Forthcoming in
Panel Data Econometrics Volume 1: Theory, Elsevier.

Panel Cointegration Techniques and Open Challenges

Peter Pedroni*
Williams College

October 5, 2018

Abstract

This chapter discusses the challenges that shape panel cointegration techniques, with an emphasis on the challenge of maintaining the robustness of cointegration methods when temporal dependencies interact with both cross sectional heterogeneities and dependencies. It also discusses some of the open challenges that lie ahead, including the challenge of generalizing to nonlinear and time varying cointegrating relationships. The chapter is written in a nontechnical style that is intended to be assessable to non-specialists, with an emphasis on conveying the underlying concepts and intuition.

JEL Classifications: C33.

Keywords: Panel Time Series, Cointegration, Nonstationary Panels, Nonlinear Panels.

*Correspondence to: ppedroni@williams.edu.

1 Introduction

In this chapter I discuss the development and status of panel cointegration techniques as well as some of the open challenges that remain. Over the past quarter century, the investigation of panel cointegration methods has involved many dozens of econometric papers that have studied and developed methodology as well as many hundreds of economic papers that have employed the techniques. This chapter is not intended to be a survey of the vast literature on the topic. Rather, it is written as a guide to some of the key aspects of the concepts and implementation of panel cointegration analysis in a manner that is intended to be intuitive and assessable to applied researchers. It is also written from the perspective of a personal assessment of the status of panel cointegration techniques and the open challenges that remain.

Notwithstanding the overall approach of the chapter, some occasional overview is instructive to understanding some of the key motivations that have helped to shape the literature and the associated challenges that remain. Indeed, personally, one of the earliest motivations for panel cointegration methods in my Ph.D. dissertation, Pedroni (1993), was the desire to import some of the remarkable features of the time series properties of cointegration into a panel data framework where they could be exploited in the context of data series that are often far too short for reliable cointegration analysis in a conventional time series context. In particular, what was at the time a relatively young field of cointegration analysis for pure time series provided considerable promise in its potential to circumvent traditional concerns regarding endogeneity of regressors due to certain forms of reverse causality, simultaneity, omitted variables, measurement error and so forth. The potential robustness with respect to these features stems fundamentally from the superconsistency properties under cointegration, which are described in the next section.

However, bringing superconsistency associated with cointegration to panel analysis naturally brought to the forefront numerous challenges for panel data analysis that became more apparent in the treatment of the type of aggregate level data that is typically used in cointegration analysis. In particular, while cointegration analysis in panels reduces the need for series to be as long as one would require for cointegration analysis in a pure time series context, it does require the panels to have moderately long length, longer than one would typically require for more conventional panel data techniques that are oriented toward micro economic data analysis. This naturally leads many of the panels that are used for cointegration analysis to be composed of aggregate level data, which are more often observed over longer periods of time and therein fall into the realm of what has come to be known as time series panels.

Typical data include formats such as multi-country panels of national level data, or multi-regional panels or panels composed of relatively aggregated industry level data to give a few examples. With these data formats, the need to address cross sectional heterogeneity becomes apparent, not just in the form of fixed effects as was typical in earlier panel data methods that were oriented toward microeconomic data, but more importantly heterogeneity in both the short run and long run dynamics. Another challenge that becomes more readily apparent from these types of data structures is that the nature of cross sectional dependency is likely to be more complex than was typical in the treatment of earlier micro oriented panel methods, particularly in the sense that the cross sectional dependency is likely to be intertwined with the temporal dependencies. In short, both the cross sectional heterogeneity and the cross sectional dependency interact with an essential feature of time series panels, namely the temporal dependence.

Of course the panel cointegration techniques discussed in this chapter can be applied equally well to microeconomic data panels given sufficient length of the panels. But by addressing the challenges that arise from the typical applications to aggregate level ”macro” panels they have helped to highlight some of the attractive features of panel time series techniques in general, which has helped to fuel the growth of the literature. One way to better appreciate this is to compare these methods in broad terms to alternative strategies for empirical analysis of aggregate level data. For example, at one end of the spectrum, one can consider simply using cross sectional methods to study aggregate country level data. While this has the attraction of providing ample variation in macroeconomic conditions along the cross sectional dimension, it runs into the usual challenges in treating endogeneity and searching for underlying structural causation. Furthermore, when the cross sections represent point in time observations, the estimation may reflect the arbitrariness of the time period, and similarly, when the cross sections represent averages over time the estimation may reflect a relationship that exists among the unconditional time averages rather than for a well defined sense of a long run steady state relationship.

Another strategy might be to use more conventional static micro panel methods for aggregate data. In fact, static micro panel methods can be viewed as essentially repeated cross sections, observed in multiple time periods. But aside from offering controls for unobserved fixed effects or random effects, in the absence of cointegration, the challenges in treating endogeneity and the issues associated with the temporal interpretation still pertain for these methods. While dynamic panel methods such as those of Holz-Eakin, Newey and Rosen (1988) and Arellano and Bond (1991) among others exist for micro data which can help give more precise meaning to the temporal interpretations, the difficulty with these approaches is that they require the dynamics to be strictly homogeneous among the individual members of the panel. When this assumption is violated, as would be typical for most aggregate data, then, as noted in

Pesaran and Smith (1995) and discussed in detail in section 3 of this chapter, this leads to inconsistent estimation, even for the average dynamic relationships, which makes these dynamic panel methods unattractive for the analysis of dynamics in aggregate level macro type data.

At the other end of the spectrum of alternatives, it is worth considering what one learns simply from time series estimation applied to the series of an individual country. In this context plenty of methods exist for treating endogeneity without the need for external instruments, and providing specific temporal interpretations is often central to these methods. However, by using the data from an individual country, the sample variation that pertains to a particular question of interest may be limited. For example, learning about the economic consequences of changing from one type of monetary policy regime to another type is difficult when the time series data from a country spans only one regime. For this, cross sectional variation that spans both regimes in the form of multi-country time series data becomes important and useful.

Viewed from this perspective, panel time series methods, which includes panel cointegration techniques, provide an opportunity to blend the attractive features of time series with potential aggregate level cross sectional variation in data settings where the time series length are moderate. Furthermore, as we will see, when the challenges posed by the interaction of temporal dependencies with cross sectional heterogeneity and cross sectional dependence are properly addressed, the techniques offer a further opportunity to study the underlying determinants of the cross sectional variation.

The remainder of this chapter is structured as follows. In the next section, I use a simple bivariate example to review the concepts behind the superconsistency result that is key to understanding the robustness properties that cointegration brings to panel data analysis. Following this, in sections 3 through 7, I describe how the challenge of addressing cross sectional heterogeneity in the dynamics has shaped testing, estimation and inference in cointegrated panels, including testing for directions of long run causality in panels. In sections 8 and 9, I discuss how addressing the interaction of both cross sectional heterogeneity and cross sectional dependencies continue to drive some of the open challenges in panel cointegration analysis, and in section 10, I conclude with a discussion of some open challenges that are being explored currently which are associated with generalizing panel cointegration analysis to allow for time varying heterogeneity and nonlinearities in the long run relationships. It should be reiterated that this chapter is not intended as a comprehensive or even partial survey, as the panel cointegration literature on the whole is vast, and there are by necessity topics that are not touched upon in detail here, including for example non-classical, Bayesian approaches, as these are reserved for another chapter.

2 Cointegration and the Motivation for Panels

In this section of the chapter I discuss the property of superconsistency and the motivation that this gives to bringing cointegration to a panel setting in order to allow for estimation that is robust to a myriad of issues typically associated with endogenous regressors. In particular, to gain some intuition, I illustrate these concepts using a simple bivariate OLS regression framework.

Toward this end, consider the following very simple and standard example taken from a classical time series perspective. Let

$$y_t = \alpha + \beta x_t + \mu_t \tag{1}$$

for $t = 1, ..., T$ be the data generating process that describes the true unknown relationship between y_t and x_t for some unknown error process μ_t. For simplicity of notation, we will work with the time demeaned versions of the variables, so that $y_t^* = y_t - T^{-1} \sum_{t=1}^{T} y_t$ and similarly $x_t^* = x_t - T^{-1} \sum_{t=1}^{T} x_t$. Then we know that the OLS estimator for β can be written as

$$\hat{\beta}_{OLS} = \frac{\frac{1}{T} \sum_{t=1}^{T} x_t^* (\beta x_t^* + \mu_t)}{\frac{1}{T} \sum_{t=1}^{T} x_t^{*2}} = \beta + R_T, \tag{2}$$

$$where\ R_T = \frac{R_{1T}}{R_{2T}},\ \ R_{1T} = \frac{1}{T} \sum_{t=1}^{T} x_t^* \mu_t,\ \ R_{2T} = \frac{1}{T} \sum_{t=1}^{T} x_t^{*2} \tag{3}$$

Thus, OLS is a consistent estimator of the true value β only when the remainder term R_T is eliminated, and much of the use and adaptation of OLS for empirical work revolves around the conditions under which this occurs.

When x_t and μ_t are both covariance stationary, and in the simplest special case are i.i.d. serially uncorrelated over time, then as we envision the sample growing large and consider $T \to \infty$, the probability limit of both the numerator and denominator go to constants, such that $R_{1,T} \to E_T[x_t^* \mu_t] = \sigma_{x,\mu}$ and $R_{2,T} \to E_T[x_t^{*2}] = \sigma_x^2$. Thus, OLS becomes consistent such that $\hat{\beta}_{OLS} \to \beta$ only when x_t and μ_t are orthogonal, such that $E_T[x_t^* \mu_t] = \sigma_{x,\mu} = 0$. When the condition is violated, one classic solution is to look for an external instrumental variable, z_t, such that $E_T[z_t^* \mu_t] = 0$ and $E_T[z_t^* x_t^*] \neq 0$, which can often be difficult to justify in practice, particularly for aggregate time series data.

However, in a different scenario, wherein x_t and μ_t are not both covariance stationary, but rather x_t is unit root nonstationary, denoted $x_t \sim I(1)$, while μ_t is covariance stationary, denoted $\mu_t \sim I(0)$, then y_t and x_t are said to be cointegrated, in which case the large sample OLS properties become very different. Specifically, in this case OLS becomes consistent in the sense that $\hat{\beta}_{OLS} \to \beta$ regardless of whether or not the regressor is orthogonal to the residual μ_t, and regardless of any serial correlation dynamics that endogenously relate the changes in x_t to μ_t. In a nutshell, this occurs due to the fact that since x_t is nonstationary, its variance is no longer finite but rather grows indefinitely with respect to the sample size, while by contrast, under cointegration, due to the stationarity of μ_t, the covariance between x_t and μ_t does not similarly diverge.

To see this more precisely, it is worth introducing a few concepts that are typically used in the analysis of nonstationary time series, which will be useful in other sections of this chapter as well. For example, to allow for fairly general vector stationary processes with jointly determined serial correlation dynamics, one typically assumes that the conditions are present for a multivariate functional central limit theorem, which essentially generalizes more standard central limit theorems to allow for time dependent processes. Specifically, if we let $\xi_t = (\mu_t, \eta_t)'$ where $\Delta x = \eta_t$ is the stochastic process which describes how x_t changes, then we can replace the standard central limit theorem for *i.i.d.* processes with one that allows for endogenous, jointly determined dependent process by writing

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \xi_t \Rightarrow B_r(\Omega) \;\; as\ T \to \infty \;\; for\;\; r \in [0,1] \, , \tag{4}$$

where $B_r(\Omega)$ is a vector of demeaned Brownian motion with long run covariance Ω. This functional central limit theorem applies for a broad class of processes for ξ_t, including for example linear time series representations such as VARs.

If we define the vector $Z_t = Z_{t-1} + \xi_t$, it is fairly straightforward to show based on (4) and what is known as the continuous mapping theorem, that

$$\frac{1}{T} \sum_{t=1}^{T} Z_t \xi_t' \Rightarrow \int_{r=0}^{1} B_r(\Omega) dB_r(\Omega) + \Lambda + \Omega_0 \;\; as\ T \to \infty \tag{5}$$

$$\frac{1}{T^2} \sum_{t=1}^{T} Z_t Z_t' \Rightarrow \int_{r=0}^{1} B_r(\Omega) B_r(\Omega)' dr \;\; as\ T \to \infty. \tag{6}$$

Note that these expressions simply indicate that the sample statistics on the left of the thick arrows converge in distribution to the expressions on the right of the thick arrow, which are multivariate stable distributions expressed in terms of Brownian motion, known as Brownian motion functionals. In the case of (5) the distribution is further uncentered by constants, which come from the decomposition of the long run covariance matrix into its forward spectrum, Λ, and standard covariance, Ω_0, components such that $\Omega = \Lambda + \Lambda' + \Omega_0$. But for our current purposes, the more important detail to notice is that the OLS remainder terms from (3) are closely related to the sample statistics on the left hand sides of (5) and (6), such that the numerator and denominator terms correspond to off-diagonal and lower diagonal elements of these matrix expressions, so that $R_{1T} = (\frac{1}{T}\sum_{t=1}^{T} Z_t\zeta_t')_{21}$ and $R_{2T} = (\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t')_{22}$. Therefore, according to (5), R_{1T} converges to a stable distribution as the sample grows large. By contrast, according to (6), R_{2T} is off by a factor of T. In order to converge to a stable distribution, one would need to divide R_{2T} by an additional factor of T. By not doing so in the construction of the OLS estimator, the implication is that R_{2T} diverges to infinity at rate T so that the remainder term $R_T = R_{2T}^{-1} R_{1T}$ collapses to zero as the sample size grows large. Therefore, under cointegration we have

$$R_{1T} \Rightarrow \left(\int_{r=0}^{1} B_r(\Omega) dB_r(\Omega) \right)_{21} + \Lambda_{21} + \Omega_{0,21} \;\; as\ T \to \infty \tag{7}$$

$$R_{2T} \to \infty\ ,\;\; R_T \to 0\ ,\;\; \hat{\beta}_{OLS} \to \beta\ ,\;\; as\ T \to \infty. \tag{8}$$

Notice that under cointegration this occurs regardless of the covariance structure between x_t and μ_t in the DGP. Furthermore, since under (4) the vector process for μ_t and $\Delta x_t = \eta_t$ is permitted to have very general forms of dynamic dependence, the parameter β can be interpreted as the relationship between x_t and y_t that is invariant to any stationary and therefore transitional dynamics associated with either changes in x_t or changes in y_t conditional on x_t. In this way, the parameter β can also be interpreted as reflecting the stable steady state relationship that exists between x_t^* and y_t^*, which under cointegration can be estimated consistently even when the transition dynamics are unknown and omitted from the estimation.

For these reasons, the presence of cointegration brings with it a form of robustness to many of the classic empirical problems that lead to the so called violation of exogeneity condition for the regressors. Obvious examples include omitted variables, measurement error, simultaneity, reverse causality, or indeed anything that leads the data generating process, for $\Delta x_t = \eta_t$ to be jointly determined with the data generating process, hereafter referred to as the DGP, for μ_t. To be clear, one must make sure that the reasons for the violation are not so extreme as to essentially break the cointegration and thereby induce μ_t to become unit root nonstationary. For example, measurement error that is stationary but unknown will not affect consistency of the OLS estimator, nor will omission of a stationary variable, nor will omission of stationary transition dynamics, and so forth. But if the measurement error is itself unit root nonstationary, or the omitted variable is unit root nonstationary and belongs in the cointegrating relationship such that without it μ_t is nonstationary, then robustness is lost. But of course this is just another way to state the fact that y_t and x_t are not cointegrated, in which case there is no claim to the robustness. In practice, one can either assert on an *a priori* basis that the cointegration is likely to hold based on economic reasoning, or more commonly, one can test whether the cointegrating relationship appears to hold empirically, as I discuss in the next section.

Of course, these arguments are based on asymptotics, and the practical question is how closely these properties hold as approximations in small samples. If the empirical interest were limited only to the actual estimation of the steady state relationship by OLS under cointegration, then on balance one could say that estimation performs reasonably well in small samples, though needless to say, precisely how well it performs depends on a myriad of details of what the regression omits relative to the DGP.

The bigger practical issue, however, pertains to the performance of the various tests typically associated with cointegration analysis. For example, as alluded to in the previous paragraph, one is often interested in confirming by empirical test whether a relationship is cointegrated, so that one has greater confidence that the robustness properties associated with cointegration are in play. Similarly, beyond simply robustly estimating the coefficients associated with the long run steady state relationship, one is interested in conducting inferential tests regarding the estimated coefficients, or simply reporting standard errors or confidence bands. In contrast to what is required in order to consistently and robustly estimate the steady state relationship, each of these inferential aspects of cointegration analysis require one to account for the stationary transitional dynamics, most commonly through estimation either parametrically or non-parametrically. The classic methods for these are also based on asymptotic arguments, and it is these methods for treating the dynamics that often require distressingly long time series in order to perform well. It is in this context that panels can help to substantially reduce the length of the series required in order for the tests to perform well and for the inference to be reliable.

However, by using cross sectional variation to substitute for temporal variation in the estimation of the transitional dynamics, this is the context in which the challenges posed by the interaction of temporal dependencies with cross sectional heterogeneity and cross sectional dependence arise, as discussed in the introductory section. This is an important theme for the next section, in which I discuss how these challenges help to shape the strategies for testing cointegration in time series panels and constructing consistent and robust methods of inference in cointegrated panels.

3 Strategies for Treating Cross Sectional Heterogeneity in Cointegration Testing and Inference

In the next several sections I discuss some the key aspects of using panels to test for the presence of cointegration and to test hypotheses about cointegrating relationships in panels. As discussed in the previous section, classic approaches to this in time series contexts invariably require the estimation of dynamics. An important challenge for panels occurs when these dynamics are cross sectionally heterogeneous, as one would expect for virtually all aggregate level data, and in this section I describe in greater detail the challenge that this creates. Specifically, cross sectional heterogeneity in the dynamics rules out standard approaches to pooling data cross sectionally as is done in tradition micro panel methods. This is due to the fact that if one pools the data when the true dynamics are heterogeneous, this leads to inconsistent estimation of all coefficients of the regression. More precisely, as pointed out in Pesaran and Smith (1995), in the presence of heterogeneity, the pooled coefficients on lagged dependent variables do not converge to any notion of the average of the underlying heterogeneous parameters as one might hope.

To see this point more clearly, consider a simple illustration for a dynamic process characterized by a first order autoregressive process. For example, imagine that for a panel y_{it} with $i = 1, ..., N$ cross sectional units, which I will often refer to as "members" of the panel to avoid ambiguity, and $t = 1, ..., T$ time periods, the data generating process for the dynamics in stationary form can be represented as

$$\Delta y_{it} = \alpha_i + \phi_i \Delta y_{it-1} + \mu_{it} \ , \tag{9}$$

$$\phi_i = \phi + \eta_i \ , \ \ \eta_i \sim iid(0, \sigma_\eta^2) \ , \ \ \sigma_\eta^2 < \infty \ , \ \ |\phi_i| < 1 \ \forall i \ , \tag{10}$$

so that the coefficient reflecting the stationary transition dynamics, ϕ_i, is heterogeneous among the members of the panel, i. But imagine that in the process of estimation the dynamic coefficient is pooled across i, so that estimation takes the form

$$\Delta y_{it} = \alpha_i + \phi \Delta y_{it-1} + v_{it} \ , \tag{11}$$

so that we have in effect imposed the homogeneity restriction $\phi_i = \phi \ \forall i$, when in truth $\phi_i = \phi + \eta_i$. This would not be a problem if the pooled estimation for $\hat{\phi}$ consistently estimated the average or some other notion of the "typical" value of ϕ_i among the members of the panel. But as noted by Pesaran and Smith (1995), this is not what happens under this scenario. To see this, notice that for the estimated residuals in (11) we have

$$v_{it} = \mu_{it} + \eta_i \Delta y_{it-1} \ , \tag{12}$$

which now consists of both the original stochastic term μ_{it} from the DGP plus a contamination term $\eta_i \Delta y_{it-1}$. Consequently, $E[(\Delta y_{it-1} - \Delta \bar{y}_{it-1})v_{it}] \neq 0$ and the usual condition for consistency is violated so that the pooled OLS estimator no longer estimates the average value for ϕ_i in the sense that $\hat{\phi}_{POLS} \nrightarrow \phi$. Most importantly, there is no easy solution to this problem when the heterogeneous coefficients are pooled, since the same value Δy_{it-1} appears in both the regressor and the residuals, so that instrumentation for this is not possible. This is a simple illustrative example, but the principle generalizes to higher order dynamics and multi-variate dynamics. Indeed, this issue is pervasive in any panel time series methods that estimate dynamics, and since both testing for the presence of cointegration and constructing consistent tests for hypotheses about cointegrating relationships typically require estimation of dynamics, this issue must be addressed in most panel cointegration techniques.

4 Treating Heterogeneity in Residual Based Tests for Cointegration

In this section I focus specifically on the challenges that cross sectional heterogeneity in the dynamics creates for testing for the presence of cointegration in panels, with an initial focus on residual based tests. Furthermore, it is important to understand that in addition to the issue of heterogeneity in the stationary transition dynamics as discussed in the previous section, the specifics of testing for the presence of cointegration also introduces another important heterogeneity issue, which is possible heterogeneity of the *long run* steady state dynamics. This was an important theme regarding cointegration testing in Pedroni (1993), as presented at the 1994 Econometric Society meetings and then most widely circulated as Pedroni (1995) and many years later published as part of Pedroni (1999, 2004). To understand this issue, which is fairly unique specifically to testing for the presence of cointegration, consider a panel version of the DGP described in (1) so that we have

$$y_{it} = \alpha_i + \beta_i x_{it} + e_{it}. \tag{13}$$

Imagine, analogous to the discussion surrounding heterogeneity of the stationary transition dynamics, that the cointegration slope of the steady state dynamics are also heterogeneous, so that by analogy

$$\beta_i = \beta + \eta_i \, , \;\; \eta_i \sim iid(0, \sigma_\eta^2) \, , \;\; \sigma_\eta^2 < \infty \, . \tag{14}$$

Again, imagine that in the process of estimation the cointegration slope coefficient is pooled across i, so that estimation takes the form

$$y_{it} = \alpha_i + \beta x_{it} + v_{it}. \tag{15}$$

so that the homogeneity restriction $\beta_i = \beta \; \forall i$ has been imposed when in truth $\beta_i = \beta + \eta_i$. Now, similar to when we were studying the consequences of ignoring heterogeneity in stationary dynamics, the regression error term in (15) becomes

$$v_{it} = e_{it} + \eta_i x_{it} \, , \tag{16}$$

which consists of both the original stochastic term e_{it} from the DGP plus a contamination term $\eta_i x_{it}$.

In this case the consequences we wish to consider are specifically for testing whether y_{it} and x_{it} are cointegrated. Specifically, if the linear combination is stationary, so that e_{it} is stationary, denoted $e_{it} \sim I(0)$, then y_{it} and x_{it} are cointegrated, whereas if the linear combination is unit root nonstationary, so that e_{it} follows a nonstationary unit root process, denoted $e_{it} \sim I(1)$, then y_{it} and x_{it} are not cointegrated. But notice now that based on (16) $v_{it} \sim I(1)$ follows a unit root process due to the fact that the contamination term $\eta_i x_{it}$ inherits a unit root from $x_{it} \sim I(1)$. This implies that $v_{it} \sim I(1)$ *regardless* of whether or not y_{it} and x_{it} are in truth cointegrated. Consequently, if the true cointegrating relationships are heterogeneous across i in the sense that $\beta_i \neq \beta \; \forall i$, then tests constructed from pooled regressions that treat $\beta_i = \beta \; \forall i$ will produce inconsistent tests in that they cannot distinguish between the presence or absence of cointegration regardless of the sample size. Furthermore, even when the degree of heterogeneity is small in the sense that η_i is small, due to the fact that this small number multiplies a unit root variable x_{it}, substantial contamination of the stationary component of v_{it} occurs even for very small deviations from a false homogeneity assumption. Thus, the relatively small possible gain in the degrees of freedom obtained from pooling is rarely worth the risk of mispecification, particularly since panel cointegration methods typically already have very high power under standard conditions. For these reasons, although there are later exceptions such as Kao (1999) which pools both the long run steady state dynamics and the stationary transition dynamics, for the most part all other methods for testing for the presence of cointegration allow for heterogeneity of both the short run transition dynamics as well as the long run steady state dynamics, as reflected in the heterogeneity of the cointegration slope.

Indeed, there are by now many different approaches proposed for constructing tests for the presence of cointegration that take into account heterogeneity in both the transition dynamics and the steady state cointegrating relationship. Rather than surveying all of the various approaches, I will focus here on conveying the central idea of treating the cross sectional heterogeneity in both the short run and long run dynamics. For this I will use examples based on residual based tests in this section, as well as ECM based tests in the next section. The first two examples are taken from Pedroni (1999, 2004). Somewhat ironically, due to the lengthy and uneven publication process, the 2004 paper is actually the published version of the original paper, with the 1999 one being the published version of follow-up paper, which reported numerical adjustment values that applied to the case in which larger numbers of variables were used in the cointegrating regressions. Both papers studied a total of seven different statistics spanning various parametric and semi-parametric approaches, but I will focus here on only the parametric ADF based test statistics that were studied in order to illustrate two different general methods for treating heterogeneity.

Specifically, I will use these to illustrate two different methods for treating the heterogeneous stationary transition dynamics. The first uses a technique that conditions out the heterogeneity in the pooled dynamics, while the second uses a simple group mean technique for accommodating heterogeneous dynamics. In keeping with the illustrative style of this chapter, I will continue to use a bivariate regression example, although of course all of the techniques generalize to multivariate regresssions.

Since all of these methods account for potential heterogeneity in the long run steady state dynamics, the first stage regression of the residual based methods always takes the form

$$y_{it} = \alpha_i + \beta_i x_{it} + e_{it}. \tag{17}$$

for the bivariate case. The only difference in the various methods of testing lies in how the estimated residuals $\hat{e}_{it}$ from this regression are treated, either semi-parametrically using long run variance estimators or parameterically using ADF principles as with the two that I discuss here.

The first of these, taken from Pedroni (1999, 2004), is based on constructing a pooled ADF regression on the estimated residuals $\hat{e}_{it}$ by conditioning out the heterogenous dynamics. Toward that end, rather then estimating the full ADF regression with lagged differences, one estimates a simple DF type of regression, but with the dynamics conditioned out individually for each member of the panel, for both the regessor and regressand. Specifically, the regression takes the form

$$\hat{v}_{it} = \rho \hat{\eta}_{i,t-1} + u_{it} \tag{18}$$

where $\hat{v}_{it}$ and $\hat{\eta}_{i,t-1}$ are obtained as the estimated residuals from the regressions

$$\Delta \hat{e}_{i,t} = \sum_{k=1}^{K_{1i}} \hat{\gamma}_{1i,k} \Delta \hat{e}_{i,t-k} + v_{it} \tag{19}$$

$$\hat{e}_{i,t-1} = \sum_{k=1}^{K_{2i}} \hat{\gamma}_{2i,k} \Delta \hat{e}_{i,t-k} + \eta_{i,t-1} \tag{20}$$

applied to each of the members of the panel individually. Notice that (19) and (20) thereby serve to condition out the member specific dynamics for the for significance of ρ in the pooled DF style regression (18).

This method for conditioning out the heterogeneous dynamics is analogous to the approach taken in Levin, Lin and Chu's (2003) panel unit root test. Indeed, a further refinement, consistent with LLC's approach, can be made for the cross sectional heteroscedasticity of the long run variances, in which case it is known as the "weighted" pooled ADF based test. But as shown in Pedroni (2004), this refinement is not necessary for the consistency of the test, even when the dynamics and thus the long run variances are heterogeneous across i, provided that the dynamics are conditioned out via regressions (19) and (20). The "unweighted" pooled version therefore simply computes the t-statistic associated with the pooled estimator for ρ in (18), which we will denote here as t_{POLS}.

The final step then is to adjust the statistic in a manner that will allow it to converge in distribution as the sample size grows large. Specifically, the adjustment takes the form

$$Z_{PADF} = \frac{t_{POLS,\rho} - \mu_{PADF,\rho}\sqrt{N}}{\sqrt{\nu_{PADF,\rho}}}. \tag{21}$$

The adjustment terms $\mu_{PADF,\rho}$ and $\nu_{PADF,\rho}$ are numerical values that are either computed analytically or simulated based on the properties of the distribution of $t_{POLS,\rho}$, and depend on the moments of the underlying Wiener functionals that describe the distributions. The particular numerical values that result from these computations or simulations differ depending on details of the hypothesized cointegrating relationship (17), such as whether intercepts or trends are included, and also on the number of regressors that are included, and are reported accordingly in Pedroni (2004) for the case of a single regressor, and in Pedroni (1999) for the case of multiple regressors. The adjusted statistic is then distributed as standard normal under the null hypothesis of no cointegration and diverges to the left under the alternative of cointegration, so that for example -1.28 and -1.64 are the 10 percent and 5 percent critical values required to reject the null in favor of cointegration.

As noted previously, conditioning out the member specific dynamics prior to pooling is just one strategy for dealing with heterogeneous transition dynamics. Another technique, which has become far more common, is to use group mean methods rather than the combination of pooling with heterogenous dynamic conditioned out of the regression. Group mean methods have become more popular in large part because they are relatively easier to implement and interpret. To illustrated this, I use a second example taken from Pedroni (1999, 2004), namely the group mean ADF residual based test. To implement this test, one begins by simply estimating the individual ADF regressions using the estimated residuals from the hypothesized cointegrating regression (17), so that one estimates

$$\Delta \hat{e}_{i,t} = \rho_i \hat{e}_{i,t-1} + \sum_{k=1}^{K_i} \gamma_{i,k} \Delta \hat{e}_{i,t-k} + u_{i,t} \tag{22}$$

by OLS individually for each member i of the panel. The group mean ADF t-statistic for the null of cointegration is then computed as $t_{GOLS,\rho} = N^{-1} \sum_{i=1}^{N} t_{i,ADF}$, where $t_{i,ADF}$ is the standard ADF t-statistic for significance of ρ_i for member i. The statistic is then adjusted to ensure it converges in distribution as the sample grows large, so that

$$Z_{GADF} = \frac{t_{GOLS,\rho}\sqrt{N} - \mu_{GADF,\rho}\sqrt{N}}{\sqrt{\nu_{GADF,\rho}}}. \tag{23}$$

where $\mu_{GADF,\rho}$ and $\nu_{GADF,\rho}$ are numerical values that are either computed analytically or simulated based on the properties of the distribution of $t_{GOLS,\rho}$, and depend on the moments of the underlying Wiener functionals that describe the distributions. While these values differ from those of $\mu_{PADF\rho}$ and $\nu_{PADF\rho}$, they also depend on the details of the hypothesized cointegrating relationship (17), such as whether intercepts or trends are included, and also on the number of regressors that are included, and are reported accordingly in Pedroni (2004) for the case of a single regressor, and in Pedroni (1999) for the case of multiple regressors. Again the statistic is distributed as standard normal under the null hypothesis of no cointegration and diverges to the left under the alternative of cointegration, so that for example -1.28 and -1.64 are the 10 percent and 5 percent critical values required to reject the null in favor of cointegration.

Monte Carlo simulation studies reported in Pedroni (2004) show that for all of the residual based test statistics studied in the paper, including the two ADF based tests described above, size distortions are low and power is extremely high even in modestly dimensioned panels. For example, even when the time series length, T, is far too short for reliable inferences in a conventional time series context, in the panel framework, panels with similarly short lengths for T and modest N dimensions can in many cases deliver close to 100 percent power with relatively small degrees of size distortion.

5 Comparison of Residual Based and Error Correction Based Testing

While residual based methods are the most common approach, there are also other methods for testing for cointegration in time series, which have been extended to heterogenous panel frameworks. One such example are error correction methods, and it is worth comparing these to residual methods in order to understand the trade-offs. For example, Westerlund (2007) studied the use of single equation ECMs in panels with heterogeneous dynamics, including a group mean version. In contrast to residual based methods, single equation ECM approaches require the assumption of weak exogeneity. The basic idea then is to exploit this assumption in order to estimate the error correction loading parameter from a single equation and use it to test for the null of no cointegration.

The first step is therefore to estimate by OLS what is known as an augmented form of the ECM equation as

$$\Delta y_{it} = c_i + \lambda_{1,i} y_{i,t-1} + \gamma_i x_{i,t-1} + \sum_{j=1}^{K_i} R_{i,j,11} \Delta y_{i,t-j} + \sum_{j=-K_i}^{K_i} R_{i,j,12} \Delta x_{i,t-j} + \epsilon_{1,it}, \tag{24}$$

where $\gamma_i = -\lambda_{1,i}\beta_i$. The equation has been augmented relative to the standard ECM equation by the inclusion of lead terms of the differences in Δx_{it}, rather than just the usual lagged terms of Δx_{it}. This allows one to loosen the exogeneity requirements on x_{it} to one of weak exogeneity, rather than stronger forms of exogeneity. As discussed later, imposing weak exogeneity in this context can be interpreted as imposing the *a priori* restriction that causality only runs in one direction in the long run cointegrating relationship, from innovations in x_{it} to y_{it}, but not the other way around. Imposing such an exogeneity restriction is contrary to the full endogeneity that is typically permitted for most panel cointegration methods, and the implications of this are discussed below.

Under the maintained assumption of weak exogeneity in the relationship between y_{it} and x_{it}, the null of no cointegration between y_{it} and x_{it} can be tested by testing whether $\lambda_{1,i} = 0$, and so the group mean test is constructed by computing the average value of the t-statistics associated with these such that $t_{GOLS,\lambda} = N^{-1} \sum_{i=1}^{N} t_{i,\lambda}$, where $t_{i,\lambda}$ are the individual t-statistics for significance of λ_{1i} for each member i. Analogous to the other tests, this statistic can then be standardized as

$$Z_{G\lambda} = \frac{t_{GOLS,\lambda}\sqrt{N} - \mu_{GOLS,\lambda}\sqrt{N}}{\sqrt{\nu_{GOLS,\lambda}}}. \tag{25}$$

where $\mu_{GOLS,\lambda}$ and $\nu_{GOLS,\lambda}$ are the numerical adjustment values based on the properties of the distribution of $t_{GOLS,\lambda}$, so that $Z_{G\lambda}$ is similarly distributed as standard normal under the null hypothesis of no cointegration and diverges to the left under the alternative of cointegration.

To better understand the motivation for the ECM based approach in relation to residual based approaches, and to see the consequences of violating the specialized weak exogeneity condition, it is worth comparing the details of the ECM estimation equation to the residual based estimation equation. In particular, consider rearranging the various terms in (24) as

$$\begin{aligned}\Delta y_{it} - \beta_i \Delta x_{it} = \lambda_{1,i}(y_{it-1} - \beta_i x_{it-1}) + \sum_{j=1}^{K_i} R_{ij,11} \Delta y_{it-j} + \sum_{j=1}^{K_i} R_{ij,11} \beta_i \Delta x_{it-j} \\ + \sum_{j=0}^{K_i} R_{ij,12} \Delta x_{it-j} - \sum_{j=0}^{K_i} R_{ij,11} \beta_i \Delta x_{it-j} + \epsilon_{1,it}\end{aligned} \tag{26}$$

where for ease of notation I have dropped the deterministics, c_i, and the leads of Δx_{it} from the equation since these are not central to the issues that I discuss next. Specifically the above form is convenient as it allows us to substitute e_{it} for $y_{it} - \beta_i x_{it}$ and similarly for $\Delta e_{it} = \Delta y_{it} - \beta_i \Delta x_{it}$ where these appear in the first line of (26). This gives us the form

$$\Delta e_{it} = \lambda_{1,i} e_{it-1} + \sum_{j=1}^{K_i} R_{ij,11} \Delta e_{it-j} + \sum_{j=0}^{K_i} R_{ij,12} \Delta x_{it-j} - \sum_{j=0}^{K_i} R_{ij,11} \beta_i \Delta x_{it-j} + \epsilon_{1,it}, \tag{27}$$

which allows us to easily compare what the ECM equation is estimating relative to what the residual based methods are estimating. In particular, for a given finite lag truncation K_i, we can see from (27), that estimating the ADF regression for the residuals e_{it} is equivalent to setting $\sum_{j=0}^{K_i} R_{ij,12} \Delta x_{it-j} = \sum_{j=0}^{K_i} R_{ij,11} \beta_i \Delta x_{it-j}$. This is the so called "common factor" restriction. One of the motivations for ECM based approaches is that residual based tests ignore information contained in fact that these two factors need not be the same, and ignoring this can add variance to the small sample distribution of the $\lambda_{1,i}$ estimator. It should be noted however, that this is not a form of misspecification that leads to inconsistency. The key is that the lag truncation is not treated as given in residual based methods, and can simply increase to absorb any additional serial correlation due to these terms. So the gain from using the ECM form for the estimation is simply a potential increase in small sample power, although it is not guaranteed to increase power, as this depends on the tradeoff between the number of lag coefficients estimated by the ADF regression versus the number of coefficients estimated by the ECM.

However, the trade-off for this potential increase in small sample power is the specialized assumption of weak exogeneity. In light of this, it is worth considering what the consequences can be when this assumption does not hold, yet the single equation ECM test is used. To see this, it is worth noting that in general for the case in which y_{it} and x_{it} are cointegrated, the VECM representation provides for a two equation system with the error correction coefficient taking the form $\lambda_i \beta_i'$ where λ_i is a loading vector with two elements such that $\lambda_i = (\lambda_{1i}, \lambda_{2i})'$. Cointegration between y_{it} and x_{it} requires that at least one of the values for λ_i is nonzero. In this context, the weak exogeneity assumption can be interpreted as an *a priori* assumption that $\lambda_{2i} = 0$. Therefore, since λ_{2i} is zero by assumption, then in order for y_{it} and x_{it} to be cointegrated, it must be the case that λ_{1i} is nonzero, and hence the test for the null of no cointegration proceeds by testing whether $\lambda_{1i} = 0$ via (24). But the risk with this strategy is that if the *a priori* maintained assumption that $\lambda_{2i} = 0$ corresponding to weak exogeneity turns out not to be true, then the test risks becoming inconsistent in the sense that it cannot distinguish the null of no cointegration from the alternative of cointegration no matter how large the sample size.

To see this, consider what the first equation of the VECM form looks like when the weak exogeneity assumption is not true, so that potentially both elements of λ_i appear in front of the error correction term, which can be written as

$$\Delta y_{it} = (\lambda_{1,i} - R_{i0,12}\lambda_{2,i})(y_{it-1} - \beta_i x_{it-1}) + \sum_{j=1}^{K_i} R_{ij,11}\Delta y_{i,t-j} + \sum_{j=1}^{K_i} R_{ij,12}\Delta x_{i,t-j} + \epsilon_{1,it}. \tag{28}$$

In this context, the single equation ECM based approach can be interpreted as testing whether $\lambda_{1,i} - R_{i0,12}\lambda_{2,i} = 0$ under the null of no cointegration. However, in general, the value for $R_{i0,12}\lambda_{2,i}$ is unrestricted under cointegration. Therefore, if we consider the scenarios in which $R_{i0,12}\lambda_{2,i} < 0$ and $\lambda_{1,i} < 0$, but $|R_{i0,12}\lambda_{2,i}| > |\lambda_{1,i}|$ then $(\lambda_{1,i} - R_{i0,12}\lambda_{2,i}) > 0$, and the test will fail to reject the null of cointegration with certainty as the sample size grows despite the fact that the null is false. In this way, in contrast to residual based tests, the test becomes inconsistent in that under this scenario it will be unable to reject a false null even for large samples if the maintained assumption of weak exogeneity is not true. The tradeoff between residual based tests therefore amounts to a tradeoff between a potential gain in small sample power at the expense of robustness in the sense that the test risks become meaningless if the weak exogeneity condition is violated. Since small sample power is already fairly large in almost all tests for the null of no cointegration in panels, for these reasons in most applications the potential gain is unlikely to be worth the risk if one is not absolutely *a priori* certain of the weak exogeneity assumption.

So far I have been discussing panel cointegration tests that are designed to test the classic null of no cointegration, against the alternative of cointegration. However, for some applications one might be interested in reversing the null hypothesis, so that the null becomes cointegration against the alternative of no cointegration. In principle it can be useful to consider both types of tests, in particular when the empirical application is likely to be such that results may be mixed in the sense that some members of the panel are best described as cointegrated while others may not be cointegrated. While doing a test for the null of cointegration does not resolve the issue of so called mixed panel applications per se, which we discuss in greater detail later in this chapter, the combination of both types of tests can sometimes serve to narrow down the fraction of individual members that are consistent with either alternative as discussed and illustrated for example in Pedroni (2007). Indeed there are many proposed tests in the literature for the null of cointegration in panels, starting with McCoskey and Kao (1998), which develops a pooled panel version of the Shin (1994) time series test for the null of cointegration. However, the difficulty with virtually all of the tests that have been proposed in the literature is that, similar to the corresponding time series based tests, they inherit the property of high size distortion and low power in finite samples, and are unable to mitigate this problem even for fairly large panels. A good reference that documents these difficulties through a series of large scale Monte Carlo simulations is Hlouskova and Wagner (2006) for the case of tests for the null of stationarity which also applies to tests for the null of cointegration. A generalized solution to this problem and the related problem associated with inference for mixed panel applications remains an open challenge, which I discuss later in this chapter.

6 Estimation and Testing of Cointegrating Relationships in Heterogeneous Panels

For panels in which cointegration has been established or is expected to hold, the typical next step is to estimate the cointegrating relationships and construct consistent tests of hypotheses pertaining to the cointegrating relationships. In what follows I discuss some simple methods for this which account for heterogeneous dynamics. To be clear, as discussed previously, if one is simply interested to obtain superconsistent estimates, then static OLS provides an immediate solution as it is robust to any features that lead to endogeneity of the regressors, including the omitted dynamics. The problem that presents itself with OLS, however, is that the associated standard errors are not consistently estimated when the regressors are endogenous, even when cointegration is present. The methods discussed here are designed to correct for this, such that both the estimates of the cointegrating relationship and the associated standard errors are consistently estimated so that standard test statistics that rely on standard error estimates, such as t-statistics or F-statistics, can be used.

There are in fact many ways to construct cointegration estimators that also produce consistent standard error estimates for the purposes of testing hypotheses about cointegrating relationships. Here I focus on two relatively easy to understand approaches that are based on the time series principles of fully modified OLS estimation and dynamic OLS estimation. In both cases, the primary strategy is to adjust for a second order bias that arises from the dynamic feedback due to the endogeneity of the regressors by using dynamics of the regressors as an internal instrument. Fully modified OLS makes these adjustments via nonparametric estimates of the autocovariances, while dynamic OLS makes these adjustments by parametric estimates using the leads and lags of the differenced regressors. Since dynamics are estimated in both of these cases, as discussed previously, an important issue for panels is to accommodate any heterogeneity in the dynamics that is likely to be present among the members of the panel. Analogous to previous discussions, one can in principle either use a pooled approach that conditions out the member specific heterogeneous dynamics, or one can use a group mean approach.

Group mean approaches are popular in that they are easy to implement, and the group mean estimates can be interpreted as the average cointegrating relationship among the members of the panel. Another attractive advantage for group mean approaches is that they produce a sample distribution of estimated cointegration relationships for the individual members of the panel which can be further exploited in order to study what characteristics of the members are associated with different values for the cointegrating relationships as illustrated in Pedroni (2007). In what follows I therefore discuss the details of the group mean fully modified OLS (FMOLS) approach developed in Pedroni (2000, 2001) and the group mean dynamic OLS (DOLS) approach introduced in Pedroni (2001).

Specifically, the group mean FMOLS approach simply makes the FMOLS adjustments to each member of the panel individually, and then computes the average of the corresponding cointegration estimator. For example, continuing with the bivariate example of this chapter, the first step is to obtain the estimated residuals $\hat{e}_{it}$ from the OLS regression for the cointegrating relationship, as described in (17). These residuals are then paired with the differences in the regressors to create the panel vector series $\hat{\xi}_{it} = (\hat{e}_{it}, \Delta x_{it})'$. From this the vector of autocovariances $\hat{\Psi}_{ij} = T^{-1} \sum_{t=j+1}^{T} \hat{\xi}_{it}\hat{\xi}'_{it-j}$ are estimated, and then weighted using for example the Bartlett kernel as per the Newey-West estimator, to estimate the various elements of the long run covariance matrix

$$\hat{\Omega}_i = \hat{\Sigma}_i + \hat{\Gamma}_i + \hat{\Gamma}'_i\,, \quad \textit{where}\;\; \hat{\Gamma}_i = \sum_{j=1}^{K_i} \left(1 - \frac{j}{K_i+1}\right) \hat{\Psi}_{ij}\,,\;\; \hat{\Sigma}_i = \hat{\Psi}_{i0} \qquad (29)$$

for each member i for some bandwidth K_i, typically set according to the sample length as $K_i = 4\left(\frac{T_i}{100}\right)^{\frac{2}{9}}$, rounded down to nearest integer. These are then used to create the modification to the usual OLS estimator such that the FMOLS estimator for each member i becomes

$$\hat{\beta}_{FMOLS,i} = \frac{\sum_{t=1}^{T} x_{it}^{*} \tilde{y}_{it}^{*} - T\hat{\gamma}_i}{\sum_{t=1}^{T} x_{it}^{*2}} \ , \tag{30}$$

where analogous to earlier in the chapter, $y_{it}^{*} = y_{it} - T^{-1}\sum_{t=1}^{T} y_{it}$ and $x_{it}^{*} = x_{it} - T^{-1}\sum_{t=1}^{T} x_{it}$ are the time demeaned versions of the variables, and the FMOLS corrections are now such that

$$\tilde{y}_{it}^{*} = y_{it}^{*} - \frac{\hat{\Omega}_{21,i}}{\hat{\Omega}_{22,i}} \Delta x_{it} \ , \ \ \hat{\gamma}_i = \hat{\Gamma}_{21,i} + \hat{\Sigma}_{21,i} - \frac{\hat{\Omega}_{21,i}}{\hat{\Omega}_{22,i}} (\hat{\Gamma}_{22,i} + \hat{\Sigma}_{22,i}) \tag{31}$$

To understand the role of these adjustment terms, it is worth pointing out that according to (7) the numerator of the OLS estimator converges to a distribution with a stochastic nonzero mean due to the feedback effect that arises from the endogeneity of the regressors. Once the adjustment terms (31) are made, the distribution for the FMOLS estimator becomes centered around zero, so that when the FMOLS t-statistic is computed based on the variance of the distribution, the t-statistic becomes asymptotically standard normal. In fact, it is worth noting that in the special case in which the regressors happen to be exogenous, the off-diagonal elements of the autocovariances between Δx_{it} and e_{it} go to zero, so that $\tilde{y}_{it}^{*} \to y_{it}^{*}$ and $\hat{\gamma}_i \to 0$, and therefore the $\hat{\beta}_{FMOLS,i}$ estimator becomes identical to the $\hat{\beta}_{OLS,i}$ estimator.

In any event, once the $\hat{\beta}_{FMOLS,i}$ estimator is computed, the associated FMOLS t-statistics is then constructed on the basis of (30) in a manner analogous to conventional t-statistics, except that in place of the usual standard deviation the standard deviation of the long run variance $\hat{\Omega}_{11,i}$ is used as estimated by (29), so that the FMOLS t-statistic becomes

$$t_{FMOLS,i} = \frac{\hat{\beta}_{FMOLS,i}^{*} - \beta_{o,i}}{\sqrt{\hat{\Omega}_{11,i}^{-1} \sum_{t=1}^{T} x_{it}^{*2}}} \ . \tag{32}$$

The group mean FMOLS estimator and group mean t-statistic are then computed as

$$\hat{\beta}_{GFMOLS} = N^{-1} \sum_{i=1}^{N} \hat{\beta}_{FMOLS,i} \ , \quad t_{GFMOLS} = N^{-1/2} \sum_{i=1}^{N} t_{FMOLS,i} \tag{33}$$

where $\hat{\beta}_{FMOLS,i}$ and $t_{FMOLS,i}$ are the individual member FMOLS estimator and t-statistics from (30) and (32) respectively. Note that since the individual t-statistics have an asymptotic distribution which is standard normal, there is no need to use the usual μ, v adjustment terms to render the group mean asymptotically normal, and indeed under the null $t_{GFMOLS} \Rightarrow N(0,1)$ and under the alternative $t_{GFMOLS} \to \pm\infty$ as a two-tailed test, so that the critical values are the familiar ± 1.96 for the 5% p-value, and so forth.

In pure time series applications, FMOLS is notorious for suffering from small sample size distortion, and this is also inherited to some degree by pooled FMOLS as documented in Pedroni (1996). On the other hand, as documented in Pedroni (2000), group mean FMOLS has remarkly high power and very little size distortion in small samples. Intuitively, this appears to be due to the fact that while the individual FMOLS t-statistic distributions have fairly fat tails that lead to size distortion in short samples, they are nevertheless fairly symmetric so that as the cross sectional dimension N increases, the group mean t-statistic converges quickly and is well approximated by a standard normal even in short panels.

DOLS also appears to behave similarly, with the pooled version inheriting the poor small sample properties, while the grouped version appears to do well. As noted previously, DOLS also makes the adjustments to OLS that are necessary in order to obtain consistent standard errors and thus produce standard tests such as t-statistics that are consistent and nuisance parameter free under the null. However, in contrast to FMOLS which uses estimated autocovariances to make the adjustments, DOLS accomplished the adjustments via a parametric strategy that uses leads and lags of Δx_{it} directly in the regression. Thus, in order to construct the group mean DOLS estimator as described in Pedroni (2001), one first estimates the individual DOLS regression for each member of the panel as

$$y_{it} = \alpha_i + \beta_i x_{it} + \sum_{j=-K_i}^{K_i} \phi_{i,j} \Delta x_{it-j} + e_{it}. \tag{34}$$

The inclusion of the leads and lags of Δx_{it} serve to center the distribution of the numerator of the estimator for β_i here, which we refer to as $\hat{\beta}_{DOLS,i}$, much in the same way that the adjustment with the autocovariances in FMOLS served to center the distribution. Again, analogous to the t-statistic for FMOLS, the DOLS t-statistics is then constructed on the basis of (34) in a manner analogous to conventional t-statistics, except that in place of the usual standard deviation the standard deviation of the long run variance $\hat{\Omega}_{11,i}$ is used, which can be estimated by (29). The corresponding group mean DOLS estimators and t-statistics then become

$$\hat{\beta}_{DOLS} = N^{-1} \sum_{i=1}^{N} \hat{\beta}_{DOLS,i}\ , \quad t_{DOLS} = N^{-1/2} \sum_{i=1}^{N} t_{DOLS,i} \tag{35}$$

and again there is no need to use μ, v adjustment terms to render the group mean asymptotically normal, and under the null $t_{DOLS} \Rightarrow N(0,1)$ while under the alternative $t_{DOLS} \to \pm\infty$ as a two-tailed test, so that the critical values here are also the familiar ± 1.96 for the 5% p-value, and so forth.

As alluded to earlier, pooled approaches are also possible, as for example the pooled FMOLS approaches studied in Pedroni (1996) and Phillips and Moon (1999) and the pooled DOLS studied in Kao and Chiang (2000). In the latter DOLS approach the dynamics are pooled, which can be problematic for reasons discussed previously, but one can easily imagine conditioning out the heterogeneous dynamics in a pooled DOLS approach. Another approach that is sometimes used is the panel autoregressive distributed lag approach of Pesaran (1999). While autoregressive distributed lag approaches are in general built around the assumption that the regressors are fully exogenous, the approach in Pesaran (1999) is able to relax the restriction to one of weak exogeneity, analogous to the assumption discussed previously for the Westerlund (2007) ECM based approach. By contrast, FMOLS and DOLS approaches allow for full endogeneity of the regressors as is typical in the panel cointegration literature. Finally, it is worth noting that rank based tests using VECM approaches are also possible, but we defer this to a more general discussion of rank based tests later in this chapter, and instead turn next to the use of the panel VECM framework for the purposes of causality testing in cointegrated panels.

7 Testing Directions of Long Run Causality in Heterogeneous Cointegrated Panels

It is worth noting that while cointegration analysis in panels is in general robust to the presence of full endogeneity of the regressors, as with any econometric method, consistent estimation in the presence of endogeneity is not synonymous with establishing a direction of causality. In order to establish causality one needs to impose further restrictions that relate the structure of the estimated relationship to exogenous processes, which in general requires additional *a priori* assumptions when the observed processes are endogenous, and is therefore not synonymous with consistency of estimation under endogeneity. I would argue that in this regard, cointegration analysis is on par with any other econometric method that treats endogeneity to establish consistency of estimation. In a nutshell, additional structure is needed to establish the nature of the causal relationships.

Fortunately, in this context cointegration can be interpreted as a type of identification which already implicitly imposes some structure on dynamic systems, so that the additional *a priori* structure that is needed to establish causal relationships can in many cases be relatively easy to come by. Specifically, as discussed previously, in dynamic systems the presence of cointegration can be interpreted to imply the existence of a long run steady state relationship among the variables. Continuing with the bivariate example of this chapter, the implication is that if y_{it} and x_{it} are cointegrated, then a long run causal effect must exist that links the two variables in their steady state. However, the long run causal effect can run in either direction. It can originate in something that induces an innovation in x which causes y to move in the long run, or it can originate in something that induces an innovation in y which causes x to move in the long run, or it can be both of these. It what follows I describe the panel VECM long run causality tests introduced originally in Canning and Pedroni (1997) and eventually published in Canning and Pedroni (2008).

Specifically, the technique relies on the panel VECM form to estimate the vector loadings and construct panel tests based on these. Both the direction of causality and the sign of the causal effect can be tested in this way. To see how this works, it is worth considering how some of the implications of cointegration lead to a natural test for these. Again, I will use a simple bivariate example to illustrate. But to economize on notation, I will use polynomial operator notation here. Specifically, cointegration has three important implications which can be used to understand the nature of the tests. First, cointegration between y_{it} and x_{it} implies that their relationship can be represented in VECM form as

$$R_i(L)\Delta Z_{it} = c_i + \lambda_i \beta_i' Z_{it-1} + \mu_{it} \ , \quad R_i(L) = I - \sum_{j=1}^{P_i} R_{i,j} L^j \tag{36}$$

where $Z_{it} = (y_{it}, x_{it})'$ is the vector of variables, $R_i(L)$ contains the coefficients for the lagged differences which reflect the heterogenous dynamics specific to member i, μ_{it} are the iid white noise innovations, and $\beta_i' Z_{it-1}$ is the error correction term. Since β_i is typically unknown, when (36) is estimated for the purposes of constructing long run causality tests, this error correction term must be estimated individually for each member, and it is important that it be estimated in a manner that has no asymptotic second order bias, such that the associated standard errors are consistently estimated. Thus, the Johansen procedure may be used to estimate the VECM, or alternatively one may use estimated residuals, computed on the basis of the FMOLS or DOLS estimator, so that for example

$$\hat{e}_{FMOLS,it} = y_{it}^* - \hat{\beta}_{i,FMOLS} x_{it}^* \tag{37}$$

is used in place of $\beta_i' Z_{it-1}$ in (36). When $\hat{e}_{FMOLS,it}$ is used in place of $\beta_i' Z_{it-1}$, then each of the equations of (36) can be estimated individually by OLS for each member i to obtain consistent estimates of the loadings λ_i, and the associated t-statistics will be asymptotically standard normal.

The second, fairly trivial implication is that a stationary vector moving average representation exists for the differenced data, ΔZ_{it}, which we write as

$$\Delta Z_{it} = c_i + F_i(L)\mu_{it} \ , \quad F_i(L) = \sum_{j=0}^{Q_i} F_{i,j} L^j \ , \ \ F_{i,0} = I \tag{38}$$

Notice that in this form, when we evaluate the polynomial $F_i(L)$ at $L = 1$, this gives us the total sum $F(1) = \sum_{j=0}^{Q_i} F_{i,j}$, which can be interpreted as the total accumulated response of ΔZ_{it} to the innovations μ_{it}, which is equivalent to the long run steady state response of the levels Z_{it} to the innovations. Therefore, the off-diagonal elements of $F_i(1)$ can be interpreted as the long run responses of the variables to each other's innovations, so that for example $F_i(1)_{21}$ represents the long run response of x_{it} to a $\mu_{it,1}$ unanticipated innovation in y_{it} and therefore can be interpreted as a measure of the causal effect from y to x.

The third, and in this context most substantial implication of cointegration, is known as the Granger representation theorem, which serves two tie together the first two implications. Specifically, it tells us that the relationship between the loadings on the error correction terms and the long run steady state responses of the levels is restricted via a singularity such that

$$F_i(1)\lambda_i = 0. \tag{39}$$

If for example we are interested to test hypotheses regarding the long run causal effect represented by $F_i(1)_{21}$, then we can use one of the characteristic equations of (39) to see the implications in terms of the loadings, λ_i. Specifically, (39) implies that

$$F_i(1)_{21}\lambda_{i,1} + F_i(1)_{22}\lambda_{i,2} = 0. \tag{40}$$

Under cointegration, both elements of λ_i cannot be zero, as in this case the error correction term would drop out of (36). If we are willing to make the fairly innocuous assumption that x causes itself to move in the long run, so that $F(1)_{i,22} \neq 0$, then (40) implies that $F_i(1)_{21} = 0$ if and only if $\lambda_{i,2} = 0$. This implies that the construction of a test for the null hypothesis that $\lambda_{i,2} = 0$ becomes a test for the null of no long run causality running from y to x. A grouped panel version of the t-statistic for this test can then be constructed as

$$Z_{GLRC} = N^{-1/2} \sum_{i=1}^{N} t_{i,\lambda_2} \tag{41}$$

where t_{i,λ_2} is the individual t-statistic for the significance of $\lambda_{i,2}$ for unit i. Under the null hypothesis of no long run causality running from y to x, the grouped test is asymptotically standard normal, while under the alternative the test diverges to positive or negative infinity. By substituting t_{i,λ_1} in place of t_{i,λ_2} in (41) one can analogously test for the null hypothesis of no long run causality running from x to y.

Of course, since these are two tailed tests, it is possible that positive and negative values for the loadings are averaging out over the i dimension, so that the test is effectively asking whether there is no long run causality ”on average”. To address the extent to which this might occur, one can use the same individual t_{i,λ_2} values to compute the corresponding Fischer style statistic, which is constructed as

$$P_\lambda = -2 \sum_{i=1}^{N} \ln p_i \ , \tag{42}$$

where $\ln p_i$ is the natural log of the p-value associated with either t_{i,λ_1} or t_{i,λ_2} depending on which causal direction one wishes to test. Under the null hypothesis of no causality, the P_λ statistic is distributed as χ^2_{2N}, i.e. a chi-square with $2N$ degrees of freedom. Since this is a one tailed test with only positive values, there is no canceling out of positive and negative values, and the test can be interpreted as a test of how ”pervasive” non causality is in the long run from y to x or x to y depending on which element of λ_i is used.

Another advantage of this general framework is that one can use the implications of (40) to test the sign of the long run causal effect. Specifically, to give an example, imagine we have rejected the null of no long run causality running from y to x so that $\lambda_{i,2} \neq 0$ and therefore $F_i(1)_{21} \neq 0$. If we are further willing to make a sign normalization such that we call an innovation to x positive if it increases x in the long run and negative if it decreases x in the long run, so that $F_i(1)_{22} > 0$, then (40) implies that the sign of $F_i(1)_{21}$ is the opposite of the sign of the ratio of the two elements of the loading vector. Specifically, if causality runs both directions in the long run, so that neither $\lambda_{i,2}$ nor $\lambda_{i,1}$ are zero, then

$$\text{sign}[F_i(1)_{21}] = \text{sign}[-\frac{\lambda_{i,2}}{\lambda_{i,1}}]. \tag{43}$$

so that this ratio can be used to test the sign of the long run causal effect. Note of course that if $\lambda_{i,1} = 0$, there is no need to compute such a ratio, since in that case causality only runs in one direction and the sign of the OLS or FMOLS estimator trivially reflects the sign of the remaining long run causal effect. Constructing the panel version of a test based on the ratio is not as straightforward as some of the other tests discussed in this chapter. This is due to the fact that the ratio in (43) is distributed as a Cauchy, which does not have a defined mean and variance. Instead the median, which is defined for the Cauchy, is used to recenter the distribution and the panel distribution is then simulated by bootstrap from the estimated version of (36).

In contrast to the other techniques discussed in this chapter, for which the bivariate examples were simple illustrations of techniques that in general work for any number of variables, the panel long run causality tests are in fact best suited for simple bivariate investigations. In this regard, they can easily be interpreted as the total derivative causal effect rather than partial derivative causal effects. If one is interested to investigate multivariate channels, it is in principle possible to generalize to larger systems of variables. But the generalizations are not trivial as they require additional restrictions beyond the fairly simple and innocuous normalization assumptions made for the bivariate case. If these are to be justified on the basis of economic restrictions, then the approach begins to take on the flavor of the heterogeneous panel SVAR approach developed in Pedroni (2013), which can also be used to test for long run directions of causality whether cointegration is present or not. Embedding an error correction term in the panel structural VAR approach of Pedroni (2013) is conceptually straightforward, although the properties of the approach specifically when the ECM term is embedded is a topic that will benefit from further study.

8 Strategies for Treating Cross Sectional Dependence in Heterogeneous Panels

The emphasis in my presentation of the techniques so far in this chapter has been on the treatment of heterogeneity in the dynamics. However, as discussed in the introduction, it is also imperative to consider how the heterogeneity of the temporal dependencies interacts with the cross sectional, or spatial, dependencies in such panels. In this section I discuss a number of approaches, each of which can in general be applied to the techniques discussed so far in this chapter.

One the earliest and simplest ways that was used for treating cross sectional dependencies initially was to use time effects, much the way fixed effects were regularly used. For this, one simply computes the time effects as $\bar{y}_t = N^{-1}\sum_{i=1}^{N} y_{it}$, $\bar{x}_t = N^{-1}\sum_{i=1}^{N} x_{it}$, which can then be used to purge the raw data of these so that $\tilde{y}_{it} = y_{it} - \bar{y}_t$, $\tilde{x}_{it} = x_{it} - \bar{x}_t$. Keeping with our bivariate example, one can then proceed to use the purged data in place of the raw data for any of the techniques discussed in this chapter. Mechanically, this treatment is entirely symmetric with the treatment of fixed effects discussed earlier, such that these were computed as the means over time for each member and subtracted from the raw data. Indeed, keeping with our bivariate example, if we account for both time effects and fixed effects, then we can represent the prototypical cointegrating regression as

$$\tilde{y}^*_{it} = \beta_i \tilde{x}^*_{it} + e_{it} \tag{44}$$

where consistent with previous discussions the $*$ denotes that fixed effects also have been extracted, so that for example $\tilde{y}^*_{it} = \tilde{y}_{it} - T^{-1}\sum_{t=1}^{T} \tilde{y}_{it}$ where $\tilde{y}_{it}$ is as defined above, and similarly for $\tilde{x}^*_{it}$. The advantage of this approach is that it is easy to implement, and it can be applied to the raw data as a stand alone solution, which after processing in this way can then be fed into any one of the techniques discussed in this chapter, as was typically done in empirical applications. Furthermore, the asymptotic properties of estimators and tests are unaffected by this.

Economically, the solution can be justified when most of the cross sectional dependency in the data derives from sources that commonly impact all members of the panel. This is a typical assumption in microeconomic applications where the members of the panel are small, and it can be a reasonable first approximation in macroeconomic applications when for example the panel consists of a large number of small open economies which are responding to the global economy, but are not individually affecting the global economy much. Similar justifications can also be used for regions of a large country, or disaggregated industries of a large economy for example.

However, in many applications time effects may not be sufficient to accommodate all of the cross sectional dependency. This can occur most obviously when the individuals that constitute the members of the panel are large enough to affect one another rather than merely being affected by a commonality. More importantly the cross sectional dependencies can be intertwined with the temporal dependencies so that one member affects another member over time. In other words, conceptually, one can think of autocovariances that run across both time and space for the cross sectional dimension, so that there is an $N \times N$ long run covariance matrix that characterizes this. Indeed, a GLS approach for cointegration and unit root testing in panels based on such a long run covariance matrix estimation was explored in a conference paper, Pedroni (1997).

While the approach studied in Pedroni (1997) allows for a generalization of the dependency structure relative to time effects, as noted in the paper, it suffers from two important shortcomings. The first is that it requires the time series dimension to be substantially longer than what one requires for time effects. The second is that it falls apart when the cross sectional dependencies that run across the members of the panel are not temporally transitory, but are permanent. In other words, it is possible that series are cointegrated not simply across variables for a given member of the panel, but also for variables from different members of the panel, sometime referred to as cross-member or cross-unit cointegration, so that for example y_{it} might be cointegrated with y_{jt} for $i \neq j$ regardless of whether or not y_{it} is cointegrated with x_{it}. In this case the long run covariance becomes singular, and the estimators used for GLS may not be good approximations of the true dependency.

A more elegant solution is what can be thought of as a generalization that is more closely related to the time effects solution, which is to model the commonalities in terms of a dynamic factor model. Specifically one can think of time effects as a special case in which there is a single common factor that drives the dependency structure in the panel. The factor model approach generalizes this in two regards. First it allows for multiple factors, and allows the individual members of the panel to respond in a heterogeneous manner to the common factors by allowing member specific loadings for the factors. Secondly, the factors themselves can be thought of as dynamic so that there is temporal dependence in the evolution of the vector of common factors. This is the approach taken for example in Bai and Ng (2004), among others.

Specifically, Bai and Ng (2004) suggest estimating the common factors by principle components and then conducting the subsequent analysis on the defactored data. Bai and Ng originally proposed the approach in the context of panel unit root testing, and showed that treating the cross sectional dependency in this manner did not impact the asymptotic properties of the subsequent panel unit root tests. In this regard, similar to time effects, one can think of this as a stand alone treatment that can be performed prior to using the data for any of the techniques discussed in this chapter. The technique works well for a small known number of factors. When the number of factors is unknown and must itself be estimated, the technique can be sensitive to misspecification of the number of factors. The practical consequence is that when the number of factors is unknown is that inference regarding unit roots and cointegration can be sensitive to the number of chosen factors.

Another related approach advocated by Pesaran in numerous papers, including initially Pesaran (2007), is to use the cross sectional averages directly in the panel regressions, in what is known as cross sectional augmentation. This is equivalent to estimating the time effects from the data as described above, but rather than extracting them from the data, one includes them in the regressions. This has the consequence of allowing the individual members of the panel to respond in a heterogeneous manner to the time effects similar to the common factor approach, but without the need to estimate principle components. Pesaran (2007) also originally proposed the method in the context of panel unit root testing, but the approach can in principle also be used in the context of any type of panel cointegration technique. However, one important implication, in contrast to other approaches, is that using the time effects in this way does affect the asymptotic distributions of the subsequent tests. This stems from the fact that member specific coefficients on the cross sectional averages must be jointly estimated within the same equation as one is estimating for the panel analysis. Thus, in contrast to the principle components based factor model approach, the cross sectional augmentation technique should not be thought of as a stand alone treatment for the data prior to analysis, but rather as a method for adapting existing techniques. Westerlund and Urbain (2015) compare the cross sectional based approach versus the principle component based approach analytically and in Monte Carlo simulations to draw out comparisons of the relative merits of the two approaches.

While simple time effects extraction, common factor extraction and conditioning regressions on cross sectional averages have econometric appeal, an important practical concern stems from the idea that their implementation has the potential to alter the economic interpretation of the results, depending on what has been extracted by these methods. For cointegration analysis this is particularly relevant when the commonality that has been extracted or conditioned out based on any of these methods potentially follows a unit root process. To give a simple empirical example, imagine that one is testing whether long run purchasing power parity holds for a panel of real exchange rates. Imagine furthermore that the truth is that the parity condition fails due to a unit root process in the common total factor productivity frontier shared by countries, which causes differential terms of trade effects in different economies in the spirit of the Balassa-Samuelson hypothesis. But the researcher is unaware of this truth, and simply wants to control for possible cross sectional dependency by extracting a common factor by principle components or conditioning out the effect of the common factor by means of a cross sectional average. In this case we expect the raw data to reject PPP as the individual real exchange rates will follow a unit root process due to the common TFP unit root, while the data that has been treated for cross sectional dependency in either of these ways will fail to reject PPP. But it would be a mistake to conclude from this that PPP holds in the data. In the name of controlling for cross sectional dependency, we would have unwittingly eliminated the very factor that is responsible for the fact that in truth PPP fails. In a nutshell, this manner of controlling for cross sectional dependency is not innocuous in that it has the potential to substantially impact the economic interpretation of the results in unknown ways if we do not know what the commonality is that has been eliminated. To avoid this, rather than working with defactored data, it would be preferable if we could work with the raw data in a way that accounted for the dependency without potentially changing the interpretation of the results.

There are several possible avenues for alternative approaches to controlling for cross sectional dependency without the need to eliminate the source that creates the dependency, that can nevertheless work in modestly dimensioned panels. One such approach is to account for the dependencies via bootstrap methods. Estimating and replicating by bootstrap very general forms of dynamic cross sectional dependency parametrically is not feasible in moderately dimensioned panels, so that sieve bootstrap methods are likely to be a nonstarter in this regard if the hope is generality in the dependence structure. By contrast, block bootstrap methods do have the potential to accommodate fairly general processes, as for example the approach developed in Palm, Urbain and Smeekes (2011). The basic idea is to sample blocks of random temporal length, $T_n < T$ for each draw n which span the entire cross sectional dimension with width N for each draw. In this way whatever form of cross sectional dependency is present in the data will be captured and replicated within the block with each draw. Performance of the bootstrap is sensitive to some of the details such as choices by which randomization of the block length occurs, and at this point the Palm, Urbain and Smeekes approach is specifically designed for panel unit root testing rather than for cointegration applications. But this remains a promising area of current and future research.

In the next two sections I discuss some other lines of research, which although not exclusively focused on the treatment of cross sectional dependency, nevertheless offer very broad alternative solutions to accounting for general unknown forms of cross sectional and temporal dependencies in a manner that does not alter the economic interpretation of the results, as potentially occurs when commonalities are extracted.

9 A Nonparametric Rank Based Approach to Some Open Challenges

In this section I discuss a method for testing cointegration rank in panels using robust methods, and its relationship to some of the challenges in the literature. In particular, the approach addresses four of the important challenges, some of which have been touched upon in earlier discussions of this chapter. Specifically, one key challenge is the ability to address the interaction of temporal dependencies with both cross sectional heterogeneities and dependencies in a very general manner that does not require the extraction of commonalities, as discussed in the previous section. A second related challenge is to do so in a way that does create sensitivity to *ad hoc* choices. Examples of potentially *ad hoc* choices include not only choices related to numbers of common factors when treating the cross sectional dependence, but also choices with respect to choosing lag length or choosing the number of autocovariances for the bandwidth when treating the cross sectionally heterogeneous temporal dependence. A third challenge discussed previously in this chapter is the problem of "mixed" panels, whereby different members of the panel may exhibit different properties with regard to cointegration and unit roots. Finally, a challenge for many of the techniques discussed so far in this chapter is that they tend not to perform well when incidental member-specific deterministic trends are present and estimated in the regressions. For all of these challenges, in the spirit of the literature, it would be good to have techniques that perform well without the need for exceedingly large panels.

As it turns out, these challenges are inter-related and can be viewed as stemming fundamentally from the over-riding challenge presented by the classic curse of dimensionality problem. To see the connection, imagine treating a panel of time series as if it were a large vector of time series to be investigated a large dimensional unrestricted VECM, with each member of the panel contributing variables and equations to the VECM. For example, imagine a panel with N members, each one of which includes $M = 2$ variables, y_{it}, x_{it}. This could be loaded into an $MN \times 1$ dimensional vector to produce a VECM of dimension $MN \times MN$. This is appropriate conceptually, since without restrictions the VECM would allow for both full heterogeneity of the dynamics among the members as well as full unrestricted dynamic "cross sectional" dependencies among the members. The dependencies could include non-transitory, permanent dependencies across the variables analogous to cross-member cointegration, which would be reflected in a reduction in the rank of the VECM.

The question of rank is also of interest here because it relates to the issue of mixed panels discussed earlier in this chapter. It is common in the literature to think of the problem of mixed panels in terms of questions about how many members of the panel are consistent with the alternative when we reject the null. For example, if we reject the null of a unit root or the null of no cointegration, if the empirical application allows for the possibility that the answer differs across members of the panel, then how many of the members are consistent with the alternative? However, there is a conceptual problem in thinking about the question in this way when one recognizes that the members of the panel might be linked through cross-member cointegration. To give a simple example, imagine a panel consisting of a hypothetical state GDP price deflator series for the 50 states of the U.S. Imagine furthermore that each of the series follows a unit root process, but that the unit root in each of these series is due to their single common link to the U.S. dollar, which creates a unit root for the U.S. national GDP deflator. In other words, the panel has a cointegration rank of 1 rather than 50. In this case, depending on ones perspective, one could argue either that 50 of the state deflators have unit roots, or, after accounting for the cross sectional dependence structure, one could argue that in effect there is really only one unit root shared among all 50. More generally, in applications with unknown forms of cross sectional dependency and unknown degrees of cross member cointegration dependencies, the answer can lie anywhere in between. I would argue that in this case, conceptually the more salient question is not how many members have unit roots but rather what is the rank of the panel. In effect one would like to know how many unit roots are responsible for determining the properties of the panel, and whether the rank is large and close to full rank, or whether the rank is low and close to zero. The same applies if we are asking about the number of members for which for example two variables within the same member appear to cointegrate.

While the VECM approach helps us to sort through these various issues conceptually, it is not feasible to apply the VECM form directly. The reason is simply that the number of parameters that would need to be estimated is far too large. Consider the example described above, where we have $N = 30$ members with $M = 2$ variables, and say $K = 8$ lags. Estimating the VECM would in this case require the estimation of $N^2 M^2 (K+1) + NM$ parameters, which comes to 32,460 parameters to be estimated in this case. If we require at least 10 data points per parameter in order to allow enough degrees of freedom, which is likely an understatement, then this would imply that we should look for panels of length $T = 10x(32460/30)$, hence panels of length $T = 10,820$. Clearly, this makes the approach infeasible and contrary to the spirit of the panel cointegration literature which attempts to find techniques that work well in panels of moderate length.

One way to think about this more broadly is that the vast majority of the parameters that would need to be estimated for such a VECM approach are parameters that are associated with nuisance features of the data which are not necessarily central to the questions of interest. A different strategy therefore is to look for approaches that do not require that the nuisance features be controlled for by estimation of the associated parameters. This is central to the approach discussed in this section as well the very different approach discussed in the next section. Specifically, in this section I discuss the approach taken in Pedroni, Vogelsang, Wagner and Westerlund (2015) to test for the cointegration rank in panels in a way that is robust to the interaction of cross sectional dynamic heterogeneity and cross sectional dynamic dependence of unknown form. The approach is based on using untruncated kernel estimation. An added advantage to the untruncated kernel estimation is that it does not require the choice of any tuning parameters such as numbers of lags or autocovariances or common factors to be estimated, and thus eliminates the sensitivity to these. Furthermore, since the dependence structure is not explicitly modeled or estimated, the method may be implemented with much shorter panels, provided that the time series dimension, T, is greater in magnitude to the cross sectional dimension, N. Finally, freeing up degrees of freedom in this way leaves enough room for the tests to perform almost as well with the inclusion of member-specific deterministic trends as without.

To gain some intuition for how the technique works, imagine at first that we interested in asking whether a single series or potentially cointegrated linear combination of series follows a unit root or is stationary. We will take the series to be μ_t to denote the idea that any deterministics such as intercepts or trends are accounted for by regressing the individual member series against an intercept and possibly also a trend. Consider then estimating the untruncated kernel for μ_t. This is equivalent to estimating (29) for a single series for a single member, but with the bandwidth K_i set to the maximum possible for the sample, so that $K_i = T$. Ordinarily this would not be done if one is interested in estimating the long run covariance, as this will lead to inconsistent estimation of the long run variance. However, in our present context, the nature of the inconsistent estimation turns out to be useful. Specifically, in this case Kiefer and Vogelsang (2002) show that when μ_t follows a unit root process

$$T^{-2}\hat{\omega}^2 \Rightarrow 2\sigma^2 D_1 \;\; as\ T \to \infty \,, \tag{45}$$

where $\hat{\omega}^2$ is the untruncated Bartlett kernel estimate of μ_t, σ^2 is the true long run variance, and D_1 is a known nuisance parameter free distribution based on a Brownian bridge. Similarly, it is well known that if one computes the standard variance for a process that follows a unit root, then

$$T^{-1}\hat{s}^2 \Rightarrow 2\sigma^2 D_2 \;\; as\ T \to \infty \,, \tag{46}$$

where s^2 is the standard variance estimate of μ_t, σ^2 is the true long run variance, and D_2 is a different but known nuisance parameter free distribution based on a Wiener functional. The implication of (45) and (46) is that for their ratio we have

$$T^{-1}\frac{\hat{\omega}^2}{\hat{s}^2} \Rightarrow 2\frac{D_1}{D_2} \;\; as\ T \to \infty \,, \tag{47}$$

so that the ratio converges to a known nuisance parameter free distribution when μ_t follows a unit root. By contrast, if μ_t is stationary, then $s^2 \to s^2$ converges to a constant given by the true standard variance, while $T^{-1}\omega^2 \to 0$, so that the ratio in (47) collapses to zero as $T \to \infty$. In this way, the ratio in (47) can be used to consistently test whether μ_t follows a unit root process against the alternative that it is stationary without the need to consistently estimate and control for the unknown dynamics associated with σ^2.

Consider now the case of a panel imagined as a large vector of variables. This may be for a univariate case, or for the case in which the variable represents a linear combination of unit root variables which are hypothesized to be cointegrated for each member i of the panel. In this case μ_t becomes an $N \times 1$ vector of variables. If we use these to compute the untruncated Bartlett kernel, we obtain the analogous symmetric matrix estimate such that

$$T^{-2}\hat{\Omega} \Rightarrow 2\Omega^{1/2} D_{1,R} \Omega'^{1/2} \;\; as\ T \to \infty \,, \tag{48}$$

where $\hat{\Omega}$ is the untruncated kernel estimate, Ω is the true unknown long run covariance structure and $D_{1,R}$ is a known nuisance parameter free vector Brownian bridge of dimension R, which will be explained shortly. Similarly for the standard covariance matrix estimator we obtain

$$T^{-1}\hat{\Sigma} \Rightarrow \Omega^{1/2} D_{2,R} \Omega'^{1/2} \;\; as \; T \to \infty \,, \tag{49}$$

where $\hat{\Sigma}$ is the standard covariance estimate, Ω is the same true unknown long run covariance structure, and $D_{2,R}$ is a different but known nuisance parameter free vector Wiener functional, also of dimension R. Note that the long run covariance matrix Ω summarizes all possible heterogeneous temporal and cross sectional dependencies and is unknown. Unfortunately, it is no longer the case that these simply cancel out if we form the ratio $\hat{\Omega}\hat{\Sigma}^{-1}$. Fortunately, however, if we perform the trace operation over the ratio, then the Ω terms do cancel out, so that

$$T^{-1}\hat{\Omega}\hat{\Sigma}^{-1} \Rightarrow D_{1,R} D_{2,R}^{-1} \;\; as \; T \to \infty \,. \tag{50}$$

Notice what this has accomplished. Since the Ω terms which contain all the information about the heterogeneous temporal and cross sectional dynamic dependencies has dropped out, there is no need to estimate any of these and we are left with a pure nuisance parameter free known distribution which can be used for testing in a manner that is robust to the temporal and cross sectional dependencies. What is furthermore very useful is that the dimensionality R of the vector distributions, and therefore the tail values of the distributions, depend on the rank of the vector μ_t, so that one can use these to test the rank of the panel. Note furthermore that in this light, conventional panel unit root tests can be viewed as testing hypotheses which are special cases of this. In the simplest interpretation of conventional panel unit root tests such that they are used in applications where the individual members either all follow a unit root or are all stationary, conventional panel unit root tests can be interpreted as a special case of the rank test of this section whereby the null of full rank $R = N$ is tested against the alternative of zero rank $R = 0$. In more nuanced "mixed" panel applications of conventional panel unit root tests, where individual members are free to follow either a unit root process or a stationary process, conventional test can be interpreted as a special case of the rank test whereby one tests the null of full rank $R = N$ against the alternative of *any* reduced rank $R < N$. By contrast, here we have a continuum of possibilities to test for the null as well as the alternative, ranging anywhere between full rank to zero rank. Indeed Pedroni, Vogelsang, Wagner and Westerlund (2015) describe a sequential step down procedure for determining the rank.

It should be noted, however, that while the test has high power even in the presence of deterministic trends to distinguish full rank from zero rank, or in general "high" rank from "low" rank, the test does not have sufficient power to reliably distinguish the exact numerical ranks in moderately dimensioned panels. On the other hand, the precise numerical rank is not likely to be of interest in most economic applications. For example, it is hard to foresee many economic hypotheses that revolve around whether a panel of dimension $N = 30$ has a rank of say 17 or 18. Instead, I would argue that what is typically of interest is whether the rank of the panel is relatively high, or whether it is relatively low so that one knows whether there are many or only a few unit roots that drive the properties of the panel. This can also be very useful as a type of empirical cross check for more conventional panel unit root and panel cointegration tests. Imagine for example that one has confirmed through panel cointegration testing that the null of no cointegration has been rejected. In "mixed" applications, if one would like confirmation that the fraction of members consistent with this rejection is high, then one can use this type of rank test to check the rank of the residuals. If the rank is low, then the fraction of the members consistent with the rejection is high. Of course, since we estimate $N \times N$ untruncated kernels, we require $T > N$ to implement the rank test. But in cases where $T < N$, it is always possible to break the panel into smaller subsets of members for the purposes of rank testing.

It should also be noted that in unpublished versions of the study, tests for the null of stationarity were also initially explored, but later dropped in order to focus on the pure rank tests, and that the general approach of using untruncated kernels also holds promise for constructing tests for the null of stationarity or the null of cointegration that have good small sample properties. In general, the testing framework of this section is an example of one in which we obtain robustness to unknown forms of temporal and cross sectional dependencies in panels of moderate sample length due to the fact that we do not need to estimate the parameters associated with this. In the next section I continue this discussion with some recent techniques that do so in a completely different manner while attempting to address further open challenges.

10 New Directions and Challenges for Nonlinear and Time Varying Long Run Relationships

In this section I discuss some new directions and their relationship to the open challenges of treating nonlinearities and time varying relationships in heterogeneous cross sectionally dependent panels. In particular, I discuss this in relation to some of the details of an approach first introduced by application in Al Masri and Pedroni (2016) and studied econometrically in terms of its asymptotic and small sample properties in Pedroni and Smeekes (2018).

The basic idea is to exploit some the desired robustness properties discussed in this chapter and to bring them to estimation of long run nonlinear relationships as well as potentially time varying long run relationships. In particular, the idea is to be able to estimate arbitrary potentially nonlinear and possibly even time varying functions of the form

$$y_{it} = f(X_{it}, Z_i) \tag{51}$$

for some vector of unit root variables $X_i t$, possibly conditional on the value of some vector of cross sectional observations Z_i. This is a challenging goal as cointegration was developed in the time series literature as a fundamentally linear concept, and while nonlinearities have been explored in the recent time series literature, it is often hard to retain the superconsistency robustness properties that come from cointegration once nonlinearities are introduced. To gain some quick intuition for this, imagine a nonlinear relationship among unit root variables naively estimated by grouped OLS in the following form

$$y_{it} = \gamma_o + \gamma_1 x_{it} + \gamma_2 x_{it}^2 + e_{it} \; . \tag{52}$$

The problem with this format relates to the way in which unit root variables contributes to the regression properties when they appear in nonlinear form. For example, imagine that y_{it} and x_{it} follow unit roots and are cointegrated. If we then square the x_{it} variable, the stochastic properties are altered and it becomes difficult to think about y_{it} being cointegrated with both x_{it} and x_{it}^2 in a way that preserves the conventional superconsistency. Conversely if we start by thinking about y_{it} being cointegrated with x_{it}^2, it is difficult to imagine that it is also cointegrated with the square root of this variable in a way that preserves the superconsistency associated with cointegration in a conventional sense.

Therefore the approach we take is not to estimate anything like the format in (52), and indeed the approach we take is entirely unrelated to existing approaches to treating nonlinearities in nonstationary time series. Rather, we take an approach that is uniquely possible only in a heterogeneous panel context. The result is an approach which can estimate a general class of functions of unknown form in a way that is robust to any of the forms of temporal and cross sectional dependency discussed in this chapter, including dependencies in the form of cross member cointegration, which we will not need to extract or identify in order to estimate the function. Intuitively, the approach works by estimating what can be interpreted as the Taylor polynomial approximation to (51) in a way that envisions different members i of the panel as being realizations along different portions of the domain of the function (51). A cross sectional sampling of a linear approximation of the polynomial is taken across these different portions of the domain that correspond to the different units of the panel, and this is then interacted with fixed point in time observations, s, of the regressors $X_i(s)$ via a second stage regression in order to approximate the Taylor polynomial.

Specifically, if we continue with the bivariate example used throughout this chapter, then we can describe the technique as composed of two key steps. The first step is to estimate a static time series regression for each unit of the panel in the form

$$y_{it} = \alpha_i + \beta_i x_{it} + \mu_{it} \ . \tag{53}$$

The second stage is to take the heterogeneous estimated slope values, $\hat{\beta}_i$ from (53), and use them in a second stage cross sectional regression as

$$\hat{\beta}_i = \sum_{j=0}^{P} c_{j,s} x_i^j(s) + v_i \tag{54}$$

where the order of the polynomial P in (54) is chosen by data dependent methods, and $x_i(s)$ is a point in time observation of x_{it} at any fixed point in time s from the observed sample. In practice, (54) can repeated for any and all available values of s. Furthermore, if the data generating process is understood to be time invariant, then the group mean values can be used to obtain the time invariant estimates $\hat{c}_j = S^{-1} \sum_{s=1}^{S} \hat{c}_{j,s}$ for any value j. If instead the data generating process is understood to be time varying, subject to smoothness constraints, then one can use individual or rolling window averages of the $\hat{c}_{j,s}$ to trace their evolution over time.

To gain some further intuition for how the technique works, consider a simple case where the polynomial being estimated is relatively low order. For example, imagine for simplicity that the chosen value for P in (54) is $P = 1$. Now if we take the fitted values from (54) and imagine plugging them into the fitted values of (53), for the case of $P = 1$ we obtain

$$y_{it} = \alpha_i + c_0 x_{it} + c_1 x_i(s) x_{it} \ , \tag{55}$$

so that by setting $P = 1$ in (54) we obtain a quadratic relationship in x for (55). However, what is important to note is that the quadratic term in (55) is specialized in that it is not x_{it}^2, but rather $x_i(s)x_{it}$. It is this detail which allows us to use variation in the domain realizations of $x_i(s)$ over the cross sectional dimension i to trace out the polynomial. Specifically, if we picture the polynomial as having the curvature of a quadratic, then if we take a fixed point $x_i(s)$ and then vary x_{it} over t around this point we obtain a line representing the tangency of the curve at that location. If we now imagine doing this at various points along the x axis corresponding to the different i realizations for $x_i(s)$, with enough variation over i, we begin to trace out the entire polynomial. In this way, we are exploiting the heterogeneity among the i realizations to map out the details of the polynomial. The same principle applies when we take higher order values for P so that we are in effect taking a higher order expansion around the linear relationship between y_{it} and x_{it} corresponding to unit i.

While the regressions (53) and (54) are both static and linearly additive, it is important to keep in mind that the data generating process y_{it} and x_{it} is permitted to be dynamic, cross sectionally dependent and potentially nonlinear, with the idea being that these regressions are able to consistently estimate the underlying nonlinear long run relationship between y and x in a way that is robust to these features, without the need to specify and estimate the dynamics and cross sectional dependencies. Indeed, the robustness properties owe much to the fact that the nonlinear panel form has been decomposed into two simple sets of regressions, the first of which is a static time series regression for each member i and the second which is an additively linear cross sectional regression for each fixed time point s. In particular, the first stage regressions (53) simply needs to estimate a linear approximation that is appropriate for the range over which the data is realized for each member i. Since these are unit root variable, stationary transition dynamics play only a second order role in this estimation and vanishes asymptotically as the number of observations for the range associated with a given i increases.

In the second stage regressions (54) the cross sectional distribution of these estimates are then related to the corresponding cross sectional distributions of observations taken at a given point in time s. Since this latter step is done as a cross sectional regression for a given period s, dynamic cross sectional dependencies do not play a role in the consistency of the estimation viewed from the perspective of the cross sectional estimation as the number of members grows large. More broadly, the fact that the interaction of the linear approximation based on the relationship between y_{it} and x_{it} and the cross sectional point in time observations on $x_i(s)$ are used to obtain the robustness properties can be interpreted as exploiting the fact that the specific historical realizations $x_i(s)$ matter in the way they interact in the incremental relationship between x_{it} and y_{it} to create the nonlinearities that we observe.

Another interesting aspect of the approach is that since for the first stage regressons (53) we do not require the variables to be cointegrated in the conventional sense of a linear combination of variables that are stationary, the technique is also robust to the omission of unit root common factors that would in more conventional setting break the cointegrating relationship between y_{it} and x_{it}. In this regard, the technique also offers the possibility of a type of robustness for mixed panel applications, since we do not require each member to be individually cointegrated in the conventional sense. Furthermore, Monte Carlo simulations for both the Al Masri and Pedroni (2016) and Pedroni and Smeekes (2018) studies show that the technique works well even when the length of the panel is relatively short, even in the presence of omitted dynamics and common factors. Furthermore, Pedroni and Smeekes (2018) study the conditions under which the distributions are asymptotically normal, and under which standard t-statistics have good size and strong power even in relatively short samples.

As alluded to earlier, I have described here a simple bivariate example, but as shown in both studies, and as applied in Al Masri and Pedroni (2016), the technique can also be used in the general case when X_{it} is an $M \times 1$ vector, and the corresponding multivariate polynomials can also be conditioned on cross sectional variables. Since the generalization is less obvious than for some of the other techniques discussed in this chapter, it is worth elaborating briefly on how this is done. Specifically, when (53) is replaced with a multivariate regression of the form

$$y_{it} = \alpha_i + \beta_i' X_{it} + \mu_{it} \tag{56}$$

where X_{it} is an $M \times 1$ vector, the second stage regressions now take the form

$$\hat{\beta}_i = \sum_{j=0}^{P} C_{j,s} X(s)_i^j + v_i \ , \tag{57}$$

which represents a system of equations, one for each estimate of the $M \times 1$ vector $\hat{\beta}_i$ from (56), where $X(s)_i$ is an $M \times 1$ vector realization of X_{it} for some fixed time period s and the C_j are the $M \times M$ estimated matrices, which is diagonal for $j = 0$, symmetric for $j = 1$ and unrestricted for $j > 1$. In this way, the form of the approximating polynomial is interacted among the various elements of the vector version of (51. For example, in Al Masri and Pedroni (2016) arguments $X_{1,it}$ and $X_{2,it}$ reflecting measures of development of financial institutions and measures of development of financial markets respectively are allowed to interact with one another in their relationship to per capita income, so that by taking time derivatives of the estimated relationships one can infer the implications of different relative rates of development in financial institutions versus financial markets for various types of countries.

Furthemore, as mentioned previously, it is also possible to condition these polynomial relationships on any vector of cross sectional observables, Z_i. In such cases (58) can optionally be extended to take the form

$$\hat{\beta}_i = \sum_{j=0}^{P} C_{j,s} X(s)_i^j + \sum_{k=1}^{K} \sum_{j=0}^{P} D_{k,j} X_i(s)^j Z_{k,i} + v_i \tag{58}$$

Z_i is $K \times 1$ vector of unit specific variables and $D_{k,j}$ are conformably dimensioned $M \times M$ matrices. In practice Z_i can take the form of static cross sectional variables, or either point in time realizations or time averaged realizations of stationary time series variables. In this way, cross sectional and stationary variables can also be made to have a role in shaping the form of the polynomials. For example Al Masri and Pedroni (2016) show how the relationships between the different types of financial development and long run economic growth depend in part on the degree of financial openness, which is incorporated as a static conditioning variable, Z_i, that reflects the financial openness of member i. Furthermore, as discussed previously, by estimating the relationship over a rolling window for s one can see the evolution of the polynomials over time.

While this general line of research on nonlinear and time varying long run relationships is in its early stages, it should be clear that the promise is fairly high for addressing some of the open challenges that remain in the literature on panel cointegration, and for having very broad empirical applicability. In that spirit, far from being an exhaustive survey of the literature on panel cointegration methods, this chapter has instead selectively touched in a simple and hopefully intuitive manner on what I believe to be some of the key challenges that have helped to shape the literature, as well as some the key challenges that I expect are likely to be a part of what continues to motivate the literature, both in its theoretical development as well as its broad empirical applicability.

References

Al Masri, D. and P. Pedroni (2016) "Nonlinearities in Financial Development; The Good, the Bad and the Beautiful," Working Paper, Williams College.

Arellano, M. and S. Bond (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," Review of Economic Studies, 58, 277 - 97.

Bai, J. and S. Ng (2004). "A PANIC Attack on Unit Roots and Cointegration," Econometrica, 72, 1127 - 77.

Canning, D., and P. Pedroni (2008). "Infrastructure, Long Run Economic Growth and Causality Test for Cointegrated Panels," The Manchester School, 76, 504 - 27.

Canning, D., and P. Pedroni (1997). "Does Infrastructure Cause Economic Growth," Working Paper, Queen's University of Belfast.

Hlouskova J. and M. Wagner (2006). "The Performance of Panel Unit Root and Stationarity Tests: Results from a Large Scale Simulation Study," Econometric Reviews, 25, 85 - 116.

Holz-Eakin, D., W. Newey and H. Rosen (1988). "Estimating Vector Autoregressions with Panel Data," Econometrica, 56, 1371 - 95.

Kao, C. (1999). "Spurious Regression and Residual-Based Tests for Cointegration in Panel Data." Journal of Econometrics, 90, 1 - 44.

Kiefer, N. and T. Vogelsang (2002) ”Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation,” Econometrica, 70, 2093-95.

Levin, A., C. Lin and C. Chu (2002). ”Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties,” Journal of Econometrics, 108, 1 - 22.

McCoskey, S. and C. Kao (1998). ”A Residual-Based Test of the Null of Cointegration in Panel Data, Econometric Reviews, 17, 57 - 84.

Palm, F., S. Smeekes and J. Urbain (2011). ”Cross-sectional Dependence Robust Block Bootstrap Panel Unit Root Tests,” Journal of Econometrics, 163, 85 - 104.

Pedroni, P. (1993). ”Panel Cointegration, Endogenous Growth and Business Cycles in Open Economies,” Ph.D. Thesis, Columbia University, University Microfilms International.

Pedroni, P. (1995). ”Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis,” Indiana University Working Paper, No. 95-013.

Pedroni, P. (1996). ”Fully Modified OLS for Heterogeneous Cointegrated Panels and the Case of Purchasing Power Parity,” Indiana University Working Paper, No. 96-020.

Pedroni, P. (1997). ”Cross Sectional Dependence in Cointegration Tests of Purchasing Power Parity in Panels,” Indiana University Working Paper.

Pedroni, P. (1999). ”Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors,” Oxford Bulletin of Economics and Statistics, 61, 653 - 70.

Pedroni, P. (2000). ”Fully Modified OLS for Heterogeneous Cointegrated Panels,” in Advances in Econometrics: Nonstationary Panels, Panel Cointegration and Dynamic Panels, 15, 93 - 130.

Pedroni, P. (2001). ”Purchasing Power Parity Tests in Cointegrated Panels,” The Review of Economics and Statistics, 83, 727 - 31.

Pedroni, P. (2004). ”Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis,” Econometric Theory, 20, 597 - 625.

Pedroni, P. (2007). ”Social Capital, Barriers to Production and Capital Shares: Implications for the Importance of Parameter Heterogeneity from a Nonstationary Panel Approach,” Journal of Applied Econometrics, Vol. 22, pp. 429 - 51.

Pedroni, P. (2013). "Structural Panel VARs," Econometrics, 1, 180 - 206.

Pedroni, P. and S. Smeekes (2018). "A New Approach to Estimating Nonlinearities in Nonstationary Panels," working paper, Williams College.

Pedroni, P., T. Vogelsang, M. Wagner and J. Westerlund (2015). "Nonparametric Rank Tests for Non-Stationary Panels," Journal of Econometrics, Vol 185, No. 2, pp 378 - 91.

Phillips, P and H. Moon (1999). "Linear Regression Limit Theory for Nonstationary Panel Data," Econometrica, 67, 1057 - 1111.

Pesaran, M. H. (2007). "A Simple Panel Unit Root Test In The Presence Of Cross Section Dependence", Journal of Applied Econometrics, 22, 265 - 312.

Pesaran, M. H., Y. Shin and R. Smith (1999). "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels", JASA, 94, 621 - 34.

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

Shin, Y. (1994). "A Residual-Based Test of the Null of Cointegration against the Alternative of No Cointegration," Econometric Theory, 10, 91 - 115.

Westerlund, J. (2007). "Testing for Error Correction in Panel Data," Oxford Bulletin of Economics and Statistics, 69, 709 48.

Westerlund J. and J. Urbain (2015). "Cross-Sectional Averages Versus Principal Components," Journal of Econometrics, 185, 372 - 77.