Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 869

August 2006

Predictive Regressions with Panel Data

Erik Hjalmarsson

Predictive Regressions with Panel Data

Erik Hjalmarsson[*]

Abstract

This paper analyzes panel data inference in predictive regressions with endogenous and nearly persistent regressors. The standard fixed effects estimator is shown to suffer from a second order bias; analytical results, as well as Monte Carlo evidence, show that the bias and resulting size distortions can be severe. New estimators, based on recursive demeaning as well as direct bias correction, are proposed and methods for dealing with cross sectional dependence in the form of common factors are also developed. Overall, the results show that the econometric issues associated with predictive regressions when using time-series data to a large extent also carry over to the panel case. However, practical solutions are more readily available when using panel data. The results are illustrated with an application to predictability in international stock indices.

Keywords: Cross-sectional dependence; Panel data; Pooled regression; Predictive regression; Stock return predictability.

JEL classification: C22, C23, G1.

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).[1]

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 derivations of the asymptotic properties of pooled estimators in forecasting equations and the proposal of new procedures to deal with the bias effects arising from the persistence and endogeneity of the regressors. New results for controlling for the effects of common factors in panels are also derived. The methods developed in the paper are used to test for stock-return predictability in a panel of international stock returns.

By pooling the data, the econometric issues encountered in the time-series case can, to some extent,

---

[1] The literature on time-series forecasting regressions is very large. Some examples are Mankiw and Shapiro (1986), Nelson and Kim (1993), Goetzman and Jorion (1993), Cavanagh et al. (1995), Stambaugh (1999), Janson and Moreira (2004), Lewellen (2004), Polk et al. (2004), and Campbell and Yogo (2005). Many of these studies are primarily concerned with tests of stock-return predictability, although the results are generally applicable to more general forecasting regressions.

be dealt with more easily. Intuitively, persistent regressors cause no problems when they are exogenous. When pooling the data, independent cross-sectional information dilutes the endogeneity effects, and thus potentially alleviates the bias effects seen in the time-series case. 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.

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, 2000, and Sul et al., 2005). 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.

Although the estimator based on recursive demeaning is asymptotically normally distributed, it gives up some efficiency by disregarding parts of the data in the demeaning process. An alternative approach to control for the bias in the standard fixed effects estimator is to directly estimate the bias term and subtract it from the original estimator. Monte Carlo simulations show that such a correction works very well in practice, and produces unbiased estimators as well as correctly sized tests with good power. This bias corrected fixed effects estimator thus provides a simple and relatively efficient way of dealing with the bias and size distortions induced by the near persistence and endogeneity of the forecasting variables.

The overall conclusion from the theoretical results and the supporting Monte Carlo simulations is that, in the typical panel data case, persistent and endogenous regressors will cause standard inference

to be biased. In the time-series case, this result is well established and the results in this paper show that equal caution is required when working with panel data. However, unlike the time-series case, bias correction methods can be implemented in a relatively straightforward manner and normally distributed test-statistics can be achieved. Unbiased point estimates are also easily calculated, in contrast to the time-series case where the popular Bonferroni bound methods lead to correctly sized tests but not unique unbiased point estimates (Cavanagh et al., 1995, and Campbell and Yogo, 2005).

Another econometric issue that arises in the analysis of panels of financial or macroeconomic data is the potential presence of common factors. The standard panel data assumption of cross-sectional independence is often too restrictive, and I show how the framework in this paper can be extended to a setting where common factors are present in the data. These methods follow the work of Pesaran (2006) and extend his methods to a setting with nearly integrated regressors.

As an illustration of the methods derived in this paper, I consider the classical issue of stock return predictability. I use an international panel of returns from 18 different stock indices and the corresponding dividend- and earnings-price ratios, as well as the book-to-market values. The empirical results from the forecasting regressions with stock returns illustrate well the theoretical results derived in the paper. Based on the results from the standard fixed effects estimator, the evidence in favour of return predictability is very strong, using either of the three predictor variables. However, when using the robust methods developed here, the evidence disappears almost completely.

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. The finite sample properties of the procedures developed in this paper are analyzed through Monte Carlo experiments in Section 5. Some generalizations of the econometric model are considered in Section 6, and Section 7 contains the empirical application to stock return predictability. Section 8 concludes and technical proofs are found in the appendix.

Following the work of Phillips and Moon (1999), results for the panel estimators are derived 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 $(T, n \to \infty)_{\text{seq}}$. Subject to potential rate restrictions, such as $n/T \to 0$, these results can generally be shown to hold as n and T go to infinity jointly, denoted $(n, T \to \infty)$, by showing that the sufficient conditions in Phillips and Moon (1999) are satisfied; proofs of such joint convergence are

not pursued here, however. Otherwise, standard notation is used. $BM(\Omega)$ denotes a Brownian motion with covariance matrix Ω, $\Rightarrow$ 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 $y_{i,t}$, $i = 1, ..., n$, $t = 1, ..., T$, and the corresponding vector of regressors, $x_{i,t}$, where $x_{i,t}$ is an $m \times 1$ vector. The behavior of $y_{i,t}$ and $x_{i,t}$ are modelled as follows,

$$y_{i,t} = \alpha_i + \beta' x_{i,t-1} + \gamma'_i f_t + u_{i,t}, \tag{1}$$

$$x_{i,t} = A x_{i,t-1} + \Gamma'_i f_t + v_{i,t}, \tag{2}$$

where $A = I + C/T$ is an $m \times m$ matrix, and f_t is a $k \times 1$ vector capturing common factors in the error terms. The factor loadings γ_i $(k \times 1)$ and Γ_i $(k \times m)$ are treated as random coefficients, as specified below.

This model is a panel analogue of the time-series models studied by Mankiw and Shapiro (1986), Cavanagh et al. (1995), Stambaugh (1999), Jansson and Moreira (2004), Lewellen (2004), and Campbell and Yogo (2005).

Assumption 1 (Innovation processes) Let $w_{i,t} = (u_{i,t}, v_{i,t}, f_t)'$ and $\mathcal{F}_t = \{w_{i,s} | s \leq t, i = 1, ..., n\}$ be the filtration generated by the innovation processes. Then, for all $i = 1, ..., n$, and $t = 1, ..., T$,

1. $E[w_{it} | \mathcal{F}_{t-1}] = 0$.
2. $E[w_{i,t} w'_{i,t} | \mathcal{F}_{t-1}] = \Omega_i = [(\Omega_{uv,i}, 0), (0, \Omega_f)]'$ where $\Omega_{uv,i} = [(\omega_{11i}, \omega_{12i}), (\omega_{21i}, \Omega_{22i})]'$ and $\Omega = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \Omega_i$.
3. $E[u_{i,t}^4] < \infty$, $E[||v_{i,t}||^4] < \infty$, and $E[||f_{i,t}||^4] < \infty$.
4. $E[(u_{i,t}, v_{i,t})(u_{j,s}, v_{j,s})'] = 0$ for all t, s and $i \neq j$.

Assumption 2 (Factor loadings) The coefficients γ_i and Γ_i are iid across i and independently distributed of the specific errors, $u_{j,t}$ and $v_{j,t}$, and the common factors f_t, for all i, j, and t, with fixed means γ and Γ, and finite variances.

4

Assumption 3 (Rank condition) $Rank(\Gamma) = k$.

Assumption 1 specifies that the innovation processes follow a martingale difference sequence (mds) with finite fourth moments. The regressor can be endogenous in the sense that $u_{i,t}$ and $v_{i,t}$ may be contemporaneously correlated. The common factors, f_t, are assumed independent of the specific error components, and Assumption 2 specifies that the factor loadings are distributed independently of other random variables in the model. At the expense of some extra notation, the model could allow for a more general time-series structure in the innovation process $v_{i,t}$; the results in the paper would carry through with virtually no changes. The mds assumption for the errors in the dependent variables, $u_{i,t}$, is standard in predictive regressions, and 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 rank condition in Assumption 3 is used for identification in the estimation procedures that control for the common factors. It essentially states that all information regarding the common factors in the data can potentially be recovered from the innovation processes of the regressors. This condition turns out to be less restrictive than it seems, since it is the factors in the regressor errors that play the key role in the asymptotic properties of the pooled estimators. That is, a common factor that is only present in the dependent variable will not affect the analysis in any fundamental way. This is analyzed in more detail later in the paper.

It is also assumed that all the time-series in the panel share the same auto-regressive root, $A = I + C/T$. This assumption is imposed due to the presence of the common factors in the regressors, which would make it difficult to allow for heterogenous persistence. The effects of relaxing this assumption are briefly discussed later in the paper. Given a common auto-regressive root, the $x_{i,t}$ process can be expressed in a convenient component form,

$$x_{i,t} = x_{i,t}^0 + \Gamma_i' z_t, \quad x_{i,t}^0 = A x_{i,t-1}^0 + u_{i,t}, \quad z_t = A z_{t-1} + f_t. \tag{3}$$

Under Assumption 1, by standard arguments (Phillips and Solo, 1992), $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} w_{i,t} \Rightarrow B_i(r) = BM(\Omega_i)(r)$, where $B_i(\cdot) = (B_{1i}(\cdot), B_{2i}(\cdot), B_f(\cdot))'$ denote a $1+m+k-$dimensional Brownian motion. Further, by the results in Phillips (1987,1988), it follows that as $T \to \infty$, $\frac{x_{i,t}}{\sqrt{T}} = \frac{x_{i,t}^0}{\sqrt{T}} + \Gamma_i' \frac{z_{it}}{\sqrt{T}} \Rightarrow J_i(r) + \Gamma_i' J_f(r)$, where $J_i(r) = \int_0^r e^{(r-s)C} dB_{2,i}(s)$ and $J_f(r) = \int_0^r e^{(r-s)C} dB_f(s)$. Analogous results hold for the time-series demeaned data, $\underline{x}_{i,t} = x_{i,t} - \frac{1}{T} \sum_{t=1}^T x_{i,t}$, with J_i replaced by $\underline{J}_i = J_i - \int_0^1 J_i$; when there is no risk of confusion, the dependence of J_i, J_f and B_i on r will be suppressed. The

following lemma summarizes the key asymptotic results used in the paper.

Lemma 1 *Under Assumptions 1-2, as* $(T, n \to \infty)_{\text{seq}}$,

(a) $n^{-1/2} \sum_{i=1}^{n} T^{-1} \sum_{t=1}^{T} u_{i,t} x_{i,t-1} \Rightarrow MN(0, \Phi_{ux} + \Phi_{uz})$ *where* $\Phi_{ux} \equiv E\left[\left(\int_0^1 dB_{1,i} J_i\right)\left(\int_0^1 dB_{1,i} J_i\right)'\right]$,

$\Phi_{uz} \equiv \Gamma' E\left[\left(\int_0^1 dB_{1,i} J_f\right)\left(\int_0^1 dB_{1,i} J_f\right)'\Big| \mathcal{C}\right]\Gamma$, *and* $\mathcal{C}$ *is the* σ*−field generated by* $\{f_t\}_{t=1}^{\infty}$.

(b) $n^{-1} \sum_{i=1}^{n} T^{-1} \sum_{t=1}^{T} \gamma_i' f_t x_{i,t-1} \Rightarrow \int_0^1 (\gamma' dB_f)(\Gamma' J_f)$.

(c) $n^{-1} \sum_{i=1}^{n} T^{-2} \sum_{t=1}^{T} x_{i,t} x_{i,t}' \Rightarrow \Omega_{xx} + \Omega_{zz}$ *where* $\Omega_{xx} \equiv E\left[\int_0^1 J_i J_i'\right]$ *and* $\Omega_{zz} \equiv \Gamma'\left(\int_0^1 J_f J_f'\right)\Gamma$.

(d) $n^{-1} \sum_{i=1}^{n} T^{-1} \sum_{t=1}^{T} u_{i,t} \underline{x}_{i,t-1} \to_p -\left(\int_0^1 \int_0^r E\left[e^{(r-s)C}\right] dsdr\right) \omega_{21}$.

(e) $n^{-1} \sum_{i=1}^{n} T^{-1} \sum_{t=1}^{T} \gamma_i' \underline{f}_t \underline{x}_{i,t-1} \Rightarrow \int_0^1 (\gamma' dB_f)(\Gamma' \underline{J}_f)$.

(f) $n^{-1} \sum_{i=1}^{n} T^{-2} \sum_{t=1}^{T} \underline{x}_{i,t} \underline{x}_{i,t}' \Rightarrow \underline{\Omega}_{xx} + \underline{\Omega}_{zz}$ *where* $\underline{\Omega}_{xx} \equiv E\left[\int_0^1 \underline{J}_i \underline{J}_i'\right]$ *and* $\underline{\Omega}_{zz} \equiv \Gamma'\left(\int_0^1 \underline{J}_f \underline{J}_f'\right)\Gamma$.

3 Cross-sectional independence

3.1 The standard pooled estimator

To understand the basic properties of the pooled estimator of β, it is instructive to start with analyzing the case when there are no common factors in the data. That is, let $\gamma_i \equiv 0$ and $\Gamma_i \equiv 0$, for all i. To estimate the parameter β consider first the traditional pooled estimator when there are no individual effects, i.e. when $\alpha_i \equiv 0$ for all i.[2] The pooled estimator is given by

$$\hat{\beta}_{Pool} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} x_{i,t-1} x_{i,t-1}'\right)^{-1}\left(\sum_{i=1}^{n} \sum_{t=1}^{T} y_{i,t} x_{i,t-1}\right), \tag{4}$$

and the following theorem gives its asymptotic properties.

Theorem 1 *Under Assumptions 1 and 2, with* $\gamma_i \equiv 0$, $\Gamma_i \equiv 0$, *and* $\alpha_i \equiv 0$ *for all* i, *as* $(T, n \to \infty)_{\text{seq}}$,

$$\sqrt{n}T\left(\hat{\beta}_{Pool} - \beta\right) \Rightarrow N\left(0, \Omega_{xx}^{-1} \Phi_{ux} \Omega_{xx}^{-1}\right). \tag{5}$$

The pooled estimator of β is thus asymptotically normally distributed and the limiting distribution depends on Ω_{xx} and Φ_{ux}. To perform inference, estimates of these parameters are required. Let $\hat{u}_{i,t} = y_{i,t} - \hat{\beta}_{n,T} x_{i,t-1}$, $\hat{\Phi}_{ux} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} (\hat{u}_{i,t} x_{i,t-1})(\hat{u}_{i,s} x_{i,s-1})'$, and $\hat{\Omega}_{xx} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} x_{i,t-1} x_{i,t-1}'$. The estimator $\hat{\Phi}_{ux}$ is thus the panel equivalent of HAC estimators for long-run variances.

---

[2] The results developed below also hold in the case with a *common* non-zero intercept α.

6

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. Let $\hat{\Sigma} = \hat{\Omega}_{xx}^{-1}\hat{\Phi}_{ux}\hat{\Omega}_{xx}^{-1}$. Using the results derived above, it follows easily that under the null-hypothesis,

$$t_k = \frac{\hat{\beta}_{k,pool} - \beta_{k,0}}{\sqrt{a'\hat{\Sigma}a}} \Rightarrow N(0,1), \tag{6}$$

in sequential limits, as $(T, n \to \infty)_{\mathrm{seq}}$, 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.

Thus, when there are no fixed effects in the pooled regression, inference in the panel case becomes trivial since the pooled estimator is asymptotically normally distributed. This is in contrast with the time-series case where the OLS estimator has a non-normal asymptotic distribution which depends on unknown nuisance parameters.

3.2 Fixed effects

In the above analysis, the individual intercepts α_i 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 fixed effects pooled estimator, which allows for individual intercepts, is then given by

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\underline{x}_{i,t-1}\underline{x}'_{i,t-1}\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\underline{y}_{i,t}\underline{x}_{i,t-1}\right), \tag{7}$$

and

$$\hat{\beta}_{FE} - \beta = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\underline{x}_{i,t-1}\underline{x}'_{i,t-1}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\underline{u}_{i,t}\underline{x}_{i,t-1}\right). \tag{8}$$

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

$$T\left(\hat{\beta}_{FE} - \beta\right) \Rightarrow \left(\frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{J}_i\underline{J}'_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\int_0^1 dB_{1,i}\underline{J}_i\right). \tag{9}$$

Let $\omega_{21} = \lim_{n\to\infty} n^{-1} \sum \omega_{21i}$, and observe that

$$
E\left[\int_0^1 dB_{1,i} \underline{J}_i\right] = E\left[\int_0^1 dB_{1,i}(r) J_i(r) - \int_0^1 dB_{1,i}(s) \int_0^1 J_i(r) dr\right] \tag{10}
$$

$$
= -\int_0^1 \int_0^1 \int_0^r e^{(r-q)C} E\left[dB_{1,i}(s) dB_{2,i}(q)\right] ds dr
$$

$$
= -\left(\int_0^1 \int_0^r e^{(r-s)C} ds dr\right) \omega_{21}, \tag{11}
$$

which is different from zero whenever $\omega_{21} \neq 0$. Thus, as $(T, n \to \infty)_{\text{seq}}$,

$$
T\left(\hat{\beta}_{FE} - \beta\right) \to_p -\underline{\Omega}_{xx}^{-1}\left(\int_0^1 \int_0^r e^{(r-s)C} ds dr\right) \omega_{21}, \tag{12}
$$

and the estimator suffers from a second order bias from the demeaning process.

3.3 Recursive demeaning

The second order bias arises because the demeaning process induces a correlation between the inno-
vation processes $u_{i,t}$ and the demeaned regressors $\underline{x}_{i,t-1}$.[3] 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. At the expense of
some efficiency, one solution is therefore to use recursive demeaning of $x_{i,t}$ and $y_{i,t}$ (e.g. Moon and
Phillips, 2000, and Sul et al., 2005). That is, define

$$
\underline{x}_{i,t}^d = x_{i,t} - \frac{1}{t}\sum_{s=1}^t x_{i,s}, \quad \underline{y}_{i,t}^{dd} = y_{i,t} - \frac{1}{T-t}\sum_{s=t}^T y_{i,s}, \quad \text{and} \quad \underline{x}_{i,t}^{dd} = x_{i,t} - \frac{1}{T-t}\sum_{s=t}^T x_{i,s}. \tag{13}
$$

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 continuous mapping theorem,
as $T \to \infty$ for a fixed i,

$$
\frac{\underline{x}_{i,t}^d}{\sqrt{T}} = \frac{x_{i,t}}{\sqrt{T}} - \left(\frac{t}{T}\right)^{-1}\frac{1}{T}\sum_{s=1}^t \frac{x_{i,s}}{\sqrt{T}} \Rightarrow J_i(r) - r^{-1}\int_0^r J_i(u) du = \underline{J}_i^d(r), \tag{14}
$$

---

[3] The phenomenon is analogous to that found by Moon and Phillips (2000), in their estimation of local-to-unity roots
in panels with incidental trends.

and

$$\frac{x_{i,t}^{dd}}{\sqrt{T}} = \frac{x_{i,t}}{\sqrt{T}} - \left(\frac{T-t}{T}\right)^{-1} \frac{1}{T} \sum_{s=t}^{T} \frac{x_{i,s}}{\sqrt{T}} \Rightarrow J_i(r) - (1-r)^{-1} \int_r^1 J_i(u)\, du = \underline{J}_i^{dd}(r). \tag{15}$$

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

$$\hat{\beta}_{RD} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \underline{x}_{i,t-1}^{dd} \underline{x}_{i,t-1}^{d\prime}\right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \underline{y}_{i,t}^{dd} \underline{x}_{i,t-1}^{d}\right). \tag{16}$$

Theorem 2 *Under Assumptions 1 and 2, as* $(T, n \to \infty)_{\mathrm{seq}}$,

$$\sqrt{n}T\left(\hat{\beta}_{RD} - \beta\right) \Rightarrow N\left(0, \left(\underline{\Omega}_{xx}^{RD}\right)^{-1} \underline{\Phi}_{ux}^{RD} \left(\underline{\Omega}_{xx}^{RD}\right)^{-1}\right), \tag{17}$$

where $\underline{\Phi}_{ux}^{RD} = E\left[\left(\int_0^1 dB_{1,i}^{dd} \underline{J}_i^d\right)\left(\int_0^1 dB_{1,i}^{dd} \underline{J}_i^d\right)'\right]$, $\underline{\Omega}_{xx}^{RD} = E\left[\int_0^1 \underline{J}_i^{rr} \underline{J}_i^{r\prime}\right]$, *and* $dB_{1,i}^{dd}(r) = dB_{1,i}(r) - (1-r)^{-1}(B_{1,i}(1) - B_{1,i}(r))$.

To perform inference, let $\hat{u}_{i,t}^{dd} = \underline{y}_{i,t}^{dd} - \hat{\beta}_{RD} \underline{x}_{i,t-1}^{dd}$, $\hat{\underline{\Phi}}_{ux}^{RD} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\hat{u}_{i,t}^{dd} \underline{x}_{i,t-1}^{d}\right)\left(\hat{u}_{i,t}^{dd} \underline{x}_{i,s-1}^{d}\right)'$, and $\hat{\underline{\Omega}}_{xx}^{RD} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^2} \sum_{t=1}^{T} \underline{x}_{i,t-1}^{dd} \underline{x}_{i,t-1}^{d\prime}$. The $t-$test and Wald-test based on $\hat{\underline{\Phi}}_{ux}^{RD}$ and $\hat{\underline{\Omega}}_{xx}^{RD}$ will satisfy the usual properties and the results follow in the same manner as above.

3.4 Direct bias correction

The estimator $\hat{\beta}_{RD}$ gives up some efficiency through the recursive demeaning process. A more efficient approach would be to directly estimate the bias term in (12) and subtract it from the standard pooled estimator. A simple biased corrected estimator is given by

$$\hat{\beta}_{FE}^{+} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \underline{x}_{i,t-1} \underline{x}_{i,t-1}'\right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \underline{y}_{i,t} \underline{x}_{i,t-1} - nT\left(\int_0^1 \int_0^r e^{(r-s)\hat{C}} ds\, dr\right) \hat{\omega}_{21}\right). \tag{18}$$

Provided that ω_{21} and C are consistently estimated, it follows easily that as $(T, n \to \infty)_{\mathrm{seq}}$,

$$\sqrt{n}T\left(\hat{\beta}_{FE}^{+} - \beta\right) \Rightarrow N\left(0, \underline{\Omega}_{xx}^{-1} \underline{\Phi}_{ux} \underline{\Omega}_{xx}^{-1}\right), \tag{19}$$

where $\underline{\Phi}_{ux}$ is defined in an analogous manner to $\underline{\Omega}_{xx}$, and estimates of $\underline{\Phi}_{ux}$ and $\underline{\Omega}_{xx}$ are obtained in an identical manner as before.

An estimate of ω_{21} is easy to obtain by averaging the estimates of ω_{21i} obtained from time-series

regressions. As discussed below, it is possible to consistently estimate C using panel data, unlike in the time-series case.

3.5 Estimation of C

The estimator proposed above relies on an estimate of C. Moon and Phillips (2000) show how C can be consistently estimated in equation (2) when C is a scalar. The diagonal of C can, of course, be estimated by the individual univariate estimates; thus, if one restricts C to be diagonal, then an estimate of a matrix C can be obtained using the methods of Moon and Phillips (2000).

Consider the case of a scalar C, and assume that $x_{i,t}$ is generated according to equation (2). Noting that $A = 1 + \frac{C}{T}$, it is natural to consider estimators of the form $\hat{C} = T\left(\hat{A} - 1\right)$. The pooled estimator of A is given by,

$$\hat{A} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{i,t-1}^2\right)^{-1} \left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{i,t}x_{i,t-1}\right),$$

and the corresponding pooled estimator of C is $\hat{C} = T\left(\hat{A} - 1\right)$. Moon and Phillips (2000) show that this estimator of C is consistent in the absence of cross-sectional dependence. Observe that the data used in estimating C is not time-series demeaned; demeaning the data in the time-series dimension will lead to a bias in the estimator.

It is beyond the scope of this paper to formally consider the effects of common factors in the data on the estimator $\hat{C}$. However, Monte Carlo simulations not reported in the paper indicate that it also remains unbiased in the presence of common factors.

4 Cross-sectional dependence

4.1 The effects of common factors

I now return to the general setup with common factors in the data. The following theorem summarizes the asymptotic properties of the standard pooled estimator, as well as the fixed effects estimator, when there are common factors.

Theorem 3 (a) Under Assumptions 1-2, with $\alpha_i \equiv 0$, as $(T, n \to \infty)_{\text{seq}}$,

$$T\left(\hat{\beta}_{Pool} - \beta\right) \Rightarrow (\Omega_{xx} + \Omega_{zz})^{-1}\left[\int_0^1 (\gamma' dB_f)(\Gamma' J_f)\right]. \tag{20}$$

10

(b) Under Assumptions 1-2, as $(T, n \to \infty)_{\text{seq}}$,

$$T\left(\hat{\beta}_{FE} - \beta\right) \Rightarrow (\underline{\Omega}_{xx} + \underline{\Omega}_{zz})^{-1} \left[\int_0^1 (\gamma' dB_f) \left(\Gamma' \underline{J}_f\right) - \left(\int_0^1 \int_0^r e^{(r-s)C} ds dr\right) \omega_{21}\right]. \tag{21}$$

Thus, in the presence of the general factor structure outlined in Assumptions 1 and 2, the standard pooled estimator exhibits a non-standard limiting distribution, although it is still consistent; standard tests can therefore not be used. Similarly, the limiting behavior of the fixed effects estimator is determined by the bias term arising from the time-series demeaning of the data, as well as an additional term that stems from the common factors in the data.

4.2 Robust estimators

Based on the methods of Pesaran (2006), I propose an estimator that is more robust to cross-sectional dependence in the data. Write the model in matrix form,

$$\underset{T \times 1}{\mathbf{Y}_i} = \underset{T \times m}{\mathbf{X}_{i,-1}} \underset{m \times 1}{\beta} + \underset{T \times k}{\mathbf{f}} \underset{k \times 1}{\gamma_i} + \underset{T \times 1}{\mathbf{u}_i}, \tag{22}$$

$$\underset{T \times m}{\mathbf{X}_i} = \underset{T \times m}{\mathbf{X}_{i,-1}} \underset{m \times m}{A} + \underset{T \times k}{\mathbf{f}} \underset{k \times m}{\Gamma_i} + \underset{T \times m}{\mathbf{v}_i}, \tag{23}$$

where $\mathbf{Y}_i$ denotes the $T \times 1$ matrix of the observations for the dependent variable and $\mathbf{X}_i$ the $T \times m$ matrix of regressor observations. The $T \times k$ matrix $\mathbf{f}$ denotes the unobserved common factors.

The idea of Pesaran (2006) is to project the data onto the space orthogonal to the common factors, thereby removing the cross sectional dependence from the data used in the estimation. However, since the factors in f_t are not observed in practice, an indirect approach is required.

Consider the following estimator of β,

$$\tilde{\beta}_{Pool} = \left(\sum_{i=1}^n \mathbf{X}'_{i,-1} \mathbf{M}_{\bar{\mathbf{H}}} \mathbf{X}_{i,-1}\right)^{-1} \left(\sum_{i=1}^n \mathbf{X}'_{i,-1} \mathbf{M}_{\bar{\mathbf{H}}} \mathbf{Y}_i\right) \tag{24}$$

where $\mathbf{M}_{\bar{\mathbf{H}}} = \mathbf{I} - \bar{\mathbf{H}} \left(\bar{\mathbf{H}}' \bar{\mathbf{H}}\right)^{-1} \bar{\mathbf{H}}'$ is a $T \times T$ matrix and $\bar{\mathbf{H}}$ is the $T \times 2m$ matrix of observations of $\bar{H}_t$, where

$$\bar{H}_t = \frac{1}{n} \sum_{i=1}^n H_{i,t} = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \Delta_C x_{i,t} \\ x_{i,t-1} \end{pmatrix} = \begin{pmatrix} \Delta_C \bar{x}_{\cdot,t} \\ \bar{x}_{\cdot,t-1} \end{pmatrix}. \tag{25}$$

Here, Δ_C denotes the quasi-differencing operator and $\Delta_C x_{i,t} = x_{i,t} - (I + C/T) x_{i,t-1} = \Gamma'_i f_t + v_{i,t}$.

Since $\mathbf{X}_i = \mathbf{X}_i^0 + \mathbf{Z}\Gamma_i$, it follows that

$$\bar{\mathbf{H}} = \left(\begin{array}{cc} \boldsymbol{\Delta}_C\bar{\mathbf{X}} & \bar{\mathbf{X}}_{\cdot,-1} \end{array} \right) = \left(\begin{array}{cc} \mathbf{f}\bar{\Gamma} + \bar{\mathbf{v}} & \bar{\mathbf{X}}_{\cdot,-1}^0 + \mathbf{Z}_{-1}\bar{\Gamma} \end{array} \right). \tag{26}$$

The estimator $\tilde{\beta}_{Pool}$ is thus obtained by applying the pooled estimator to the residuals from a projection of the original data onto the cross-sectional averages of the regressors and the innovations in the regressors. The intuition behind this is that the cross-sectional averages of $\Delta_C x_{i,t}$ and $x_{i,t}$ are close to the innovations in the common factors, f_t, and the common stochastic trend z_t, respectively, since the cross-sectional averages of the cross-sectionally independent data may be expected to be close to zero. In practice, a panel estimate of C is needed to quasi-difference the data, but since that will not affect the asymptotic properties of the estimator, we ignore this in the analysis below.

To form a better understanding behind the functioning of the above estimator, let $\mathbf{Q} = \left(\begin{array}{cc} \mathbf{f}\Gamma & \mathbf{Z}_{-1}\Gamma \end{array} \right)$, and $\mathbf{M_Q} = \mathbf{I} - \mathbf{Q}\left(\mathbf{Q'Q}\right)^{-1}\mathbf{Q'}$. Also, let $\mathbf{G} = \left(\begin{array}{cc} \mathbf{f} & \mathbf{Z}_{-1} \end{array} \right)$ and $\mathbf{M_G} = \mathbf{I} - \mathbf{G}\left(\mathbf{G'G}\right)^{-1}\mathbf{G'}$. Observe that,

$$\sqrt{n}T\left(\tilde{\beta}_{Pool} - \beta\right) = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{X}_{i,-1}'\mathbf{M_{\bar{H}}}\mathbf{X}_{i,-1}}{T^2}\right)^{-1}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{X}_{i,-1}'\mathbf{M_{\bar{H}}}\mathbf{f}\gamma_i}{T}\right) + \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{X}_{i,-1}'\mathbf{M_{\bar{H}}}\mathbf{u}_i}{T}\right)\right]. \tag{27}$$

Since $\mathbf{f} \subset \mathbf{G}$, $\mathbf{M_G}\mathbf{f} = \mathbf{0}$, and under the rank condition on Γ in Assumption 3, it follows that $\mathbf{M_Q}\mathbf{f} = \mathbf{M_G}\mathbf{f}$. Also, since $\mathbf{Z}_{-1} \subset \mathbf{G}$, $\mathbf{M_G}\mathbf{X}_{i,-1} = \mathbf{X}_{i,-1}^0$. Thus, to the extent that $\mathbf{M_{\bar{H}}}$ is close to $\mathbf{M_Q}$, the projection onto the compliment of the cross-sectional means will remove most of the effects of the common factors in the data. The following theorem states this formally.

Theorem 4 *Under Assumptions 1-3, with $\alpha_i \equiv 0$, as $(T, n \to \infty)_{\text{seq}}$, with $n/T \to 0$,*

$$\sqrt{n}T\left(\tilde{\beta}_{Pool} - \beta\right) \Rightarrow MN\left(0, \Omega_{xx}^{-1}E\left[\left(\int_0^1 J_{i\cdot f\Gamma}dB_{1,i}\right)\left(\int_0^1 J_{i\cdot f\Gamma}dB_{1,i}\right)'\Bigg| \mathcal{C}\right]\Omega_{xx}^{-1}\right), \tag{28}$$

where $J_{i\cdot f\Gamma} = J_i - \left(\int_0^1 J_i J_f'\Gamma\right)\left(\Gamma'\int_0^1 J_f J_f'\Gamma\right)^{-1}\left(\int_0^1 \Gamma'J_f\right)$ is the residual from the orthogonal projection of J_i onto $\Gamma'J_f$.

The estimator $\tilde{\beta}_{Pool}$ thus achieves a $\sqrt{n}T-$convergence rate and an asymptotic mixed normal distribution. The mixed normality in this case arises from the common factors, which leads to a mixed normal rather than a normal distribution as in the case above with no common factors; a similar result

is also noted in Jin (2004). Andrews (2005) provides an extensive discussion on convergence with common shocks. This theorem also extends the results for stationary data in Pesaran (2006) to the nearly integrated case, although using a somewhat different definition of the $\bar{H}_t$ matrix. Allowing for fixed effects in the arguments, it is easy to show the following result.

Corollary 1 *Let*

$$\tilde{\beta}_{FE} = \left(\sum_{i=1}^{n} \underline{\mathbf{X}}'_{i,-1} \mathbf{M}_{\bar{\mathbf{H}}} \underline{\mathbf{X}}_{i,-1} \right)^{-1} \left(\sum_{i=1}^{n} \underline{\mathbf{X}}'_{i,-1} \mathbf{M}_{\bar{\mathbf{H}}} \underline{\mathbf{Y}}_i \right), \tag{29}$$

and

$$\tilde{\beta}_{FE}^{+} = \left(\sum_{i=1}^{n} \underline{\mathbf{X}}'_{i,-1} \mathbf{M}_{\bar{\mathbf{H}}} \underline{\mathbf{X}}_{i,-1} \right)^{-1} \left(\sum_{i=1}^{n} \underline{\mathbf{X}}'_{i,-1} \mathbf{M}_{\bar{\mathbf{H}}} \underline{\mathbf{Y}}_i - nT \left(\int_0^1 \int_0^r e^{(r-s)\hat{C}} ds dr \right) \hat{\omega}_{21} \right) \tag{30}$$

where $\underline{\mathbf{X}}_{i,-1}$ and $\underline{\mathbf{Y}}_i$ represent the time-series demeaned data. Then, under Assumptions 1-3, as $(T, n \to \infty)_{\text{seq}}$, with $n/T \to 0$,

$$T \left(\tilde{\beta}_{FE} - \beta \right) \to_p -\underline{\Omega}_{xx}^{-1} \left(\int_0^1 \int_0^r e^{(r-s)C} ds dr \right) \omega_{21}, \tag{31}$$

and

$$\sqrt{n} T \left(\tilde{\beta}_{FE}^{+} - \beta \right) \Rightarrow MN \left(0, \underline{\Omega}_{xx}^{-1} E \left[\left(\int_0^1 \underline{J}_{i \cdot f\Gamma} dB_{1,i} \right) \left(\int_0^1 \underline{J}_{i \cdot f\Gamma} dB_{1,i} \right)' \middle| C \right] \underline{\Omega}_{xx}^{-1} \right). \tag{32}$$

The fixed effects transformation thus has an identical bias effect on the estimator that controls for common factors, and can also be corrected in an identical manner. Similarly, it could also be shown that using recursive demeaning on the data projected onto the cross-sectional means would be asymptotically mixed normally distributed, although these results are omitted here. $\tilde{\beta}_{Pool}$ and $\tilde{\beta}_{FE}$ thus provide pooled estimators for predictive regressions that are asymptotically mixed normally distributed in the presence of common factors and with the allowance for fixed effects. Standard $t-$tests and Wald-tests can therefore be used; the variance-covariance matrix of $\tilde{\beta}_{FE}$ or $\tilde{\beta}_{Pool}$ can be estimated in an analogous manner as described above for the case with no common factors, by simply using the defactored data. The practical implementation of $\tilde{\beta}_{FE}^{+}$, or $\tilde{\beta}_{Pool}$, is thus very simple: Premultiply the data by $\mathbf{M}_{\bar{\mathbf{H}}}$, and use the resulting variables in the original procedures for $\hat{\beta}_{FE}^{+}$ and $\hat{\beta}_{Pool}$. Note that this also automatically facilitates the correct estimation of ω_{21}, which now represents the correlation between the cross-sectionally independent errors, $u_{i,t}$ and $v_{i,t}$, rather than the correlation between the total innovation processes in the regressors and regressand.

4.3   More general factor structures

The key assumption in deriving the results in Theorem 4 was the rank condition in Assumption 3. It essentially allows for the whole factor structure to be revealed through the innovations in the regressor variables alone. While this is a convenient assumption, since it does not require first stage estimates of the innovations in the regressand, it is also potentially limiting since it effectively restricts the factors in the dependent variable to be a subset of the factors in the regressors. However, it is easy to show that additional factors in the regressand do not fundamentally alter the above results.

Consider the following model, which is a generalization of the original model in equations (1) and (2),

$$y_{i,t} = \alpha_i + \beta' x_{i,t-1} + \gamma_i' f_t + \eta_i' g_t + u_{i,t}, \tag{33}$$

$$x_{i,t} = A x_{i,t-1} + \Gamma_i' f_t + v_{i,t}, \tag{34}$$

where g_t is a $l \times 1$ vector of additional common factors that only appear in the dependent variable, and η_i are the corresponding factor loadings. Assume that g_t and η_i are independent of all other random elements in the model and satisfy the same conditions as f_t and γ_i.

Corollary 2 *Suppose the data is generated according to equations (33) and (34) with $\alpha_i \equiv 0$, and that Assumptions 1-3 hold. Then, as $(T, n \to \infty)_{\mathrm{seq}}$, with $n/T \to 0$,*

$$\sqrt{n} T \left(\tilde{\beta}_{Pool} - \beta \right) \Rightarrow MN \left(0, \Omega_{xx}^{-1} E \left[\left(\int_0^1 J_{i \cdot f\Gamma} \left(dB_{1,i} + \eta' dB_g \right) \right) \left(\int_0^1 J_{i \cdot f\Gamma} \left(dB_{1,i} + \eta' dB_g \right) \right)' \middle| \mathcal{C} \right] \Omega_{xx}^{-1} \right), \tag{35}$$

where B_g is the Brownian motion such that $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} g_t \Rightarrow B_g(r)$ as $T \to \infty$, and $\mathcal{C}$ now represents the $\sigma-$field generated by $\{f_t, g_t\}_{t=1}^{\infty}$.

The limiting distribution changes but remains mixed normal and inference can be performed in an identical manner using standard test-statistics. The procedures derived in the previous subsection are thus robust to additional factors in the dependent variable.

5   Finite sample evidence

5.1   No cross-sectional dependence

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 parameter β is set equal to 0.05 and the local-to-unity parameter C is set to -10. The sample size is given by $T = 100$, $n = 20$. The small value of β 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 α_i are set equal to zero but individual effects are still fitted for all the estimators, except the standard pooled one, in order to evaluate the second-order bias effects arising from demeaning. All results are based on 10,000 repetitions.

Four different estimators are considered: the pooled estimator with no fixed effects, $\hat{\beta}_{Pool}$, the fixed effects estimator using standard demeaning, $\hat{\beta}_{FE}$, the recursively demeaned pooled estimator, $\hat{\beta}_{RD}$, and the bias corrected estimator $\hat{\beta}_{FE}^+$. The bias correction in the estimator $\hat{\beta}_{FE}^+$ is estimated by $-\left(\int_0^1 \int_0^r e^{(r-s)\hat{C}} ds\, dr\right) \hat{\omega}_{21}$, where $\hat{C}$ is the panel estimate of the local-to-unity parameter and $\hat{\omega}_{21}$ is estimated as $n^{-1} \sum_{i=1}^n \hat{\omega}_{21i}$ with $\hat{\omega}_{21i}$ the covariance between the residuals from a time-series estimation of equation (1) and the quasi-differenced regressors, $\Delta_{\hat{C}} x_{i,t}$. In general, the standard pooled estimator does not work well when the $\alpha_i s$ differ across i, but is used as a comparison here.

The results are shown in Figure 1. All estimators, except $\hat{\beta}_{FE}$, are virtually unbiased. The estimator $\hat{\beta}_{FE}$, which uses standard demeaning to account for individual effects, exhibits a rather substantial bias when the absolute value of the correlation δ is large. The recursively demeaned estimator, $\hat{\beta}_{RD}$, 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, in order to calculate the power of the tests, the slope coefficient β now varies between -0.05 and 0.05. Figure 2 shows the average rejection rates of the 5 percent two-sided $t-$tests, evaluating a null of $\beta = 0$; that is, the power curves of the tests. Panel A in Table 1 shows the average sizes of the nominal 5 percent tests under the null hypothesis of $\beta = 0$ for the two sided $t-$tests corresponding to the four different estimators considered above. Again, the results are based

on 10,000 repetitions.

Apart from the test based on the standard fixed effects estimator, all tests perform well in terms of size, although they all tend to over reject the null hypothesis somewhat. Table 1 and the power curves in Figure 2 clearly show the effects of the second order bias in the fixed effects estimator. The three other tests all exhibit decent power properties although the test based on $\hat{\beta}_{RD}$ has lower power than the bias corrected estimator.

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 appears to do well and results in test-statistics with correct size and decent power properties. The bias correction of the fixed effects estimator also appears to work well, producing nearly unbiased results and correctly sized tests with good power. Overall, the simulations confirm the analytical results previously derived.

5.2   Common factors

In this section, I repeat the Monte Carlo experiments above, with the exception that there is now a common factor in the innovations. In particular, equations (1) and (2) are now simulated with a single regressor and a single common factor f_t, drawn from a standard normal distribution. The factor loadings, γ_i and Γ_i, are also normally distributed with means of minus one and plus one, respectively, and standard deviations equal to $2^{-1/2}$ in both cases. The innovations in the returns and regressor processes are formed as $2^{-1/2}\left(\gamma_i' f_t + u_{i,t}\right)$ and $2^{-1/2}\left(\Gamma_i' f_t + v_{i,t}\right)$, respectively, where $(u_{i,t}, v_{i,t})$ are drawn from standard normal distributions; the scaling by $2^{-1/2}$ is performed in order to achieve an approximate unit variance in the innovations which enables easier comparison with the cross-sectionally independent case. As before, the correlation between $u_{i,t}$ and $v_{i,t}$ is set to $\delta = 0, -0.4, -0.7$, and -0.95. Note that δ no longer represents the overall correlation between the innovations, but rather that between the cross-sectionally independent parts of the innovations. In addition, I allow α_i to vary across i according to a normal distribution with mean and standard deviation equal to 0.05. Otherwise, the setup is identical to that used in the case with no common factors. Again, all results are based on 10, 000 repetitions.

The results are shown in Panels B and C of Table 1 and in Figures 3-6. Panel B in Table 1 and Figures 3 and 4 show the outcomes of the Monte Carlo experiments when the model generated with

common factors is estimated using the standard estimators $\hat{\beta}_{FE}, \hat{\beta}_{RD}$, and $\hat{\beta}_{FE}^{+}$, that do not control for cross-sectional dependence; since α_i now varies across the panel, the standard pooled estimator without fixed effects is not considered. Figure 3 shows that $\hat{\beta}_{RD}$ and $\hat{\beta}_{FE}^{+}$ are still fairly unbiased, in accordance with the asymptotic result in Theorem 3, but they are much more variable than in the case with no common factors and exhibit a non-normal distribution. The real downside from not controlling for the cross-sectional dependence is seen in the size of the $t-$tests displayed in Panel B of Table 1 and in the power curves in Figure 3. It is clear that when the common factors are ignored in the estimation process, the actual size of the corresponding $t-$tests is very far from the nominal size of 5 percent, with rejection rates between 30 and 40 percent under the null.

Panel C in Table 1 and Figures 5 and 6 show the same results for the estimators $\tilde{\beta}_{FE}, \tilde{\beta}_{RD}$, and $\tilde{\beta}_{FE}^{+}$, which control for the common factors; $\tilde{\beta}_{RD}$ represents the estimator with recursive demeaning using defactored data. The distributions of the point estimates, as seen in Figure 5, are now much better behaved and normal looking, as well as much more peaked than in Figure 3. The $t-$tests also possess better size and power properties, although there is a slight tendency to over-reject when δ is large in absolute numbers. As before, the standard fixed effects estimator exhibits a finite sample bias and extremely poor size properties.

In summary, the estimator $\tilde{\beta}_{RD}$, and $\tilde{\beta}_{FE}^{+}$, which were designed to control for cross-sectional dependence, appear to work reasonably well in small samples, although there is a small bias and over-rejection when δ is large. Compared to the estimators that do not take the common factors into account, there is a huge improvement, however.

6 Heterogenous auto-regressive roots

One dimension along which the model in this paper can be generalized is to allow the auto-regressive root A to vary across i, such that each time-series i possesses a root $A_i = I + C_i/T$, where C_i may be modelled as a random variable. Most of the analysis will carry through in this setting, with the limiting expressions being expectations also over the $C_i s$. Problems arise in the bias correcting exercises, however. For instance, the bias-corrected fixed effects estimator, $\hat{\beta}_{FE}^{+}$, relies on an estimate of C, but when the $C_i s$ are no longer identical, panel methods cannot be used to form estimates of the individual $C_i s$. In particular, an estimate of $E\left[e^{(r-s)C_i}\right]$ is needed. An approximate estimate of $E\left[e^{(r-s)C_i}\right]$ is given by $e^{(r-s)E[C_i]}$. Hjalmarsson (2005) shows that, in general, the pooled estimator of

Moon and Phillips (2000) will be an upward biased estimator of $E[C_i]$ when the $C_i s$ are non-identical. By Jensen's inequality, $e^{(r-s)E[C_i]} \leq E\left[e^{(r-s)C_i}\right]$, so the bias term will be underestimated by this approximation. The upward bias in the estimate of $E[C_i]$, based on the Moon and Phillips (2000) estimator, will thus to some extent counteract the downward bias induced by Jensen's inequality. Alternatively, the estimator $\hat{\beta}_{RD}$, which is based on recursive demeaning, could be used instead since it requires no knowledge of the $C_i s$.

The heterogenous $A_i s$ also cause some difficulties when dealing with the common factors. When the $A_i s$ are not identical, it is no longer possible to write $x_{i,t} = x_{i,t}^0 + \Gamma_i' z_t$ and the subsequent analysis becomes more involved and is outside the scope of this paper.

7   Empirical application

To illustrate the methods developed in this paper, I consider the question of stock-return predictability in an international data set. The data are obtained from the MSCI database and consist of a panel of total returns for stock markets in 18 different countries and three corresponding forecasting variables: the dividend- and earnings-price ratios as well as the book-to-market values. With varying success, all three of these variables have been used extensively in tests of stock-return predictability for U.S. data (e.g. Campbell and Shiller, 1988, Fama and French 1988, Lewellen, 2004, and Campbell and Yogo, 2005), and to a lesser degree in international data (e.g. Ang and Bekaert, 2003, and Campbell, 2003). All three of these forecasting variables are highly persistent, and since they are all valuation ratios, their innovations are likely to be highly correlated with the innovations to the returns process. The data are on a monthly basis and the returns data span the period 1970.1 to 2002.12, whereas not all forecasting variables are available for this whole time-period, or for all countries. In particular, I have data for stock indices in the following countries: Australia, Austria, Belgium, Canada, Denmark, France, Germany, Hong Kong, Italy, Japan, the Netherlands, Norway, Singapore, Spain, Sweden, Switzerland, the UK, and the USA.[4] The dividend price ratio $(d-p)$ is available for all countries except Hong Kong and for the entire sample period from 1970.1 onwards. The earnings price ratio $(e-p)$ is available for all countries except Italy and Switzerland, from 1974.12 onwards. The book-to-market value $(b-p)$ is available for all countries from 1974.12 onwards. The forecasting variables are the valuation ratios provided by MSCI, with earnings representing cash earnings.

---

[4] Hong Kong is, of course, not a country.

All returns and forecasting variables are expressed in U.S. dollars, and excess returns over the 1-month U.S. T-bill rate are calculated. The dependent variable in all regressions is thus excess returns over the U.S. short rate. Finally, all data are log-transformed.

The results from the pooled forecasting regressions are shown in Table 2. Panel A displays the results when there is no control for common factors, whereas the results in Panel B are based on the methods developed for dealing with cross-sectional dependence.

The estimates of C and the estimates of the correlation between the innovations in the returns and predictor processes show that the forecasting variables are clearly near-unit processes and highly endogenous. Starting with the results in Panel A, the standard pooled fixed effects estimator, $\hat{\beta}_{FE}$, delivers highly significant estimates and clearly rejects the null-hypothesis of no predictability. Given the high persistence and endogeneity found in the data, however, these results are likely to be upward biased and as seen from the estimates based on the estimator using recursive demeaning, $\hat{\beta}_{RD}$, and the bias corrected fixed effects estimator, $\hat{\beta}_{FE}^{+}$, the significance disappears when controlling for the bias induced by the time-series demeaning in the fixed effects estimator.

In Panel B, the regression estimates are based on the methods controlling for common factors across the different markets. For the dividend price ratio, the standard fixed effects estimate is now even more significant, but the $t-$statistics for the two other estimates, $\tilde{\beta}_{RD}$ and $\tilde{\beta}_{FE}^{+}$, are in fact both negative and insignificant. For the earnings-price ratio and the book-to-market ratio, the results from the fixed effects estimator become less strong when controlling for common factors and the two other estimators are both negative for both variables. Overall, the case for stock-return predictability in this international data set, using either of the three predictor variables, must be considered very weak.

The empirical results shown here again illustrate the difficulties of performing inference in regressions with persistent and endogenous variables, and that these difficulties also prevail when a panel of data, rather than a single time-series, is available. Indeed, judging by the vast difference between the estimates and test statistics resulting from the standard fixed effects estimator and those from the robust estimators, it is clear that the bias effects can be as large in panel estimations as in time-series regressions.

8   Conclusion

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. Alternatively, a bias corrected version of the standard fixed effects estimator is also proposed. Following the work of Peseran (2006), I also extend the results to a setting that allows for cross-sectional dependence in the form of common factors in the panel.

Monte Carlo evidence suggests that the proposed estimators have 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. An application to predictability in international stock-returns also illustrates that failure to account for the endogeneity and persistence in the regressors can lead to highly biased inference.

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

A Preliminary results

Lemma 2 *The following orders of magnitudes hold:*

1. $\frac{\bar{\mathbf{v}}'\bar{\mathbf{v}}}{T}$, $\frac{\bar{\mathbf{v}}'\bar{\mathbf{X}}^0_{.,-1}}{T}$, *and* $\frac{\bar{\mathbf{X}}^{0'}_{.,-1}\bar{\mathbf{X}}^0_{.,-1}}{T^2}$ *are of order* $O_p\left(\frac{1}{n}\right)$.

2. $\frac{\bar{\mathbf{v}}'\mathbf{Z}_{-1}}{T}$, $\frac{\mathbf{X}^{0'}_{i,-1}\bar{\mathbf{X}}^0_{.,-1}}{T^2}$, $\frac{\mathbf{Z}'_{-1}\bar{\mathbf{X}}^0_{.,-1}}{T^2}$, $\frac{\mathbf{f}'\bar{\mathbf{X}}^0_{.,-1}}{T}$, *and* $\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\mathbf{X}^{0'}_{i,-1}\bar{\mathbf{X}}^0_{.,-1}}{T^2}$ *are of order* $O_p\left(\frac{1}{\sqrt{n}}\right)$.

3. $\frac{\bar{\mathbf{v}}'\mathbf{f}}{T}$, $\frac{\bar{\mathbf{v}}'\mathbf{X}^0_{i,-1}}{T^2}$, *and* $\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\bar{\mathbf{v}}'\mathbf{X}^0_{i,-1}}{T^2}$ *are of order* $O_p\left(\frac{1}{\sqrt{nT}}\right)$.

Proof of Lemma 1. *(a)* By standard results (Phillips and Solo, 1992) and the continuous mapping theorem (CMT), for a fixed i as $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^T u_{i,t}x_{i,t-1} = \frac{1}{T}\sum_{t=1}^T \left(u_{i,t}x^0_{i,t-1} + u_{i,t}\Gamma'_i z_{t-1}\right) \Rightarrow \int_0^1 dB_{1,i}J_i + \Gamma'_i\left(\int_0^1 dB_{1,i}J_f\right).$$

Since $u_{i,t}$, and hence $B_{1,i}$, are cross-sectionally independent, by the Lindeberg-Levy central limit theorem (CLT) as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^1 dB_{1,i}J_i \Rightarrow N\left(0, E\left[\left(\int_0^1 dB_{1,i}J_i\right)\left(\int_0^1 dB_{1,i}J_i\right)'\right]\right).$$

Conditional on J_f (or $\mathcal{C}$), $\int_0^1 dB_{1,i}J_f$ is *iid* with mean zero, and using similar arguments as in Andrews (2005) it follows that,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \Gamma'_i\left(\int_0^1 dB_{1,i}J_f\right) \Rightarrow MN\left(0, E\left[\Gamma'_i\left(\int_0^1 dB_{1,i}J_f\right)\left(\int_0^1 dB_{1,i}J_f\right)'\Gamma_i\Bigg|\mathcal{C}\right]\right).$$

(b) Similarly, as $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^T \gamma'_i f_t x_{i,t-1} = \gamma'_i\left(\frac{1}{T}\sum_{t=1}^T \left(f_t x^0_{i,t-1} + f_t \Gamma'_i z_{t-1}\right)\right) \Rightarrow \gamma'_i \int_0^1 dB_f J_i + \gamma'_i \int_0^1 dB_f \Gamma'_i J_f,$$

and conditional on $\mathcal{C}$, $\int_0^1 dB_f J_i$ is *iid* with mean zero, and by the weak law of large numbers (WLLN), as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^n \gamma'_i \int_0^1 dB_f J_i \to_p \gamma' E\left[\int_0^1 dB_f J_i\Bigg|\mathcal{C}\right] = 0, \text{ and } \frac{1}{n}\sum_{i=1}^n \gamma'_i \int_0^1 dB_f \Gamma'_i J_f \Rightarrow \gamma' \int_0^1 dB_f \Gamma' J_f.$$

21

(c) Observe that

$$\frac{1}{T^2}\sum_{t=1}^{T}x_{i,t}x'_{i,t} = \frac{1}{T^2}\sum_{t=1}^{T}\left(x^0_{i,t-1}x^{0\prime}_{i,t-1} + \Gamma'_i z_{t-1}x^{0\prime}_{i,t-1} + x^0_{i,t-1}z'_{t-1}\Gamma_i + \Gamma'_i z_{t-1}z'_{t-1}\Gamma_i\right)$$

$$\Rightarrow \int_0^1 J_i J'_i + \Gamma'_i\left(\int_0^1 J_f J'_i\right) + \left(\int_0^1 J_i J'_f\right)\Gamma_i + \Gamma'_i\left(\int_0^1 J_f J'_f\right)\Gamma_i$$

as $T \to \infty$. By the WLLN, as $n \to \infty$, $\frac{1}{n}\sum_{i=1}^{n}\int_0^1 J_i J'_i \to_p E\left[\int_0^1 J_i J'_i\right]$, and $\frac{1}{n}\sum_{i=1}^{n}\Gamma'_i\left(\int_0^1 J_f J'_i\right) \to_p 0$, since, conditional on $\mathcal{C}$, $\int_0^1 J_f J'_i$ is *iid* with mean zero. Finally, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}\Gamma'_i\left(\int_0^1 J_f J'_f\right)\Gamma_i \Rightarrow \Gamma'\left(\int_0^1 J_f J'_f\right)\Gamma.$$

(d) As before, as $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^{T}u_{i,t}\underline{x}_{i,t-1} = \frac{1}{T}\sum_{t=1}^{T}\left(u_{i,t}\underline{x}^0_{i,t-1} + u_{i,t}\Gamma'_i\underline{z}_{t-1}\right) \Rightarrow \int_0^1 dB_{1,i}\underline{J}_i + \left(\int_0^1 dB_{1,i}\Gamma'_i\underline{J}_f\right).$$

Now, $E\left[\left(\int_0^1 dB_{1,i}\Gamma'_i\underline{J}_f\right)\Big|\mathcal{C}\right] = 0$ given the independence between $u_{i,t}$ and f_t, and by the results derived in the main text it follows that, as $(T,n \to \infty)_{\text{seq}}$,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}u_{i,t}\underline{x}_{i,t-1} \to_p -\left(\int_0^1\int_0^r E\left[e^{(r-s)C}\right]dsdr\right)\omega_{21}.$$

(e) As $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^{T}\gamma'_i\underline{f}_t\underline{x}_{i,t-1} = \gamma'_i\left(\frac{1}{T}\sum_{t=1}^{T}\left(f_t\underline{x}^0_{i,t-1} + f_t\Gamma'_i\underline{z}_{t-1}\right)\right) \Rightarrow \gamma'_i\int_0^1 dB_f\underline{J}_i + \gamma'_i\int_0^1 dB_f\Gamma'_i\underline{J}_f.$$

By the independence of $u_{i,t}$ and f_t, $E\left[\gamma'_i\int_0^1 dB_f\underline{J}_i\Big|\mathcal{C}\right] = 0$ and by the WLLN as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}\gamma'_i\int_0^1 dB_f\Gamma'_i\underline{J}_f \Rightarrow \gamma'\int_0^1 dB_f\Gamma'\underline{J}_f.$$

(f) The result follows in an identical manner to that in *(c)*. ∎

Proof of Lemma 2. *1.* The result for $\frac{\bar{\mathbf{v}}'\bar{\mathbf{v}}}{T}$ follows directly from Lemma 2 in Pesaran (2006). Further,

as $T \to \infty$,

$$\frac{\bar{\mathbf{v}}'\bar{\mathbf{X}}^0_{\cdot,-1}}{T} = \frac{1}{T}\sum_{t=1}^{T}\bar{v}_{\cdot,t}\bar{x}^{0\prime}_{\cdot,t-1} = \frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T}\sum_{t=1}^{T}v_{i,t}x^{0\prime}_{j,t-1} \Rightarrow \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_0^1 dB_{2i}J'_j\right) = O_p\left(\frac{1}{n}\right)O_p(1),$$

where the last equality follows since $\int_0^1 dB_{2i}J'_j$ is *iid* with mean zero for all i and j, and thus satisfies a CLT. Similarly, as $T \to \infty$,

$$\frac{\bar{\mathbf{X}}^{0\prime}_{\cdot,-1}\bar{\mathbf{X}}^0_{\cdot,-1}}{T^2} = \frac{1}{T^2}\sum_{t=1}^{T}\bar{x}^0_{\cdot,t-1}\bar{x}^{0\prime}_{\cdot,t-1} = \frac{1}{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}x^0_{i,t-1}x^{0\prime}_{j,t-1} \Rightarrow \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\int_0^1 J_i J'_j\right) = O_p\left(\frac{1}{n}\right)O_p(1),$$

since $\int_0^1 J_i J'_k$ is *iid* with mean zero for all $i \neq j$.

2. Similar to above, as $T \to \infty$,

$$\frac{\bar{\mathbf{v}}'\mathbf{Z}_{-1}}{T} = \frac{1}{T}\sum_{t=1}^{T}\bar{v}_{\cdot,t}z'_{t-1} = \frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}v_{i,t}z'_{t-1} \Rightarrow \frac{1}{n}\sum_{i=1}^{n}\int_0^1 dB_{2i}J'_f = O_p\left(\frac{1}{\sqrt{n}}\right),$$

since conditional of J_f, $\int_0^1 dB_{2i}J'_f$ is *iid* with mean zero. The rest follow in an analogous manner.

3. $\frac{\bar{\mathbf{v}}'\mathbf{f}}{T} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ follows directly from Lemma 2 in Pesaran (2006). Further, as $T \to \infty$,

$$\frac{\bar{\mathbf{v}}'\mathbf{X}^0_{i,-1}}{T^2} = \frac{1}{T}\sum_{t=1}^{T}\bar{v}_{\cdot,t}x^{0\prime}_{i,t-1} = \frac{1}{T}\frac{1}{n}\sum_{j=1}^{n}\frac{1}{T}\sum_{t=1}^{T}v_{j,t}x^{0\prime}_{i,t-1} \Rightarrow \frac{1}{T}\left(\frac{1}{n}\sum_{i=1}^{n}\int_0^1 dB_{1,j}J'_i\right) = O_p\left(\frac{1}{\sqrt{nT}}\right),$$

since $\int_0^1 dB_{1,j}J'_i$ is *iid* with mean zero. Similarly, as $T \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\bar{\mathbf{v}}'\mathbf{X}^0_{i,-1}}{T} = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}v_{i,t}x^{0\prime}_{j,t-1} \Rightarrow \frac{1}{\sqrt{n}}\left(\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\int_0^1 dB_{2,i}J_j\right) = O_p\left(\frac{1}{\sqrt{n}}\right).$$

∎

B   Proofs of main results

Proof of Theorem 1.   The result follows directly by *(a)* and *(c)* in Lemma 1, with $\Phi_{uz} = 0$ and $\Omega_{zz} = 0$, and the CMT. ∎

Proof of Theorem 2. Observe that

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

For fixed n, as $T \to \infty$,

$$\sqrt{n}T\left(\hat{\beta}_{RD} - \beta\right) = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T}\underline{x}_{i,t-1}^{dd}\underline{x}_{i,t-1}^{r\prime}\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\underline{u}_{i,t}^{dd}\underline{x}_{i,t-1}^{r}\right)$$

$$\Rightarrow \left(\frac{1}{n}\sum_{i=1}^{n}\int_0^1 \underline{J}_i^{dd}\underline{J}_i^{d\prime}dr\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^1 dB_{1,i}^{dd}\underline{J}_i^{d}\right),$$

since

$$\frac{1}{T}\sum_{t=1}^{T}\underline{u}_{i,t}^{dd}\underline{x}_{i,t-1}^{d} = \frac{1}{T}\sum_{t=1}^{T}\left(u_{i,t} - \frac{1}{T-t}\sum_{s=t}^{T}u_{i,s}\right)\underline{x}_{i,t-1}^{d}$$

$$= \frac{1}{T}\sum_{t=1}^{T}u_{i,t}\underline{x}_{i,t-1}^{d} - \frac{1}{T}\sum_{t=1}^{T}\left(\frac{T-t}{T}\right)^{-1}\left(\frac{1}{\sqrt{T}}\left(\sum_{s=1}^{T}u_{i,s} - \sum_{s=1}^{t-1}u_{i,s}\right)\right)\frac{\underline{x}_{i,t-1}^{d}}{\sqrt{T}}$$

$$\Rightarrow \int_0^1 dB_{1,i}(r)\underline{J}_i^{d}(r) - \int_0^1 (1-r)^{-1}(B_{1,i}(1) - B_{1,i}(r))\underline{J}_i^{d}(r)dr$$

$$= \int_0^1 dB_{1,i}^{dd}(r)\underline{J}_i^{d}(r), \tag{36}$$

by standard arguments. By the independent increments property of the Brownian motion, it follows that the expectation of (36) is equal to zero and the result follows from similar arguments as before. ∎

Proof of Theorem 3. The results follow immediately from Lemma 1 and the CMT. ∎

Proof of Theorem 4. Define the $2m \times 2m$ matrix $D_T = diag\left(\sqrt{T}, ..., \sqrt{T}, T, ..., T\right)$, such that

$$\bar{\mathbf{H}}D_T^{-1} = \left(\begin{array}{cc} \boldsymbol{\Delta}_C\bar{\mathbf{X}} & \bar{\mathbf{X}}_{\cdot,-1} \end{array}\right)\left(\begin{array}{cc} I_m\sqrt{T} & 0 \\ 0 & I_m T \end{array}\right) = \left(\begin{array}{cc} \frac{\boldsymbol{\Delta}_C\bar{\mathbf{X}}}{\sqrt{T}} & \frac{\bar{\mathbf{X}}_{\cdot,-1}}{T} \end{array}\right).$$

Note that,

$$\sqrt{n}T\left(\tilde{\beta}_{Pool} - \beta\right) = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{X}_{i,-1}'\mathbf{M}_{\bar{\mathbf{H}}}\mathbf{X}_{i,-1}}{T^2}\right)^{-1}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{X}_{i,-1}'\mathbf{M}_{\bar{\mathbf{H}}}\mathbf{f}\gamma_i}{T}\right) + \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{X}_{i,-1}'\mathbf{M}_{\bar{\mathbf{H}}}\mathbf{u}_i}{T}\right)\right],$$

and consider first

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{X}'_{i,-1}\mathbf{M}_{\bar{\mathbf{H}}}\mathbf{f}}{T}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{X}'_{i,-1}\mathbf{f}}{T}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{X}'_{i,-1}\bar{\mathbf{H}}D_T^{-1}}{T}\left(D_T^{-1}\bar{\mathbf{H}}'\bar{\mathbf{H}}D_T^{-1}\right)^{-1}\left(D_T^{-1}\bar{\mathbf{H}}'\mathbf{f}\right).$$

Observe that by the rank condition in Assumption 3, $\mathbf{M}_{\mathbf{Q}}\mathbf{f}=\mathbf{M}_{\mathbf{G}}\mathbf{f}=\mathbf{0}$, since $\mathbf{f}\subset\mathbf{G}$. From Lemma 2, it follows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{X}'_{i,-1}}{T}\bar{\mathbf{H}}D_T^{-1}\left(D_T^{-1}\bar{\mathbf{H}}'\bar{\mathbf{H}}D_T^{-1}\right)^{-1}D_T^{-1}\bar{\mathbf{H}}'\mathbf{f}\gamma_i$$

$$=\quad\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{X}'_{i,-1}}{T}\left(\begin{array}{cc}\frac{\mathbf{f}\bar{\Gamma}+\bar{\mathbf{v}}}{\sqrt{T}} & \frac{\bar{\mathbf{X}}^0_{.,-1}+\mathbf{Z}_{-1}\bar{\Gamma}}{T}\end{array}\right)$$

$$\times\left(\begin{array}{cc}\frac{\bar{\Gamma}'\mathbf{f}'\mathbf{f}\bar{\Gamma}+\bar{\Gamma}'\mathbf{f}'\bar{\mathbf{v}}+\bar{\mathbf{v}}'\mathbf{f}\bar{\Gamma}+\bar{\mathbf{v}}'\bar{\mathbf{v}}}{T} & \frac{\bar{\Gamma}'\mathbf{f}'\bar{\mathbf{X}}^0_{.,-1}+\bar{\Gamma}'\mathbf{f}'\mathbf{Z}_{-1}\bar{\Gamma}+\bar{\mathbf{v}}'\bar{\mathbf{X}}^0_{.,-1}+\bar{\mathbf{v}}'\mathbf{Z}_{-1}\bar{\Gamma}}{T^{3/2}} \\ \frac{\bar{\mathbf{X}}^{0'}_{.,-1}\mathbf{f}\bar{\Gamma}+\bar{\mathbf{X}}^{0'}_{.,-1}\bar{\mathbf{v}}+\bar{\Gamma}'\mathbf{Z}'_{-1}\mathbf{f}\bar{\Gamma}+\bar{\Gamma}'\mathbf{Z}'_{-1}\bar{\mathbf{v}}}{T^{3/2}} & \frac{\bar{\mathbf{X}}^{0'}_{.,-1}\bar{\mathbf{X}}^0_{.,-1}+\bar{\mathbf{X}}^{0'}_{.,-1}\mathbf{Z}_{-1}\bar{\Gamma}+\bar{\Gamma}'\mathbf{Z}'_{-1}\bar{\mathbf{X}}^0_{.,-1}+\bar{\Gamma}'\mathbf{Z}'_{-1}\mathbf{Z}_{-1}\bar{\Gamma}}{T^2}\end{array}\right)^{-1}\left(\begin{array}{c}\frac{\bar{\Gamma}'\mathbf{f}'+\bar{\mathbf{v}}'}{\sqrt{T}} \\ \frac{\bar{\mathbf{X}}^{0'}_{.,-1}+\bar{\Gamma}'\mathbf{Z}'_{-1}}{T}\end{array}\right)\mathbf{f}\gamma_i$$

$$=\quad\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\mathbf{X}'_{i,-1}\mathbf{f}\bar{\Gamma}}{T}\right)\left(\frac{\bar{\Gamma}'\mathbf{f}'\mathbf{f}\bar{\Gamma}}{T}\right)^{-1}\left(\frac{\bar{\Gamma}'\mathbf{f}'\mathbf{f}}{T}\right)\gamma_i$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\mathbf{X}'_{i,-1}\mathbf{Z}_{-1}\bar{\Gamma}}{T^2}\right)\left(\frac{\bar{\Gamma}'\mathbf{Z}'_{-1}\mathbf{Z}_{-1}\bar{\Gamma}}{T^2}\right)^{-1}\left(\frac{\bar{\Gamma}'\mathbf{Z}'_{-1}\mathbf{f}\gamma_i}{T}\right)+O_p\left(\frac{1}{\sqrt{n}}\right)$$

$$=\quad\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{X}'_{i,-1}}{T}\mathbf{Q}D_T^{-1}\left(D\mathbf{Q}'\mathbf{Q}D_T^{-1}\right)^{-1}D_T^{-1}\mathbf{Q}'\mathbf{f}\gamma_i+O_p\left(\frac{1}{\sqrt{n}}\right).$$

Thus,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{X}'_{i,-1}\mathbf{M}_{\bar{\mathbf{H}}}\mathbf{f}\gamma_i}{T}=O_p\left(\frac{1}{\sqrt{n}}\right).$$

Next, observe that

$$\frac{\mathbf{X}'_{i,-1}\mathbf{M}_{\bar{\mathbf{H}}}\mathbf{u}_i}{T}=\frac{\mathbf{X}'_{i,-1}\left(\mathbf{M}_{\bar{\mathbf{H}}}-\mathbf{M}_{\mathbf{Q}}\right)\mathbf{u}_i}{T}+\frac{\mathbf{X}'_{i,-1}\mathbf{M}_{\mathbf{Q}}\mathbf{u}_i}{T}=\frac{\mathbf{X}'_{i,-1}\left(\mathbf{M}_{\bar{\mathbf{H}}}-\mathbf{M}_{\mathbf{Q}}\right)\mathbf{u}_i}{T}+\frac{\mathbf{X}^{0'}_{i,-1}\mathbf{M}_{\mathbf{Q}}\mathbf{u}_i}{T},$$

and,

$$\frac{\mathbf{X}_{i,-1}^{0\prime}\mathbf{M_Q}\mathbf{u}_i}{T}$$

$$= \mathbf{X}_{i,-1}^{0\prime}\left(\mathbf{I} - \frac{\mathbf{Q}D_T^{-1}}{T}\left(D_T^{-1}\mathbf{Q}'\mathbf{Q}D_T^{-1}\right)^{-1}D_T^{-1}\mathbf{Q}'\right)\mathbf{u}_i$$

$$= \mathbf{X}_{i,-1}^{0\prime}\left(\mathbf{I} - \begin{pmatrix} \frac{\mathbf{f}\bar{\Gamma}}{\sqrt{T}} & \frac{\mathbf{Z}_{-1}\bar{\Gamma}}{T} \end{pmatrix}\begin{pmatrix} \frac{\bar{\Gamma}'\mathbf{f}'\mathbf{f}\bar{\Gamma}}{T} & \frac{\bar{\Gamma}'\mathbf{f}'\mathbf{Z}_{-1}\bar{\Gamma}}{T^{3/2}} \\ \frac{\bar{\Gamma}'\mathbf{Z}_{-1}'\mathbf{f}\bar{\Gamma}}{T^{3/2}} & \frac{\bar{\Gamma}'\mathbf{Z}_{-1}'\mathbf{Z}_{-1}\bar{\Gamma}}{T^2} \end{pmatrix}^{-1}\begin{pmatrix} \frac{\bar{\Gamma}'\mathbf{f}'}{\sqrt{T}} \\ \frac{\bar{\Gamma}'\mathbf{Z}_{-1}'}{T} \end{pmatrix}\right)\mathbf{u}_i$$

$$= \frac{\mathbf{X}_{i,-1}^{0\prime}\mathbf{u}_i}{T} - \left(\frac{\mathbf{X}_{i,-1}^{0\prime}\mathbf{f}\bar{\Gamma}}{T}\right)\left(\frac{\bar{\Gamma}'\mathbf{f}'\mathbf{f}\bar{\Gamma}}{T}\right)^{-1}\frac{\bar{\Gamma}'\mathbf{f}'\mathbf{u}_i}{T} - \frac{\mathbf{X}_{i,-1}^{0\prime}\mathbf{Z}_{-1}\bar{\Gamma}}{T^2}\left(\frac{\bar{\Gamma}'\mathbf{Z}_{-1}'\mathbf{Z}_{-1}\bar{\Gamma}}{T^2}\right)^{-1}\left(\frac{\bar{\Gamma}'\mathbf{Z}_{-1}'\mathbf{u}_i}{T}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

$$\Rightarrow \int_0^1 J_i dB_{1,i} - \left(\int_0^1 J_i J_f'\bar{\Gamma}\right)\left(\bar{\Gamma}'\int_0^1 J_f J_f'\bar{\Gamma}\right)^{-1}\left(\bar{\Gamma}'\int_0^1 J_f dB_{1,i}\right) \equiv \int_0^1 J_{i\cdot f\bar{\Gamma}}dB_{1,i},$$

where $J_{i\cdot f\bar{\Gamma}}$ is the residual from projecting J_i onto $f\bar{\Gamma}$ and the $O_p\left(1/\sqrt{T}\right)$ term follows from $T^{-3/2}\mathbf{Z}_{-1}'\mathbf{f} = O_p\left(1/\sqrt{T}\right)$. Using Lemma 2 again, it follows from similar calculations as above that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\mathbf{X}_{i,-1}'\left(\mathbf{M_{\bar{H}}} - \mathbf{M_Q}\right)\mathbf{u}_i}{T} = O_p\left(\frac{1}{\sqrt{n}}\right).$$

Thus,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\mathbf{X}_{i,-1}'\mathbf{M_{\bar{H}}}\mathbf{u}_i}{T} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\mathbf{X}_{i,-1}'\mathbf{M_Q}\mathbf{u}_i}{T} + O_p\left(\frac{1}{\sqrt{n}}\right),$$

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

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\mathbf{X}_{i,-1}'\mathbf{M_Q}\mathbf{u}_i}{T} \Rightarrow MN\left(0, E\left[\left(\int_0^1 J_{i\cdot f\Gamma}dB_{1,i}\right)\left(\int_0^1 J_{i\cdot f\Gamma}dB_{1,i}\right)'\Bigg|\mathcal{C}\right]\right),$$

since $n/T = o(1)$. Finally,

$$\frac{1}{n}\sum_{i=1}^n \frac{\mathbf{X}_{i,-1}'\mathbf{M_{\bar{H}}}\mathbf{X}_{i,-1}}{T^2} = \frac{1}{n}\sum_{i=1}^n \frac{\mathbf{X}_{i,-1}'\mathbf{M_Q}\mathbf{X}_{i,-1}}{T^2} + O_p\left(\frac{1}{\sqrt{n}}\right) = \frac{1}{n}\sum_{i=1}^n \frac{\mathbf{X}_{i,-1}^{0\prime}\mathbf{X}_{i,-1}^0}{T^2} + O_p\left(\frac{1}{\sqrt{n}}\right) \to_p E\left[\int_0^1 J_i J_i'\right],$$

as $(T, n \to \infty)_{\text{seq}}$. ■

Proof of Corollary 1. The result follows in an identical manner to above, using the results in Lemma 1 and Theorem 3. ■

Proof of Corollary 2. By the same arguments as before,

$$\frac{\mathbf{X}'_{i,-1}\mathbf{M_Q g}\eta_i}{T} \Rightarrow \int_0^1 J_i \eta'_i dB_g - \left(\int_0^1 J_i J'_f \bar{\Gamma}\right)\left(\bar{\Gamma}'\int_0^1 J_f J'_f \bar{\Gamma}\right)^{-1}\left(\bar{\Gamma}'\int_0^1 J_f \eta'_i dB_g\right) \equiv \int_0^1 J_{i \cdot f\bar{\Gamma}}\eta'_i dB_g,$$

and since B_g is independent of J_i and J_f, the result follows. ∎

References

[1] Andrews, D.W.K., 2005. Cross-section Regression with Common Shocks, Econometrica 73, 1551-1585.

[2] Ang, A., and G. Bekaert, 2003. Stock Return Predictability: Is it There? Working paper, Columbia University.

[3] Campbell, J.Y., 2003. Consumption-Based Asset Pricing, in: Constantinides, G.M., Harris M., and Stulz R. eds., Handbook of the Economics of Finance, Vol. 1B (North-Holland, Amsterdam) 803-888.

[4] Campbell, J.Y., and M. Yogo, 2005. Efficient Tests of Stock Return Predictability, forthcoming Journal of Financial Economics.

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

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

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

[8] Hjalmarsson, E., 2005. Estimation of average local-to-unity roots in heterogenous panels, International Finance Discussion Paper 852, Federal Reserve Board.

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

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

[11] Lewellen, J., 2004. Predicting returns with financial ratios, Journal of Financial Economics, 74, 209-235.

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

[13] Moon, H.R., and P.C.B. Phillips, 2000. Estimation of Autoregressive Roots near Unity using Panel Data, Econometric Theory 16, 927-998.

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

[15] Pagan, A., and A. Ullah, 1999. Nonparametric Econometrics, Cambridge University Press.

[16] Pesaran, M.H., 2006. Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure, Econometrica, forthcoming.

[17] 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.

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

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

[20] Phillips, P.C.B., and D. Sul, 2004. Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence, Cowles Foundation Discussion Paper 1438.

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

[22] Stambaugh, R., 1999. Predictive regressions, Journal of Financial Economics 54, 375-421.

[23] Sul, D., P.C.B. Phillips, and C.Y. Choi, 2005. Prewhitening bias in HAC estimation, Oxford Bulletin of Economics and Statistics 67, 517-546.

Table 1: Size results from the Monte Carlo study. The table shows the average rejection rates under the null of $\beta = 0$, for the $t-$tests corresponding to the respective estimators; the nominal size of the tests are 5 percent. The differing values of δ are given in the top row of the table and the results are based on $10,000$ repitions. The sample size used is $T = 100$ and $n = 20$, and the local-to-unity parameter, C, is set equal to -10. Panel A shows the results when no common factors are included in the data generating process, Panel B shows the effects of including common factors in the data but ignoring them in the estimation, and Panel C shows the results when common factors are included and accounted for in the estimation.

Estimator	$\delta = 0.0$	$\delta = -0.4$	$\delta = -0.7$	$\delta = -0.95$
Panel A: No common factors				
$\hat{\beta}_{POOL}$	0.076	0.078	0.074	0.079
$\hat{\beta}_{FE}$	0.076	0.168	0.321	0.488
$\hat{\beta}_{RD}$	0.074	0.077	0.070	0.074
$\hat{\beta}_{FE}^{+}$	0.076	0.076	0.075	0.085
Panel B: Common factors with no correction				
$\hat{\beta}_{FE}$	0.440	0.498	0.556	0.595
$\hat{\beta}_{RD}$	0.395	0.386	0.381	0.388
$\hat{\beta}_{FE}^{+}$	0.319	0.311	0.309	0.313
Panel C: Common factors using correction				
$\tilde{\beta}_{FE}$	0.074	0.199	0.436	0.627
$\tilde{\beta}_{RD}$	0.069	0.084	0.118	0.144
$\tilde{\beta}_{FE}^{+}$	0.074	0.077	0.091	0.100

Table 2: Results from the empircal regressions. The table shows the point estimates and corresponding $t-$statistics (in parentheses below the estimates) from the pooled regressions of excess stock returns onto either the dividend price ratio $(d-p)$, the earnings price ratio $(e-p)$, or the book-to-market value $(b-p)$. Panel A shows the results when there was no correction for cross-sectional dependence and Panel B shows the results when using estimators that are robust to common factors in the data. The first colum in the table indicates which of the three forecasting variables is used and the second and third columns give the size of the panel used in the regression. The next three columns give the results for the standard fixed effects estimator, the estimator using recursively demeaned data and the bias corrected fixed effects estimator, respectively. The final two columns give the estimate of the local-to-unity paramater in the regressors and the average correlation between the innovations to the returns and the regressors, respectively.

Panel A: No correction for common factors							
Variable	n	T	$\hat{\beta}_{FE}$	$\hat{\beta}_{RD}$	$\hat{\beta}_{FE}^{+}$	$\hat{C}_{pool}$	$\hat{\delta}$
$d-p$	17	396	0.007	-0.045	-0.002	-0.004	-0.771
			(3.791)	(-2.740)	(-0.800)		
$e-p$	16	337	0.011	-0.183	0.000	0.091	-0.697
			(5.240)	(-0.591)	(-0.051)		
$b-p$	18	337	0.008	-0.141	0.002	-1.538	-0.835
			(4.893)	(-0.895)	(0.989)		

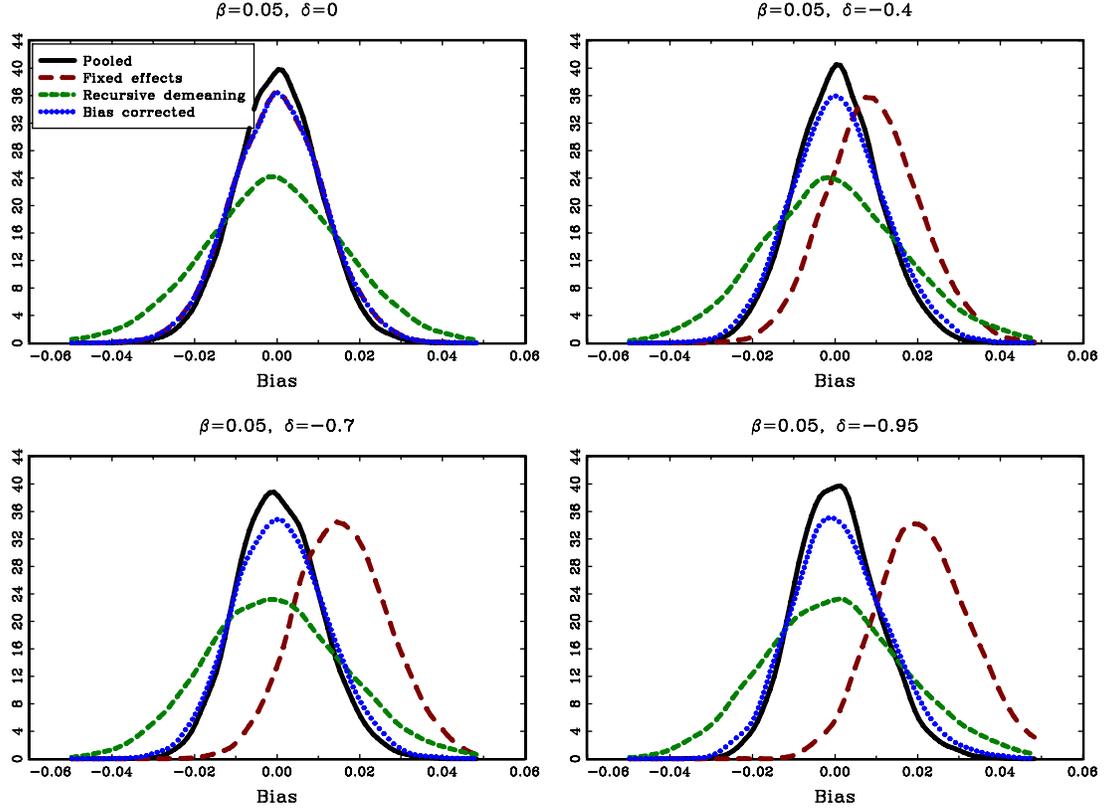| Panel B: Correction for common factors							
Variable	n	T	$\tilde{\beta}_{FE}$	$\tilde{\beta}_{RD}$	$\tilde{\beta}_{FE}^{+}$		
$d-p$	17	396	0.010	-0.008	-0.003		-0.498
			(4.643)	(-1.588)	(-0.770)		
$e-p$	16	337	0.007	-0.009	-0.005		-0.440
			(1.643)	(-1.137)	(-1.170)		
$b-p$	18	337	0.007	-0.010	-0.001		-0.547
			(2.755)	(-1.710)	(-0.194)		

Figure 1: Estimation results from the Monte Carlo study without cross-sectional dependence. The graphs show the kernel density estimates of the estimated slope coefficients, for samples with $T = 100$ and $n = 20$. The automatic bandwidth selection rules described in Pagan and Ullah (1999) were used in the kernel density estimation. The solid lines, labeled Pooled in the legend, show the results for the standard pooled estimator without individual intercepts, $\hat{\beta}_{Pool}$; the long dashed lines, labeled Fixed effects, show the results for the standard fixed effects estimator, $\hat{\beta}_{FE}$; the short dashed lines, labeled Recursive demeaning, show the results for the estimator based on recursive demeaning, $\hat{\beta}_{RD}$; the dotted lines, labeled Bias-corrected, show the results for the bias corrected fixed effects estimator, $\tilde{\beta}_{FE}^{+}$. All results are based on $10,000$ repetitions.
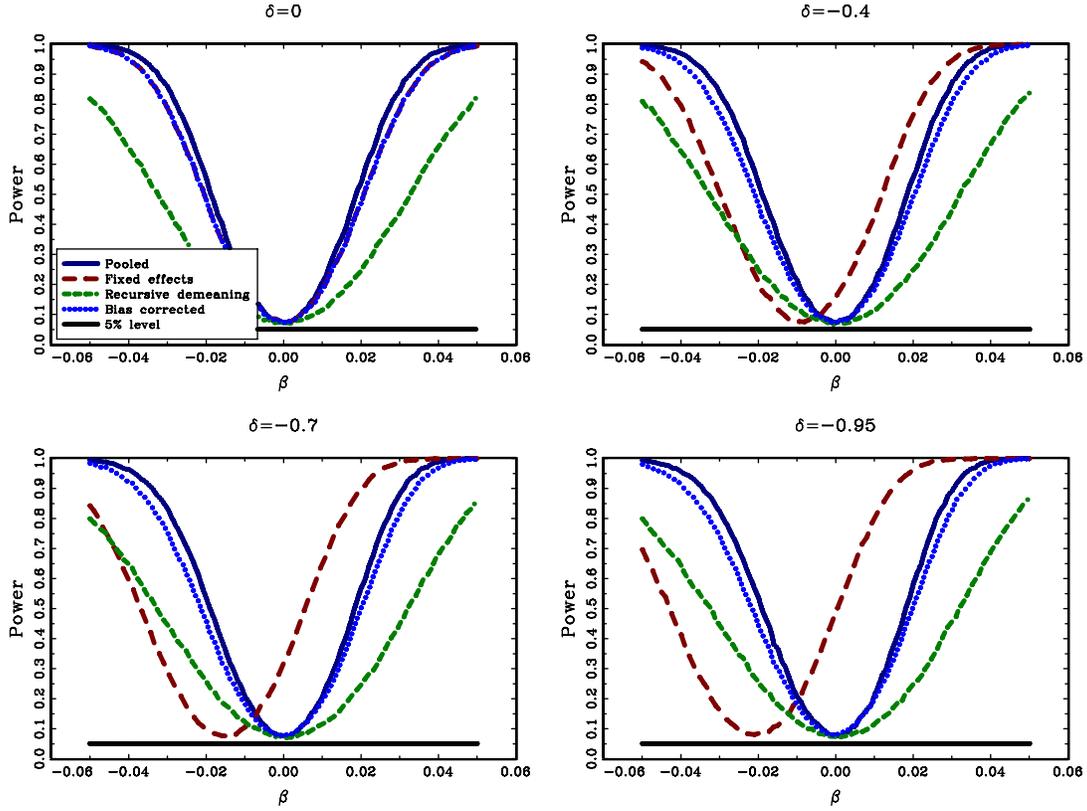
Figure 2: Size and power results from the Monte Carlo study without cross-sectional dependence. The graphs show the average rejection rates for a two-sided 5 percent $t-$test of the null hypothesis of $\beta = 0$, for samples with $T = 100$, and $n = 20$. The $x-$axis shows the true value of the parameter β, and the $y-$axis indicates the average rejection rate. The solid lines, labeled Pooled, give the results for the $t-$test corresponding to the standard pooled estimator without individual intercepts, β_{Pool}; the long dashed lines, labeled Fixed effects, show the results for the $t-$test corresponding to the standard fixed effects estimator, $\hat{\beta}_{FE}$; the short dashed lines, labeled Recursive demeaning, show the results for the $t-$test corresponding to the estimator based on recursive demeaning, $\hat{\beta}_{RD}$; the dotted lines, labeled Bias corrected, show the results for the $t-$test corresponding to the bias corrected fixed effects estimator, $\hat{\beta}^{+}_{FE}$. The flat lines indicate the 5% rejection rate. All results are based on $10,000$ repetitions.
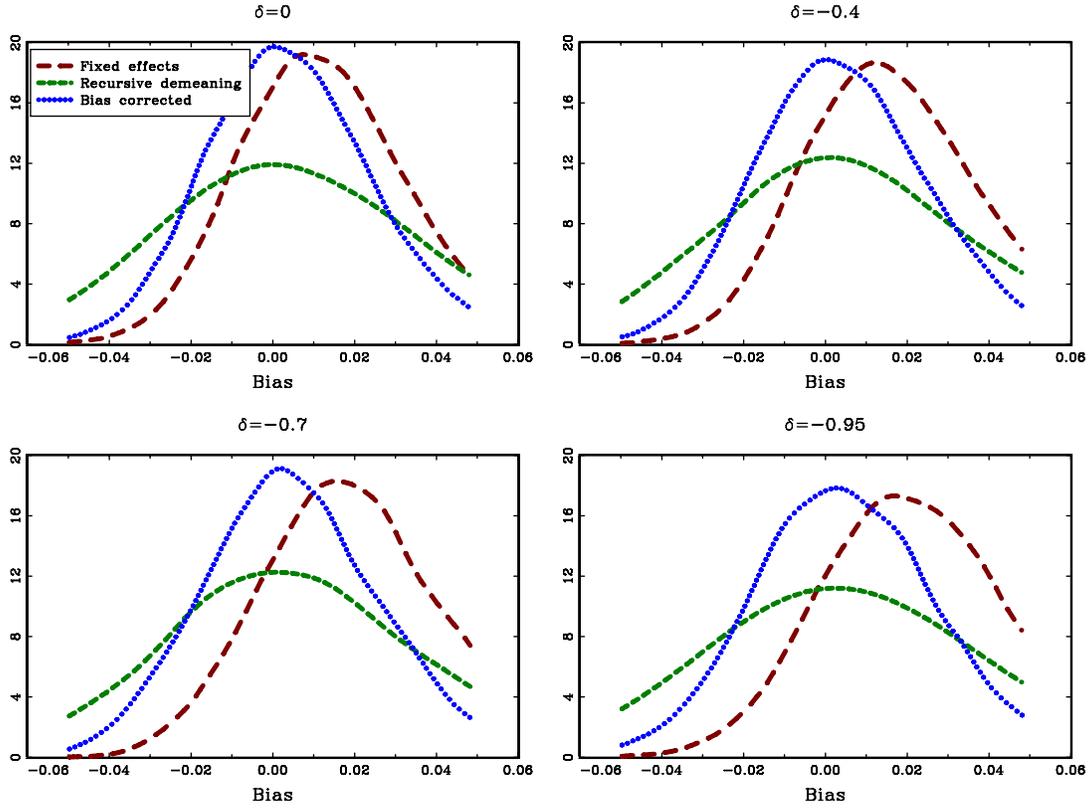
Figure 3: Estimation results from the Monte Carlo study with cross-sectional dependence that is ignored in the estimation. The graphs show the kernel density estimates of the estimated slope coefficients, for samples with $T = 100$ and $n = 20$. The automatic bandwidth selection rules described in Pagan and Ullah (1999) were used in the kernel density estimation. The long dashed lines, labeled Fixed effects, show the results for the standard fixed effects estimator, $\hat{\beta}_{FE}$; the short dashed lines, labeled Recursive demeaning, show the results for the estimator based on recursive demeaning, $\hat{\beta}_{RD}$; the dotted lines, labeled Bias-corrected, show the results for the bias corrected fixed effects estimator, $\tilde{\beta}_{FE}^{+}$. All results are based on $10,000$ repetitions.
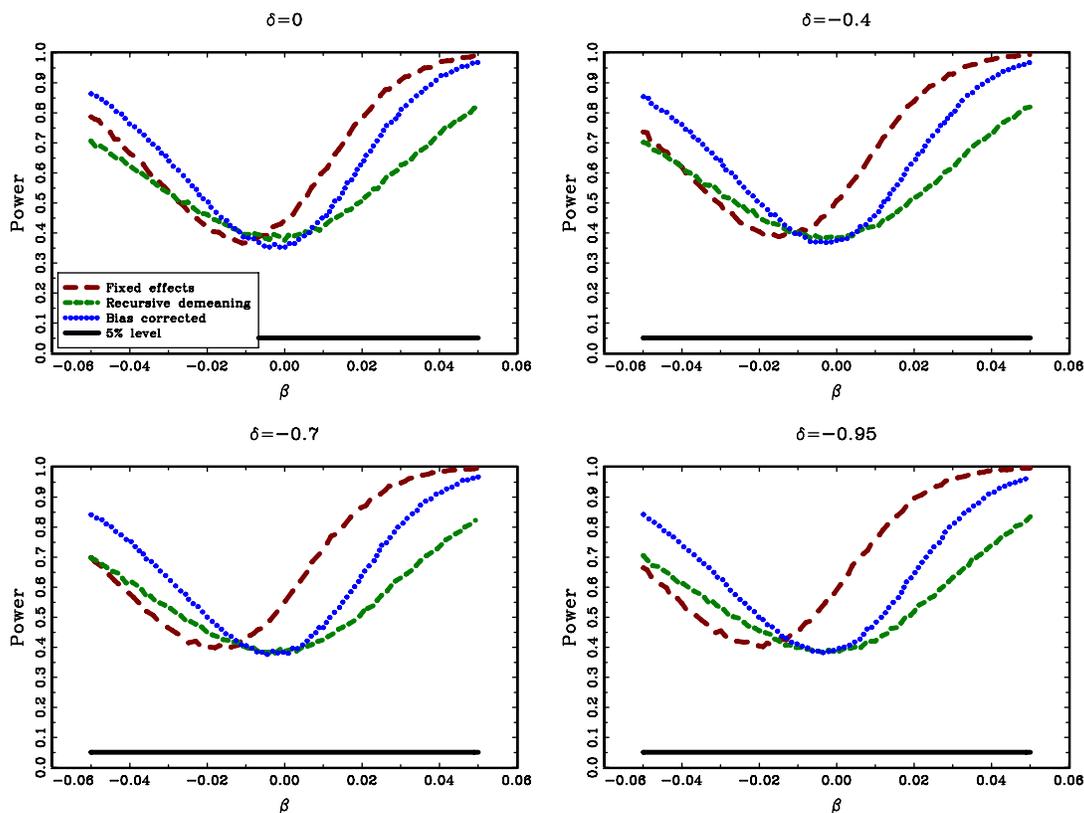
Figure 4: Size and power results from the Monte Carlo study with cross-sectional dependence that is ignored in the estimation. The graphs show the average rejection rates for a two-sided 5 percent $t-$test of the null hypothesis of $\beta = 0$, for samples with $T = 100$, and $n = 20$. The $x-$axis shows the true value of the parameter β, and the $y-$axis indicates the average rejection rate. The long dashed lines, labeled Fixed effects, show the results for the $t-$test corresponding to the standard fixed effects estimator, $\hat{\beta}_{FE}$; the short dashed lines, labeled Recursive demeaning, show the results for the $t-$test corresponding to the estimator based on recursive demeaning, $\hat{\beta}_{RD}$; the dotted lines, labeled Bias corrected, show the results for the $t-$test corresponding to the bias corrected fixed effects estimator, $\hat{\beta}_{FE}^{+}$. The flat lines indicate the 5% rejection rate. All results are based on $10,000$ repetitions.
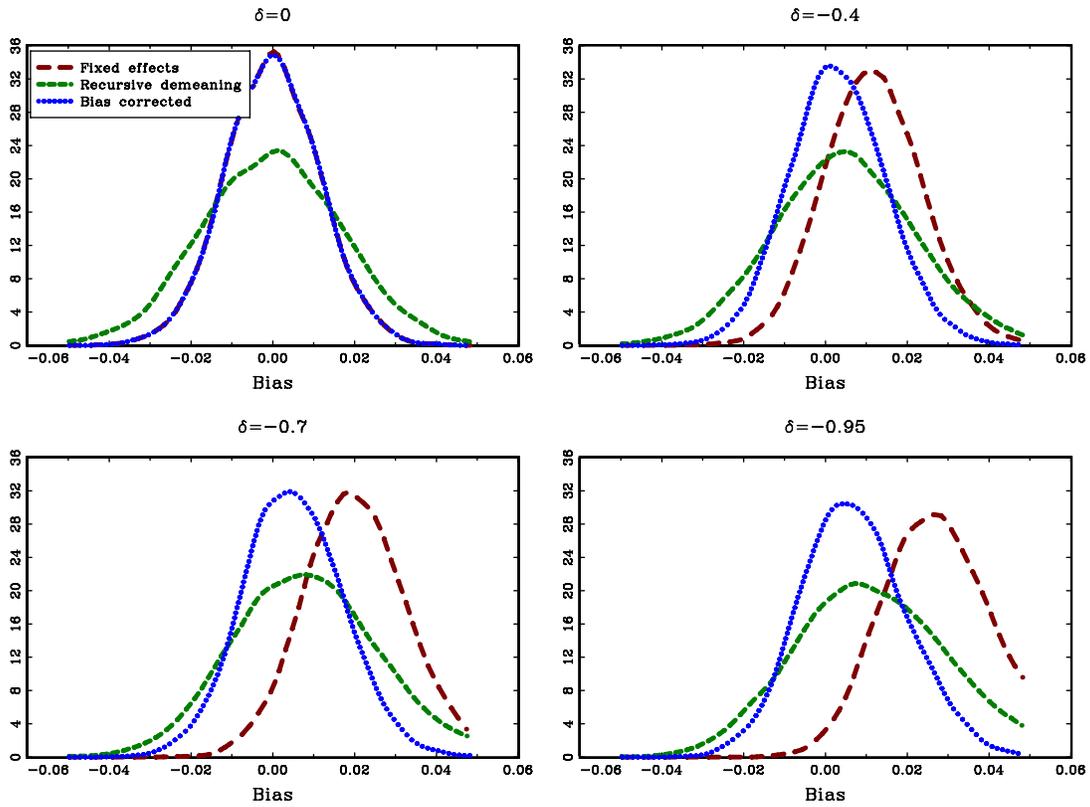
Figure 5: Estimation results from the Monte Carlo study with cross-sectional dependence, when using the estimators that take this into account. The graphs show the kernel density estimates of the estimated slope coefficients, for samples with $T = 100$ and $n = 20$. The automatic bandwidth selection rules described in Pagan and Ullah (1999) were used in the kernel density estimation. The long dashed lines, labeled Fixed effects, show the results for the fixed effects estimator, $\tilde{\beta}_{FE}$; the short dashed lines, labeled Recursive demeaning, show the results for the estimator based on recursive demeaning, $\tilde{\beta}_{RD}$; the dotted lines, labeled Bias-corrected, show the results for the bias corrected fixed effects estimator, $\tilde{\beta}_{FE}^{+}$. All results are based on $10,000$ repetitions.
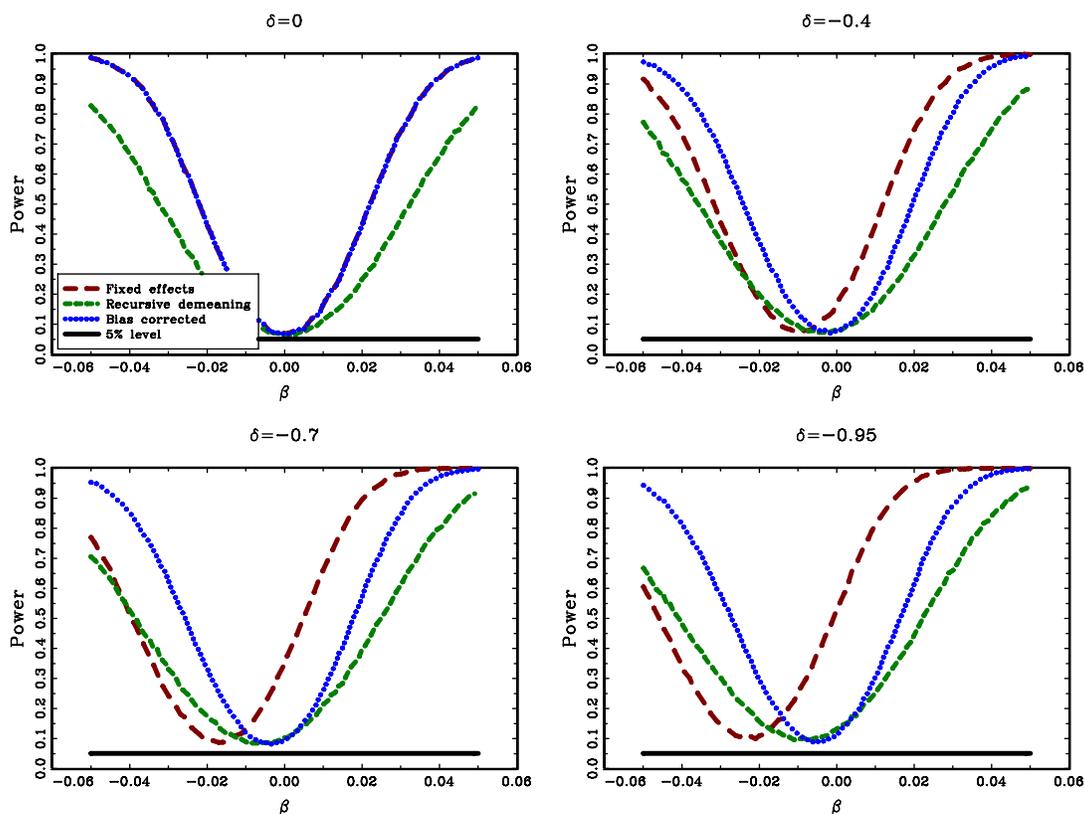
Figure 6: Size and power results from the Monte Carlo study with cross-sectional dependence, when using the estimators that take this into account. The graphs show the average rejection rates for a two-sided 5 percent $t-$test of the null hypothesis of $\beta = 0$, for samples with $T = 100$, and $n = 20$. The $x-$axis shows the true value of the parameter β, and the $y-$axis indicates the average rejection rate. The long dashed lines, labeled Fixed effects, show the results for the $t-$test corresponding to the fixed effects estimator, $\tilde{\beta}_{FE}$; the short dashed lines, labeled Recursive demeaning, show the results for the $t-$test corresponding to the estimator based on recursive demeaning, $\tilde{\beta}_{RD}$; the dotted lines, labeled Bias corrected, show the results for the $t-$test corresponding to the bias corrected fixed effects estimator, $\tilde{\beta}_{FE}^{+}$. The flat lines indicate the 5% rejection rate. All results are based on $10,000$ repetitions.