is GNI per capita of recipient country i at time t. For country i at time t, X_{it} is the weighted average of host countries' GNI per capita. E_{it} is the real exchange rate for country i at time t; μ_i is country-specific constant; δ_t is time-specific constant; e_{it} is the error term.

In the regression model, the subscript i refers to remittance receiving country. However, countries that are the origin of large numbers of emigrants should not have same weights as small countries. Therefore, for each observation (recipient country-year) this study uses

$$W_{it} = \frac{Stock\ of\ emigrants\ with\ origin\ country\ i\ at\ year\ t}{Total\ number\ of\ emigrants\ from\ all\ countries\ at\ year\ t} \tag{3.2}$$

as the weight. Therefore, a country with large numbers of emigrants would have higher weight than a country with fewer emigrants.

The data set consists of panel data for 141 developing countries and 51 developed countries and up to 25 years for each category.

To overcome autocorrelation and hetroskedasticity, Newey-West heteroskedasticity and autocorrelation consistent (HAC) standard errors are used. The number of lag periods should be chosen as $T^{1/4}$ or $T^{1/3}$ (Pesaran, 2015, p. 114). Since T=25 for the panel data, three lag periods are chosen for autocorrelation. I also check whether the variables used in the regressions for both all countries and developing countries are stationary or not. In both cases the null hypothesis is that the panel data has unit root. Using three different tests (ADF Rho, ADF Tau, and ADF F) I fail to reject the null even at the 10% significance level, which means some variables are nonstationary. However, by using the same three tests, at the 1% level of significance, I reject the null hypothesis of the unit root for the first differences of the variables, which means first differences of the variables are stationary. The next step is to check for spurious regression. Maddala-Wu (1999) panel data cointegration tests are used here (both Fisher PP and Fisher ADF). For both all countries and developing countries, in all panel stationarity tests, cointegration exists, which means the error terms are stationary and the regressions are not spurious.

3.2 Findings

The result of the main regression is provided in column 1 of Table 3.1 for all countries and column 6 of Table 3.1 for developing countries. It might take time for explanatory variables to affect remittances. Therefore, in the second regression model, lead of RPM is used as the dependent variable. The result of the main regression is provided in column 2 of Table 3.1 for all countries and column 7 of Table 3.1 for developing countries. For some of the developing countries with high remittances/GNI ratio, not only home country's GNI per capita affects remittances, but also remittances might affect home country's GNI per capita. This might cause simultaneous equations endogeneity. Remittances cannot affect home country's past GNI per capita. However, lag of home country's GNI per capita correlates with home country's GNI per capita. Therefore, this study uses one and two and three period lags of Home GNI as instruments for the Home GNI. The result of this regression (using instrumental variables) is provided in column 3 of Table 3.1 for all countries and column 8 of Table 3.1 for developing countries. Other methods used in this research are dynamic OLS (DOLS) and fully modified OLS (FMOLS) suggested by Kao and Chiang (2001). These methods might be more accurate when variables are nonstationary but cointegrated, as is the case in this study. The result of the DOLS regression is provided in column 4 of Table 3.1 for all countries and column 1 of Table 3.1 for developing countries. The result of the FMOLS regression is provided in column 5 of Table 3.1 for all countries and column 10 of Table 3.1 for developing countries. The results of these two regressions are very similar to the main regression.

Table 3.1: Fixed Effect Estimates

	All Countries					Developing Countries				
Dependent	RPM	Lead RPM	RPM	RPM	RPM	RPM	Lead RPM	RPM	RPM	RPM
Home GNI	-0.114***	-0.113***	-0.094***	-0.114***	-0.115***	-0.111***	-0.096***	-0.087***	-0.112***	-0.112***
	(0.0187)	(0.0184)	(0.017)	(0.028)	(0.0188)	(0.0298)	(0.0288)	(0.0231)	(0.0357)	(0.0354)
Host GNI	0.057***	0.055***	0.061***	0.057***	0.057***	0.092***	0.09***	0.086***	0.091***	0.092***
	(0.0166)	(0.0162)	(0.0174)	(0.0184)	(0.0172)	(0.0186)	(0.0176)	(0.0196)	(0.0226)	(0.0222)
Exchange Rates	-27.8	51.9	127	-25.7	-24.6	-65.7	60.89	188	-62	-59.7
	(270)	(255.3)	(306)	(418)	(236.4)	(340)	(324.4)	(368)	(303)	(299)
Interest Rates	3.43	3.4	1.77	3.69	3.76	3.26	3.09	3.32	3.5	3.58
	(4.6)	(4.06)	(5.82)	(11.14)	(5.15)	(4.83)	(4.2)	(6.013)	(5.99)	(5.86)
R-Squared	0.75	0.76	0.80	0.75	0.75	0.75	0.76	0.81	0.75	0.75
Observations	2116	2065	1960	2116	2116	1635	1591	1539	1635	1635
Instrument			Yes					Yes		
Method	FE	FE	FE	DOLS	FMOLS	FE	FE	FE	DOLS	FMOLS

HAC standard errors are in parentheses. Exchange rates are in logarithm. All regressions include time dummies and country dummies. Variable “RPM” stands for remittances per migrant.
*** significant at 1% level.

Since variable Interest Rates is insignificant, as a robustness check, I excluded this variable from the regression. Data on interest rates include many missing values and, dropping interest rates adds 506 observations to the data set and is good practice for the robustness check. Since variable Exchange Rates is insignificant, as a robustness check, I also excluded this variable from the regression.

Table 3.2: Fixed Effect Estimates

	All Countries					Developing Countries				
Dependent	RPM	Lead RPM	RPM	RPM	RPM	RPM	Lead RPM	RPM	RPM	RPM
Home GNI	-0.113***	-0.113**	-0.101**	-0.114***	-0.114***	-0.133***	-0.122***	-0.111***	-0.133***	-0.133***
	(0.0158)	(0.0158)	(0.0153)	(0.0208)	(0.0153)	(0.0278)	(0.0292)	(0.0254)	(0.0279)	(0.0278)
Host GNI	0.0641***	0.064***	0.061***	0.065***	0.065***	0.085***	0.086***	0.079***	0.085***	0.085***
	(0.021)	(0.0119)	(0.0141)	(0.0119)	(0.013)	(0.0136)	(0.0129)	(0.0162)	(0.0134)	(0.0158)
Exchange Rates	-181	-85	-62.3	-193	-192.3	-273	-141.7	-138	-262.2	-261
	(205.7)	(195)	(255)	(301)	(192.08)	(233.8)	(243)	(298)	(382)	(234)
R-Squared	0.73	0.74	0.78	0.73	0.73	0.76	0.75	0.79	0.74	0.74
Observations	2622	2551	2390	2622	2622	2036	1983	1870	2036	2036
Instrument			Yes					Yes		
Method	FE	FE	FE	DOLS	FMOLS	FE	FE	FE	DOLS	FMOLS

HAC standard errors are in parentheses. Exchange rates are in logarithm. All regressions include time dummies and country dummies. Variable "RPM" stands for remittances per migrant.

*** significant at 1% level.

In the results shown in Table 3.2 and Table 3.3, in all regressions variable Home GNI is negative and significant at the 1% level and variable Host GNI is positive and significant at the 1% level. Variables Exchange Rates and Interest Rates seem to have no effect on remittances.

Table 3.3: Fixed Effect Estimates

	All Countries					Developing Countries				
Dependent	RPM	Lead RPM	RPM	RPM	RPM	RPM	Lead RPM	RPM	RPM	RPM
Home GNI	-0.108***	-0.109***	-0.098***	-0.106***	-0.109***	-0.116***	-0.11***	-0.1***	-0.115***	-0.117***
	(0.0152)	(0.0154)	(0.0152)	(0.0196)	(0.015)	(0.0246)	(0.0249)	(0.0227)	(0.0198)	(0.0256)
Host GNI	0.065***	0.064***	0.061***	0.0643***	0.065***	0.085***	0.087***	0.079***	0.085***	0.085***
	(0.0122)	(0.0117)	(0.0131)	(0.0115)	(0.013)	(0.0136)	(0.0128)	(0.0152)	(0.013)	(0.0156)
R-Squared	0.73	0.74	0.78	0.73	0.73	0.74	0.75	0.79	0.74	0.74
Observations	2639	2569	2405	2639	2639	2053	2001	1885	2053	2053
Instrument			Yes					Yes		
Method	FE	FE	FE	DOLS	FMOLS	FE	FE	FE	DOLS	FMOLS

HAC standard errors are in parentheses. All regressions include time dummies and country dummies. Variable “RPM” stands for remittances per migrant.
*** significant at 1% level.

As another robustness check, in the following I use the first difference method. In this model, the first difference of remittances per migrant is regressed on the first differences of independent variables. Results are provided in Table 3.4.

$$\Delta RPM_{it} = \beta_0 + \beta_1 \Delta Y_{it} + \beta_2 \Delta X_{it} + \beta_3 \Delta E_{it} + \beta_4 \Delta R_{it} + \delta_t + e_{it} \tag{3.3}$$

The coefficient for Change in Home GNI is negative but in two regressions is significant at the 1% level, in two regressions it is significant at the 5% level, and in the last two columns is significant only at the 10% level. In contrast, the coefficient for Change in Host GNI is positive and always significant at the 1% level.

The last method that this paper uses is error correction model (ECM). In this model, the first difference of remittances is regressed on the first difference of independent variables as well as on “Error Correction Term” (Pesaran, 2015 p. 124). Since the variables are cointegrated and error term is stationary, I therefore use the following regression model:

Table 3.4: First Difference Estimates

	All Countries			Developing Countries		
Dependent	Δ RPM	Δ RPM	Δ RPM	Δ RPM	Δ RPM	Δ RPM
Change in Home GNI	-0.059***	-0.039**	-0.035**	-0.073***	-0.048*	-0.04*
	(0.018)	(0.0163)	(0.015)	(0.0275)	(0.0271)	(0.0271)
Change in Host GNI	0.028***	0.024***	0.023***	0.034***	0.028***	0.027***
	(0.0088)	(0.0077)	(0.0074)	(0.0101)	(0.0087)	(0.0085)
Change in Exchange Rates	-83	-126.5		-108	-143	
	(94)	(92)		(105.4)	(102)	
Change in Interest Rates	-1.84			-1.95		
	(2.03)			(2.15)		
R-Squared	0.07	0.06	0.06	0.08	0.07	0.06
Observations	1950	2432	2450	1515	1899	1917

HAC standard errors are in parentheses. Exchange rates are in logarithm. All regressions include time dummies. Variable "RPM" stands for remittances per migrant.
*** significant at 1% level.
** significant at 5% level.
* significant at 10% level.

$$\Delta RPM_{it} = \beta_0 + \beta_1 ECT_{it-1} + \beta_2 \Delta Y_{it} + \beta_3 \Delta X_{it} + \beta_4 \Delta E_{it} + \beta_5 \Delta R_{it} + \delta_t + e_{it} \tag{3.4}$$

where

$$ECT_{it-1} = RPM_{it-1} - \beta_0 - \beta_1 Y_{it-1} - \beta_2 X_{it-1} - \beta_3 E_{it-1} - \beta_4 R_{it-1} - \mu_i - \delta_{t-1} \tag{3.5}$$

Coefficients $\beta_2, \beta_3, \beta_4$, and β_5 in equation 3.4 show the short-term effects of independent variables on remittances per migrant. Coefficient β_1 shows the long-term effect of deviation from trend for remittances. The results of the error correction model regression are provided in Table 3.5.

The coefficients and the standard errors of all variables in the error correction model and the first difference model are very similar. Coefficient of Error Correction Term is between -0.1 and -0.11 in all six columns of Table 3.5. This analysis suggests that, even if remittances deviate from the equilibrium relationship because of common shocks, they will eventually revert. If remittances are above equilibrium, they will tend to fall relative to independent variables, and if they are below equilibrium they will tend to rise (Pesaran, 2015, p. 846).

Table 3.5: Error Correction Model Estimates

	All Countries			Developing Countries		
Dependent	Δ RPM	Δ RPM	Δ RPM	Δ RPM	Δ RPM	Δ RPM
Error Correction Term	-0.1098**	-0.1029**	-0.102**	-0.111**	-0.1047**	-0.103**
	(0.047)	(0.0414)	(0.0414)	(0.0502)	(0.0443)	(0.0477)
Change in Home GNI	-0.058***	-0.039**	-0.035**	-0.075***	-0.052*	-0.0423*
	(0.0179)	(0.0166)	(0.0158)	(0.0284)	(0.0291)	(0.0256)
Change in Host GNI	0.027***	0.024***	0.024***	0.034***	0.028***	0.027***
	(0.0087)	(0.0077)	(0.0075)	(0.0099)	(0.0086)	(0.0083)
Change in Exchange Rates	-88	-130		-109	-149	
	(89)	(86)		(99)	(95)	
Change in Interest Rates	-1.47			-1.57		
	(1.87)			(1.97)		
R-Squared	0.125	0.111	0.11	0.1373	0.119	0.11
Observations	1950	2432	2450	1515	1899	1917

HAC standard errors are in parentheses. Exchange rates are in logarithm. All regressions include time dummies. Variable "RPM" stands for remittances per migrant. *** significant at 1% level.

** significant at 5% level.

* significant at 10% level.

CHAPTER 4

EMPIRICAL EXAMPLE: CYCLICALITY OF REMITTANCES

4.1 Introduction

The problem of cyclicality of remittances has been under scrutiny recently. The classical way to check the cyclicality of remittances with respect to the recipient (home) country or sender (host) country is to use the Hodrick-Prescott (H-P) filter to decompose data into a trend component and a deviation from the trend component (cyclical component). If the correlation coefficient between the cyclical component of remittances and the cyclical component of GNI per capita is significant and positive, remittances are procyclical, and if it is significant and negative, remittances are countercyclical

4.2 Data and Model

Remittance data is from the International Monetary Fund (defined as personal transfers). GNI per capita data is from the United Nations. This paper uses Hodrick-Prescott filter with $\lambda = 100$ as a penalty parameter to decompose each time series into a trend component and a cyclical component (Pesaran, 2015, p. 358). Then the correlation coefficient between the cyclical component of the remittances and the cyclical component of the home and host countries' GNI per capita is calculated. The positive (negative) amount of this correlation coefficient indicates that remittances are procylical (countercyclical) with respect to the home and host countries' GNI per capita. All calculations are done twice: once for all countries with available data and once for just those countries with remittance inflow to GNI ratio of greater than 1%. The reason is that the reasearch finds the cyclicality of remittances important when they are non-negligible in comparison to the total size of the country's economy. Also the number of observations falls from 2806 to 1607 and this is a good robustness check. Since it might take some time for remittances to respond to change in host or home countries' GNI per capita, the correlation coefficients between cyclical component of remittances and cyclical component of two and one lag of Host GNI and Home GNI are also explored. The results are provided in Table 5.1.

Table 4.1: Correlation Coefficients

	All Countries			Big Remittance Receivers		
	2lags	1lag	level	2lags	1lag	level
Remittances and Home GNI	0.0288	0.045**	0.062***	0.01915	0.04	0.094***
Remittances and Host GNI	0.0645***	0.09***	0.125***	0.04224*	0.084***	0.141***
Home GNI and Host GNI	0.323***	0.32***	0.36***	0.3695***	0.376***	0.42***
Observations	2516	2595	2806	1518	1537	1607

*** significant at 1% level.

** significant at 5% level.

* significant at 10% level.

The positive and significant correlation coefficient between the cyclical component of remittances and cyclical component of host countries and home countries indicate that remittances are procyclical with respect to both home and host countries. Procyclicality of remittances with respect to the host countries is intuitive. When migrants earn more than the historical trends in the host countries, they have more money available to remit and therefore they remit more. On the other hand, procyclicality of remittances with respect to the home countries is unexpected. Some studies based on the procyclicality of remittances with respect to the home countries have claimed that the main motivation behind remittances is investment. However, note that to generate Table 5.1 correlation was used rather than regression. High correlation between the host country's GNI per capita and the home country's GNI per capita shows that when families of migrants in the home countries earn below the trend component of their income, migrants probably earn below the trend component of their income too. Therefore migrants have less money available to remit and they remit less. This does not necessarily mean that their incentive to remit is investment.

To control for cyclical component of the host country, regression must be used. The regression model is

$$dev.rem_{it} = \beta_0 + \beta_1 dev.host_{it} + \beta_2 dev.home_{it} + e_{it} \tag{4.1}$$

The results of the regression are shown in columns 1, 2, and 3 of Table 5.2 for all countries and for countries with remittance to GNI ratio of greater than 1 in columns 4, 5, and 6 of Table 5.2.

Table 4.2: Regression Results: H-P Filtered Data

	All Countries			Big Remittance Receivers		
	2 lag	1lag	level	2 lag	1lag	level
Home GNI	0.0044	0.009	0.008	0.0058	0.0124	0.0413
	(0.0104)	(0.01)	(0.0079)	(0.039)	(0.0333)	(0.0269)
Host GNI	0.019***	0.026***	0.034***	0.0149	0.03***	0.0445***
	(0.0066)	(0.0063)	(0.0058)	(0.0101)	(0.0101)	(0.00977)
R-Squared	0.004	0.008	0.0159	0.0018	0.007	0.0215
Observations	2516	2595	2806	1518	1537	1607

Standard errors are in parentheses.

*** significant at 1% level.

The results indicate that the coefficient for cyclical component of the host country is positive and statistically significant at 1% level. The coefficient for the deviation from the trend of the home country is statistically insignificant. The cyclicality of remittances with respect to recipients' countries is not determinant of the motivations behind remittances. The fact that remittances are weakly procyclical with respect to recipients' countries does not imply that investment is the main motivation behind remittances. Remittances are procylical with respect to the host countries and there is a strong correlation between the host countries and home countries. Therefore, remittances seem to be procyclical with respect to the home countries as well. In order to discover the motivations behind remittances, unfiltered data should be used (rather than the H-P filtered data). The regression equation is

$$rem_{it} = \beta_0 + \beta_1 host_{it} + \beta_2 home_{it} + \mu_i + \delta_t + e_{it} \tag{4.2}$$

Where rem_{it} is remittance inflow to the country i at year t, $host_{it}$ is the weighted average GNI per capita of all remittance sending countries to the home country i at year t, and $home_{it}$ is GNI per capita of remittance receiving country i at year t. The variables used in the regression are not stationary. However, their first differences are. With using Maddala-Wu (1999) panel data cointegration tests (both Fisher PP and Fisher ADF), I conclude the error term is stationary, which means the panel data cointegration exists and the regressions are not spurious. Since autocorrelation and hetroskedasticity exist, Knewly-West HAC standard errors are used. Other methods used are dynamic OLS (DOLS) and fully modified OLS (FMOLS) suggested by Kao and Chiang (2001). The results of this regression are shown in Table 5.3:

Table 4.3: Regression Results

	All Countries					Big Remittance Receivers				
	2 lag	1lag	level	FMOLS	DOLS	2 lag	1lag	level	FMOLS	DOLS
Home GNI	-0.1302***	-0.1107***	-0.1145***	-0.123***	-0.114***	-0.354***	-0.26812***	-0.255***	-0.286***	-0.292***
	(0.0298)	(0.0255)	(0.0274)	(0.028602)	(0.025)	(0.1264)	(0.0803)	(0.0788)	(0.0814)	(0.085)
Host GNI	0.08969***	0.07749***	0.0796***	0.0897***	0.0796	0.115***	0.11062***	0.1124***	0.116***	0.117***
	(0.0243)	(0.0217)	(0.023)	(0.0247)	(0.021)	(0.0363)	(0.0357)	(0.037)	(0.0369)	(0.038)
R-Squared	0.648	0.628	0.6128	0.61	0.61	0.6874	0.66	0.6582	0.654	0.68
Observations	2523	2737	2806	2806	2806	1525	1589	1607	1607	1607

HAC standard errors are in parentheses. All regressions include time dummies and country dummies.

*** significant at 1% level.

The coefficient for the variable Home GNI is negative and statistically significant in all the regressions. The coefficient for the variable Host GNI is positive and statistically significant in all the regressions.

Investment theory and altruism theory are two main theories related to the motivations behind remittances. Based on the investment theory, migrants remit more if they earn more money in the host country and they remit more if their home country's economy is prospering. In this case, both the coefficients of Home GNI and Host GNI are expected to be positive. Based on the altruism theory, migrants remit more if they earn more money in the host country, and they remit less if their families earn more in the home country. In this case the coefficient of Host GNI is expected to be positive and the coefficient of Home GNI is expected to be negative. The results provided in Table 5.3 confirm that the primary motivation of remittances is altruism (Azizi, 2016).

REFERENCES

Azizi, S. (2016). Altruism: primary motivation of remittances. *Applied Economics Letters*, 1-4.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. New York: John Wiley and Sons.

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton: Princeton University Press.

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115(1), 53-74.

Kao, C., & Chiang, M. H. (2001). On the estimation and inference of a cointegrated regression in panel data. In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels* (pp. 179-222). Emerald Group Publishing Limited.

Maddala, G. S., & 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(S1), 631-652.

Pesaran, M. H. (2015). *Time Series And Panel Data Econometrics*. Oxford University Press.

APPENDIX: SAS CODES AND RESULTS

```
*At first I check for unit root test (stationarity);
title 'panel unit root test';
proc panel data=last6;
id id year;
model col1=col12 col3 col4 col5/ stationarity(Fisher);
run;
title;
*Then I check for unit root test (stationarity) for differences;
title 'panel unit root test';
proc panel data=Fd;
id id year;
model col1=col12 col3 col4 col5/ stationarity(Fisher);
run;
title;
*Finaly I check for cointegration;
title 'cointegration test';
proc panel data=Fd;
id id year;
model col1=col12 col3 col4 col5/ fixtwo stationarity(fisher(test=pp));
run;
title;
*or;
```

```
proc panel data=Fd;
id id year;
model col1=col12 col3 col4 col5/ fixtwo;
stationarity(fisher(test=ADF));
run;
```

Here are the SAS results for stationarity tests for the level variables:

The PANEL Procedure
Panel Stationarity Tests

Dependent Variable: COL1 (RPM)

Combination Test Results with ADF Rho								
	Fisher Test		Asymptotic Fisher Test		Inverse Normal Test		Logit Test	
Deterministic Variables	Chi-Square	Pr > ChiSq	FisherI	Pr > FisherI	Z	Pr < Z	L*	Pr < L*
Zero Mean	78.09	1.0000	-5.45	1.0000	7.58	1.0000	7.93	1.0000
CS Fixed	95.31	1.0000	-3.97	1.0000	7.58	1.0000	8.23	1.0000
CS Fixed, Time	118.59	0.9979	-2.60	0.9954	5.04	1.0000	5.46	1.0000
TS Fixed	109.02	1.0000	-3.83	0.9999	3.60	0.9998	3.46	0.9997
CS, TS Fixed	115.06	0.9994	-2.89	0.9981	5.59	1.0000	5.67	1.0000

Combination Test Results with ADF Tau								
	Fisher Test		Asymptotic Fisher Test		Inverse Normal Test		Logit Test	
Deterministic Variables	Chi-Square	Pr > ChiSq	FisherI	Pr > FisherI	Z	Pr < Z	L*	Pr < L*
Zero Mean	127.66	0.9992	-2.85	0.9978	8.30	1.0000	9.09	1.0000
CS Fixed	113.66	0.9996	-2.96	0.9985	6.81	1.0000	7.49	1.0000
CS Fixed, Time	152.38	0.7679	-0.75	0.7725	3.76	0.9999	3.96	1.0000
TS Fixed	151.32	0.9528	-1.61	0.9461	3.60	0.9998	3.94	1.0000
CS, TS Fixed	90.64	1.0000	-4.22	1.0000	6.47	1.0000	6.95	1.0000

Since p-values are greater than 0.05, the null hypothesis of unit root test (nonstationarity) of variables cannot be rejected.

Here are the SAS results for stationarity tests for the first difference of variables:

panel unit root test

The PANEL Procedure
Panel Stationarity Tests

Dependent Variable: COL1 (RPM)

Combination Test Results with ADF Rho								
	Fisher Test		Asymptotic Fisher Test		Inverse Normal Test		Logit Test	
Deterministic Variables	Chi-Square	Pr > ChiSq	FisherI	Pr > FisherI	Z	Pr < Z	L*	Pr < L*
Zero Mean	829.98	<.0001	36.11	<.0001	-19.80	<.0001	-23.74	<.0001
CS Fixed	714.92	<.0001	30.42	<.0001	-17.20	<.0001	-20.37	<.0001
CS Fixed, Time	477.38	<.0001	18.66	<.0001	-11.54	<.0001	-13.03	<.0001
TS Fixed	807.38	<.0001	34.88	<.0001	-19.80	<.0001	-23.47	<.0001
CS, TS Fixed	677.92	<.0001	28.38	<.0001	-16.90	<.0001	-19.48	<.0001

Combination Test Results with ADF Tau								
	Fisher Test		Asymptotic Fisher Test		Inverse Normal Test		Logit Test	
Deterministic Variables	Chi-Square	Pr > ChiSq	FisherI	Pr > FisherI	Z	Pr < Z	L*	Pr < L*
Zero Mean	845.11	<.0001	36.94	<.0001	-20.31	<.0001	-24.40	<.0001
CS Fixed	576.72	<.0001	22.79	<.0001	-13.91	<.0001	-15.77	<.0001
CS Fixed, Time	428.25	<.0001	15.84	<.0001	-9.76	<.0001	-10.93	<.0001
TS Fixed	811.22	<.0001	35.09	<.0001	-20.30	<.0001	-23.91	<.0001
CS, TS Fixed	524.23	<.0001	19.89	<.0001	-13.25	<.0001	-14.46	<.0001

Since p-values are less than 0.05, the null hypothesis of unit root test (nonstationarity) of the first differences of variables will be rejected.

Here are the SAS results for cointegration tests:

cointegration test

The PANEL Procedure
Panel Stationarity Tests

Dependent Variable: COL1 (RPM)

Combination Test Results with PP Rho								
	Fisher Test		Asymptotic Fisher Test		Inverse Normal Test		Logit Test	
Deterministic Variables	Chi-Square	Pr > ChiSq	FisherI	Pr > FisherI	Z	Pr < Z	L*	Pr < L*
Zero Mean	1295.8	<.0001	44.90	<.0001	-25.97	<.0001	-30.58	<.0001
CS Fixed	1110.6	<.0001	36.84	<.0001	-22.76	<.0001	-25.91	<.0001
CS Fixed, Time	795.03	<.0001	23.46	<.0001	-16.30	<.0001	-17.78	<.0001
TS Fixed	1247.0	<.0001	42.78	<.0001	-24.94	<.0001	-29.16	<.0001
CS, TS Fixed	1038.0	<.0001	33.68	<.0001	-21.48	<.0001	-24.16	<.0001

Combination Test Results with PP Tau								
	Fisher Test		Asymptotic Fisher Test		Inverse Normal Test		Logit Test	
Deterministic Variables	Chi-Square	Pr > ChiSq	FisherI	Pr > FisherI	Z	Pr < Z	L*	Pr < L*
Zero Mean	1422.4	<.0001	50.41	<.0001	-28.07	<.0001	-33.72	<.0001
CS Fixed	1011.3	<.0001	32.52	<.0001	-20.99	<.0001	-23.48	<.0001
CS Fixed, Time	777.25	<.0001	22.68	<.0001	-15.84	<.0001	-17.24	<.0001
TS Fixed	1339.8	<.0001	46.82	<.0001	-26.84	<.0001	-31.70	<.0001
CS, TS Fixed	890.78	<.0001	27.28	<.0001	-18.38	<.0001	-20.22	<.0001

Since p-values are less than 0.05, the null hypothesis of no cointegration between variables (nonstationarity residuals) will be rejected.