

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 914

November 2007

The Stambaugh Bias in Panel Predictive Regressions

Erik Hjalmarsson

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at www.federalreserve.gov/pubs/ifdp/.

The Stambaugh Bias in Panel Predictive Regressions

Erik Hjalmarsson*

Division of International Finance

Federal Reserve Board, Mail Stop 20, Washington, DC 20551, USA

November 29, 2007

Abstract

This paper analyzes predictive regressions in a panel data setting. The standard fixed effects estimator suffers from a small sample bias, which is the analogue of the Stambaugh bias in time-series predictive regressions. Monte Carlo evidence shows that the bias and resulting size distortions can be severe. A new bias-corrected estimator is proposed, which is shown to work well in finite samples and to lead to approximately normally distributed t-statistics. 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. The results are illustrated with an application to predictability in international stock indices.

JEL classification: C22, C23, G1.

Keywords: Panel data; Pooled regression; Predictive regression; Stock return predictability.

*Helpful comments have been provided by Lennart Hjalmarsson, Randi Hjalmarsson, and George Korniotis, as well as seminar participants at the European Summer Meeting of the Econometric Society in Vienna. Tel.: +1-202-452-2426; fax: +1-202-263-4850; email: erik.hjalmarsson@frb.gov. The views in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

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; for instance, Cohen et al. (2003) and Polk et al. (2006) rely on predictive panel data regressions in some of their analyses.

It is well known 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 suffer from a small sample bias and normal t -tests to have the wrong size; this is the so-called Stambaugh (1999) bias in predictive regressions.

In the panel case, with pooled regressions, it turns out that as long as no fixed effects are included, the pooled estimator is unbiased. However, once one allows for fixed effects in the pooled regression, an analogue of the Stambaugh bias is also present in the panel case. This result can be understood in light of the representation of the bias in predictive regressions derived by Stambaugh. Under the assumption that the predictor variable follows an auto-regressive process, he shows that the bias in the OLS estimate of the slope coefficient in the predictive regression is a function of the bias of the OLS estimate of the auto-regressive coefficient in the predictor variable. It is well known that the bias in the OLS estimator of auto-regressive coefficients is more severe if an intercept is included in the regression equation. Therefore, in the time-series case, the Stambaugh bias is less severe if no intercept is included in the predictive regression. This is, of course, mostly of theoretical interest since in almost all empirical applications an intercept is required. The same idea holds in the panel case; but, rather than differentiating between the case of intercept or no intercept, the relevant cases are now a common intercept or individual intercepts, i.e. fixed effects.

In this paper, I propose a simple bias correction to the fixed effects estimator in pooled predictive regressions. An analogue representation of the time-series Stambaugh bias is also derived. It is shown that the asymptotic bias in the fixed effects estimator in the predictive regression can be expressed as a function of the bias in the pooled fixed effects estimator of the auto-regressive coefficient in the predictor variable. The results in this paper complement those of Hjalmarsson (2007), which also studies the bias in the fixed effects

estimator but does not explicitly analyze the connection with the Stambaugh bias in time-series regressions or the direct bias correction procedure suggested here.

The bias-corrected estimator is straightforward to implement. The key parameter on which the bias depends is the auto-regressive root in the regressor variable. The practical implementation of the bias-corrected fixed effects estimator is therefore facilitated by the fact that even though the fixed effects estimator of the auto-regressive coefficient is biased, an alternative unbiased and consistent estimator is readily available. Since the bias-corrected estimator is approximately asymptotically normal, it becomes trivial to perform inference on the slope parameter.

Simulation results show that the bias-corrected estimator works well in finite samples. These simulations also show the importance of controlling for the bias in the panel case. The average rejection rates for the t -test corresponding to the standard fixed effects estimator exceed 75 percent in some cases, under the null hypothesis of no predictability, and for a nominal five percent test.

As an illustration of the methods derived in this paper, I test for stock return predictability in an international panel of returns from 18 different stock indices using the corresponding dividend- and earnings-price ratios, as well as the book-to-market values, as predictors. The empirical results clearly illustrate the theoretical results 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. Thus, both the simulation results and the empirical results clearly show that the Stambaugh bias is at least as important in panel regressions as it is in time-series regressions.

The rest of the paper is organized as follows. Section 2 outlines the panel model and shows the Stambaugh bias in panel predictive regressions. Section 3 describes the bias-corrected estimator and Section 4 illustrates the small sample properties of this estimator, as well as those of the pooled estimator without fixed effects. Section 5 shows the results from the empirical application to stock return predictability and offers a brief conclusion. The Appendix outlines the derivations of the main results.

2 The Stambaugh bias

2.1 Model and assumptions

Consider a panel model with dependent variable $y_{i,t}$, $i = 1, \dots, n$, $t = 1, \dots, T$, and corresponding regressor, $x_{i,t}$. Here, i represents the cross-sectional dimension (e.g. firm or country) and t represents the time-series

dimension. The behavior of $y_{i,t}$ and $x_{i,t}$ are modelled as follows,

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

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

where $\rho = 1 + c/T$. The auto-regressive root of the regressor is parameterized as being local-to-unity, which captures the near unit-root behavior of many predictor variables but is less restrictive than a pure unit-root assumption (e.g. Cavanagh et al. 1995, and Campbell and Yogo, 2006). The model can be seen as a panel analogue of the time-series models studied by Mankiw and Shapiro (1986), Cavanagh et al. (1995), Stambaugh (1999), Lewellen (2004), and Campbell and Yogo (2006), among others.

The innovation processes are assumed to satisfy martingale difference sequences with finite fourth order moments and the regressor $x_{i,t}$ is generally endogenous in the sense that $u_{i,t}$ and $v_{i,t}$ are contemporaneously correlated. That is, let $w_{i,t} = (u_{i,t}, v_{i,t})'$ and $\mathcal{F}_t = \{w_{i,s} | s \leq t, i = 1, \dots, n\}$ be the filtration generated by the innovation processes. Then, for all $i = 1, \dots, n$, and $t = 1, \dots, T$, $E[w_{it} | \mathcal{F}_{t-1}] = 0$, $E[w_{i,t} w'_{i,t} | \mathcal{F}_{t-1}] = \Omega_i = [(\omega_{11i}, \omega_{12i}), (\omega_{12i}, \omega_{22i})]'$ and $\Omega = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Omega_i$. Finally, it is assumed that the innovations are cross-sectionally independent.¹

Let $J_i(r)$ denote the limiting process of the scaled regressor $x_{i,t}$. That is, as $T \rightarrow \infty$, $\frac{x_{i,t=[Tr]}}{\sqrt{T}} \Rightarrow J_i(r)$, where $J_i(r)$, defined in the Appendix, is the standard asymptotic process for a near unit-root variable. Also, let $\underline{J}_i = J_i - \int_0^1 J_i$ be the demeaned version of J_i and let $\Omega_{xx} \equiv E \left[\int_0^1 J_i^2 \right]$, and $\underline{\Omega}_{xx} \equiv E \left[\int_0^1 \underline{J}_i^2 \right]$.

Following the work of Phillips and Moon (1999), results for the panel estimators are derived using sequential limits, which implies first keeping n fixed and letting T go to infinity, and then letting n go to infinity. Such sequential convergence is denoted $(T, n \rightarrow \infty)_{\text{seq}}$.²

2.2 The bias in the fixed effects estimator

Let $\tilde{y}_{i,t} = y_{i,t} - \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T y_{i,t}$ denote the overall demeaned data and let $\underline{y}_{i,t} = y_{i,t} - \frac{1}{T} \sum_{t=1}^T y_{i,t}$ denote the time-series demeaned data. Define $\tilde{x}_{i,t}$ and $\underline{x}_{i,t}$ analogously. The pooled estimator of β when there are no individual effects, i.e. when $\alpha_i \equiv \alpha$ for all i , is given by

$$\hat{\beta}_{Pool} = \left(\sum_{i=1}^n \sum_{t=1}^T \tilde{x}_{i,t-1}^2 \right)^{-1} \left(\sum_{i=1}^n \sum_{t=1}^T \tilde{y}_{i,t} \tilde{x}_{i,t-1} \right). \quad (3)$$

¹In order to highlight the effects of the Stambaugh bias in panel regressions, the effects of cross-sectional dependence are not considered. In certain applications it may be desirable to allow for clustering of the errors either across time for a given individual i , or across individuals (i.e. cross-sectional correlation). As shown by Thompson (2006), it is straightforward to construct standard error estimators that control for such clustering across both time and individuals. His framework could easily be used in the current context and the details are omitted.

²Subject to potential rate restrictions, such as $n/T \rightarrow 0$, these results can generally be shown to hold as n and T go to infinity jointly; technical proofs of such joint convergence is not pursued in the current study, however.

The fixed effects estimator, allowing for individual effects is given by

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^n \sum_{t=1}^T \underline{x}_{i,t-1}^2 \right)^{-1} \left(\sum_{i=1}^n \sum_{t=1}^T y_{i,t} \underline{x}_{i,t-1} \right). \quad (4)$$

As shown by Hjalmarrsson (2007), and outlined in the Appendix, as $(T, n \rightarrow \infty)_{\text{seq}}$,

$$\sqrt{nT} \left(\hat{\beta}_{Pool} - \beta \right) \Rightarrow N \left(0, \omega_{11} \Omega_{xx}^{-1} \right), \quad (5)$$

under the assumption that $\alpha_i \equiv \alpha$ for all i , whereas

$$T \left(\hat{\beta}_{FE} - \beta \right) \rightarrow_p -\omega_{12} \left(\int_0^1 \int_0^r e^{(r-s)c} ds dr \right) \underline{\Omega}_{xx}^{-1}, \quad (6)$$

whenever $\omega_{12} \neq 0$.³ Thus, the estimator without individual effects is asymptotically unbiased and normally distributed; summing up over the cross-section in the pooled estimator eliminates the usual near unit-root asymptotic distributions found in the time-series case. The fixed effects estimator, on the other hand, suffers from a second order bias; in practice, this means that the estimator will exhibit a small sample bias and test statistics will not have standard distributions.

The intuition behind these results is that 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; persistent regressors that are exogenous do not cause any inferential issues. This intuition holds when no individual intercepts are allowed in the specification. The bias in the fixed effects estimation arises because the time-series demeaning induces a correlation between the innovation processes $u_{i,t}$ and the demeaned regressors $\underline{x}_{i,t-1}$; intuitively, this happens because information available after time $t - 1$ is used in the demeaning of $x_{i,t-1}$.⁴

2.3 An alternative representation of the fixed effects bias

In the case of a predictive time-series regression, Stambaugh (1999) shows that the bias in the OLS estimator of the slope coefficient β in equation (1) is a function of the bias in the OLS estimator of the AR -coefficient ρ in equation (2). Here we derive an analogue result for the fixed effects estimator in the panel case.

Note that $-\int_0^1 \left(\int_0^r e^{(r-s)c} ds \right) dr = -(e^c - c - 1)/c^2$ and let $\theta(c) \equiv -(e^c - c - 1)/c^2$. The limiting bias

³In the special case of $\omega_{12} = 0$, it follows easily that $\hat{\beta}_{FE}$ is also asymptotically normally distributed with convergence rate \sqrt{nT} .

⁴Polk et al. (2006) make the same conjecture regarding inference in pooled predictive regressions, namely that independent cross-sectional information dilutes the endogeneity effects, but do not recognize that this intuition fails in the presence of fixed effects. Their regressor is nearly exogenous however, and their empirical conclusions should therefore still be fairly accurate.

of $\hat{\beta}_{FE}$ is thus given by $T^{-1}(\omega_{12}\theta(c)/\underline{\Omega}_{xx})$. Let the fixed effects estimator of ρ be given by

$$\hat{\rho}_{FE} = \left(\sum_{i=1}^n \sum_{t=1}^T x_{i,t-1}^2 \right)^{-1} \left(\sum_{i=1}^n \sum_{t=1}^T x_{i,t} x_{i,t-1} \right). \quad (7)$$

As shown in Moon and Phillips (2000), the bias in $\hat{\rho}_{FE}$ is in fact equal to $T^{-1}(\omega_{22}\theta(c)/\underline{\Omega}_{xx})$. Thus, the limiting bias in the pooled fixed effects estimator of β can be written as a function of the limiting bias in the fixed effects estimator of the auto-regressive parameter ρ . That is,

$$\text{p-lim}_{(T,n \rightarrow \infty)_{\text{seq}}} T \left(\hat{\beta}_{FE} - \beta \right) = \text{p-lim}_{(T,n \rightarrow \infty)_{\text{seq}}} \frac{\omega_{12}}{\omega_{22}} T \left(\hat{\rho}_{FE} - \rho \right). \quad (8)$$

This is the analogue of the expression for the bias in the time-series estimator of β given by Stambaugh (1999). The Stambaugh bias thus carries over directly to pooled regressions, once fixed effects are included.

In the time-series case, it is well known that standard least squares estimates of auto-regressive coefficients close to unity are much more biased when there is an intercept included in the regression. These effects also carry over to a predictive regression with persistent regressors; if no intercept is included in the regression, the Stambaugh bias will be much smaller. Of course, an intercept is required in almost all time-series applications. The panel data therefore gets us halfway: if only a common intercept is included in the pooled regression, the resulting estimator is well behaved, but once individual intercepts are included the bias shows up in the panel case as well.

3 A bias-corrected estimator

3.1 The infeasible estimator

For a known c , a bias-corrected fixed effects estimator is given by

$$\hat{\beta}_{FE}^+ = \left(\sum_{i=1}^n \sum_{t=1}^T x_{i,t-1}^2 \right)^{-1} \left(\sum_{i=1}^n \sum_{t=1}^T y_{i,t} x_{i,t-1} - nT\omega_{12}\theta(c) \right). \quad (9)$$

As shown in the Appendix, as $(T, n \rightarrow \infty)_{\text{seq}}$,

$$\sqrt{nT} \left(\hat{\beta}_{FE}^+ - \beta \right) \Rightarrow N \left(0, \omega_{11}\underline{\Omega}_{xx}^{-1} - (\omega_{12}\theta(c))^2 \underline{\Omega}_{xx}^{-2} \right). \quad (10)$$

Thus, the bias-corrected estimator $\hat{\beta}_{FE}^+$ is asymptotically normally distributed and converges at the same \sqrt{nT} -rate as the standard pooled estimator.

3.2 The feasible estimator

In order to implement $\hat{\beta}_{FE}^+$ in practice, estimates of c and ω_{12} are required. The parameter ω_{12} is the average covariance between the error terms $u_{i,t}$ and $v_{i,t}$ and can be estimated by averaging the estimates of the individual covariances ω_{12i} ; estimates of ω_{12i} can be formed using fitted residuals from either the pooled or time-series estimates of equations (1) and (2).⁵ In practice, the implementation of $\hat{\beta}_{FE}^+$ will not be very sensitive to the exact way of estimating ω_{12} . Rather, the crucial parameter is c , which is more difficult to estimate consistently.

In the time-series case, consistent estimation of c is not possible. That is, ρ can be estimated consistently, but not with enough precision to identify $c = T(\rho - 1)$. This is also the reason why the Stambaugh bias is difficult to correct in practice in time-series regressions; for instance, given the lack of precise knowledge of ρ , Lewellen (2004) suggests a bias correction that leads to conservative tests by imposing a maximum value on the bias under the assumption that $\rho \leq 1$. In the panel data case, it is possible to estimate c consistently.

As discussed previously, the pooled estimate of ρ is biased when including fixed effects. This bias naturally carries over to the estimator of $c = T(\rho - 1)$ as well. However, as discussed in Moon and Phillips (2000), even when there are fixed effects in equation (2), a consistent estimator of c is obtained by simply using the plain pooled estimator without any demeaning of the variables. The estimator of ρ *with* fixed effects is biased for reasons similar to those of the fixed effects estimator of β ; by not demeaning the data, the bias is no longer present. Intuitively, the fixed effects, or intercepts, in equation (2) can be ignored in the estimation of ρ because when the root $\rho = 1 + c/T$ is close to unity, there is enough variation in $x_{i,t}$ that these intercepts are of negligible importance. Therefore, let $\hat{\rho}_{Pool}$ be the plain pooled estimator of ρ ,

$$\hat{\rho}_{Pool} = \left(\sum_{i=1}^n \sum_{t=1}^T x_{i,t-1}^2 \right)^{-1} \left(\sum_{i=1}^n \sum_{t=1}^T x_{i,t} x_{i,t-1} \right), \quad (11)$$

and define the corresponding estimator of c as $\hat{c} = T(\hat{\rho}_{Pool} - 1)$. Moon and Phillips (2000) show that this estimator of c is consistent; again, observe that the data used in estimating c is not time-series demeaned and that demeaning the data in the time-series dimension will lead to a bias in the estimator. A feasible version of $\hat{\beta}_{FE}^+$ is thus given by substituting $\omega_{12}\theta(c)$ with $\hat{\omega}_{12}\theta(\hat{c})$ in equation (9).

Formally, the asymptotic normality of $\hat{\beta}_{FE}^+$ is shown only for the infeasible version of the estimator, which is based on the true (unknown) value of c . Although it is outside the scope of this paper to derive the exact limiting distribution of the feasible version of the estimator, the simulation results below show that inference based on the assumption of normality works well also in this case.

⁵Recall that although both the time-series and pooled fixed effects estimators of β and ρ are generally biased in finite samples, they are still consistent estimators. Estimators of the covariance ω_{12i} based on the fitted residuals will therefore be consistent.

3.3 Practical inference

Finally, in order to perform feasible inference using either $\hat{\beta}_{Pool}$ or $\hat{\beta}_{FE}^+$, one only needs estimates of Ω_{xx} , $\underline{\Omega}_{xx}$, and ω_{11} . Natural estimators of Ω_{xx} and $\underline{\Omega}_{xx}$ are given by $\hat{\Omega}_{xx} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \tilde{x}_{i,t-1}^2$ and $\hat{\underline{\Omega}}_{xx} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \underline{x}_{i,t-1}^2$, respectively. Let $\hat{u}_{i,t}$ be the fitted residuals and $\hat{\omega}_{11} = \frac{1}{n} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T \hat{u}_{i,t}^2$.⁶ An estimate of the variance of $\hat{\beta}_{Pool}$ is thus given by $\hat{\omega}_{11} \hat{\Omega}_{xx}^{-1}$.

Similarly, the natural estimator of $\underline{\Phi}_{ux}$ is given by $\hat{\underline{\Phi}}_{ux}^+ = \hat{\omega}_{11} \hat{\underline{\Omega}}_{xx} - (\hat{\omega}_{12} \theta(\hat{c}))^2$. However, this estimator of $\underline{\Phi}_{ux}$ suffers from the drawback that it is not necessarily positive. Furthermore, subtracting off the term coming from the bias correction of the estimator, without controlling for the possibility that the feasible bias correction induces additional variance in the estimator of β through the sampling error in \hat{c} , may lead to too low an estimate of the variance of the feasible version of $\hat{\beta}_{FE}^+$. That is, as mentioned above, the exact limiting distribution of the feasible estimator is unknown, and it therefore seems reasonable to use an estimator that is more robust. Thus, I propose to use the estimator $\hat{\underline{\Phi}}_{ux}^+ = \hat{\omega}_{11} \hat{\underline{\Omega}}_{xx}$, and estimate the variance of $\hat{\beta}_{FE}^+$ by $\hat{\omega}_{11} \hat{\underline{\Omega}}_{xx}^{-1}$; this will result in a more conservative, i.e. larger, estimate of the variance. In both the pooled and fixed effects cases, therefore, standard estimators can be used to estimate the variance of the estimators.

Since the distributions of $\hat{\beta}_{Pool}$ and $\hat{\beta}_{FE}^+$ are (approximately) asymptotically normal, implementing tests on the slope coefficient becomes trivial. The t -statistic for the estimator $\hat{\beta}_{Pool}$, for instance, will satisfy

$$t_{Pool} = \frac{\hat{\beta}_{Pool} - \beta_0}{\sqrt{\hat{\omega}_{11} \hat{\Omega}_{xx}^{-1} / (nT^2)}} \Rightarrow N(0, 1). \quad (12)$$

The t -statistic t_{FE}^+ corresponding to $\hat{\beta}_{FE}^+$ is constructed in an analogous manner using $\hat{\underline{\Omega}}_{xx}$ instead of $\hat{\Omega}_{xx}$. In the simulations and empirical illustrations below, we also consider the properties of the t -statistic corresponding to the standard fixed effects estimator, t_{FE} , which again is identical to t_{Pool} with $\hat{\Omega}_{xx}$ replaced by $\hat{\underline{\Omega}}_{xx}$; given the above discussion, t_{FE} will not be standard normally distributed unless $\omega_{12} = 0$. However, inference using $\hat{\beta}_{FE}$ and t_{FE} under the normality assumption provides a useful illustration of the biases that occur if one ignores the issues resulting from the endogeneity and persistence of the regressor.

4 Simulation 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

⁶Alternatively, the estimator of ω_{11} could be replaced by a robust variance estimator to allow for heteroskedasticity in the error terms.

normal distributions with mean zero, unit variance, and correlations $\delta = 0, -0.4, -0.7,$ and -0.95 . The slope parameter β is set equal to 0.05 and the local-to-unity parameter c is set to -5 . 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 all set equal to zero. All results are based on 10,000 repetitions.

Three different estimators are considered: the pooled estimator with no fixed effects, $\hat{\beta}_{Pooled}$, the fixed effects estimator, $\hat{\beta}_{FE}$, and the bias-corrected fixed effects estimator, $\hat{\beta}_{FE}^+$. The bias-correction term in the estimator $\hat{\beta}_{FE}^+$ is estimated by $\hat{\omega}_{12}\theta(\hat{c})$, where \hat{c} is the panel estimate of the local-to-unity parameter and $\hat{\omega}_{12}$ is estimated as $n^{-1} \sum_{i=1}^n \hat{\omega}_{12i}$ with $\hat{\omega}_{12i}$ the covariance between the residuals from a time-series estimation of equation (1) and the residuals from the pooled estimation of equation (2).⁷

The results are shown in Figure 1. $\hat{\beta}_{Pooled}$ and $\hat{\beta}_{FE}^+$ are virtually unbiased whereas $\hat{\beta}_{FE}$ exhibits a rather substantial bias when the absolute value of the correlation δ is large. The bias-corrected estimator, $\hat{\beta}_{FE}^+$, has a slightly less peaked distribution than the standard pooled estimator, $\hat{\beta}_{Pooled}$, but overall the bias correction appears to produce good point estimates.

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 . The tests are evaluated under the assumption that the limiting distributions are standard normal; i.e. the null is rejected for absolute test values greater than 1.96. Panel A in Table 1 shows the average sizes of the nominal five percent tests under the null hypothesis of $\beta = 0$ for the two sided t -tests corresponding to the three different estimators considered above. Figure 2 shows the average rejection rates of the five percent two-sided t -tests, evaluating a null of $\beta = 0$ for different values of the true β ; that is, the power curves of the tests. Again, the results are based on 10,000 repetitions.

Apart from the test based on the standard fixed effects estimator, the other two tests perform very well in terms of size, with actual rejection rates very close to five percent in the nominal five percent test. Table 1 and the power curves in Figure 2 clearly show the effects of the second order bias in the fixed effects estimator. The test based on the bias-corrected fixed effects estimator has similar power properties to the test based on the standard pooled estimator.

In practice, the assumption that ρ is identical for all i , i.e. that the regressors all have the same persistence, may seem restrictive. I therefore briefly analyze the robustness of the bias-correction method proposed here to deviations from this assumption. In particular, identical size simulations as those reported in Panel A of Table 1 are shown in Panel B when the individual local-to-unity parameters, c_i , are drawn from a uniform distribution with support $[-20, -2]$. As is seen, the results are very similar, and the bias correction

⁷As mentioned before, the exact estimation procedure for the ω_{12i} s does not play a crucial role in the properties of $\hat{\beta}_{FE}^+$.

appears fairly robust to this generalization. The standard pooled estimator should not be affected, since the assumption of a common parameter c is not needed in deriving its asymptotic result.

In summary, the simulation evidence shows the importance of controlling for the bias arising from fitting individual intercepts in the pooled regression. The bias correction of the fixed effects estimator appears to work well, producing nearly unbiased results and correctly sized tests with good power. In cases where individual effects are not present, the pooled estimator performs well also when the regressors are highly endogenous, as the theory would predict.

5 Empirical illustration and conclusion

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. Lewellen, 2004, and Campbell and Yogo, 2006), and to a lesser degree in international data (e.g. Ang and Bekaert, 2007). 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, though not all forecasting variables or all countries are available for this whole time-period. 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.⁸ 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. All returns are expressed in U.S. dollars, and the dependent variable in the predictive regressions is given by the excess return over the 1-month U.S. T-bill rate. Finally, all data are log-transformed.

The results from the pooled forecasting regressions are shown in Table 2. The estimates of c and the correlation between the innovations in the returns and predictor processes show that the forecasting variables are clearly near unit-root processes and highly endogenous. 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.

⁸Hong Kong is, of course, not a country.

As seen from the estimates based on the bias-corrected fixed effects estimator, $\hat{\beta}_{FE}^+$, significance disappears when controlling for the bias induced by the time-series demeaning in the fixed effects estimator; $\hat{\beta}_{FE}^+$ is implemented in a manner identical to that described in the simulation section above. Overall, the case for stock return predictability in this international data set, using either of the three predictor variables, must be considered very weak.

These results are in line with the extensive study of stock return predictability in international data by Hjalmarrsson (2007), which suggests that the predictive power of valuation ratios is typically weak in international data. Ang and Bekaert (2007) also find in a smaller international sample, covering France, Germany, the UK and the USA over a similar sample period, that the predictive ability of the dividend-price ratio is very weak in all four of these countries.⁹

The empirical results illustrate well 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.

A The asymptotic properties of $\hat{\beta}_{Pool}$, $\hat{\beta}_{FE}$, and $\hat{\beta}_{FE}^+$

Hjalmarrsson (2007) derives the asymptotic properties of $\hat{\beta}_{Pool}$ and $\hat{\beta}_{FE}$ in a similar setting to the one considered here, although he does not consider the bias-corrected estimator $\hat{\beta}_{FE}^+$. The following derivations therefore primarily recollect those found in Hjalmarrsson (2007).

Given the conditions on $u_{i,t}$ and $v_{i,t}$, as $T \rightarrow \infty$, $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} w_{i,t} \Rightarrow B_i(r) = BM(\Omega_i)(r)$, where $B_i(\cdot) = (B_{1i}(\cdot), B_{2i}(\cdot))'$ denotes a two-dimensional Brownian motion. As $T \rightarrow \infty$, $\frac{x_{i,t=[Tr]}}{\sqrt{T}} \Rightarrow J_i(r)$, where $J_i(r) = \int_0^r e^{(r-s)c} dB_{2,i}(s)$. Analogous results hold for the time-series demeaned data, $\underline{x}_{i,t}$, with J_i replaced by \underline{J}_i .

First note that $\frac{\tilde{x}_{i,t=[Tr]}}{\sqrt{T}} = \frac{x_{i,t}}{\sqrt{T}} + O_p\left(\frac{1}{\sqrt{n}}\right) \Rightarrow J_i(r)$, so that the overall demeaning has no asymptotic effects. By standard results as $T \rightarrow \infty$, $\frac{1}{T} \sum_{t=1}^T u_{i,t} \tilde{x}_{i,t-1} \Rightarrow \int_0^1 dB_{1,i} J_i$ and $\frac{1}{T^2} \sum_{t=1}^T \tilde{x}_{i,t-1}^2 \Rightarrow \int_0^1 J_i^2$. Since $B_{1,i}$ and J_i are *iid* across i and $E\left[\int_0^1 dB_{1,i} J_i\right] = 0$, it follows that as $n \rightarrow \infty$, $\frac{1}{n} \sum_{i=1}^n \int_0^1 J_i^2 \rightarrow_p \Omega_{xx}$ and $\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^1 dB_{1,i} J_i \Rightarrow N(0, \Phi_{ux})$ where $\Phi_{ux} \equiv E\left[\left(\int_0^1 dB_{1,i} J_i\right)^2\right]$, by the weak law of large numbers and the central limit theorem, respectively. Thus, as $(T, n \rightarrow \infty)_{\text{seq}}$, $\frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \tilde{x}_{i,t-1}^2 \rightarrow_p \Omega_{xx}$ and $\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T u_{i,t} \tilde{x}_{i,t-1} \Rightarrow N(0, \Phi_{ux})$. It follows that $\sqrt{n}T \left(\hat{\beta}_{Pool} - \beta\right) \Rightarrow N(0, \Phi_{ux} \Omega_{xx}^{-2})$. By the Itô isometry, $\Phi_{ux} = \omega_{11} E\left[\int_0^1 J_i^2\right] = \omega_{11} \Omega_{xx}$ and thus $\Phi_{ux} \Omega_{xx}^{-2} = \omega_{11} \Omega_{xx}^{-1}$.

⁹Ang and Bekaert (2007) do argue that the dividend-price ratio has some predictive ability when considered jointly with the short interest rate.

Similarly, $\frac{1}{n} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \underline{x}_{i,t-1}^2 \rightarrow_p \underline{\Omega}_{xx}$, as $(T, n \rightarrow \infty)_{\text{seq}}$. However, simple calculations yield that $E \left[\int_0^1 dB_{1,i} \underline{J}_i \right] = -\omega_{12} \left(\int_0^1 \int_0^r e^{(r-s)C} ds dr \right) \neq 0$, and it follows that $\frac{1}{n} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T u_{i,t} \underline{x}_{i,t-1} \rightarrow_p -\omega_{12} \left(\int_0^1 \int_0^r e^{(r-s)C} ds dr \right)$, as $(T, n \rightarrow \infty)_{\text{seq}}$. Thus, $T \left(\hat{\beta}_{FE} - \beta \right) \rightarrow_p -\omega_{12} \left(\int_0^1 \int_0^r e^{(r-s)C} ds dr \right) \underline{\Omega}_{xx}^{-1}$ as $(T, n \rightarrow \infty)_{\text{seq}}$.

By removing the mean of the term $\int_0^1 dB_{1,i} \underline{J}_i$ in the bias-corrected estimator $\hat{\beta}_{FE}^+$, the central limit theorem once more applies and, by the same arguments as for $\hat{\beta}_{Pool}$, $\sqrt{n}T \left(\hat{\beta}_{FE}^+ - \beta \right) \Rightarrow N \left(0, \underline{\Phi}_{ux} \underline{\Omega}_{xx}^{-2} \right)$ where $\underline{\Phi}_{ux} = E \left[\left(\int_0^1 dB_{1,i} \underline{J}_i - E \left[\int_0^1 dB_{1,i} \underline{J}_i \right] \right)^2 \right] = E \left[\left(\int_0^1 dB_{1,i} \underline{J}_i - \omega_{12} \theta(c) \right)^2 \right]$. By the Itô isometry, it follows that $\underline{\Phi}_{ux} = \omega_{11} \underline{\Omega}_{xx} - (\omega_{12} \theta(c))^2$ and $\underline{\Phi}_{ux} \underline{\Omega}_{xx}^{-2} = \omega_{11} \underline{\Omega}_{xx}^{-1} - (\omega_{12} \theta(c))^2 \underline{\Omega}_{xx}^{-2}$.

References

- [1] Ang, A., and G. Bekaert, 2007. Stock Return Predictability: Is it There? *Review of Financial Studies* 20, 651-707.
- [2] Campbell, J.Y., and M. Yogo, 2006. Efficient Tests of Stock Return Predictability, *Journal of Financial Economics* 81, 27-60.
- [3] Cavanagh, C., G. Elliot, and J. Stock, 1995. Inference in Models with Nearly Integrated Regressors, *Econometric Theory* 11, 1131-1147.
- [4] Cohen, R., C. Polk, and T. Vuolteenaho, 2003. The Value Spread, *Journal of Finance* 58, 609-641.
- [5] Hjalmarsson, E., 2007. Predicting Global Stock Returns, Working Paper, Federal Reserve Board.
- [6] Lewellen, J., 2004. Predicting Returns with Financial Ratios, *Journal of Financial Economics*, 74, 209-235.
- [7] 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.
- [8] Moon, H.R., and P.C.B. Phillips, 2000. Estimation of Autoregressive Roots near Unity using Panel Data, *Econometric Theory* 16, 927-998.
- [9] Phillips, P.C.B., and H.R. Moon, 1999. Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica* 67, 1057-1111.
- [10] Polk, C., S. Thompson, and T. Vuolteenaho, 2006. Cross-Sectional Forecasts of the Equity Premium, *Journal of Financial Economics* 81, 101-141.
- [11] Stambaugh, R., 1999. Predictive Regressions, *Journal of Financial Economics* 54, 375-421.
- [12] Thompson, S.B., 2006. Simple Formulas for Standard Errors that Cluster by Both Firm and Time, Working Paper, Harvard University.

Table 1: Size results from the Monte Carlo study. The table shows the average rejection rates under the null of $\beta = 0$, for the two-sided 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 repetitions. The sample size used is $T = 100$ and $n = 20$. In Panel A, the local-to-unity parameter, c , is set equal to -5 . In Panel B, separate local-to-unity parameters c_i are drawn for each i from a uniform distribution with support $[-20, -2]$.

Estimator	$\delta = 0.0$	$\delta = -0.4$	$\delta = -0.7$	$\delta = -0.95$
Panel A: $c = -5$				
$\hat{\beta}_{POOL}$	0.050	0.051	0.054	0.050
$\hat{\beta}_{FE}$	0.052	0.211	0.546	0.807
$\hat{\beta}_{FE}^+$	0.054	0.052	0.056	0.054
Panel B: $c_i \sim U[-20, -2]$				
$\hat{\beta}_{POOL}$	0.053	0.051	0.053	0.053
$\hat{\beta}_{FE}$	0.056	0.150	0.362	0.584
$\hat{\beta}_{FE}^+$	0.056	0.054	0.059	0.064

Table 2: Results from the empirical regressions. The table shows the point estimates and corresponding t -statistics (in parentheses) 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 ($b - p$). The first column 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 two columns give the results for the standard fixed effects estimator and the bias-corrected fixed effects estimator, respectively. The final two columns give the estimate of the local-to-unity parameter in the regressors and the average correlation between the innovations to the returns and the regressors, respectively.

Variable	n	T	$\hat{\beta}_{FE}$	$\hat{\beta}_{FE}^+$	\hat{c}_{pool}	$\hat{\delta}$
$d - p$	17	396	0.007 (3.840)	-0.002 (-1.254)	-0.004	-0.771
$e - p$	16	337	0.011 (4.924)	0.000 (-0.089)	0.091	-0.697
$b - p$	18	337	0.008 (4.016)	0.002 (0.948)	-1.538	-0.835

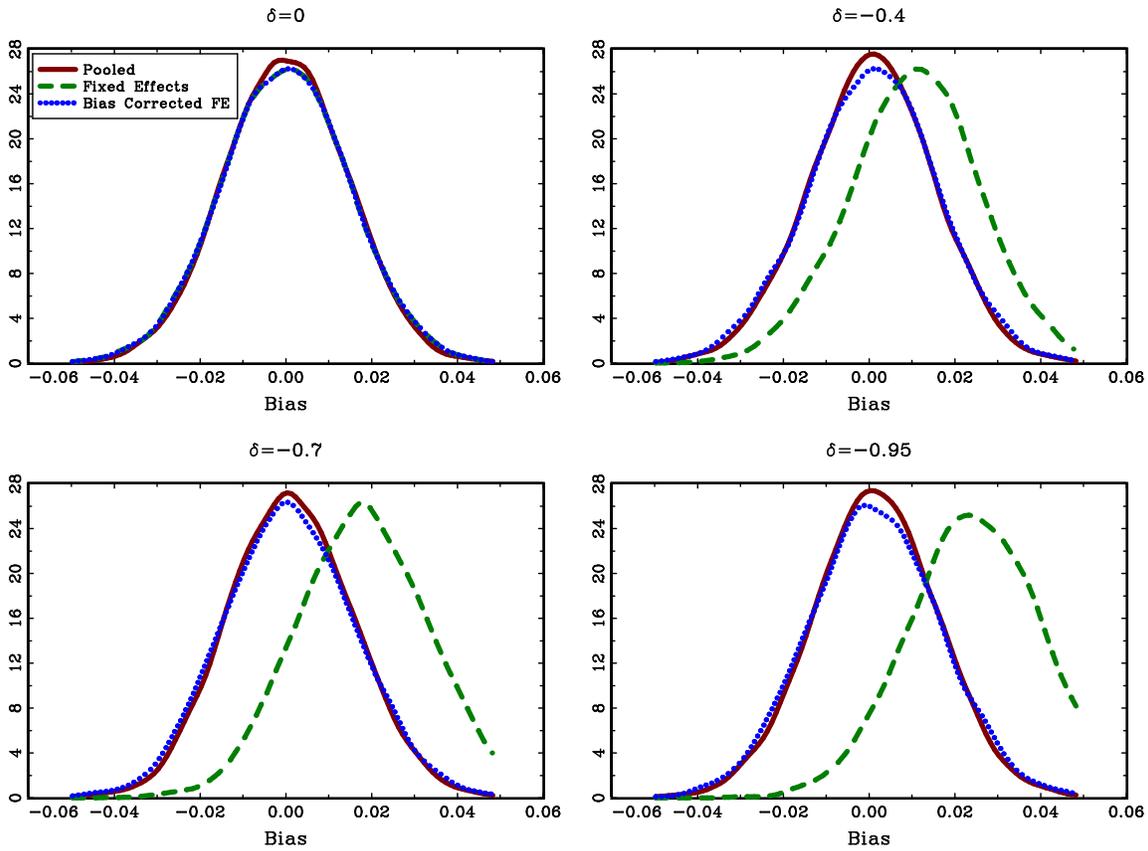


Figure 1: Estimation results from the Monte Carlo study. The graphs show the kernel density estimates of the estimated slope coefficients, for samples with $T = 100$ and $n = 20$. The solid lines, labeled Pooled in the legend, show the results for the standard pooled estimator without individual intercepts, $\hat{\beta}_{Pool}$; the long dashed lines, labeled Fixed Effects, show the results for the standard fixed effects estimator, $\hat{\beta}_{FE}$; the dotted lines, labeled Bias Corrected FE, show the results for the bias-corrected fixed effects estimator, $\hat{\beta}_{FE}^+$. All results are based on 10,000 repetitions.

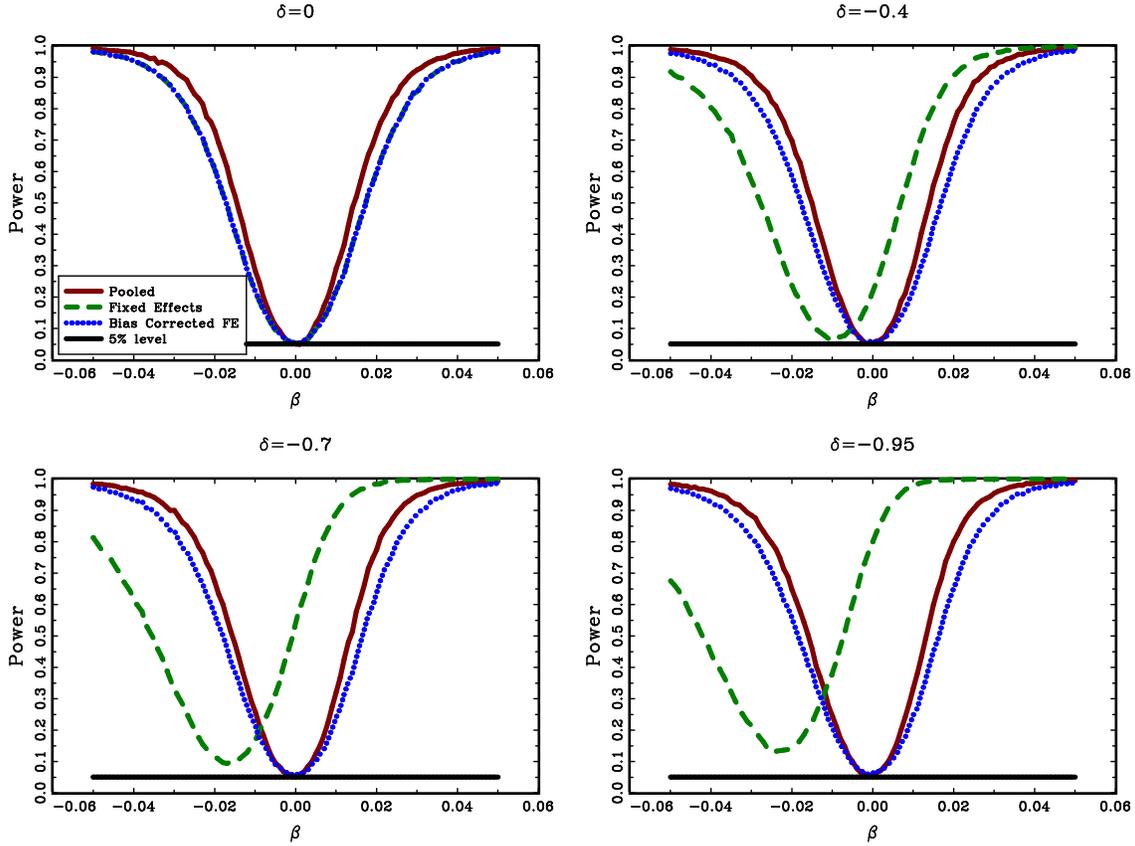


Figure 2: Power results from the Monte Carlo study. The graphs show the average rejection rates for a two-sided 5 percent t -test of the null hypothesis of $\beta = 0$, for samples with $T = 100$, and $n = 20$. The x -axis shows the true value of the parameter β , and the y -axis indicates the average 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 dotted lines, labeled Bias Corrected FE, 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.