

# Predictive regressions with panel data

Erik Hjalmarsson\*

Department of Economics, Yale University

December 7, 2004

## Abstract

This paper analyzes econometric inference in predictive regressions in a panel data setting. In a traditional time-series framework, estimation and testing are often made difficult by the endogeneity and near persistence of many forecasting variables; tests of whether the dividend-price ratio predicts stock returns is a prototypical example. I show that, by pooling the data, these econometric issues can be dealt with more easily. When no individual intercepts are included in the pooled regression, the pooled estimator has an asymptotically normal distribution and standard tests can be performed. However, when fixed effects are included in the specification, a second order bias in the fixed effects estimator arises from the endogeneity and persistence of the regressors. A new estimator based on recursive demeaning is proposed and its asymptotic normality is derived; the procedure requires no knowledge of the degree of persistence in the regressors and thus sidesteps the main inferential problems in the time-series case. Since forecasting regressions are typically applied to financial or macroeconomic data, the traditional panel data assumption of cross-sectional independence is likely to be violated. New methods for dealing with common factors in the data are therefore also developed. The analytical results derived in the paper are supported by Monte Carlo evidence.

*JEL Classification:* C22, C23

*Keywords:* Predictive regression, panel data, pooled regression, cross-sectional dependence, stock return predictability, fully modified estimation, local-to-unity

---

\*Comments from my advisors Peter Phillips and Robert Shiller have been very useful. I am also grateful for the advice from Don Andrews, Yuichi Kitamura, Alex Maynard, Taisuke Otsu, Kevin Song, Randi Pintoff, Vadim Marmer, participants in the summer workshop and econometrics seminar at Yale University, as well as the finance seminar participants at Göteborg University.

# 1 Introduction

Predictive regressions are important tools for evaluating and testing economic models. Although tests of stock return predictability, and the related market efficiency hypothesis, are probably the most common application, many rational expectations models can be tested in a similar manner (Mankiw and Shapiro, 1986). Traditionally, forecasting regressions have been evaluated in time-series frameworks. However, with the increased availability of data, in particular international financial and macroeconomic data, it becomes natural to extend the single time-series framework to a panel data setting.

It has gradually been discovered that the apparently simple linear regression model most often used for evaluating predictability in fact raises some very tough econometric issues. The high degree of persistence found in many predictor variables, such as the earnings- or dividend-price ratios in the prototypical stock return forecasting regression, is at the root of most econometric problems associated with predictive regressions. The near persistence of the regressors, coupled with a strong contemporaneous correlation between the innovations in the regressor and the regressand, causes standard OLS estimates to be inefficient and normal  $t$ -tests to have the wrong size. If the regressor is a unit-root process, the predictive regression becomes a cointegrating relationship and well established methods for dealing with endogenous regressors can be used. However, if the regressor is not a pure unit-root process, but rather a so called near-unit-root process, standard cointegration methods can yield misleading results (c.f. Cavanagh et al., 1995 and Elliot, 1998).<sup>1</sup>

In this paper, I analyze econometric inference in predictive regressions in a panel data setting, when the regressors are nearly persistent and endogenous. The main contributions are the development of an asymptotic theory for pooled estimators in forecasting equations and the proposal of new procedures for dealing with the bias effects arising from the persistence and endogeneity of the regressors. Some new results for controlling for the effects of common factors in panels are also derived.

By pooling the data, the econometric issues encountered in the time-series case can be dealt with more easily. Intuitively, persistent regressors cause no problems when they are exogenous. When pooling the data, independent cross-sectional information makes the regressors behave as if they were, on average, exogenous. This intuition holds when no individual intercepts, or fixed effects, are allowed in the specification. In this case, the standard pooled estimator has an asymptotically normal distribution; the summing up over the cross-section in the pooled estimator eliminates the usual near unit-root asymptotic distributions found in the time-series case. It follows immediately that test statistics have standard distributions and normal inference can be performed.

---

<sup>1</sup>The literature on time-series forecasting regressions is very large. Some early examples are Mankiw and Shapiro (1986), Nelson and Kim (1993), and Goetzman and Jorion (1993). Recent work include Cavanagh et al. (1995), Stambaugh (1999), Lewellen (2003), Campbell and Yogo (2003), Janson and Moreira (2004), and Polk et al. (2004). Many of these studies are primarily concerned with tests of stock-return predictability, although the results are generally applicable to more general forecasting regressions.

However, when fixed effects are allowed for, the asymptotic properties of the pooled estimator change. The time-series demeaning of the data, which is implicit in a fixed effects estimation, causes the fixed effects estimator to suffer from a second order bias that invalidates inference from standard test-statistics. To correct for this bias, I develop an estimator based on the idea of recursive demeaning (e.g. Moon and Phillips, 1999, and Sul et al., 2003). When demeaning each time-series in the panel, information after time  $t$  is used to form the time  $t$  regressor; this induces a correlation between the lagged value of the demeaned regressor, used in the estimation of the predictive regression, and the error term in the forecasting equation, which gives rise to the second order bias in the fixed effects estimator. By using information only up till time  $t$  in the demeaning of the regressor and only information after time  $t$  in the demeaning of the dependent variable, the distortive effects arising from standard demeaning are eliminated. The estimator based on recursively demeaned data is shown to have an asymptotically normal distribution and standard inference can again be performed.

The standard panel data assumption of cross-sectional independence is often too restrictive. This is especially true for panels of financial or macroeconomic data that are most often used in forecasting regressions. Unfortunately, there is, as of yet, no well developed econometric theories for dealing with general forms of dependence in the cross-section. However, factor models are often used to capture the co-movement in financial and macroeconomic time-series and I show that the methods developed above can be extended to a setting where common factors are present in the data. Using first-stage individual time-series regressions, the factors in the residuals of the forecasting regression can be consistently estimated. The original data can then be ‘de-factored’, and the estimators described above can be used, but with the de-factored data instead of the original data.

A separate application of the methods developed in this paper is provided in Hjalmarsson (2004), where predictability in international stock returns is analyzed.

The rest of the paper is organized as follows. Section 2 describes the model while Sections 3 and 4 derive the main asymptotic properties of the pooled estimators. Section 5 derives a fully modified pooled estimator (Phillips and Hansen, 1990) for the near-unit root case. This estimator is unfeasible in general since the modification relies on the exact degree of persistence in the time-series, which is typically unknown. However, in the special case of identical persistence in all of the time-series, the estimator becomes practically feasible since the common degree of persistence can then be estimated by panel methods (Moon and Phillips, 1999). In this case, the fully modified pooled estimator provides a more efficient alternative to the standard pooled estimator when no individual intercepts are present and to the recursively demeaned pooled estimator, described above, when fixed effects are included. In Section 6, methods for controlling for common factors are described. Some results for unbalanced panels are given in Section 7, and finite sample properties of the procedures developed in this paper are analyzed through Monte Carlo simulations in Section 8. Section 9 concludes and all technical proofs are found in the appendices.

A brief word on notation. Following the work of Phillips and Moon (1999), results for the panel estimators are first derived heuristically using sequential limits, which usually implies first keeping the cross-sectional dimension,  $n$ , fixed and letting the time-series dimension,  $T$ , go to infinity, and then letting  $n$  go to infinity. Such sequential convergence is denoted  $(T, n \rightarrow \infty)_{\text{seq}}$ . These results are then shown to hold as  $n$  and  $T$  go to infinity jointly, denoted  $(n, T \rightarrow \infty)$ . Otherwise, standard notation is used.  $BM(\Omega)$  denotes a Brownian motion with covariance matrix  $\Omega$ ,  $\Rightarrow$  signifies weak convergence, and  $\rightarrow_p$  denotes convergence in probability.

## 2 Model and assumptions

### 2.1 The data generating process

Consider a panel model with dependent variables  $y_{i,t}$ ,  $i = 1, \dots, n$ , and the corresponding vector of regressors,  $x_{i,t}$ , where  $x_{i,t}$  is an  $m \times 1$  vector and  $t = 1, \dots, T$ . The behavior of  $y_{i,t}$  and  $x_{i,t}$  are modelled as follows,

$$y_{i,t} = \alpha_i + \beta_i x_{i,t-1} + u_{i,t}, \quad (1)$$

$$x_{i,t} = A_i x_{i,t-1} + v_{i,t}, \quad (2)$$

where  $A_i = I + C_i/T$  is an  $m \times m$  matrix, with diagonal elements  $1 + c_{k,i}/T$ , and off-diagonal elements  $c_{kl,i}/T$ ,  $k, l = 1, \dots, m$ ,  $k \neq l$ .

The error processes are assumed to satisfy the following assumption.

**Assumption 1** Let  $w_{i,t} = (u_{i,t}, \epsilon_{i,t})'$  and  $\mathcal{F}_t = \{w_{i,s} | s \leq t, i = 1, \dots, n\}$  be the filtration generated by  $w_{i,t}$ ,  $i = 1, \dots, n$ . Then, for all  $i = 1, \dots, n$ ,

1.  $v_{i,t} = D_i(L) \epsilon_{i,t} = \sum_{j=0}^{\infty} D_{i,j} \epsilon_{i,t-j}$ .
2.  $D_{i,j}$  are sequences of real numbers with  $\bar{D}_j \equiv \sup_i \|D_{i,j}\| < \infty$  and  $\sum_{j=0}^{\infty} j^3 \|\bar{D}_j\| < \infty$ .
3.  $E[w_{it} | \mathcal{F}_{t-1}] = 0$ .
4.  $E[w_{i,t} w'_{i,t} | \mathcal{F}_{t-1}] = \Sigma_i = [(\sigma_{11i}, \sigma_{12i}), (\sigma_{21i}, I)]$ .
5.  $\sup_t E[u_{i,t}^4] < \infty$  and  $\sup_t E[\|\epsilon_{i,t}\|^4] < \infty$ .
6.  $E[w_{i,t} w'_{j,s}] = 0$  for all  $t, s$  and  $i \neq j$ .

Denote  $D_i = D_i(1)$ ,  $\bar{D} = \sup_i \|D_i\|$ , and  $\Omega_{22i} = D_i^2$ .

**Assumption 2**  $\Omega_{22} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Omega_{22i}$ .

Assumption 1 allows for a linear time-series structure in the errors of the predictor variables and imposes the usual martingale difference (mds) assumption for the errors in the dependent variables.

The mds assumption, which is standard in predictive regressions, is often based on some orthogonality condition from an underlying rational expectations model. For instance, in financial forecasting regressions the mds assumption is motivated by the efficient markets hypothesis. The time-series aspects of the model formulated by (1) and (2) and Assumption 1 is similar to those of the models studied by Mankiw and Shapiro (1986), Nelson and Kim (1993), Cavanagh et al. (1995), Stambaugh (1999), Campbell and Yogo (2003), Lewellen (2003), Jansson and Moreira (2004), and Polk et al. (2004). The final condition of Assumption 1 states the standard panel data assumption that the innovations are uncorrelated in the cross-section. Some effects of cross-sectional dependence arising from common factors are discussed later in the paper.

Assumption 3 introduces heterogeneity in the persistence of the regressor variables, but rules out processes that are too explosive; this is not much of a limitation in practice.

**Assumption 3 (*Heterogeneous persistence*)**

1. *The local-to-unity parameters  $C_i$  are iid across  $i$ .*
2.  *$\max C_i < \bar{C}$  for all  $i$ .*

Two different assumptions on the slope coefficients  $\beta_i$  will be considered. The first assumption allows the individual  $\beta_i$ s to differ from each other in a general fashion, whereas the second assumption imposes an identical coefficient  $\beta$  across the whole panel. Clearly, the first assumption is more general, but many useful results also arise from the study of the homogeneous model.

**Assumption 4 (*Heterogeneous slope coefficients*)**

1.  *$\beta_i = \beta + \theta_i$ .*
2.  *$\{\theta_i\}_{i=1}^n$  is iid with mean zero and variance  $\Omega_{\theta\theta}$ .*
3.  *$\{\theta_i\}_{i=1}^n$  is independent of  $w_{i,t}$  and  $C_i$  for all  $i$  and  $t$ .*

**Assumption 5 (*Homogeneous slope coefficients*)**

$$\beta_i = \beta \text{ for all } i.$$

The focus of this paper is inference on the parameter  $\beta$ , which in the heterogeneous case represents the average relationship in the panel. In the homogeneous case, the  $\beta_i$ s are all identical and the estimate of  $\beta$  thus gives an estimate of the slope coefficient in each individual forecasting equation. This is typically not the case, though, and it is less obvious what the advantages are of using panel methods to estimate the average parameter  $\beta$ , instead of using time-series estimators to estimate each individual  $\beta_i$ . There are, however, good reasons why panel methods might be useful, even in the heterogeneous case.

Since the panel based inference analyzed in this paper does not require any knowledge of the parameters  $C_i$ , either for estimation or testing, the nuisance parameter problem arising from  $C_i$  in

time-series inference is no longer an issue. It therefore becomes easy to test hypotheses for  $\beta$ . Of course, rejecting the null hypothesis of  $\beta = 0$  does not reveal whether the variable  $x_{i,t-1}$  predicts  $y_{i,t}$  for a specific  $i$ , but it does say that on average there is a predictive relationship in the panel. The interpretation of  $\beta$  as an average relationship in the panel has the advantage of resolving evidence from individual time-series regressions. It is often the case that, say, stock returns are found predictable in some countries, but not in others. The interpretations of such results is not straightforward but the results from a pooled regression provide an answer; if the average slope coefficient  $\beta$  is significant, then *on average* there is a significant relationship in the panel.

Predictive regressions are usually employed to test a null of no predictive relationship, and any deviations from this null hypothesis are often small. Thus, even if the slope coefficients in the panel are non-identical, they are still likely to be similar to each other in absolute terms. Under such circumstances, the pooled estimate of the average parameter might provide a point estimate of the individual intercepts that is at least as useful as the point estimate provided by time-series methods. Evidence of this is provided by the out-of-sample exercises in Hjalmarsson (2004), where the forecasts based on the pooled estimates often outperform those based on the time-series estimates.

## 2.2 Preliminary results

Before analyzing the panel estimator for  $\beta$ , it useful to establish some preliminary results that will be used repeatedly. Let  $E_{i,t} = (u_{i,t}, v_{i,t})'$  be the joint innovations process. Under Assumption 1, by standard arguments (Phillips and Solo, 1992),

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} E_{i,t} \Rightarrow B_i(r) = BM(\Omega_i)(r), \quad (3)$$

where

$$\Omega_i = \begin{pmatrix} \sigma_{11i} & \omega_{12i} \\ \omega_{21i} & \Omega_{22i} \end{pmatrix}, \quad (4)$$

and  $\omega_{21i} = D_i(1)\sigma_{12i}$ ,  $\omega_{12i} = \omega'_{21i}$ , and  $B_i(\cdot) = (B_{1i}(\cdot), B_{2i}(\cdot))'$  denote a  $1 + m$ -dimensional Brownian motion. By the continuous mapping theorem, it follows that as  $T \rightarrow \infty$ , for all  $i$ ,

$$\frac{1}{T^2} \sum_{t=1}^T x_{i,t} x'_{i,t} \Rightarrow \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr, \quad (5)$$

where

$$J_{i,C_i}(r) = \int_0^r e^{(r-s)C_i} dB_{2,i}(s) = D_i \int_0^r e^{(r-s)C_i} dW_{2,i}(s). \quad (6)$$

**Lemma 1** Under Assumptions 1 and 3,

$$E \left[ \left\| \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right\|^2 \right] < \infty, \quad (7)$$

for all  $i$ .

From Lemma 1 and the cross-sectional independence, it follows that as  $n \rightarrow \infty$ ,

$$\frac{1}{n} \sum_{i=1}^n \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \rightarrow_p E \left[ \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right]. \quad (8)$$

### 3 The pooled estimator

#### 3.1 Heterogeneous slope coefficients

To estimate the average parameter  $\beta$  under Assumption 4, I consider the traditional pooled estimator. Its asymptotic properties are derived heuristically here, using sequential limit arguments. Formal proofs of the results for joint limits are found in the Appendix. To keep the arguments more transparent, I consider first the case where there are no individual effects in equation (1). That is, assume for now that  $\alpha_i = 0$  for all  $i$ .<sup>2</sup> The case with individual intercepts will be dealt with later.

The pooled estimator of  $\beta$  is now given by

$$\hat{\beta}_{n,T} = \left( \sum_{i=1}^n \sum_{t=1}^T y_{i,t} x'_{i,t-1} \right) \left( \sum_{i=1}^n \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \right)^{-1} \quad (9)$$

$$\begin{aligned} &= \beta + \left( \frac{1}{n} \sum_{i=1}^n \theta_i \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \right)^{-1} \\ &\quad + \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T u_{it} x'_{i,t-1} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \right)^{-1}. \end{aligned} \quad (10)$$

---

<sup>2</sup>The results developed below also hold in the case with a *common* non-zero intercept  $\alpha$ .

Using sequential limits, for fixed  $n$  as  $T \rightarrow \infty$ ,

$$\begin{aligned}
& \sqrt{n} \left( \hat{\beta}_{n,T} - \beta \right) \\
&= \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \theta_i \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \right)^{-1} \\
&\quad + \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T u_{it} x'_{i,t-1} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \right)^{-1} \\
&\Rightarrow \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \theta_i \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right) \left( \frac{1}{n} \sum_{i=1}^n \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right)^{-1}. \tag{11}
\end{aligned}$$

By Lemma 1 and Assumption 4

$$E \left[ \left\| \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right\|^2 \right] < \infty \text{ and } E \left[ \theta_i \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right] = 0. \tag{12}$$

Thus, as  $n \rightarrow \infty$ , by the weak law of large numbers (WLLN)

$$\frac{1}{n} \sum_{i=1}^n \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \xrightarrow{a.s.} E \left[ \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right], \tag{13}$$

and by the Lindeberg-Levy central limit theorem (CLT)

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^n \theta_i \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \\
&\Rightarrow N \left( 0, E \left[ \left( \theta_i \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right)' \left( \theta_i \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right) \right] \right). \tag{14}
\end{aligned}$$

To simplify the notation, let

$$\Omega_{xx} = E \left[ \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right] \tag{15}$$

and

$$\Phi_{xx}^\theta = E \left[ \left( \theta_i \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right)' \left( \theta_i \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right) \right]. \tag{16}$$

In summary, as  $(T, n \rightarrow \infty)_{\text{seq}}$ ,

$$\sqrt{n} \left( \hat{\beta}_{n,T} - \beta \right) \Rightarrow N \left( 0, \Omega_{xx}^{-1} \Phi_{xx}^\theta \Omega_{xx}^{-1} \right). \tag{17}$$



In general, calculating explicit expressions for  $\Omega_{xx}$  and  $\Phi_{xx}^\theta$  would require a distributional assumption for the  $C_i$ s. Under the assumption that  $C_i = C$  for all  $i$ , however, results can be obtained. For simplicity, consider the case when the regressors are scalar. It is shown in the Appendix that under these circumstances,

$$\Omega_{xx} = \frac{1}{4}\Omega_{22}C^{-2}(e^{2C} - 2C - 1), \quad (18)$$

and

$$\Phi_{xx}^\theta = \frac{1}{16}\Omega_{\theta\theta}\Omega_{22}^2C^{-4}(4C^2 - 4C + 6e^{2C} + 3e^{4C} - 20Ce^{2C} - 9). \quad (19)$$

Observe that the limiting distribution is driven by the heterogeneity in the  $\beta_i$ s and is not directly affected by the error terms  $u_{i,t}$ . To understand this result, write equation (1) as

$$y_{i,t} = \alpha_i + \beta x_{i,t-1} + \theta_i x_{i,t-1} + u_{i,t} = \alpha_i + \beta x_{i,t-1} + U_{i,t} \quad (20)$$

where  $U_{i,t} = \theta_i x_{i,t-1} + u_{i,t}$ . The process  $U_{i,t}$  is a near integrated process, and the nearly non-stationary part of it,  $\theta_i x_{i,t-1}$ , will thus determine its asymptotic properties. The stationary part,  $u_{i,t}$ , will therefore have no direct effect on the asymptotic properties of the pooled estimator  $\hat{\beta}_{n,T}$ .

The fact that  $U_{i,t}$  is nearly integrated also explains the  $\sqrt{n}$ -convergence rate of the pooled estimator in the completely heterogeneous case. As is well known, typical time-series estimators of  $\beta_i$ , such as the standard OLS estimator, have a convergence rate  $T$ . In a homogeneous panel, where  $\beta_i = \beta$ , for all  $i$ , a  $\sqrt{nT}$ -convergence rate can be obtained for the pooled estimator, as will be shown below. However, in the case when the  $\beta_i$ s are heterogeneous, the residuals in the fitted regression (20) will be nearly non-stationary. This slows down the rate of convergence since the expression for  $\hat{\beta}_{n,T} - \beta$  now involves the sum  $\sum_{t=1}^T x_{i,t-1}x'_{i,t-1}$  in addition to the standard sum  $\sum_{t=1}^T u_{it}x'_{i,t-1}$ . The former sum requires a standardization of  $1/T^2$  while the latter only requires a  $1/T$  standardization. The division by  $T^2$ , as opposed to  $T$ , explains the slow down from  $\sqrt{nT}$  in the homogeneous case to  $\sqrt{n}$  in the heterogeneous case.

The results derived heuristically above, using sequential limit arguments, continue to hold under joint limits.

**Theorem 1** *Under Assumptions 1-4,*

(i)  $\hat{\beta}_{n,T} \rightarrow_p \beta$  as  $(n, T \rightarrow \infty)$ , where  $\beta$  is the average slope coefficient in the panel.

(ii)

$$\sqrt{n}(\hat{\beta}_{n,T} - \beta) \Rightarrow N(0, \Omega_{xx}^{-1}\Phi_{xx}^\theta\Omega_{xx}^{-1}).$$

as  $(n, T \rightarrow \infty)$  with  $n/T \rightarrow 0$ .

The limiting distribution of  $\hat{\beta}_{n,T}$  depends on  $\Omega_{xx}$  and  $\Phi_{xx}^\theta$ . To perform inference, estimates of these

parameters are required. Let  $\hat{u}_{i,t} = y_{i,t} - \hat{\beta}_{n,T}x_{i,t-1}$ , and define

$$\hat{\Phi}_{xx}^\theta = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^4} \sum_{t=1}^T \sum_{s=1}^T (\hat{u}_{i,t}x'_{i,t-1})' (\hat{u}_{i,s}x'_{i,s-1}), \quad (21)$$

and

$$\hat{\Omega}_{xx} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1}x'_{i,t-1}. \quad (22)$$

The estimators  $\hat{\Phi}_{xx}^\theta$  and  $\hat{\Omega}_{xx}$  are thus the panel equivalents of standard HAC estimators for long-run variances.

Standard tests can now be performed. For instance, the null hypothesis  $\beta_k = \beta_{k,0}$ , for some  $k$ , can be tested using a  $t$ -test. Using the results derived above, it follows easily that under the null-hypothesis,

$$t_k = \frac{\sqrt{n} \left( \hat{\beta}_{k,n,T} - \beta_{k,0} \right)}{\sqrt{a' \hat{\Omega}_{xx}^{-1} \hat{\Phi}_{xx}^\theta \hat{\Omega}_{xx}^{-1} a}} \Rightarrow N(0, 1), \quad (23)$$

in sequential limits, as  $(T, n \rightarrow \infty)_{\text{seq}}$ , where  $a$  is an  $m \times 1$  vector with the  $k$ 'th component equal to one and zero elsewhere. More general linear hypotheses can be evaluated using a Wald test. Suppose the null hypothesis is  $H_0 : Q\beta = b$ , where  $\text{rank}(Q) = q$ . Then, under the null

$$W_{Q,b} = n \left( Q\hat{\beta}_{n,T} - b \right)' \left( Q\hat{\Omega}_{xx}^{-1} \hat{\Phi}_{xx}^\theta \hat{\Omega}_{xx}^{-1} Q' \right)^{-1} \left( Q\hat{\beta}_{n,T} - b \right) \Rightarrow \chi_q^2, \quad (24)$$

as  $(T, n \rightarrow \infty)_{\text{seq}}$ . The  $t$ -test and Wald test are complete analogues of their time-series correspondents when robust HAC estimators for the covariance matrix are used. These results can also be verified for joint limits.

**Theorem 2** *Suppose Assumptions 1-4 hold. Then, as  $(n, T \rightarrow \infty)$  with  $n/T \rightarrow 0$ ,*

- (i)  $t_k \Rightarrow N(0, 1)$  under the null hypothesis  $\beta_k = \beta_{k,0}$ , and
- (ii)  $W_{Q,b} \Rightarrow \chi_q^2$  under the null hypothesis  $Q\beta = b$ , where  $\text{rank}(Q) = q$ .

### 3.2 homogeneous slope coefficients

Consider again the case with  $\alpha_i = 0$ , for all  $i$ . Then, under Assumption 5, the pooled estimator  $\hat{\beta}_{n,T}$  satisfies

$$\sqrt{nT} \left( \hat{\beta}_{n,T} - \beta \right) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T u_{i,t}x'_{i,t-1} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1}x'_{i,t-1} \right)^{-1}. \quad (25)$$

Using sequential limits and arguments similar to those in the heterogeneous case, as  $(T, n \rightarrow \infty)_{\text{seq}}$ ,

$$\sqrt{n}T \left( \hat{\beta}_{n,T} - \beta \right) \Rightarrow N \left( 0, \Omega_{xx}^{-1} \Phi_{ux} \Omega_{xx}^{-1} \right). \quad (26)$$

where

$$\Phi_{ux} = E \left[ \left( \int_0^1 dB_{i,1} J_{C_i}(r)' \right)' \left( \int_0^1 dB_{i,1} J_{C_i}(r)' \right) \right]. \quad (27)$$

The limiting covariance matrix of  $\hat{\beta}_{n,T}$  is similar to that found in the heterogeneous case, although the rate of convergence is different. In the case of scalar regressors, with  $C_i = C$  for all  $i$ , it is shown in the Appendix that,

$$\Phi_{ux} = \frac{1}{4} \sigma_{11} \Omega_{22} C^{-2} (e^{2C} - 2C - 1). \quad (28)$$

Combined with the result for  $\Omega_{xx}$  derived above, it follows that in this case the limiting variance satisfies

$$\Omega_{xx}^{-1} \Phi_{ux} \Omega_{xx}^{-1} = 4 \sigma_{11} \Omega_{22}^{-1} C^2 (e^{2C} - 2C - 1)^{-1}. \quad (29)$$

Analogous to the heterogeneous case, let

$$\hat{\Phi}_{ux} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T (\hat{u}_{i,t} x'_{i,t-1})' (\hat{u}_{i,s} x'_{i,s-1}). \quad (30)$$

A  $t$ -test for  $H_0 : \beta_k = \beta_{k,0}$ , can be formed, and as  $(T, n \rightarrow \infty)_{\text{seq}}$ ,

$$t'_k = \frac{\sqrt{n}T \left( \hat{\beta}_{k,n,T} - \beta_{k,0} \right)}{\sqrt{a' \hat{\Omega}_{xx}^{-1} \hat{\Phi}_{ux} \hat{\Omega}_{xx}^{-1} a}} \Rightarrow N(0, 1). \quad (31)$$

Since  $\hat{\Phi}_{ux} = T^2 \hat{\Phi}_{xx}^\theta$ , it follows that

$$t'_k = \frac{\sqrt{n}T \left( \hat{\beta}_{k,n,T} - \beta_{k,0} \right)}{\sqrt{a' \hat{\Omega}_{xx}^{-1} \hat{\Phi}_{ux} \hat{\Omega}_{xx}^{-1} a}} = \frac{\sqrt{n} \left( \hat{\beta}_{k,n,T} - \beta_{k,0} \right)}{\sqrt{a' \hat{\Omega}_{xx}^{-1} \hat{\Phi}_{xx}^\theta \hat{\Omega}_{xx}^{-1} a}} = t_k \quad (32)$$

and the  $t$ -tests in the homogeneous and heterogeneous cases are identical. The same also holds for

the more general Wald test,

$$\begin{aligned}
W'_{Q,b} &= nT^2 \left( Q\hat{\beta}_{n,T} - b \right)' \left( Q\hat{\Omega}_{xx}^{-1}\hat{\Phi}_{ux}\hat{\Omega}_{xx}^{-1}Q' \right)^{-1} \left( Q\hat{\beta}_{n,T} - b \right) \\
&= n \left( Q\hat{\beta}_{n,T} - b \right)' \left( Q\hat{\Omega}_{xx}^{-1}\hat{\Phi}_{xx}^{\theta}\hat{\Omega}_{xx}^{-1}Q' \right)^{-1} \left( Q\hat{\beta}_{n,T} - b \right) \\
&= W_{Q,b}.
\end{aligned} \tag{33}$$

Thus, although the rates of convergence are different in the heterogeneous and homogeneous case, there is no practical difference when performing inference. This is a convenient result, since one need not make any assumptions on the degree of homogeneity in the panel before performing inference. Of course, the interpretation of the results are somewhat different depending on whether the panel is homogeneous or heterogeneous.

The following two theorems show that these results also hold in joint limits.

**Theorem 3** *Under Assumptions 1-3, and 5,*

(i)  $\hat{\beta}_{n,T} \rightarrow_p \beta$  as  $(n, T \rightarrow \infty)$ .

(ii)

$$\sqrt{nT} \left( \hat{\beta}_{n,T} - \beta \right) \Rightarrow N \left( 0, \Omega_{xx}^{-1} \Phi_{ux} \Omega_{xx}^{-1} \right).$$

as  $(n, T \rightarrow \infty)$  with  $n/T \rightarrow 0$ .

**Theorem 4** *Suppose Assumptions 1-3, and 5 hold. Then, as  $(n, T \rightarrow \infty)$  with  $n/T \rightarrow 0$ ,*

(i)  $t'_k = t_k \Rightarrow N(0, 1)$  under the null hypothesis  $\beta_k = \beta_{k,0}$ , and

(ii)  $W'_{Q,b} = W_{Q,b} \Rightarrow \chi^2_q$  under the null hypothesis  $Q\beta = b$ , where  $\text{rank}(Q) = q$ .

## 4 Individual effects

In the above analysis, the individual intercepts in equation (1) were all assumed to be equal to zero. This section considers the effects on the pooled estimator when the  $\alpha_i$ s are no longer zero and are allowed to vary across the panel.

Let  $\underline{y}_{i,t}$  and  $\underline{x}_{i,t}$  denote the time-series demeaned data. That is,  $\underline{x}_{i,t} = x_{i,t} - \frac{1}{T} \sum_{t=1}^T x_{i,t-1}$  and  $\underline{y}_{i,t} = y_{i,t} - \frac{1}{T} \sum_{t=1}^T y_{i,t}$ . The pooled estimator, allowing for individual intercepts, is then given by

$$\tilde{\beta}_{n,T} = \left( \sum_{i=1}^n \sum_{t=1}^T \underline{y}_{i,t} \underline{x}'_{i,t-1} \right) \left( \sum_{i=1}^n \sum_{t=1}^T \underline{x}_{i,t-1} \underline{x}'_{i,t-1} \right)^{-1}. \tag{34}$$

The next two sub-sections analyze the asymptotic properties of  $\tilde{\beta}_{n,T}$  in the homogeneous and heterogeneous cases.

## 4.1 The heterogeneous case

Under Assumption 4, as  $(T, n \rightarrow \infty)_{\text{seq}}$ , using the same arguments as previously,

$$\sqrt{n} \left( \tilde{\beta}_{n,T} - \beta \right) \Rightarrow N \left( 0, \underline{\Omega}_{xx}^{-1} \underline{\Phi}_{xx}^{\theta} \underline{\Omega}_{xx}^{-1} \right), \quad (35)$$

where

$$\underline{\Omega}_{xx} = E \left[ \int_0^1 \underline{J}_{i,C_i}(r) \underline{J}_{i,C_i}(r)' dr \right], \quad (36)$$

$$\underline{\Phi}_{xx}^{\theta} = E \left[ \left( \theta_i \int_0^1 \underline{J}_{i,C_i}(r) \underline{J}_{i,C_i}(r)' dr \right)' \left( \theta_i \int_0^1 \underline{J}_{i,C_i}(r) \underline{J}_{i,C_i}(r)' dr \right) \right], \quad (37)$$

and

$$\underline{J}_{i,C_i}(r) = J_{i,C_i}(r) - \int_0^1 J_{i,C_i}(r) dr. \quad (38)$$

Estimates of  $\underline{\Omega}_{xx}$  and  $\underline{\Phi}_{xx}^{\theta}$  can be formed in the same manner as above, using the demeaned data. The test statistics using these estimates are still valid with the same asymptotic distributions under the null hypothesis. Apart from using demeaned data, there are no practical changes to the inference. As in the case with no intercepts, these results can also be shown to hold in joint limits, although the results are not repeated here.

## 4.2 The homogeneous case

As was just shown in the heterogeneous case, using demeaned data causes no changes for practical inference with the pooled estimator. However, some caution is required since the same result does not hold for the homogeneous case. Under Assumption 5, the pooled estimator with individual effects satisfies

$$\tilde{\beta}_{n,T} - \beta = \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \underline{u}_{i,t} \underline{x}'_{i,t-1} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \underline{x}_{i,t-1} \underline{x}'_{i,t-1} \right)^{-1}. \quad (39)$$

Clearly, the estimator is still consistent. Its asymptotic distribution, however, will be affected by the demeaning. For fixed  $n$ , as  $T \rightarrow \infty$ ,

$$T \left( \tilde{\beta}_{n,T} - \beta \right) \Rightarrow \left( \frac{1}{n} \sum_{i=1}^n \int_0^1 dB_{1,i} \underline{J}_{C_i}(r)' \right) \left( \frac{1}{n} \sum_{i=1}^n \int_0^1 \underline{J}_{C_i}(r) \underline{J}_{C_i}(r)' dr \right)^{-1}. \quad (40)$$

Let  $\omega_{12} = \lim_{n \rightarrow \infty} n^{-1} \sum \omega_{12i}$ , and observe that

$$\begin{aligned}
E \left[ \int_0^1 dB_{1,i} \underline{J}_{C_i}(r)' \right] &= E \left[ \int_0^1 dB_{1,i} J_{i,C_i}(r)' - \int_0^1 dB_{1,i} \int_0^1 J_{i,C_i}(r)' dr \right] \\
&= -E \left[ \int_0^1 \int_0^1 dB_{1,i}(s) J_{i,C_i}(r)' dr \right] \\
&= -E \left[ \int_0^1 \int_0^1 \int_0^r E [dB_{1,i}(s) dB_{2,i}(q)'] e^{(r-q)C_i} dr \middle| C_i \right] \\
&= -\omega_{12} \int_0^1 \int_0^r E [e^{(r-s)C_i}] ds dr \\
&\neq 0,
\end{aligned} \tag{41}$$

whenever  $\omega_{12} \neq 0$ . Thus, as  $(T, n \rightarrow \infty)_{\text{seq}}$ ,

$$T \left( \tilde{\beta}_{n,T} - \beta \right) \rightarrow_p - \left( \omega_{12} \int_0^1 \int_0^r E [e^{(r-s)C_i}] ds dr \right) \underline{\Omega}_{xx}^{-1}, \tag{42}$$

and the estimator suffers from a second order bias from the demeaning process. The second order bias arises because the demeaning process induces a correlation between the innovation processes  $u_{i,t}$  and the demeaned regressors  $\underline{x}_{i,t-1}$ .<sup>3</sup> Intuitively,  $u_{i,t}$  and  $\underline{x}_{i,t-1}$  are correlated because, in the demeaning of  $x_{i,t-1}$ , information available after time  $t-1$  is used. One solution is to use recursive demeaning of  $x_{i,t}$  and  $y_{i,t}$  (e.g. Moon and Phillips, 1999, and Sul et al., 2003). That is, define

$$\underline{x}_{i,t}^d = x_{i,t} - \frac{1}{t} \sum_{s=1}^t x_{i,s}, \tag{43}$$

and

$$\underline{y}_{i,t}^{dd} = y_{i,t} - \frac{1}{T-t} \sum_{s=t}^T y_{i,s}, \quad \underline{x}_{i,t}^{dd} = x_{i,t} - \frac{1}{T-t} \sum_{s=t}^T x_{i,s}. \tag{44}$$

The process  $\underline{x}_{i,t-1}^d$  now only relies on information up till time  $t-1$ , and  $\underline{y}_{i,t}^{dd}$  only depends on information from  $t$  to  $T$ ; the recursive demeaning will not induce a correlation between  $u_{i,t}$  and  $\underline{x}_{i,t-1}^d$ . The process  $\underline{x}_{i,t}^{dd}$  is used to properly balance the estimator, as shown below. By the CMT, as  $T \rightarrow \infty$  for a fixed  $i$ ,

$$\frac{\underline{x}_{i,t}^d}{\sqrt{T}} = \frac{x_{i,t}}{\sqrt{T}} - \left( \frac{t}{T} \right)^{-1} \frac{1}{T} \sum_{s=1}^t \frac{x_{i,s}}{\sqrt{T}} \Rightarrow J_{i,C_i}(r) - r^{-1} \int_0^r J_{i,C_i}(u) du = \underline{J}_{C_i}^d(r), \tag{45}$$

---

<sup>3</sup>The phenomenon is analogous to that found by Moon and Phillips (1999), when analyzing estimation of local-to-unity roots in panels with incidental trends.

and

$$\frac{\underline{x}_{i,t}^{dd}}{\sqrt{T}} = \frac{x_{i,t}}{\sqrt{T}} - \left(\frac{T-t}{T}\right)^{-1} \frac{1}{T} \sum_{s=t}^T \frac{x_{i,s}}{\sqrt{T}} \Rightarrow J_{i,C_i}(r) - (1-r)^{-1} \int_r^1 J_{i,C_i}(u) du = \underline{J}_{C_i}^{dd}(r). \quad (46)$$

Consider the following pooled estimator, using the recursively demeaned data,

$$\tilde{\beta}_{n,T}^{rd} = \left( \sum_{i=1}^n \sum_{t=1}^T \underline{y}_{i,t}^{dd} \underline{x}_{i,t-1}^{dt} \right) \left( \sum_{i=1}^n \sum_{t=1}^T \underline{x}_{i,t-1}^{dd} \underline{x}_{i,t-1}^{dt} \right)^{-1}. \quad (47)$$

**Theorem 5** *Under Assumptions 1-3, and 5,*

- (i)  $\tilde{\beta}_{n,T}^{rd} \rightarrow_p \beta$  as  $(n, T \rightarrow \infty)$ .
- (ii) As  $(n, T \rightarrow \infty)$  with  $n/T \rightarrow 0$ ,

$$\sqrt{n}T \left( \tilde{\beta}_{n,T}^{rd} - \beta \right) \Rightarrow N \left( 0, \left( \underline{\Omega}_{xx}^{rd} \right)^{-1} \underline{\Phi}_{ux}^{rd} \left( \underline{\Omega}_{xx}^{rd} \right)^{-1} \right), \quad (48)$$

where

$$\underline{\Phi}_{ux}^{rd} = E \left[ \left( \int_0^1 dB_{1,i}^{dd}(r) \underline{J}_{C_i}^d(r)' \right)' \left( \int_0^1 dB_{1,i}^{dd}(r) \underline{J}_{C_i}^d(r)' \right) \right], \quad (49)$$

$$dB_{1,i}^{dd}(r) = dB_{1,i}(r) - (1-r)^{-1} (B_{1,i}(1) - B_{1,i}(r)), \quad (50)$$

and

$$\underline{\Omega}_{xx}^{rd} = E \left[ \int_0^1 \underline{J}_{C_i}^{rr}(r) \underline{J}_{C_i}^r(r)' dr \right]. \quad (51)$$

The usual test-statistics can now be formed. Specifically, let

$$\tilde{u}_{i,t}^{dd} = \underline{y}_{i,t}^{dd} - \tilde{\beta}_{n,T}^{rd} \underline{x}_{i,t-1}^{dd}, \quad (52)$$

and form

$$\tilde{\underline{\Phi}}_{ux}^{rd} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T (\tilde{u}_{i,t}^{dd} \underline{x}_{i,t-1}^{dt})' (\tilde{u}_{i,t}^{dd} \underline{x}_{i,s-1}^{dt}), \quad (53)$$

and

$$\tilde{\underline{\Omega}}_{xx}^{rd} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \underline{x}_{i,t-1}^{dd} \underline{x}_{i,t-1}^{dt}. \quad (54)$$

The  $t$ -test and Wald-test based on  $\tilde{\underline{\Phi}}_{ux}^{rd}$  and  $\tilde{\underline{\Omega}}_{xx}^{rd}$  will satisfy the usual properties. The results follow in the same manner as above and, as previously, these tests will also work for the heterogeneous case. As evidenced by the Monte Carlo simulation in Section 8, the second order bias effect arising from

demeaning the data is also present in panels that exhibit moderate degrees of heterogeneity in the slope coefficients. Thus, the estimator  $\tilde{\beta}_{n,T}^{rd}$  seems, in general, to be a safer choice when fitting fixed effects than the plain fixed effects estimator,  $\tilde{\beta}_{n,T}$ , since the former works well in both the homogeneous and heterogeneous cases.

## 5 A fully modified pooled estimator

One of the main advantages of the pooled estimators described above is their asymptotic normal distributions, which enables correct inference without the knowledge of the nuisance parameters  $\{C_i\}_{i=1}^n$ . However, when there is some knowledge available about  $\{C_i\}_{i=1}^n$ , efficiency gains can be made.

The fully modified OLS (FM-OLS) estimator for cointegrated time-series regressions was developed by Phillips and Hansen (1990) to correct for the presence of endogeneity and serial correlation in the OLS estimator. Phillips and Moon (1999) extend this idea to a panel data setting and derive a pooled fully modified (PFM) estimator for the pure unit-root case, when  $C_i = 0$  for all  $i$ . In this section, I generalize the PFM estimator to the case of nearly integrated regressors.

As pointed out before, the individual  $C_i$ s are not estimable in general, so an estimator based on the knowledge of  $C_i$  is typically unfeasible in practice. However, as shown by Moon and Phillips (1999), when the  $C_i$ s are identical across the panel, then  $C_i = C$  can be estimated. This estimator is thus feasible when the  $C_i$ s are all identical. The modifications to the pooled estimator will have no asymptotic effects under the heterogeneous slope coefficients of Assumption 4; only the homogeneous case described by Assumption 5 is therefore considered. As in the case of individual effects, however, for panels that exhibit only a limited degree of heterogeneity, the modifications are still likely to have some effect.

Define the quasi-differencing operator

$$\Delta_{C_i} x_{i,t} = x_{i,t} - x_{i,t-1} - \frac{C_i}{T} x_{i,t-1} = v_{i,t}, \quad (55)$$

and let

$$y_{i,t}^+ = y_{i,t} - \hat{\omega}_{12i} \hat{\Omega}_{22i}^{-1} \Delta_{C_i} x_t, \quad \hat{\Lambda}_{12i}^+ = -\hat{\omega}_{12i} \hat{\Omega}_{22i}^{-1} \hat{\Lambda}_{22i}, \quad (56)$$

where  $\Lambda_{22i} = \sum_{k=1}^{\infty} E(v_{i,k} v'_{i,0})$ , and  $\hat{\omega}_{12i}$ ,  $\hat{\Omega}_{22i}$ , and  $\hat{\Lambda}_{22i}$  are estimates of  $\omega_{12i}$ ,  $\Omega_{22i}$ , and  $\Lambda_{22i}$ , respectively. The pooled FM (PFM) estimator is given by

$$\hat{\beta}_{PFM} = \left( \sum_{i=1}^n \sum_{t=1}^T \left( y_{i,t}^+ x'_{i,t-1} - \hat{\Lambda}_{12i}^+ \right) \right) \left( \sum_{i=1}^n \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \right)^{-1}. \quad (57)$$



As  $T \rightarrow \infty$ , for fixed  $n$ ,

$$\sqrt{n}T \left( \hat{\beta}_{PFM} - \beta \right) \Rightarrow \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \int_0^1 dB_{i,1.2} J_{i,C_i}(r)' \right) \right) \left( \frac{1}{n} \sum_{i=1}^n \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right)^{-1},$$

where  $B_{i,1.2} = B_{i,1} - \omega_{12i} \Omega_{22i}^{-1} B_{i,2} = BM(\sigma_{11.2,i})$  and  $\sigma_{11.2,i} = \sigma_{11i} - \omega_{12i} \Omega_{22i}^{-1} \Omega_{21i}$ . As  $(T, n \rightarrow \infty)_{\text{seq}}$ ,

$$\sqrt{n}T \left( \hat{\beta}_{PFM} - \beta \right) \Rightarrow N(0, \Omega_{xx}^{-1} \Phi_{u \cdot x} \Omega_{xx}^{-1}) \quad (58)$$

with

$$\Phi_{u \cdot x} = E \left[ \left( \int_0^1 dB_{i,1.2} J_{C_i}(r)' \right)' \left( \int_0^1 dB_{i,1.2} J_{C_i}(r)' \right) \right]. \quad (59)$$

As before, let  $\hat{u}_{i,t} = y_{i,t} - \hat{\beta}_{n,T} x_{i,t-1}$  and define  $\hat{u}_{i,2,t} = \hat{u}_{i,t} - \hat{\omega}_{12i} \hat{\Omega}_{22i}^{-1} v_{i,t}$ . Then

$$\hat{\Phi}_{u \cdot x} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T (\hat{u}_{i,2,t} x'_{i,t-1})' (\hat{u}_{i,2,s} x'_{i,s-1}) \rightarrow_p \Phi_{u \cdot x}, \quad (60)$$

as  $(T, n \rightarrow \infty)_{\text{seq}}$ . The PFM estimator is more efficient than the standard pooled estimator since  $\Phi_{u \cdot x} < \Phi_{u \cdot u}$ . Observe also that for  $\beta_{PFM}$ , using demeaned data will not cause a second order bias effect. The fully modified pooled estimator thus provides a more efficient alternative also in the case when individual effects are included. These results could also be shown to hold in joint limits, with some extra rate restrictions, but the results are omitted here for brevity.

## 6 Cross-sectional dependence

In many cases, the assumption of cross-sectional independence is restrictive. It is imposed because there is yet no satisfactory model or theory for dealing with general cross-sectional dependence. However, factor models are often used as a proxy for more general cross-sectional dependencies. The effects of common factors on panel data methods are relatively straightforward to analyze and this section provides some results along these lines. For ease of exposition, assume that  $\alpha_i = 0$ , for all  $i$ .

To study the effects of common factors in the returns process, suppose that equation (1) is replaced by

$$y_{i,t} = \beta_i x_{i,t-1} + \gamma_i \Lambda_t + u_{i,t}, \quad (61)$$

where the idiosyncratic error terms  $u_{i,t}$  still satisfy Assumption 1, but the common factor  $\Lambda_t$  is now a

part of the data generating process. Further, let the regressor process satisfy

$$x_{i,t} = z_{i,t} + \delta_i \Pi_t, \quad (62)$$

where

$$z_{i,t} = A_i z_{i,t} + v_{i,t}, \quad (63)$$

and

$$\Pi_t = G_\Pi \Pi_{t-1} + \eta_t. \quad (64)$$

The following high-level assumption states the properties of the common factors  $\Lambda_t$  and  $\Pi_t$ .

**Assumption 6**  $\frac{1}{\sqrt{T}} \sum_{s=1}^t \Lambda_s \Rightarrow B_\Lambda(r)$  and  $\frac{\Pi_t}{\sqrt{T}} \Rightarrow J_\Pi(r) = \int_0^r e^{(r-s)C_\Pi} dB_\Pi(r)$ , for  $t = [Tr]$ .

That is,  $\Lambda_t$  is assumed to be stationary whereas  $\Pi_t$  is a nearly integrated process. Consider first the pooled estimator in the case of homogeneous slope coefficients and no intercepts,

$$\begin{aligned} \sqrt{nT} (\hat{\beta}_{n,T} - \beta) &= \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T u_{i,t} x'_{i,t-1} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \right)^{-1} \\ &\quad + \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T \gamma_i \Lambda_t x'_{i,t-1} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \right)^{-1}. \end{aligned} \quad (65)$$

By Assumption 6, the first term satisfies,

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T u_{i,t} x'_{i,t-1} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T u_{i,t} z'_{i,t-1} + \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \frac{1}{T} \sum_{t=1}^T u_{i,t} \Pi'_{t-1} \\ &\Rightarrow \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \int_0^1 dB_{1,i} J'_{i,C_i} + \delta_i \int_0^1 dB_{1,i} J'_\Pi \right), \end{aligned} \quad (66)$$

as  $T \rightarrow \infty$ , for a fixed  $n$ . Conditional on the common shock,  $\Pi$ , the random variables  $\int_0^1 dB_{1,i} J'_\Pi$  are *iid* with mean zero, and thus

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \int_0^1 dB_{1,i} J'_{i,C_i} + \delta_i \int_0^1 dB_{1,i} J'_\Pi \right) \Rightarrow MN(0, \Phi_{ux} + \Phi_{u\Pi} + \Phi_{ux\Pi} + \Phi'_{ux\Pi})$$

where

$$\Phi_{u\Pi} = E \left[ \left( \delta_i \int_0^1 dB_{1,i} J'_{\Pi} \right)' \left( \delta_i \int_0^1 dB_{1,i} J'_{\Pi} \right) \middle| \Pi \right], \quad (67)$$

and

$$\Phi_{ux\Pi} = E \left[ \left( \int_0^1 dB_{1,i} J_{i,C_i} \right) \delta_i \left( \int_0^1 dB_{1,i} J'_{\Pi} \right) \middle| \Pi \right]. \quad (68)$$

Further, it follows that under Assumption 6, as  $(T, n \rightarrow \infty)_{\text{seq}}$ ,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \\ \rightarrow & {}_p E \left[ \int_0^1 J_{i,C_i} J'_{i,C_i} + \delta_i \int_0^1 J_{\Pi} J'_{i,C_i} + \left( \delta_i \int_0^1 J_{\Pi} J'_{i,C_i} \right)' + \delta_i \left( \int_0^1 J_{\Pi} J'_{\Pi} \right) \delta'_i \right] \\ \equiv & \Omega_{z\Pi}. \end{aligned} \quad (69)$$

However, as  $T \rightarrow \infty$ , for a fixed  $n$ ,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T \gamma_i \Lambda_t x'_{i,t-1} \\ = & \frac{1}{n} \sum_{i=1}^n \gamma_i \frac{1}{T} \sum_{t=1}^T \Lambda_t z'_{i,t-1} + \frac{1}{n} \sum_{i=1}^n \gamma_i \delta_i \frac{1}{T} \sum_{t=1}^T \Lambda_t \Pi'_{t-1} \\ \Rightarrow & \frac{1}{n} \sum_{i=1}^n \gamma_i \int_0^1 dB_{\Lambda} J_{i,C_i} + \frac{1}{n} \sum_{i=1}^n \gamma_i \delta_i \left( \int_0^1 dB_{\Lambda} J'_{\Pi} \right). \end{aligned} \quad (70)$$

Define

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \gamma_i \delta_i = \gamma \delta, \quad (71)$$

and it follows that

$$T \left( \hat{\beta}_{n,T} - \beta \right) \rightarrow_p \gamma \delta \left( \int_0^1 dB_{\Lambda} J'_{\Pi} \right) \Omega_{z\Pi}^{-1}. \quad (72)$$

With common factors in the error terms and the regressors, the pooled estimator thus suffers from a second order asymptotic bias. This can be controlled for in the following manner: (i) Estimate  $\beta_i$  with time-series OLS and estimate the residuals,  $\gamma_i \Lambda_t + u_{i,t}$  for all  $i$ . (ii) From the estimated residuals, estimate the common factor  $\gamma_i \Lambda_t$ . (iii) Subtract the common factor from  $y_{i,t}$ , and create

the ‘de-factored’ data,

$$y_{i,t}^{df} = y_{i,t} - \hat{\gamma}_i \hat{\Lambda}_t = \beta_i x_{i,t-1} + \gamma_i \Lambda_t - \hat{\gamma}_i \hat{\Lambda}_t + u_{i,t} = \beta_i x_{i,t-1} + u_{i,t} + o_p(1). \quad (73)$$

(iv) Use  $y_{i,t}^{df}$  instead of  $y_{i,t}$  in the pooled estimation.

It follows that

$$\hat{\beta}_{n,T}^{df} = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T y_{i,t}^{df} x'_{i,t-1} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \right)^{-1}, \quad (74)$$

satisfies

$$\sqrt{nT} \left( \hat{\beta}_{n,T}^{df} - \beta \right) \Rightarrow MN \left( 0, \Omega_{z\Pi}^{-1} (\Phi_{ux} + \Phi_{u\Pi} + \Phi_{ux\Pi} + \Phi'_{ux\Pi}) \Omega_{z\Pi}^{-1} \right), \quad (75)$$

as  $(T, n \rightarrow \infty)_{\text{seq}}$ . As also noted in Jin (2004), in the presence of common shocks the limiting distribution becomes mixed normal but the rate of convergence remains identical, although only after a de-factoring of the data in the case considered above. Again, these results can also be shown to hold in joint limits, but the details are omitted.

Results for the case with heterogeneous slope coefficients can be obtained in the same manner. However, it is easy to see that no de-factoring is necessary in this case, given that the asymptotic results are driven by the nearly non-stationary residuals arising from the heterogeneous slope coefficients, as discussed previously.

## 7 Unbalanced panels

In applied work, it is often the case that data for the individual time-series that make up the panel are available over different periods of time. If all the available data in such a panel is to be utilized, estimators dealing with the unbalanced nature of the panel is needed.

Suppose time-series  $i$  has  $T_i$  observations and let  $T = \max_i T_i$ . Further, assume that for all  $i$ ,  $T_i/T \rightarrow r_i \in (0, 1]$  as  $T_i \rightarrow \infty$  for all  $i$ . Define the unbalanced pooled estimator as follows,

$$\hat{\beta}_{n,T}^{ub} = \left( \sum_{i=1}^n \sum_{t=1}^{T_i} y_{i,t} x'_{i,t-1} \right) \left( \sum_{i=1}^n \sum_{t=1}^{T_i} x_{i,t-1} x'_{i,t-1} \right)^{-1}. \quad (76)$$

Consider first the heterogeneous case under Assumption 4, with no intercepts. Using sequential limits,

as  $(T, n \rightarrow \infty)_{\text{seq}}$ ,

$$\begin{aligned} \sqrt{n} \left( \hat{\beta}_{n,T}^{ub} - \beta \right) &= \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^{T_i} \theta_i x_{i,t-1} x'_{i,t-1} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^{T_i} x_{i,t-1} x'_{i,t-1} \right)^{-1} \\ &\quad + \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^{T_i} u_{i,t} x'_{i,t-1} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^{T_i} x_{i,t-1} x'_{i,t-1} \right)^{-1} \\ &\Rightarrow N \left( 0, (\Omega_{xx}^{ub})^{-1} \Phi_{xx}^{ub,\theta} (\Omega_{xx}^{ub})^{-1} \right), \end{aligned} \quad (77)$$

where

$$\Omega_{xx}^{ub} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \left[ \int_0^{r_i} J_{i,C_i}(r) J_{i,C_i}(r)' dr \right], \quad (78)$$

and

$$\Phi_{xx}^{ub,\theta} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \left[ \left( \theta_i \int_0^{r_i} J_{i,C_i}(r) J_{i,C_i}(r)' dr \right)' \left( \theta_i \int_0^{r_i} J_{i,C_i}(r) J_{i,C_i}(r)' dr \right) \right]. \quad (79)$$

As in the balanced case, these can be easily estimated using the following estimators,

$$\hat{\Omega}_{xx}^{ub} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^{T_i} x_{i,t-1} x'_{i,t-1}, \quad (80)$$

and

$$\hat{\Phi}_{xx}^{ub,\theta} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^4} \sum_{t=1}^{T_i} \sum_{s=1}^{T_i} (\hat{u}_{i,t} x'_{i,t-1})' (\hat{u}_{i,s} x'_{i,s-1}). \quad (81)$$

The homogeneous case follows in a completely analogous manner and is not detailed here.

## 8 Finite sample evidence

To evaluate the small sample properties of the panel data estimators proposed in this paper, a Monte Carlo study is performed. In the first experiment, the properties of the point estimates are considered. Equations (1) and (2) are simulated for the case with a single regressor. The innovations  $(u_{i,t}, v_{i,t})$  are drawn from normal distributions with mean zero, unit variance, and correlations  $\delta = 0, -0.4, -0.7$ , and  $-0.95$ ; there is no cross-sectional dependence. The slope parameters,  $\beta_i$ , are drawn from a normal distribution with mean  $\beta = 0.05$ , and with a standard deviation that is also equal to 0.05. The sample size is set to  $T = 100$ ,  $n = 20$ , and the local-to-unity parameters  $C_i$  are drawn from a uniform distribution with support  $[-10, 0]$ . The small value of  $\beta$  is chosen in order to reflect the fact that

most forecasting regressions are used to test a null of  $\beta = 0$ , and any plausible alternative is often close to zero. The intercepts  $\alpha_i$  are set equal to zero but individual effects are still fitted in each regression, to evaluate the second-order bias effects arising from demeaning. All results are based on 10,000 repetitions.

Three different estimators are considered: the recursively demeaned pooled estimator,  $\tilde{\beta}_{n,T}^{rd}$ , the unfeasible fully modified pooled estimator,  $\hat{\beta}_{PFM}$ , and the pooled estimator using standard demeaning,  $\tilde{\beta}_{n,T}$ . The estimator  $\hat{\beta}_{PFM}$  is formed using the fact that there is no serial correlation in the data, so that simple sample variance and sample covariance estimates are used rather than their long-run counterparts, and relies on knowledge of the true values of the  $C_i$ s. The results are shown in Table 1 and Figure 1. By the results in Panel A of Table 1, all estimators, except  $\tilde{\beta}_{n,T}$ , are virtually unbiased. The estimator  $\tilde{\beta}_{n,T}$ , which uses standard demeaning to account for individual effects, exhibits a rather substantial bias when the absolute value of the correlation  $\delta$  is large, however. From Panel B of Table 1 and the graphs in Figure 1, it is apparent that the unfeasible estimator  $\hat{\beta}_{PFM}$  outperforms the other estimators. The recursively demeaned estimator,  $\tilde{\beta}_{n,T}^{rd}$ , suffers from a lack of efficiency, but it is well centered around the true value.

The second part of the Monte Carlo study concerns the size and power of the pooled  $t$ -tests. The same setup as above is used but the average slope coefficient  $\beta$  now varies between  $-0.05$  and  $0.05$ ; the slope coefficients,  $\beta_i$ , are drawn from a normal distribution with mean *and* standard deviation equal to  $\beta$ . Figure 2 shows the average rejection rates of the  $t$ -tests resulting from the recursively demeaned and fully modified pooled estimators, evaluating a null of  $\beta = 0$ . Since  $\tilde{\beta}_{n,T}$  does not have a proper asymptotic distribution, no test is provided for this estimator. Again, the results are based on 10,000 repetitions. The power curves from the two panel data based  $t$ -tests are compared to that of the time-series  $t$ -test corresponding to the unfeasible maximum likelihood (ML) estimator of  $\beta_i$  described in Campbell and Yogo (2003); the ML time-series estimator and test are unfeasible since they rely on knowledge of the true value of  $C_i$ , but the resulting test provides a good benchmark against which the panel tests can be compared.<sup>4</sup> In terms of size, all tests perform well, although the test from the fully modified pooled estimator tends to overreject somewhat. The test based on the estimator using recursively demeaned data,  $\tilde{\beta}^{rd}$ , has good size for all values of  $\delta$  and also exhibit acceptable power properties, as compared to the unfeasible time-series test.

In summary, the simulation evidence shows the importance of controlling for the second order bias arising from fitting individual intercepts in the pooled regression; the estimator based on recursive demeaning appear to do so well and result in test-statistics with correct size and decent power properties. Substantial efficiency gains can also be achieved by considering a fully modified estimator. Although this estimator is unfeasible in the simulation exercise just described, it is feasible in the special case of identical persistence in the time-series.

---

<sup>4</sup>By Campbell and Yogo, this  $t$ -test is referred to as the unfeasible  $Q$ -test.

## 9 Conclusions

A panel data extension of the traditional linear forecasting model is considered. I analyze a setup where the regressors are nearly persistent processes and potentially endogenous, which captures the essential characteristics of many empirical situations. It is shown that when no fixed effects are present the standard pooled estimator is asymptotically normal and standard inference can be performed; the cross-sectional information effectively dilutes the endogeneity effects that are present in the standard time-series case and as the cross-sectional dimension grows large, these effects disappear altogether. However, when individual intercepts, or fixed effects, are estimated, the endogeneity of the regressors cause the pooled estimator to have a second order bias. To control for these effects, an alternative pooled estimator based on the concept of recursive demeaning is proposed. Monte Carlo evidence suggests that this estimator has good finite sample properties and also shows that in a typical setup, the distortions to the standard fixed effects estimator can be quite severe when the regressors are endogenous. Some new methods for dealing with cross-sectional dependence, in the form of common factors, in panels are also developed.

The results in this paper provide an important extension to the existing literature on time-series methods for predictive regressions and shows that also in the panel data case a careful analysis of the impact of nearly persistent and endogenous regressors is required.

## A Joint limit theory

The following two theorems, from Phillips and Moon (1999), provide useful conditions for establishing joint convergence.

**Theorem 6** *Suppose that  $Y_{i,T} = C_i Q_{i,T}$ , where the  $(m \times 1)$  random vectors  $Q_{i,T}$  are iid across  $i = 1, \dots, n$  for all  $T$ , and the  $C_i$  are  $(m \times m)$  nonrandom matrices for all  $i$ . Assume that*

- (i)  $Q_{i,T} \Rightarrow Q_i$  as  $T \rightarrow \infty$  for all  $i$ ;
  - (ii)  $\|Q_{i,T}\|$  is uniformly integrable in  $T$  for all  $i$ ; and
  - (iii)  $\sup_i \|C_i\| < \infty$ ,  $\inf_i \|C_i\| > 0$ , and  $C = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n C_i$ .
- Then  $\frac{1}{n} \sum_{i=1}^n Y_{i,T} \rightarrow_p CE[Q_i]$  as  $(n, T \rightarrow \infty)$ .

**Theorem 7** *Suppose that  $Y_{i,T} = C_i Q_{i,T}$ , where the  $(m \times 1)$  random vectors  $Q_{i,T}$  are iid  $(0, \Sigma_T)$  across  $i = 1, \dots, n$  for all  $T$ , and the  $C_i$  are  $(m \times m)$  nonzero and nonrandom matrices. Assume the following conditions hold:*

- (i) Let  $\sigma_T^2 = \lambda_{\min}(\Sigma_T)$  and  $\liminf_T \sigma_T^2 > 0$ ,
- (ii)  $\frac{\max_{i \leq n} \|C_i\|^2}{\lambda_{\min}(\sum_{i=1}^n C_i C_i')} = O\left(\frac{1}{n}\right)$  as  $n \rightarrow \infty$ ,
- (iii)  $\|Q_{i,T}\|^2$  are uniformly integrable in  $T$ , and
- (iv)  $\lim_{n,T} \frac{1}{n} \sum_{i=1}^n C_i \Sigma_T C_i' = \Omega > 0$ .

Then,

$$X_{n,T} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_{i,t} \Rightarrow N(0, \Omega) \text{ as } n, T \rightarrow \infty.$$

Given the sequential limit results derived earlier, establishing the joint limit results is done by verifying the conditions in Theorem (6) or (7). Typically the main challenge lies in establishing the uniform integrability condition. Three useful results for establishing uniform integrability are given next.

- (i) If  $X_T \Rightarrow X$  as  $T \rightarrow \infty$ , then uniform integrability of  $\|X_T\|$  is equivalent to  $E\|X_T\| \rightarrow E\|X\|$  as  $T \rightarrow \infty$ .
- (ii) If there exists a sequence of random variables  $U_T$ , such that  $U_T \geq \|X_T\|$  almost surely, then uniform integrability of  $U_T$  implies uniform integrability of  $\|X_T\|$ .
- (iii) If  $X_T = W_T Z_T$  and  $\|W_T\|^2$  and  $\|Z_T\|^2$  are uniformly integrable, then  $\|X_T\|$  is uniformly integrable.



## B A BN decomposition

Assumption 1 allows for the following BN decomposition of the innovations  $v_{i,t}$ ,

$$v_{i,t} = D_i \epsilon_{i,t} + \tilde{v}_{i,t-1} - \tilde{v}_{i,t}, \quad (82)$$

where  $\tilde{v}_{i,t} = \sum_{j=0}^{\infty} \tilde{D}_{i,j} \epsilon_{i,t-j}$  and  $\tilde{D}_{i,j} = \sum_{k=j+1}^{\infty} D_{i,k}$  (Phillips and Solo, 1992). By the definition of a nearly integrated process,  $x_{i,t}$  can be written as

$$x_{i,t} = \sum_{s=1}^t e^{\frac{t-s}{T} C_i} v_{i,s} + e^{\frac{t}{T} C_i} x_{i,0}. \quad (83)$$

Using the BN decomposition of  $v_{i,t}$ , the following decomposition of  $x_{i,t}$  now holds,

$$x_{i,t} = z_{i,t} + R_{i,t}, \quad (84)$$

where

$$z_{i,t} = \sum_{s=1}^t e^{\frac{(t-s)}{T} C_i} D_i \epsilon_{i,s} \quad (85)$$

and

$$\begin{aligned} R_{i,t} &= \sum_{s=1}^t e^{\frac{t-s}{T} C_i} \tilde{v}_{i,s-1} - \sum_{s=1}^t e^{\frac{t-s}{T} C_i} \tilde{v}_{i,s} + e^{\frac{t}{T} C_i} x_{i,0} \\ &= e^{\frac{t-1}{T} C_i} \tilde{v}_{i,0} + \sum_{s=1}^{t-1} \left( e^{\frac{t-s-1}{T} C_i} - e^{\frac{t-s}{T} C_i} \right) \tilde{v}_{i,s} - \tilde{v}_{i,t} + e^{\frac{t}{T} C_i} x_{i,0}. \end{aligned}$$

**Lemma 2** *Let  $M$  denote a generic constant. Under Assumptions 1-3, the following hold,*

(a)

$$\frac{1}{T} E \|z_{i,t}\|^2 < M,$$

(b)

$$E \|R_{i,t}\|^2 < M.$$

**Proof.** The result follows by applying the same arguments as in Moon and Phillips (1999), and the fact that  $C_i \leq \bar{C}$ . ■

## C Proofs

**Proof of Lemma 1.** By Assumptions 1 and 3,

$$E \left[ \left\| \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right\|^2 \right] = E \left[ \left\| D_i \left( \int_0^1 J_{i,C_i}^W(r) J_{i,C_i}^W(r)' dr \right) D_i' \right\|^2 \right] < \infty.$$

■

**Derivation of  $\Phi_{ux}^\theta$ ,  $\Omega_{xx}$  and  $\Phi_{ux}$  in the scalar case, with  $C_i = C$  for all  $i$ .** (i) For scalar regressors,

$$\Phi_{xx}^\theta = \Omega_{\theta\theta} \int_0^1 \int_0^1 E \left[ J_{i,C_i}(r)^2 J_{i,C_i}(p)^2 \right] dpdr.$$

By Moon and Phillips (1999),

$$\begin{aligned} & E \left[ J_{i,C_i}(r)^2 J_{i,C_i}(p)^2 \right] \\ &= \Omega_{22}^2 \left( \int_0^r e^{C(2r-2x)} dx \int_0^p e^{C(2p-2x)} dx + \int_0^{r \wedge p} e^{C(r+p-2x)} dx \int_0^{r \wedge p} e^{C(r+p-2x)} dx \right. \\ & \quad \left. + \int_0^{r \wedge p} e^{C(r+p-2x)} dx \int_0^{r \wedge p} e^{C(r+p-2x)} dx \right) \\ &= \Omega_{22}^2 \left( \int_0^r \int_0^p e^{2C(r+p-x-y)} dx dy + 2 \int_0^{r \wedge p} \int_0^{r \wedge p} e^{2C(r+p-x-y)} dx dy \right). \end{aligned}$$

By simple calculations

$$\int_0^1 \int_0^1 \int_0^r \int_0^p e^{2C(r+p-x-y)} dy dx dp dr = \frac{(-1 - 2C + e^{2C})^2}{16C^4},$$

and,

$$2 \int_0^1 \int_0^1 \int_0^{r \wedge p} \int_0^{r \wedge p} e^{2C(r+p-x-y)} dx dy dp dr = 4 \left( \frac{-5 + 4e^{2C} + e^{4C} - 4C(1 + 2e^{2C})}{32C^4} \right).$$

Summing up,

$$\Phi_{xx}^\theta = \Omega_{\theta\theta} \Omega_{22}^2 \frac{1}{16C^4} (4C^2 - 4C + 6e^{2C} + 3e^{4C} - 20Ce^{2C} - 9).$$

(ii) By Phillips (1987),

$$\Omega_{xx} = \int_0^1 E \left[ J_{i,C_i}(r)^2 \right] dr = \Omega_{22} \int_0^1 \left( \frac{1}{2C} (e^{2rC} - 1) \right) dr = \Omega_{22} \frac{1}{4C^2} (e^{2C} - 2C - 1).$$

(iii) Since,

$$dB_{i,1}(r) = dB_{i,1.2}(r) + \omega_{12i}\Omega_{22i}^{-1}dB_{i,2}(r),$$

it follows that,

$$\begin{aligned}\Phi_{ux} &= E \left[ \int_0^1 \int_0^1 J_{C_i}(r) J_{C_i}(s) dB_{i,1}(r) dB_{i,1}(s) \right] \\ &= E \left[ \int_0^1 \int_0^1 J_{C_i}(r) J_{C_i}(s) dB_{i,1.2}(r) dB_{i,1.2}(s) \right] \\ &\quad + \omega_{12}^2 \Omega_{22}^{-2} E \left[ \int_0^1 \int_0^1 J_{C_i}(r) J_{C_i}(s) dB_{i,2}(r) dB_{i,2}(s) \right].\end{aligned}$$

By Lemma 5(b) in Moon and Phillips (1999),

$$\omega_{12}^2 \Omega_{22}^{-2} E \left[ \int_0^1 \int_0^1 J_{C_i}(r) J_{C_i}(s) dB_{i,2}(r) dB_{i,2}(s) \right] = \omega_{12}^2 \int_0^1 \int_0^r e^{2C(r-s)} ds dr = \omega_{12}^2 \frac{1}{4C^2} (e^{2C} - 2C - 1).$$

Further,

$$\begin{aligned}& E \left[ \int_0^1 \int_0^1 J_{C_i}(r) J_{C_i}(s) dB_{i,1.2}(r) dB_{i,1.2}(s) \right] \\ &= \int_0^1 \int_0^1 E [J_{C_i}(r) J_{C_i}(s)] E [dB_{i,1.2}(r) dB_{i,1.2}(s)] \\ &= (\sigma_{11} - \omega_{12}\Omega_{22}^{-1}\omega_{21}) \int_0^1 \int_0^1 E [J_{C_i}(r)^2] dr \\ &= (\sigma_{11}\Omega_{22} - \omega_{12}^2) \frac{1}{4C^2} (e^{2C} - 2C - 1).\end{aligned}$$

Summing up,

$$\Phi_{ux} = \sigma_{11i}\Omega_{22} \frac{1}{4C^2} (e^{2C} - 2C - 1).$$

It follows that

$$\Omega_{xx}^{-1} \Phi_{ux}^\theta \Omega_{xx}^{-1} = \Omega_{\theta\theta} (e^{2C} - 2C - 1)^{-2} (4C^2 - 4C + 6e^{2C} + 3e^{4C} - 20Ce^{2C} - 9),$$

and

$$\Omega_{xx}^{-1} \Phi_{ux} \Omega_{xx}^{-1} = \sigma_{11i}\Omega_{22}^{-1} \left( \frac{1}{4C^2} (e^{2C} - 2C - 1) \right)^{-1}.$$

■

The following lemmas, proved at the end of this Appendix, help establish the main results.

**Lemma 3** *Under Assumptions 1-3, as  $(n, T \rightarrow \infty)$ ,*

(a)

$$\hat{\Omega}_{xx} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T x_{i,t} x'_{i,t} \rightarrow_p \Omega_{xx}. \quad (86)$$

**Lemma 4** Under Assumptions 1-3, as  $(n, T \rightarrow \infty)$  with  $n/T \rightarrow 0$ , then

(a) if Assumption 4 also holds,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \theta_i \frac{1}{T^2} \sum_{t=1}^T x_{i,t-1} x'_{i,t-1} \Rightarrow N(0, \Phi_{xx}^\theta), \quad (87)$$

(b)

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T u_{i,t} x'_{i,t-1} \Rightarrow N(0, \Phi_{xx}). \quad (88)$$

**Lemma 5** Under Assumptions 1-3, as  $(n, T \rightarrow \infty)$ , then

(a) if Assumption 4 also holds,

$$\frac{1}{n} \sum_{i=1}^n \theta_i \frac{1}{T^2} \sum_{t=1}^T x_{i,t} x'_{i,t} \rightarrow_p 0, \quad (89)$$

(b)

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T u_{i,t} x'_{i,t-1} \rightarrow_p 0. \quad (90)$$

**Lemma 6** Under Assumptions 1-3, as  $(n, T \rightarrow \infty)$ , then

(a) if Assumption 4 also holds,

$$\hat{\Phi}_{xx}^\theta = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^4} \sum_{t=1}^T \sum_{s=1}^T (\hat{u}_{i,t} x'_{i,t-1})' (\hat{u}_{i,t} x'_{i,s-1}) \rightarrow_p \Phi_{xx}^\theta, \quad (91)$$

(b) if Assumption 5 also holds,

$$\hat{\Phi}_{ux} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T (\hat{u}_{i,t} x'_{i,t-1})' (\hat{u}_{i,t} x'_{i,s-1}) \rightarrow_p \Phi_{ux}. \quad (92)$$

**Proof of Theorem 1.** Part (i) follows directly from Lemma 3, Lemma 5(a), and Slutsky's theorem. By combining the results of Lemma 3 and Lemma 4(a), part (ii) follows. ■

**Proof of Theorem 2.** The results follow by Theorem 1, Lemma 3 and Lemma 6(a). ■

**Proof of Theorem 3.** Lemma 3 and Lemma 5(b) provide the result in (i). Lemma 3 and Lemma 4(b) give (ii). ■

**Proof of Theorem 4.** The results follow by Theorem 3, Lemma 3 and Lemma 6(b). ■

**Proof of Theorem 5.** I first establish the result in sequential limits.

(i) Observe that

$$\begin{aligned}
\underline{y}_{i,t}^{dd} &= y_{i,t} - \frac{1}{T-t} \sum_{s=t}^T y_{i,s} \\
&= \beta \left( x_{i,t} - \frac{1}{T-t} \sum_{s=t}^T x_{i,s} \right) + u_{i,t} - \frac{1}{T-t} \sum_{s=t}^T u_{i,s} \\
&= \beta \underline{x}_{i,t}^{dd} + \underline{u}_{i,t}^{dd}.
\end{aligned}$$

Using sequential limits, as  $(T, n \rightarrow \infty)_{\text{seq}}$ ,

$$\begin{aligned}
\tilde{\beta}_{n,T}^{rd} &= \left( \sum_{i=1}^n \sum_{t=1}^T \underline{y}_{i,t}^{dd} \underline{x}_{i,t-1}^{d'} \right) \left( \sum_{i=1}^n \sum_{t=1}^T \underline{x}_{i,t-1}^{dd} \underline{x}_{i,t-1}^{d'} \right)^{-1} \\
&= \left( \sum_{i=1}^n \sum_{t=1}^T (\beta \underline{x}_{i,t}^{dd} \underline{x}_{i,t-1}^{d'} + \underline{u}_{i,t}^{dd} \underline{x}_{i,t-1}^{d'}) \right) \left( \sum_{i=1}^n \sum_{t=1}^T \underline{x}_{i,t-1}^{dd} \underline{x}_{i,t-1}^{d'} \right)^{-1} \\
&= \beta + \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \underline{u}_{i,t}^{dd} \underline{x}_{i,t-1}^{d'} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \underline{x}_{i,t-1}^{dd} \underline{x}_{i,t-1}^{d'} \right)^{-1} \rightarrow_p \beta.
\end{aligned}$$

(ii) For fixed  $n$ , as  $T \rightarrow \infty$ ,

$$\begin{aligned}
\sqrt{n}T \left( \tilde{\beta}_{n,T}^{rd} - \beta \right) &= \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T \underline{u}_{i,t}^{dd} \underline{x}_{i,t-1}^{r'} \right) \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \underline{x}_{i,t-1}^{dd} \underline{x}_{i,t-1}^{r'} \right)^{-1} \\
&\Rightarrow \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^1 dB_{1,i}^{dd}(r) \underline{J}_{C_i}^d(r)' \right) \left( \frac{1}{n} \sum_{i=1}^n \int_0^1 \underline{J}_{C_i}^{dd}(r) \underline{J}_{C_i}^d(r)' dr \right)^{-1},
\end{aligned}$$

since

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^T u_{i,t}^{dd} \underline{x}_{i,t-1}^{dt} &= \frac{1}{T} \sum_{t=1}^T \left( u_{i,t} - \frac{1}{T-t} \sum_{s=t}^T u_{i,s} \right) \underline{x}_{i,t-1}^{dt} \\
&= \frac{1}{T} \sum_{t=1}^T u_{i,t} \underline{x}_{i,t-1}^{dt} - \frac{1}{T} \sum_{t=1}^T \left( \frac{T}{T-t} \frac{1}{T} \sum_{s=t}^T u_{i,s} \right) \underline{x}_{i,t-1}^{dt} \\
&= \frac{1}{T} \sum_{t=1}^T u_{i,t} \underline{x}_{i,t-1}^{dt} - \frac{1}{T} \sum_{t=1}^T \left( \frac{T-t}{T} \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{s=t}^T u_{i,s} \right) \frac{\underline{x}_{i,t-1}^{dt}}{\sqrt{T}} \\
&= \frac{1}{T} \sum_{t=1}^T u_{i,t} \underline{x}_{i,t-1}^{dt} - \frac{1}{T} \sum_{t=1}^T \left( \frac{T-t}{T} \right)^{-1} \left( \frac{1}{\sqrt{T}} \left( \sum_{s=1}^T u_{i,s} - \sum_{s=1}^{t-1} u_{i,s} \right) \right) \frac{\underline{x}_{i,t-1}^{dt}}{\sqrt{T}} \\
&\Rightarrow \int_0^1 dB_{1,i}(r) \underline{J}_{C_i}^d(r)' - \int_0^1 (1-r)^{-1} (B_{1,i}(1) - B_{1,i}(r)) \underline{J}_{C_i}^d(r)' dr \\
&= \int_0^1 dB_{1,i}(r) \left( J_{i,C_i}(r) - r^{-1} \int_0^r J_{i,C_i}(u) du \right)' \\
&\quad - \int_0^1 (1-r)^{-1} (B_{1,i}(1) - B_{1,i}(r)) \left( J_{i,C_i}(r) - r^{-1} \int_0^r J_{i,C_i}(u) du \right)' dr \\
&= \int_0^1 dB_{1,i}(r) J_{i,C_i}(r)' - \int_0^1 dB_{1,i}(r) \left( r^{-1} \int_0^r J_{i,C_i}(u) du \right)' \\
&\quad - \int_0^1 (1-r)^{-1} (B_{1,i}(1) - B_{1,i}(r)) J_{i,C_i}(r)' dr \\
&\quad + \int_0^1 (1-r)^{-1} (B_{1,i}(1) - B_{1,i}(r)) \left( r^{-1} \int_0^r J_{i,C_i}(u) du \right)' dr \\
&= \int_0^1 dB_{1,i}^{dd}(r) \underline{J}_{C_i}^d(r)'. \tag{93}
\end{aligned}$$

by standard arguments. By the independent increments property of the Brownian motion, it follows that the expectation of (93) is equal to zero. By similar arguments as in the main text, as  $(T, n \rightarrow \infty)_{\text{seq}}$ ,

$$\sqrt{n}T \left( \tilde{\beta}_{n,T}^{rd} - \beta \right) \Rightarrow N \left( 0, \left( \underline{\Omega}^{rd} \right)^{-1} \underline{\Phi}_{ux}^{rd} \left( \underline{\Omega}^{rd} \right)^{-1} \right).$$

The result in joint limits now follows in the same manner as in the previous theorems; the details are omitted. ■

**Proof of Lemma 3.** (a) Use the decomposition in (84) to write

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T x_{i,t} x'_{i,t} \\
&= \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T (z_{i,t} + R_{i,t}) (z_{i,t} + R_{i,t})' \\
&= \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T (z_{i,t} z'_{i,t} + z_{i,t} R'_{i,t} + R_{i,t} z'_{i,t} + R_{i,t} R'_{i,t}) \\
&= \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T z_{i,t} z'_{i,t} + \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T z_{i,t} R'_{i,t} \\
&\quad + \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T R_{i,t} z'_{i,t} + \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T R_{i,t} R'_{i,t} \\
&= \frac{1}{n} \sum_{i=1}^n Q_{i,T} + \frac{1}{n} \sum_{i=1}^n P_{1,T} + \frac{1}{n} \sum_{i=1}^n P_{2,T} + \frac{1}{n} \sum_{i=1}^n P_{3,T}.
\end{aligned}$$

To prove (86), I show that  $\frac{1}{n} \sum_{i=1}^n Q_{i,T} \rightarrow_p \Omega_{xx} = E \left[ \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right]$  and that  $\frac{1}{n} \sum_{i=1}^n P_{k,T} \rightarrow_p 0$ ,  $k = 1, 2, 3$ , as  $(n, T \rightarrow \infty)$ . Start with  $\frac{1}{n} \sum_{i=1}^n Q_{i,T}$ . By Theorem 6, it is sufficient to show that  $\|Q_{i,T}\|$  is uniformly integrable in  $T$  for all  $i$ . By the triangle inequality

$$\|Q_{i,T}\| \leq \frac{1}{T^2} \sum_{t=1}^T \|z_{i,t}\|^2.$$

As  $T \rightarrow \infty$ ,

$$\frac{1}{T^2} \sum_{t=1}^T \|z_{i,t}\|^2 \Rightarrow \int_0^1 \|J_{i,C_i}\|^2,$$

and

$$E \left[ \frac{1}{T^2} \sum_{t=1}^T \|z_{i,t}\|^2 \right] = \text{tr} \left( \frac{1}{T^2} \sum_{t=1}^T E [z_{i,t} z'_{i,t}] \right) \rightarrow E \left[ \int_0^1 \|J_{i,C_i}\|^2 \right].$$

Thus  $\|Q_{i,T}\|$  is uniformly integrable. Next, to show that  $\frac{1}{n} \sum_{i=1}^n P_{k,T} \rightarrow_p 0$  as  $(n, T \rightarrow \infty)$ , it is

sufficient to show that  $E \|P_{k,T}\| \rightarrow_p 0$  as  $(n, T \rightarrow \infty)$ , for  $k = 1, 2, 3$ . Start with  $k = 1$ ,

$$\begin{aligned}
E \|P_{1,T}\| &= E \left\| \frac{1}{T^2} \sum_{t=1}^T z_{i,t} R'_{i,t} \right\| \\
&\leq \frac{1}{T^2} \sum_{t=1}^T E [\|z_{i,t}\| \|R_{i,t}\|] \\
&\leq \frac{1}{\sqrt{T}} \frac{1}{T} \sum_{t=1}^T \sqrt{E \left\| \frac{z_{i,t}}{\sqrt{T}} \right\|^2 E \|R_{i,t}\|^2} \\
&= O\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}$$

where the first inequality follows from the triangle inequality and  $\|AB\| \leq \|A\| \|B\|$ , the second inequality follows from the Cauchy-Schwarz inequality, and the last step follows from Lemma 2. An identical argument holds for  $k = 2$ . For  $k = 3$ , by the triangle inequality and Lemma 2,

$$E \|P_{3,T}\| = E \left\| \frac{1}{T^2} \sum_{t=1}^T R_{i,t} R'_{i,t} \right\| \leq \frac{1}{T^2} \sum_{t=1}^T E [\|R_{i,t}\|^2] = O\left(\frac{1}{T}\right).$$

■

**Proof of Lemma 4.** (a) Similar to above, write

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^n \theta_i \frac{1}{T^2} \sum_{t=1}^T x_{i,t} x'_{i,t} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \theta_i \frac{1}{T^2} \sum_{t=1}^T (z_{i,t} + R_{i,t}) (z_{i,t} + R_{i,t})' \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \theta_i \frac{1}{T^2} \sum_{t=1}^T (z_{i,t} z'_{i,t} + z_{i,t} R'_{i,t} + R_{i,t} z'_{i,t} + R_{i,t} R'_{i,t}) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n (Q_{i,T} + P_{1,i,T} + P_{2,i,T} + P_{3,i,T})
\end{aligned}$$

where

$$Q_{i,T} = \theta_i \frac{1}{T^2} \sum_{t=1}^T z_{i,t} z'_{i,t},$$

$$P_{1,i,T} = \theta_i \frac{1}{T^2} \sum_{t=1}^T z_{i,t} R'_{i,t},$$



$$P_{2,i,T} = \theta_i \frac{1}{T^2} \sum_{t=1}^T R_{i,t} z'_{i,t},$$

and

$$P_{3,i,T} = \theta_i \frac{1}{T^2} \sum_{t=1}^T R_{i,t} R'_{i,t}.$$

I show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n Q_{i,T} \Rightarrow N(0, \Phi_{xx}),$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n P_{k,i,T} \rightarrow_p 0$$

as  $(n, T \rightarrow \infty)$  with  $n/T \rightarrow 0$ . By the independence assumption of  $\theta_i$ , it immediately holds that

$$E[Q_{i,T}] = 0.$$

Also,

$$\begin{aligned} E[Q'_{i,T} Q_{i,T}] &= E \left[ \left( \theta_i \frac{1}{T^2} \sum_{t=1}^T z_{i,t} z'_{i,t} \right)' \left( \theta_i \frac{1}{T^2} \sum_{t=1}^T z_{i,t} z'_{i,t} \right) \right] \\ &= E \left[ \frac{1}{T^2} \sum_{t=1}^T z_{i,t} z'_{i,t} (\theta'_i \theta_i) \frac{1}{T^2} \sum_{t=1}^T z_{i,t} z'_{i,t} \right] \\ &= E \left[ \frac{1}{T^2} \sum_{t=1}^T z_{i,t} z'_{i,t} \Omega_{\theta\theta} \frac{1}{T^2} \sum_{t=1}^T z_{i,t} z'_{i,t} \right] \\ &= \Phi_{xx,T}, \end{aligned}$$

where it easily follows that  $\Phi_{xx,T} \rightarrow \Phi_{xx}$  as  $T \rightarrow \infty$ . The  $Q_{i,T}$  are thus *iid* sequence with mean zero and covariance  $\Phi_{xx,T}$ . Conditions (i), (ii), and (iv) of Theorem 7 are clearly satisfied. To prove that  $\|Q_{i,T}\|^2$  is U.I. note that by the CMT, as  $T \rightarrow \infty$ ,

$$\|Q_{i,T}\|^2 \Rightarrow \|Q_i\|^2 = \left\| \theta_i \int_0^1 J_{i,C_i}(r) J_{i,C_i}(r)' dr \right\|^2$$

and

$$\begin{aligned} E\|Q_{i,T}\|^2 &= \text{tr}(E[Q'_{i,T} Q_{i,T}]) = \text{tr}(\Phi_{xx,T}) \rightarrow \text{tr}(\Phi_{xx}) \\ &= \text{tr}(E[Q'_i Q_i]) = E\|Q_i\|^2. \end{aligned}$$

It follows that  $\|Q_{i,T}\|^2$  is U.I. It remains to show that  $\frac{1}{\sqrt{n}} \sum_{i=1}^n P_{k,i,T} \rightarrow_p 0$ , for  $k = 1, 2, 3$ , as  $(n, T \rightarrow \infty)$  with  $n/T \rightarrow 0$ . I do this by showing that  $\sqrt{n}E\|P_{k,i,T}\| \rightarrow_p 0$ . For  $k = 1$

$$\begin{aligned}
\sqrt{n}E\|P_{1,i,T}\| &= \sqrt{n}E \left\| \theta_i \frac{1}{T^2} \sum_{t=1}^T z_{i,t} R'_{i,t} \right\| \\
&\leq \sqrt{n}E \|\theta_i\| E \left\| \frac{1}{T^2} \sum_{t=1}^T z_{i,t} R'_{i,t} \right\| \\
&\leq \sqrt{n} \sqrt{\text{tr}(\Omega_{\theta\theta})} \frac{1}{T^2} \sum_{t=1}^T E \|z_{i,t}\| \|R_{i,t}\| \\
&\leq \sqrt{n} \sqrt{\text{tr}(\Omega_{\theta\theta})} \frac{1}{\sqrt{T}} \frac{1}{T} \sum_{t=1}^T \sqrt{E \left\| \frac{z_{i,t}}{\sqrt{T}} \right\|^2} E \|R_{i,t}\|^2 \\
&= \sqrt{\text{tr}(\Omega_{\theta\theta})} \sqrt{\frac{n}{T}} O(1) \\
&= o(1)
\end{aligned}$$

where the first inequality follows from the independence of  $\theta_i$ , the second from the triangle inequality and  $\|AB\| \leq \|A\| \|B\|$ , the third from the Cauchy-Schwarz inequality. The last two equalities follow from Lemma 2 and the assumption that  $n/T \rightarrow 0$ . The same argument holds for  $P_{2,i,T}$  also. Using similar arguments,

$$\begin{aligned}
\sqrt{n}E\|P_{3,i,T}\| &= \sqrt{n}E \left\| \theta_i \frac{1}{T^2} \sum_{t=1}^T R_{i,t} R'_{i,t} \right\| \\
&\leq \sqrt{n}E \|\theta_i\| E \left\| \frac{1}{T^2} \sum_{t=1}^T R_{i,t} R'_{i,t} \right\| \\
&\leq \sqrt{n} \sqrt{\text{tr}(\Omega_{\theta\theta})} \frac{1}{T^2} \sum_{t=1}^T E \|R_{i,t}\|^2 \\
&= O\left(\frac{\sqrt{n}}{T}\right) \\
&= o(1),
\end{aligned}$$

where the first inequality holds by  $\|AB\| \leq \|A\| \|B\|$  and the independence of  $\theta_i$ , the second inequality follows from the triangle inequality and the last two equalities follows from Lemma 2 and  $n/T \rightarrow 0$ .

(b) By the BN decomposition,

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T u_{i,t} x'_{i,t-1} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T (u_{i,t} z'_{i,t-1} + u_{i,t} R'_{i,t-1}) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n (Q_{i,T} + P_{i,T})
\end{aligned}$$

where

$$Q_{i,T} = \frac{1}{T} \sum_{t=1}^T u_{i,t} z'_{i,t}, \quad \text{and} \quad P_{i,T} = \frac{1}{T} \sum_{t=1}^T u_{i,t} R'_{i,t-1}.$$

As in the proof of (a), I show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n Q_{i,T} \Rightarrow N(0, \Phi_{ux}), \quad \text{and} \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n P_{i,T} \rightarrow_p 0,$$

as  $(n, T \rightarrow \infty)$  with  $n/T \rightarrow 0$ . Observe first that

$$E[Q_{i,T}] = E \left[ \frac{1}{T} \sum_{t=1}^T u_{i,t} z'_{i,t-1} \right] = 0,$$

by the martingale property of  $u_{i,t}$ . Further,

$$E[Q'_{i,T} Q_{i,T}] = E \left[ \left( \frac{1}{T} \sum_{t=1}^T u_{i,t} z'_{i,t-1} \right)' \left( \frac{1}{T} \sum_{t=1}^T u_{i,t} z'_{i,t-1} \right) \right] = \Phi_{ux,T}$$

where  $\Phi_{ux,T} \rightarrow \Phi_{ux}$  as  $T \rightarrow \infty$ . The  $Q_{i,T}$  are thus *iid* sequences with mean zero and covariance  $\Phi_{ux,T}$ . Conditions (i), (ii), and (iv) of Theorem 7 are again satisfied and it remains to show that  $\|Q_{i,T}\|^2$  is U.I. By the CMT,

$$\|Q_{i,T}\|^2 \Rightarrow \|Q_i\|^2 = \left\| \int_0^1 dB_{i,1} J_{C_i}(r)' \right\|^2,$$

and

$$\begin{aligned}
E\|Q_{i,T}\|^2 &= \text{tr}(E[Q'_{i,T} Q_{i,T}]) = \text{tr}(\Phi_{ux,T}) \rightarrow \text{tr}(\Phi_{ux}) \\
&= \text{tr}(Q'_i Q_i) = E\|Q_i\|^2.
\end{aligned}$$

It follows that  $\|Q_{i,T}\|^2$  is U.I. Since  $E[P_{i,T}] = 0$ ,

$$\begin{aligned}
E \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n P_{i,T} \right\|^2 &= \text{tr} (E [P'_{i,T} P_{i,T}]) \\
&= \text{tr} \left( \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T E [u_{i,t} u_{i,s} R_{i,t-1} R'_{i,s-1}] \right) \\
&= \text{tr} \left( \frac{1}{T^2} \sum_{t=1}^T E [u_{i,t} u_{i,t}] E [R_{i,t-1} R'_{i,t-1}] \right) \\
&\leq \text{tr} \left( \frac{1}{T^2} \sum_{t=1}^T \sigma_{11i} M \right) \\
&= O \left( \frac{1}{T} \right).
\end{aligned}$$

where the inequality follows from the martingale property of  $u_{i,t}$ . ■

**Proof of Lemma 5.** The results follow directly from Lemma 4. ■

**Proof of Lemma 6.**

(a) By definition,

$$\begin{aligned}
\hat{u}_{i,t} &= y_{i,t} - \hat{\beta}_{n,T} x_{i,t-1} \\
&= (\beta + \theta_i) x_{i,t-1} + u_{i,t} - \left( \beta + O_p \left( \frac{1}{\sqrt{n}} \right) \right) x_{i,t-1} \\
&= \theta_i x_{i,t-1} + u_{i,t} + O_p \left( \frac{1}{\sqrt{n}} \right) x_{i,t-1}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\hat{\Phi}_{xx}^\theta &= \frac{1}{n} \sum_{i=1}^n \frac{1}{T^4} \sum_{t=1}^T \sum_{s=1}^T (\hat{u}_{i,t} x'_{i,t-1})' (\hat{u}_{i,s} x'_{i,s-1}) \\
&= \frac{1}{n} \sum_{i=1}^n \frac{1}{T^4} \sum_{t=1}^T \sum_{s=1}^T \left( \theta_i x_{i,t-1} x'_{i,t-1} + u_{i,t} x'_{i,t-1} + O_p(n^{-1/2}) x_{i,t-1} x'_{i,t-1} \right)' \\
&\quad \times \left( \theta_i x_{i,s-1} x'_{i,s-1} + u_{i,s} x'_{i,s-1} + O_p(n^{-1/2}) x_{i,s-1} x'_{i,s-1} \right) \\
&= \frac{1}{n} \sum_{i=1}^n \frac{1}{T^4} \sum_{t=1}^T \sum_{s=1}^T \left( x_{i,t-1} x'_{i,t-1} \theta_i' + x_{i,t-1} u_{i,t} + x_{i,t-1} x'_{i,t-1} O_p(n^{-1/2}) \right)' \\
&\quad \times \left( \theta_i x_{i,s-1} x'_{i,s-1} + u_{i,s} x'_{i,s-1} + O_p(n^{-1/2}) x_{i,s-1} x'_{i,s-1} \right) \\
&= \frac{1}{n} \sum_{i=1}^n \frac{1}{T^4} \sum_{t=1}^T \sum_{s=1}^T Q_{i,T} + \frac{1}{n} \sum_{i=1}^n \frac{1}{T^4} \sum_{t=1}^T \sum_{s=1}^T (P_{1,i,T} + \dots + P_{8,i,T})
\end{aligned}$$

where

$$Q_{i,T} = x_{i,t-1} x'_{i,t-1} \theta_i' \theta_i x_{i,s-1} x'_{i,s-1},$$

and

$$\begin{aligned}
P_{1,i,T} &= x_{i,t-1} x'_{i,t-1} \theta_i' u_{i,s} x'_{i,s-1}, \\
P_{2,i,T} &= x_{i,t-1} x'_{i,t-1} \theta_i' O_p(n^{-1/2}) x_{i,s-1} x'_{i,s-1}, \\
P_{3,i,T} &= x_{i,t-1} u_{i,t} \theta_i x_{i,s-1} x'_{i,s-1}, \\
P_{4,i,T} &= x_{i,t-1} u_{i,t} u_{i,s} x'_{i,s-1}, \\
P_{5,i,T} &= x_{i,t-1} u_{i,t} O_p(n^{-1/2}) x_{i,s-1} x'_{i,s-1}, \\
P_{6,i,T} &= x_{i,t-1} x'_{i,t-1} O_p(n^{-1/2}) \theta_i x_{i,s-1} x'_{i,s-1}, \\
P_{7,i,T} &= x_{i,t-1} x'_{i,t-1} O_p(n^{-1/2}) u_{i,s} x'_{i,s-1}, \\
P_{8,i,T} &= x_{i,t-1} x'_{i,t-1} O_p(n^{-1/2}) O_p(n^{-1/2}) x_{i,s-1} x'_{i,s-1}.
\end{aligned}$$

The proof now proceeds in a manner analogous to those above and is not detailed here; likewise for part (b). ■

## References

Campbell, J., and M. Yogo, 2003. Efficient Tests of Stock Return Predictability, Working Paper, Harvard University.

Cavanagh, C., G. Elliot, and J. Stock, 1995. Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131-1147.

Elliot, G., 1998. On the Robustness of Cointegration Methods When Regressors Almost Have Unit Roots, *Econometrica* 66, 149-158;

Goetzman W.N., and P. Jorion, 1993. Testing the Predictive Power of Dividend Yields, *Journal of Finance* 48, 663-679.

Hjalmarsson, E., 2004. On the Predictability of Global Stock Returns, Mimeo, Yale University.

Jansson, M., and M.J. Moreira, 2004. Optimal Inference in Regression Models with Nearly Integrated Regressors, Working Paper, Harvard University.

Jin, S., 2004. Discrete Choice Modeling with Nonstationary Panels Applied to Exchange Rate Regime Choice, Mimeo, Yale University.

Lewellen, J., 2003. Predicting returns with financial ratios, *Journal of Financial Economics*, forthcoming.

Mankiw, N.G., and M.D. Shapiro, 1986. Do we reject too often? Small sample properties of tests of rational expectations models, *Economic Letters* 20, 139-145.

Moon, H.R., and P.C.B. Phillips, 1999. Estimation of Autoregressive Roots near Unity using Panel Data, Cowles Foundation Discussion Paper 1224.

Nelson, C.R., and M.J. Kim, 1993. Predictable Stock Returns: The Role of Small Sample Bias, *Journal of Finance* 48, 641-661.

Pagan, A., and A. Ullah, 1999. *Nonparametric Econometrics*, Cambridge University Press.

Phillips, P.C.B, and B. Hansen, 1990. Statistical Inference in Instrumental Variables Regression with I(1) Processes, *Review of Economic Studies* 57, 99-125.

Phillips, P.C.B., and H.R. Moon, 1999. Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica* 67, 1057-1111.

Phillips, P.C.B., and V. Solo, 1992. Asymptotics for Linear Processes, *Annals of Statistics*, 20, 971-1001.

Polk, C., S. Thompson, and T. Vuolteenaho, 2004. Cross-sectional Forecasts of the Equity Premium, Working Paper, Department of Economics, Harvard University.

Stambaugh, R., 1999. Predictive regressions, *Journal of Financial Economics* 54, 375-421.

Sul, D., P.C.B. Phillips, and C.Y. Choi, 2003. Prewhitening bias in HAC estimation, Cowles Foundation Discussion Paper 1436.

Table 1: Bias results from the Monte Carlo study. Panel A shows the average bias of the estimator based on recursive demeaning, the fully modified pooled estimator, and the plain fixed effects estimator. Panel B shows the corresponding results for the standard deviation of the bias.

Estimator	$\delta = 0.0$	$\delta = -0.4$	$\delta = -0.7$	$\delta = -0.95$
Panel A: Mean of bias				
$\tilde{\beta}_{n,T}^{rd}$	-0.00082	-0.00060	0.00069	0.00171
$\hat{\beta}_{PFM}$	-0.00017	0.00061	0.00079	0.00095
$\tilde{\beta}_{n,T}$	-0.00019	0.01152	0.01992	0.02696
Panel B: Std. dev. of bias				
$\tilde{\beta}_{n,T}^{rd}$	0.0506	0.0474	0.0297	0.0499
$\hat{\beta}_{PFM}$	0.0158	0.0157	0.0148	0.0138
$\tilde{\beta}_{n,T}$	0.0158	0.0161	0.0162	0.0165



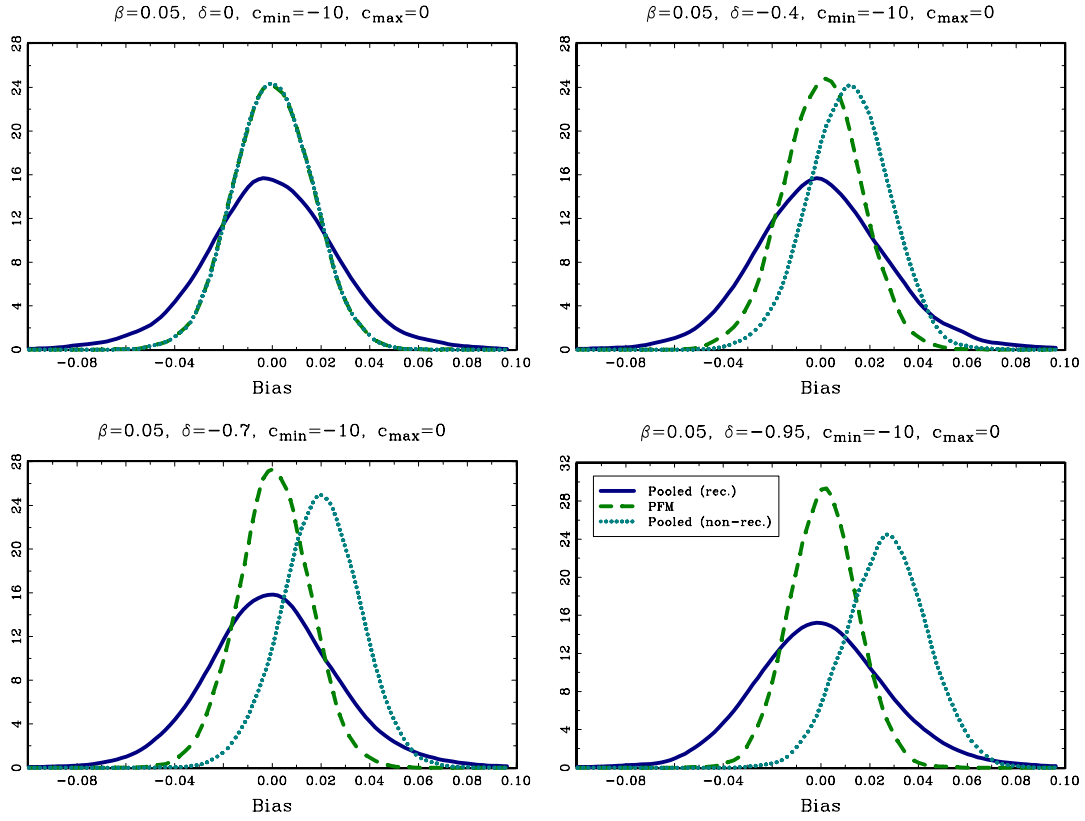


Figure 1: Estimation results from the Monte Carlo study. The graphs show the kernel density estimates of the estimated slope coefficients. The automatic bandwidth selection rules described in Pagan and Ullah (1999) were used in the kernel density estimation. The solid lines, labeled Pooled (rec.) in the legend, show the results for the estimator based on recursive demeaning,  $\tilde{\beta}_{n,T}^{rd}$ ; the dashed lines, labeled PFM, show the results for the fully modified pooled estimator  $\hat{\beta}_{PFM}$ ; and the dotted lines, labeled Pooled (non-rec.), show the results for the standard pooled estimator with individual intercepts, i.e. the standard fixed effects estimator,  $\tilde{\beta}_{n,T}$ .

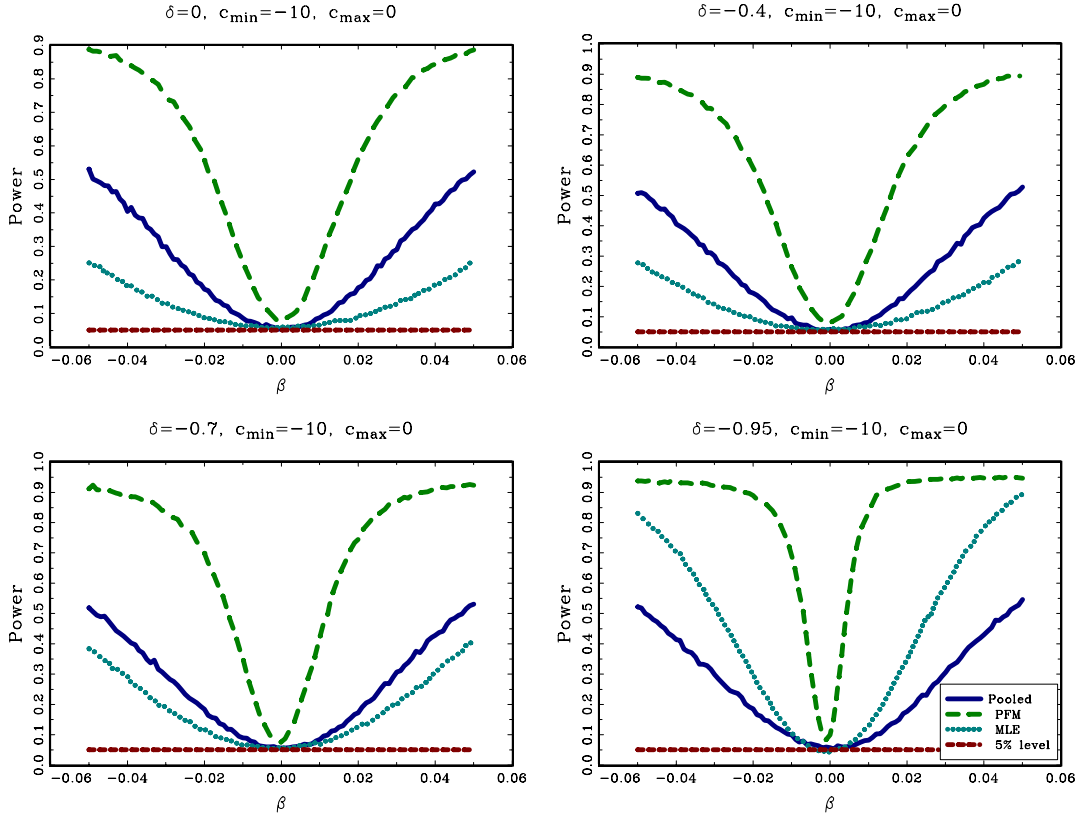


Figure 2: Size and power results from the Monte Carlo study. The graphs show the average rejection rates for a two-sided 5%  $t$ -test of the null hypothesis of  $\beta = 0$ . The  $x$ -axis shows the true value of the average parameter  $\beta$ , and the  $y$ -axis indicates the average rejection rate. The solid lines, labeled Pooled, give the results for the  $t$ -test corresponding to the estimator based on recursive demeaning,  $\tilde{\beta}_{n,T}^{rd}$ ; the dashed lines give the results for the  $t$ -test corresponding to the (unfeasible) fully modified pooled estimator,  $\hat{\beta}_{PFM}$ ; and the dotted lines give the results for the  $t$ -test corresponding to the (unfeasible) time-series maximum likelihood estimator described in Campbell and Yogo (2003). (They call this  $t$ -test the  $Q$ -test.) The flat lines indicate the 5% rejection rate.