Panel Data Inference in Finance: Least-Squares vs Fama-MacBeth

Georgios Skoulakis*
University of Maryland

November 15, 2006

ABSTRACT

Empirical research in finance frequently involves analysis of panel data sets. In corporate finance, we typically encounter panels with large cross sections ("large N"), while in asset pricing, panels with long time series ("large T") are more common. For each case, we examine four estimators: the Least-Squares (LS) and Fama-MacBeth (FM) estimators and their generalized versions. In particular, we offer a rigorous econometric analysis of the FM estimation procedure in the context of panel data. This covers the traditional FM method that is suitable for the large T case, as well as a novel modification of the method appropriate for the large N case. The generalized versions are more efficient, but the corresponding standard errors may be poorly estimated resulting in unreliable t-statistics. An extensive simulation study demonstrates that the estimators under consideration perform remarkably well in moderately small samples. In particular, we provide evidence that both estimation procedures (LS and FM), when properly applied, have comparable performance in the sense that they produce equally reliable t-statistics. Since the two approaches are justified under very similar assumptions, researchers are encouraged to use *both* approaches in their empirical work to ensure the validity of their results.

*Finance Department, Robert H. Smith School of Business, University of Maryland. I thank Daniel Cohen, Thomas George, Caroline Sasseville, Richard Spady, and seminar participants at Northwestern University and the University of Houston for comments and suggestions. I am especially grateful to Ravi Jagannathan, Mitchell Petersen and Costis Skiadas for their support, feedback, and encouragement. Any errors are my responsibility. Comments and suggestions are welcome at gskoulak@rhsmith.umd.edu.

Panel Data Inference in Finance: Least-Squares vs Fama-MacBeth

ABSTRACT

Empirical research in finance frequently involves analysis of panel data sets. In corporate finance, we typically encounter panels with large cross sections ("large N"), while in asset pricing, panels with long time series ("large T") are more common. For each case, we examine four estimators: the Least-Squares (LS) and Fama-MacBeth (FM) estimators and their generalized versions. In particular, we offer a rigorous econometric analysis of the FM estimation procedure in the context of panel data. This covers the traditional FM method that is suitable for the large T case, as well as a novel modification of the method appropriate for the large N case. The generalized versions are more efficient, but the corresponding standard errors may be poorly estimated resulting in unreliable t-statistics. An extensive simulation study demonstrates that the estimators under consideration perform remarkably well in moderately small samples. In particular, we provide evidence that both estimation procedures (LS and FM), when properly applied, have comparable performance in the sense that they produce equally reliable t-statistics. Since the two approaches are justified under very similar assumptions, researchers are encouraged to use *both* approaches in their empirical work to ensure the validity of their results.

1 Introduction

Empirical researchers in finance are frequently confronted with panel data sets. Corporate finance applications include studies of dividend and leverage policies (see, for example, the recent paper by Fama and French (2002)), while asset pricing applications include studies of the cross section of equity returns and their relation to firm characteristics (see, for example, Brennan, Chordia, and Subrahmanyam (1998)). In analyzing financial panel data sets, researchers focus on the significance of specific effects. For instance, a typical exercise in empirical asset pricing amounts to testing whether firm characteristics such as size, book-to-market, and momentum can explain the cross-sectional dispersion in expected equity returns beyond what is explained by a factor asset pricing model. Such a test examines whether the characteristics rewards are significantly different from zero. Often financial economists seek to assess the economic as well as statistical significance of an effect. One can argue that sometimes it is not easy to disentangle the two notions. For example, when analyzing panel data, the empirical evidence is summarized almost exclusively in terms of t-statistics of the coefficients of interest in the panel regression. Quite frequently, t-statistics provide the only means for gauging the importance of various effects. Therefore, empirical researchers must use robust inference methods when computing t-statistics. Accurate computation of t-statistics depends on accurate computation of standard errors. As pointed out by Petersen (2005), despite the common use of panel data sets in finance, there is a wide variation in the methods used to estimate standard errors.

The most crucial aspect for the computation of standard errors is the cross-sectional and serial correlation of the residuals and/or the independent variables in the panel regressions. Using both simulated and real data, Petersen (2005) illustrates the importance of properly taking into account the correlations that might exist in the data. Overlooking such correlations might result in gross underestimation of the standard errors, which, in turn, can cause spurious inflation of the corresponding t-statistics. This very point is forcefully made by Fama and French (2002) who write:

> In our view, however, the most serious problem in the empirical leverage literature is understated standard errors that cloud inferences. Previous work uses either cross-section regressions or panel (pooled time-series and cross-section) regressions. When cross-section regressions are used, the inference problem due to correlation of the residuals across firms is almost always ignored. The articles that use panel regressions ignore both the cross-correlation problem and the bias in the standard errors of regression slopes that arise because the residuals are correlated across years.

Explicit modelling of potential cross-sectional and serial correlations can be a daunting or even hopeless task for the econometrician. This probably explains why fully parametric methods, whether frequentist or Bayesian, are very restrictive and not widely used. An additional drawback of parametric methods is that strong and sometimes unrealistic distributional assumptions, such as normality, need to be made. Instead, researchers opt for methods that are robust with respect to general forms of correlation and require no explicit modelling. Regression-based methods, justified by asymptotic arguments, seem to be the tool of choice for theoretical econometricians, well as empiricists. Detailed textbook accounts of panel data econometric methods can be found in Arellano (2003), Baltagi (2001), Hsiao (2003), and Wooldridge (2002).

A method widely used in the empirical analysis of financial panel data sets is the two-step approach of Fama and MacBeth (1973) (FM henceforth). Occasionally referred to as the cross-sectional regression (CSR) two-step method, the FM method has been used extensively in the analysis of the cross section of stocks returns. Specifically, the FM method is used for the estimation of factor risk premia in the analysis of linear factor models. The first rigorous econometric treatment of the FM approach in the context of factor pricing models is given by Shanken (1985) and Shanken (1992). Jagannathan and Wang (1998) extend the work of Shanken to allow for conditional heteroscedasticity of the time-series regression residuals. Skoulakis (2005) uses the CSR method to provide econometric tests of asset pricing factor models and shows that the method is as efficient as the GMM approach in estimating factor risk premiums. The traditional FM approach essentially consists of two steps. In the first step, for each time period, cross-sectional regressions are used to obtain estimates of the parameters of interest. Then, in the second step, the time series of these estimates are used to obtain final estimates for the parameters and standard errors so that t-statistics can be computed. While the same idea is used in the analysis of both factor models and panel data sets, there are important differences between the two cases. In the context of a factor model, the regressors are the betas (or factor loadings) which are unknown, and thus have to be estimated, and are typically assumed to be time-invariant. The fact that the betas have to be estimated, using time series regressions, gives rise to the well-known error-in-variables problem. On the contrary, in the context of panel data, the regressors are time-varying but directly observable and, thus, there is no error-in-variables problem. The main consideration in the analysis of panel data is to properly take into account the cross-sectional and serial correlations. It is worth emphasizing that the FM method is not analyzed in the econometrics literature. More surprisingly, the method is not even mentioned in standard panel data econometric texts. The only textbook reference for the FM approach that we are aware of is Cochrane (2001). He demonstrates that, under the assumption that the explanatory variables do not vary with time, the FM procedure is essentially equivalent to using OLS. However, the assumption of time-invariant regressors is extremely restrictive and rather unrealistic from an application point of view. One of the contributions of this paper is to provide a rigorous econometric analysis of the FM method under mild conditions in the context of panel data.

In a typical financial panel data set, the cross section consists of individual firms or portfolios while the time-series frequency can be either monthly or yearly. In this paper, following Petersen (2005), we refer to the cross-sectional units as firms. The data sets typically used in empirical work in finance fall into two categories: (i) large cross section, and (ii) long time series. We use N to denote the size of the cross section and T to denote the length of the time series. We therefore focus on the two cases of primary interest: large N and large T. Data availability typically dictates which case is more appropriate for analyzing a given data set. For instance, the large T case is suitable when monthly data are available. This case is more common in the empirical asset pricing literature where researchers analyze return data consisting of a small cross section but long time series. The large N case is more frequently encountered in empirical corporate finance where there is a large number of firms, but data are available only at the yearly frequency. Our approach, following most of the econometric panel data literature, is based on asymptotic arguments as N or T approach infinity. We take this route to provide tools for inference without imposing strong distributional assumptions. For each case, we consider the LS and FM estimators, along with their generalized versions that incorporate the use of general weighting matrices other than the identity matrix. Under flexible assumptions about the dependence structure of the regressors and the disturbances, allowing for both firm and time effects, we show consistency and asymptotic normality of all four estimators, and we describe how to construct asymptotically valid t-statistics by estimating their covariances consistently.

In the large T case with generic serial dependence, we demonstrate that both LS and FM methods yield accurate asymptotic approximations. In particular, standard econometric tools for dealing with serial correlation, namely heteroscedasticity and autocorrelation consistent (HAC) estimators (Newey-West (1987)), are shown in our simulations to produce reliable t-statistics. In the large N case, we work under the assumption of cross-sectional independence besides the presence of firm-invariant time effects. Under this assumption, both LS and FM provide accurate asymptotic inference. We use an intermediate *demeaning* step to eliminate any firm-invariant time effects before constructing the various estimators. Our simulations show that this demeaning step turns out to be crucial if such effects are indeed present in the cross section.

The remaining issue for the large N case, which is beyond the scope of this paper, is that the data might exhibit other forms of cross-sectional dependence not captured by the firm-invariant time effects. This case appears to be of importance for the analysis of panels with a large number of firms but a rather small number of time-series observations (sometimes as small as 10) as encountered in corporate finance studies. Regression-based methods, such as the ones considered in this paper, do not seem adequate for dealing with generic cross-sectional dependence in panel data. Spatial methods appear to be a promising alternative for dealing with this issue. Examples of such methods include the spatial GMM approach pioneered by Conley (1999), which requires user-defined distances between cross-sectional units, and the more recent work by Kapoor, Kelejian, and Prucha (2002), which requires a user-defined contiguity matrix for determining cross-sectional correlations[1]. Although spatial methods have a sound theoretical foundation, their application is not straightforward. The implementation is data and context specific, requires a lot of input from the researcher, and might not be robust with respect to the input choice. Despite these difficulties, we should emphasize that spatial methods provide a sound econometric framework for dealing with cross-sectional dependence, and their usefulness in empirical finance is worth exploring.

Another approach for dealing with spatially dependent panel data has been developed by Driscoll and Kraay (1998) in a GMM context. Their method amounts to collapsing the entire panel into a single time series by taking cross-sectional averages. They provide formal asymptotic inference results when T goes to infinity and for arbitrarily large N. The LS estimator we use for the large T case is essentially equivalent to the Driscoll and Kraay estimator as applied to our data generating process. Our simulation evidence shows that the behavior of the t-statistics based on this estimator indeed does not depend on the cross-section size. The results are almost identical for N ranging from 250 to 1000. However, the empirical distribution of t-statistics can deviate substantially from the standard normal if T is too small. Therefore, this type of estimator is not appropriate for panel data sets with a small number of time series observations such as those typically used in empirical corporate finance with yearly observations.

This paper contributes to the existing literature in a number of ways. First, we provide a rigorous econometric justification for the use of the FM two-pass procedure in the context of panel data. A natural, although novel, modification of the FM method is developed to make it applicable to the large N case. We explain theoretically and show through simulation the importance of demeaning in the presence of firm-invariant time effects. Moreover, we offer a comparison of the estimators under examination in terms of efficiency. Finally, we provide ample simulation evidence demonstrating that both LS and FM estimators perform exceptionally well in small samples in the sense that the distribution of the corresponding t-statistics is a good approximation of the limiting standard normal distribution. Empirical researchers dealing with panel data are strongly encouraged to use both approaches to ensure the validity of their results.

The rest of the paper is organized as follows. In Section 2 we discuss estimation in the large N case, which is of particular interest in corporate finance. Section 3 is devoted to the large T case,

[1]Detailed accounts of spatial methods can be found in Anselin (1988) and Cliff and Ord (1981).

which is more commonly encountered in asset pricing. In both Sections 2 and 3, we report the findings from extensive simulation experiments examining the small-sample behavior of the various estimators. In the last section, we offer some concluding remarks.

2 Large Cross Section

Before proceeding with our formal analysis, it is instructive to provide some intuition why it is very important to properly take into account potential cross-sectional and serial correlations when constructing t-statistics in the analysis of panel data. Ignoring such correlations typically leads to underestimation of standard errors and subsequent spurious inflation of t-statistics. The following quote from Fama and French (2002) illustrates the importance of the issue:

> Cross-correlation almost always inflates the standard errors of the average slopes in the dividend and leverage regressions by a factor of more than two and often more than five. Autocorrelation sometimes produces an additional increase of about 250% in the standard errors of the average slopes. In short, the standard errors in most previous capital structure tests are almost surely understated by large unknown amounts. In our view, this means that inferences from previous tests (the things we think we know about capital structure) lack credibility until they are confirmed by robust methods.

To quantify the effect of standard error underestimation, we offer the following discussion, which, although informal, captures the main idea. Suppose we wish to estimate the parameter θ and compute standard errors so that we can derive t-statistics. In a typical situation, the estimator θ_N based on a sample of size N satisfies

$$\sqrt{N}(\theta_N - \theta_0) \simeq \frac{1}{\sqrt{N}} \sum_{n=1}^{N} x_n = \sqrt{N}\bar{x}_N$$

where $\simeq$ indicates approximation in a loose sense, θ_0 is the true value of the parameter, and $\{x_n\}$ is a suitable random sequence. Given the above expression, we proceed by applying a version of the central limit theorem for x_n to compute asymptotically valid t-statistics and confidence intervals. To show the potential pitfalls due to the presence of correlations, we consider the following hypothetical, and admittedly extreme, scenario. Assume that $x_n = y_n + u$, where y_n are i.i.d. with mean zero and variance σ_y^2, for $n = 1, \ldots, N$, and u is normally distributed with mean zero and variance σ_u^2. Under the assumption that u is independent of the sequence y_n, we can apply the central limit theorem on y_n and obtain that, for large N, $\sqrt{N}\bar{x}_N = \sqrt{N}\bar{y}_N + \sqrt{N}u \simeq N(0, \sigma_y^2 + N\sigma_u^2)$ or $\theta_N \simeq N\left(\theta_0, s_c^2\right)$, where $s_c^2 = \left(\sigma_y^2 + N\sigma_u^2\right)/N$ (the subscript c stands for *correct*). If we ignored the correlation in x_n due to the presence of the common shock u and treat x_n as i.i.d., we would conclude that $\theta_N \simeq N\left(\theta_0, s_i^2\right)$, where $s_i^2 = \left(\sigma_y^2 + \sigma_u^2\right)/N$ (the subscript i stands for *incorrect*). If we define the variance ratio $\pi = \sigma_y^2/\sigma_u^2$, it follows that the standard error ratio s_i/s_c equals $\sqrt{(\pi+1)/(\pi+N)}$, which is smaller than 1. Therefore, unless the ratio π is much larger than N, the t-statistic is spuriously inflated if we use the incorrect standard error s_i instead of the correct standard error s_c. In what follows we show how to construct asymptotically valid t-statistics using both the LS and FM estimators under suitable assumptions.

2.1 Model description

The standard linear regression model used in the context of panel data is

$$y_{it} = x_{it}'b + \varepsilon_{it},\ i = 1, \ldots, N,\ t = 1, \ldots, T$$

where y_{it} is the independent variable, and x_{it} is a vector of explanatory variables. There are available data for N cross-sectional units and T time periods. For the purposes of empirical analysis, we are interested in methods for computing t-statistics for the parameter b. It is well-known that the presence of cross-sectional and/or serial dependence of the regressors and the disturbances can lead to erroneous inferences if traditional OLS standard errors are used (see, for instance, Moulton (1990) and Wooldridge (2003)). Therefore, it is important to identify procedures for obtaining t-statistics that are robust with respect to different dependence structures. Throughout, we maintain the assumption that x_{it} and ε_{it} contain both time and firm effects. Specifically, we follow Petersen (2005) and postulate the following decomposition of regressors and disturbances

$$\begin{aligned} x_{it} &= \xi_t + \mu_i + w_{it}, \\ \varepsilon_{it} &= \phi_t + \omega_i + v_{it}. \end{aligned}$$

The terms ξ_t and ϕ_t capture firm-invariant time effects, while the terms μ_i and ω_i capture time-invariant firm effects. Note that the structure is rather general and covers important cases of practical interest. In particular, it covers all cases considered in Petersen (2005). Besides the presence of μ_i and ω_i, the terms w_{it} and v_{it} account for other sources of serial correlation for a specific firm. To simplify the notation, let us define

$$\begin{aligned} z_{it} &= \mu_i + w_{it}, \\ u_{it} &= \omega_i + v_{it}. \end{aligned}$$

In other words, the terms z_{it} and u_{it} absorb the time-invariant firm effect terms μ_i and ω_i, respectively. In the following analysis, we allow z_{it} and u_{it} to have serial correlation and heteroscedasticity of unspecified form; thus, there is no loss of generality in working with z_{it} and u_{it}. To facilitate a concise presentation, we resort to standard vector-matrix notation. We group observations by firm, and for each firm i we let

$$y^i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix} \ (T \times 1), \ X^{i\prime} = \begin{bmatrix} x'_{i1} \\ \vdots \\ x'_{iT} \end{bmatrix} \ (T \times K), \ \varepsilon^i = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix} \ (T \times 1)$$

so that the regression equation is written as

$$y^i = X^{i\prime} b + \varepsilon^i, \ i = 1, \ldots, N. \tag{1}$$

Furthermore, the following decomposition holds

$$X^{i\prime} = \Xi' + Z^{i\prime}, \ \varepsilon^i = \phi + u^i,$$

where

$$\Xi' = \begin{bmatrix} \xi'_1 \\ \vdots \\ \xi'_T \end{bmatrix}, \ Z^{i\prime} = \begin{bmatrix} z'_{i1} \\ \vdots \\ z'_{iT} \end{bmatrix}, \ \phi = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_T \end{bmatrix}, \ u^i = \begin{bmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{bmatrix}.$$

Without loss of generality, we assume that $E\left[Z^i\right] = 0_{K \times T}$ since any nonzero mean can be absorbed by the term Ξ. Similarly, we can assume $E[u^i] = 0_T$. Given the structure above, the challenge is to be able to handle the multiple sources of dependence that might exist. We propose the following simple strategy. The first goal is to eliminate the firm-invariant time effects Ξ and ϕ. This can be achieved by demeaning, which is a standard intermediate step in the econometric analysis of panel data. Specifically, for each time period t, we subtract the corresponding cross-sectional mean from each observation. Using the generic notation

$$\bar{r}_N = \frac{1}{N} \sum\nolimits_{i=1}^{N} r^i$$

we obtain

$$y^i - \bar{y}_N = (X^i - \bar{X}_N)' b + (\varepsilon^i - \bar{\varepsilon}_N) = (Z^i - \bar{Z}_N)' b + (u^i - \bar{u}_N)$$

since

$$X^i - \bar{X}_N = \left(\Xi + Z^i\right) - \left(\Xi + \bar{Z}_N\right) = Z^i - \bar{Z}_N$$

and

$$\varepsilon^i - \bar{\varepsilon}_N = \left(\phi + u^i\right) - \left(\phi + \bar{u}_N\right) = u^i - \bar{u}_N.$$

We can succinctly rewrite the last equation as

$$\tilde{y}^i = \tilde{X}^{i\prime} b + \tilde{\varepsilon}^i = \tilde{Z}^{i\prime} b + \tilde{u}^i, \; i = 1, \ldots, N, \tag{2}$$

where we employ the generic notation

$$\tilde{r}^i = r^i - \bar{r}_N.$$

Equation (2) will be the starting point in obtaining various estimators of the parameter b. At this point, we should stress the importance of the demeaning step. Clearly, if there are no firm-invariant time effects, the demeaning step is redundant. However, in the presence of firm-invariant time effects, our simulation evidence shows that omitting the demeaning step might result in severe deterioration in the performance of all estimators under consideration.

2.2 Least-squares estimators

The least-squares (LS) estimator obtained from equation (2) is

$$b_N^{\mathtt{LS}} = \left(\frac{1}{N} \sum\nolimits_{i=1}^N \tilde{X}^i \tilde{X}^{i\prime} \right)^{-1} \left(\frac{1}{N} \sum\nolimits_{i=1}^N \tilde{X}^i \tilde{y}^i \right).$$

Note that we deviate from the standard terminology. The LS estimator here *does not* refer to the least-squares estimator obtained from the original regression equation (1); rather, it refers to the least-squares estimator resulting from equation (2). We are interested in identifying conditions, preferably not very restrictive, under which the LS estimator possesses the desired properties of consistency and asymptotic normality. Note that the regression equation (2) implies the expression

$$b_N^{\mathtt{LS}} = b + \left(\frac{1}{N} \sum\nolimits_{i=1}^N \tilde{Z}^i \tilde{Z}^{i\prime} \right)^{-1} \left(\frac{1}{N} \sum\nolimits_{i=1}^N \tilde{Z}^i \tilde{u}^i \right).$$

It turns out that, under suitable conditions, we have

$$\frac{1}{N} \sum\nolimits_{i=1}^N \tilde{Z}^i \tilde{Z}^{i\prime} = \frac{1}{N} \sum\nolimits_{i=1}^N Z^i Z^{i\prime} + o_p(1)$$

and

$$\frac{1}{\sqrt{N}} \sum\nolimits_{i=1}^N \tilde{Z}^i \tilde{u}^i = \frac{1}{\sqrt{N}} \sum\nolimits_{i=1}^N Z^i u^i + o_p(1),$$

where by $o_p(1)$ we denote terms that converge to zero in probability as N tends to infinity. It follows that we can write

$$\sqrt{N}(b_N^{\mathtt{LS}} - b) = \left(\frac{1}{N} \sum\nolimits_{i=1}^N Z^i Z^{i\prime} \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum\nolimits_{i=1}^N Z^i u^i \right) + o_p(1).$$

Therefore, we observe that the asymptotic covariance of $b_N^{\mathtt{LS}}$ is given by

$$Avar\left[b_N^{\mathtt{LS}}\right] = (\Sigma^{\mathtt{Z}})^{-1}\Sigma^{\mathtt{Zu}}(\Sigma^{\mathtt{Z}})^{-1} \tag{3}$$

where

$$\Sigma^{\mathtt{Z}} = E\left[Z^i Z^{i\prime}\right] \tag{4}$$

and

$$\Sigma^{\mathtt{Zu}} = Var\left[Z^i u^i\right] = E\left[Z^i u^i u^{i\prime} Z^{i\prime}\right]. \tag{5}$$

Formally, the preceding steps are justified under the following assumption:

Condition LSLN. (i) The random sequence $\{s_i = (Z^i, u^i),\ i = 1, 2, \ldots\}$ is i.i.d., and the central limit theorem applies to functions of s_i. (ii) The matrix $E\left[Z^i Z^{i\prime}\right]$ has full rank equal to K. (iii) The orthogonality condition $E\left[Z^i u^i\right] = 0_K$ holds.

Under Condition LSLN, $b_N^{\mathtt{LS}}$ is consistent and asymptotically normal as $N \to \infty$. A comment about Condition LSLN is in order. The i.i.d. assumption LSLN(i) is stronger than necessary. One can obtain essentially equivalent results under the assumption that (Z^i, u^i), $i = 1, 2, \ldots$ are independent (but not identically distributed) and that suitable moment conditions are satisfied. This approach has been developed by White (1980, 2001). We do not pursue it here since it would not change the nature of the results but only increase notational complexity.

In order to compute t-statistics, we need a consistent estimator of the asymptotic covariance matrix $Avar[b_N^{\mathtt{LS}}] = (\Sigma^{\mathtt{Z}})^{-1}\Sigma^{\mathtt{Zu}}(\Sigma^{\mathtt{Z}})^{-1}$. Since $\tilde{X}^i = \tilde{Z}^i$ and $\tilde{\varepsilon}^i = \tilde{u}^i$, it follows that such an estimator is provided by

$$V_{\mathtt{b},N}^{\mathtt{LS}} = (\Sigma_N^{\mathtt{X}})^{-1}\Sigma_N^{\mathtt{X}\varepsilon}(\Sigma_N^{\mathtt{X}})^{-1},$$

where

$$\Sigma_N^{\mathtt{X}} = \frac{1}{N}\sum\nolimits_{i=1}^N \tilde{X}^i \tilde{X}^{i\prime},$$

$$\Sigma_N^{\mathtt{X}\varepsilon} = \frac{1}{N}\sum\nolimits_{i=1}^N \tilde{X}^i \hat{\varepsilon}^i \hat{\varepsilon}^{i\prime} \tilde{X}^{i\prime}$$

$$\hat{\varepsilon}^i = y^i - X^{i\prime} b_N,$$

and b_N is a consistent estimator of b. Note that we have made no assumptions regarding the covariance matrix of ε^i (u^i) or conditional homoscedasticity of ε^i (u^i) given X^i (Z^i). Thus, obtaining t-statistics using $b_N^{\mathtt{LS}}$ and $V_{\mathtt{b},N}^{\mathtt{LS}}$ is robust to serial autocorrelation and conditional heteroscedasticity for a specific firm. The estimator $V^{\mathtt{LS}}_{\mathtt{b},N}$ is a robust (or sometimes referred to as clustered) covariance estimator of the type first introduced by Arellano (1987). Using this type of robust covariance estimator is also referred to in the empirical corporate finance literature as using Rogers standard errors (see Petersen (2005)).

The LS estimator is a special case of a generalized least-squares (GLS) estimator when the weighting matrix is simply the identity matrix. In general, one can use a $T \times T$ weighting matrix Q, which is assumed to be symmetric and positive definite. The GLS estimator is feasible when Q is known. If Q is not known, one needs to use a consistent estimator Q_N of Q. In the following discussion, we assume that such an estimator Q_N is available and that Q_N is symmetric and positive definite. The GLS estimator is expressed as

$$b^{\mathtt{GLS}}_N = \left(\frac{1}{N}\sum\nolimits_{i=1}^{N} \tilde{X}^i Q_N \tilde{X}^{i\prime}\right)^{-1} \left(\frac{1}{N}\sum\nolimits_{i=1}^{N} \tilde{X}^i Q_N \tilde{y}^i\right).$$

It follows that the asymptotic covariance of $b^{\mathtt{GLS}}_N$ is given by

$$Avar\left[b^{\mathtt{GLS}}_N\right] = (\Sigma^{\mathtt{Z}}_{\mathtt{Q}})^{-1}\Sigma^{\mathtt{Zu}}_{\mathtt{Q}}(\Sigma^{\mathtt{Z}}_{\mathtt{Q}})^{-1}, \tag{6}$$

where

$$\Sigma^{\mathtt{Z}}_{\mathtt{Q}} = E\left[Z^i Q Z^{i\prime}\right] \tag{7}$$

and

$$\Sigma^{\mathtt{Zu}}_{\mathtt{Q}} = Var\left[Z^i Q u^i\right] = E\left[Z^i Q u^i (Z^i Q u^i)'\right]. \tag{8}$$

Establishing the asymptotic properties of the GLS estimator requires strengthening the assumptions made so far. Specifically, one has to replace Condition LSLN with the following slightly stronger assumption:

Condition GLSLN. (i) Condition LSLN(i) holds. (ii) The matrix $E\left[Z^i Q Z^{i\prime}\right]$ has full rank equal to K. (iii) The orthogonality condition $E\left[Z^i \otimes u^i\right] = 0_{TK\times T}$ holds.

Formally, we summarize the asymptotic properties of the LS and GLS estimators in the following theorem. Its proof is omitted as it is similar to the proof of Theorems 7.1 and 7.2 in Wooldridge (2002). The details are available upon request.

Theorem 1 *(a) Under Condition LSLN, the LS estimator $b^{\mathtt{LS}}_N$ is consistent and asymptotically normal as $N \to \infty$, with asymptotic covariance matrix given by expression (3).*

(b) Under Condition GLSLN, the GLS estimator $b^{\mathtt{GLS}}_N$ is consistent and asymptotically normal as $N \to \infty$, with asymptotic covariance matrix given by expression (6).

In general, it is not clear how to choose the weighting matrix Q optimally in terms of asymptotic efficiency. Nevertheless, there is one case of interest in which this is feasible. Under the cross-sectional homoscedasticity condition $E\left[u^i u^{i\prime}|Z^i\right] = \Sigma^{\mathtt{u}}$, the optimal choice of Q equals $(\Sigma^{\mathtt{u}})^{-1}$.[2] In this case, we can use the law of iterated expectations to obtain

$$Avar[b_N^{\mathtt{GLS}}] = (E\left[Z^i(\Sigma^{\mathtt{u}})^{-1}Z^{i\prime}\right])^{-1}.$$

A consistent estimator of $Avar[b_N^{\mathtt{GLS}}]$ in this case is

$$V_{\mathtt{b},N}^{\mathtt{GLS}} = \left(\frac{1}{N}\sum\nolimits_{i=1}^{N} \tilde{X}^i(\Sigma_N^{\mathtt{u}})^{-1}\tilde{X}^{i\prime}\right)^{-1},$$

where

$$\Sigma_N^{\mathtt{u}} = \frac{1}{N}\sum\nolimits_{i=1}^{N} \hat{\varepsilon}^i\hat{\varepsilon}^{i\prime}.$$

Under conditional homoscedasticity, the GLS estimator will be more efficient than LS. However, higher efficiency does not come without a cost. Condition GLSLN needs to hold instead of the weaker Condition LSLN. Moreover, under conditional heteroscedasticity, the form of the optimal weighting matrix is, generally speaking, unknown. Finally, another very important issue from a practical perspective is the small sample behavior of the GLS estimator. As demonstrated in subsection 2.5, in finite samples with T being relatively large, the t-statistics based on the GLS estimator exhibit very unstable behavior resulting in serious size distortions. It appears that the source of this instability is the poor estimation of the disturbance covariance matrix. Therefore, one has to be cautious when using the GLS procedure, unless the time-series length is significantly smaller than the size of the cross section.

2.3 Fama-MacBeth estimators

The original Fama-MacBeth two-step method was developed for data sets with long time series. However, it turns out that the FM procedure can easily be modified so that it can be used for panel data sets with large N. The modified FM estimator is shown to be consistent and asymptotically normal as $N \to \infty$. Although the idea behind the modified estimator is simple, it has neither been used in applications nor analyzed econometrically. To our knowledge, the only exception is the empirical application in the recent paper by Coval and Shumway (2005), where they "conduct trader-by-trader regressions and average the coefficients across traders"as well as "day-by-day regressions and average the coefficients across days." We provide a more detailed description of the

[2]For a detailed discussion, see Section 7.5 in Wooldridge (2002).

modified FM procedure in what follows. Instead of running a cross-sectional regression for a fixed time t as originally done by FM, we run a time-series regression for a fixed firm i and obtain an estimate of b based only on observations of firm i. Recall that, after demeaning, we have

$$\tilde{y}^i = \tilde{X}^{i\prime} b + \tilde{\varepsilon}^i = \tilde{Z}^{i\prime} b + \tilde{u}^i.$$

Therefore, an estimate of b based only on observations of firm i can be obtained as follows:

$$\hat{b}^i = (\tilde{X}^i \tilde{X}^{i\prime})^{-1} \tilde{X}^i \tilde{y}^i.$$

Note that $\hat{b}^i = b + (\tilde{Z}^i \tilde{Z}^{i\prime})^{-1} \tilde{Z}^i \tilde{u}^i$. The FM estimator is then obtained by averaging over i

$$b_N^{\mathtt{FM}} = \frac{1}{N} \sum\nolimits_{i=1}^N \hat{b}^i = \frac{1}{N} \sum\nolimits_{i=1}^N (\tilde{X}^i \tilde{X}^{i\prime})^{-1} \tilde{X}^i \tilde{y}^i.$$

Although the LS and FM estimators are obtained through different procedures, there is a clear link between the two. Note that we can express the LS estimator as

$$b_N^{\mathtt{LS}} = \frac{1}{N} \sum\nolimits_{i=1}^N \left(\frac{1}{N} \sum\nolimits_{i=1}^N \tilde{X}^i \tilde{X}^{i\prime} \right)^{-1} \tilde{X}^i \tilde{y}^i = \frac{1}{N} \sum\nolimits_{i=1}^N W_N^i \hat{b}^i,$$

where

$$W_N^i = \left(\frac{1}{N} \sum\nolimits_{i=1}^N \tilde{X}^i \tilde{X}^{i\prime} \right)^{-1} (\tilde{X}^i \tilde{X}^{i\prime}), \; i = 1, \ldots, N.$$

It follows that $\frac{1}{N} \sum_{i=1}^N W_N^i = I_K$ ($K \times K$ identity matrix). Therefore, the LS estimator is a particular weighted average of the individual FM estimates $\hat{b}^i$ with weights given by W_N^i. On the other hand, the FM estimator is simply the equally weighted average of the $\hat{b}^i$'s.

To guarantee the desired asymptotic properties for the FM estimator, we need the following assumption:

Condition FMLN. (i) Condition LSLN(i) holds. (ii) There exists a $\delta > 0$ arbitrarily small, such that, for all $i = 1, 2, \ldots$, the smallest eigenvalue of the random matrix $Z^i Z^{i\prime}$, denoted by $m(Z^i Z^{i\prime})$, is greater than δ. (iii) The orthogonality condition $E\left[u^i | Z^i\right] = 0_T$ holds.

Condition FMLN(ii) is slightly stronger that Condition LSLN(ii) (that the matrix $E[Z^i Z^{i\prime}]$ has full rank equal to K) which is equivalent to the assumption that the smallest eigenvalue of $E[Z^i Z^{i\prime}]$ is positive. Condition FMLN(ii) guarantees that the individual FM estimates are feasible. Under Condition FMLN, the FM estimator $b_N^{\mathtt{FM}}$ is consistent and asymptotically normal. The asymptotic covariance matrix of $b_N^{\mathtt{FM}}$ is given by

$$Avar[b_N^{\mathtt{FM}}] = E[(P^i - E[P^i]) u^i u^{i\prime} (P^i - E[P^i])'], \tag{9}$$

where

$$P^i = (Z^i Z^{i\prime})^{-1} Z^i. \tag{10}$$

If $E\left[P^i\right] = 0_{K\times T}$, the formula for the FM variance becomes

$$Avar[b_N^{\mathtt{FM}}] = E\left[(P^i u^i)(P^i u^i)'\right].$$

Note that the condition $E\left[P^i\right] = 0_{K\times T}$ is satisfied when Z^i is symmetric around its zero mean. Following the original suggestion of FM, an estimate of the asymptotic covariance $Avar[b_N^{\mathtt{FM}}]$ is obtained by treating the individual estimates $\hat{b}_i$ as an i.i.d. sample. This gives rise to the variance estimator

$$V_{\mathtt{b},N}^{\mathtt{FM}} = \frac{1}{N}\sum\nolimits_{i=1}^{N}(\hat{b}^i - b_N^{\mathtt{FM}})(\hat{b}^i - b_N^{\mathtt{FM}})'.$$

Since $b_N^{\mathtt{FM}} \xrightarrow{p} b$, we have

$$\begin{aligned} V_{\mathtt{b},N}^{\mathtt{FM}} &= \frac{1}{N}\sum\nolimits_{i=1}^{N}\hat{b}^i\hat{b}^{i\prime} - b_N^{\mathtt{FM}} b_N^{\mathtt{FM}\prime} \\ &= bb' + b\left[\frac{1}{N}\sum\nolimits_{i=1}^{N}(\tilde{Z}^i\tilde{Z}^{i\prime})^{-1}\tilde{Z}^i\tilde{u}^i\right]' + \left[\frac{1}{N}\sum\nolimits_{i=1}^{N}(\tilde{Z}^i\tilde{Z}^{i\prime})^{-1}\tilde{Z}^i\tilde{u}^i\right]b' \\ &\quad + \frac{1}{N}\sum\nolimits_{i=1}^{N}(\tilde{Z}^i\tilde{Z}^{i\prime})^{-1}\tilde{Z}^i\tilde{u}^i\tilde{u}^{i\prime}\tilde{Z}^{i\prime}(\tilde{Z}^i\tilde{Z}^{i\prime})^{-1} - bb' + o_p(1). \end{aligned}$$

Arguing as in the proof of Theorem 2 (see Appendix), we obtain

$$\begin{aligned} &\frac{1}{N}\sum\nolimits_{i=1}^{N}(\tilde{Z}^i\tilde{Z}^{i\prime})^{-1}\tilde{Z}^i\tilde{u}^i = o_p(1), \\ &\frac{1}{N}\sum\nolimits_{i=1}^{N}(\tilde{Z}^i\tilde{Z}^{i\prime})^{-1}\tilde{Z}^i\tilde{u}^i\tilde{u}^{i\prime}\tilde{Z}^{i\prime}(\tilde{Z}^i\tilde{Z}^{i\prime})^{-1} = \frac{1}{N}\sum\nolimits_{i=1}^{N}P^i u^i(P^i u^i)' + o_p(1) \end{aligned}$$

and so

$$V_{\mathtt{b},N}^{\mathtt{FM}} \xrightarrow{p} E\left[P^i u^i (P^i u^i)'\right].$$

It follows that, under the assumption $E\left[P^i\right] = 0_{K\times T}$, the Fama-MacBeth variance estimator $V_{\mathtt{b},N}^{\mathtt{FM}}$ is a consistent estimator of $Avar[b_N^{\mathtt{FM}}]$. As mentioned above, this assumption is satisfied when Z^i is symmetric. The simulation evidence in subsection 2.5 demonstrates that, even when Z^i has a reasonable degree of skewness, the FM variance estimator $V_{\mathtt{b},N}^{\mathtt{FM}}$ is still a very reliable estimator of the true asymptotic covariance matrix $Avar[b_N^{\mathtt{FM}}]$.

The generalized Fama-MacBeth (GFM) estimator is obtained by using a weighting matrix Q_N in the first-step time-series regression. The resulting estimator is

$$b_N^{\mathtt{GFM}} = \frac{1}{N}\sum\nolimits_{i=1}^{N}(\tilde{X}^i Q_N \tilde{X}^{i\prime})^{-1}\tilde{X}^i Q_N \tilde{y}^i.$$

The weighting matrix Q_N is assumed to converge in probability to a symmetric and positive definite matrix Q. To formally establish the asymptotic properties of the GFM estimator, we require a condition slightly different from that used for the FM estimator. To this end, we modify Condition FMLN as follows:

Condition GFMLN. (i) Condition LSLN(i) holds. (ii) There exists a $\delta > 0$ arbitrarily small, such that, for all $i = 1, 2, \ldots$, the smallest eigenvalue of the random matrix $Z^i Q Z^{i\prime}$, denoted by $m(Z^i Q Z^{i\prime})$, is greater than δ. (iii) Condition FMLN(iii) holds.

The asymptotic covariance matrix of the GFM estimator $b_N^{\texttt{GFM}}$ is given by

$$Avar[b_N^{\texttt{GFM}}] = E\left[\left[(P_{\texttt{Q}}^i - E[P_{\texttt{Q}}^i])u^i\right]\left[(P_{\texttt{Q}}^i - E[P_{\texttt{Q}}^i])u^i\right]'\right], \tag{11}$$

where

$$P_{\texttt{Q}}^i = (Z^i Q Z^{i\prime})^{-1} Z^i Q. \tag{12}$$

Note that, when Z^i is symmetric around its zero mean, we have $E[P_{\texttt{Q}}^i] = 0_{K\times T}$. If, in addition, the homoscedasticity assumption $E\left[u^i u^{i\prime}|Z^i\right] = \Sigma^{\texttt{u}}$ holds, we have

$$Avar[b_N^{\texttt{GFM}}] = E\left[P_{\texttt{Q}}^i \Sigma^{\texttt{u}} (P_{\texttt{Q}}^i)'\right]$$

as a consequence of the law of iterated expectations. Under this scenario, it follows from fact (F2) stated in the Appendix that the optimal choice for the weighting matrix is

$$Q = (\Sigma^{\texttt{u}})^{-1}$$

for which we obtain

$$Avar[b_N^{\texttt{GFM}}] = E\left[(Z^i(\Sigma^{\texttt{u}})^{-1} Z^{i\prime})^{-1}\right].$$

The following theorem summarizes the asymptotic properties of the FM and GFM estimators. The proof can be found in the Appendix.

Theorem 2 *(a) Under Condition FMLN, the FM estimator $b_N^{\texttt{FM}}$ is consistent and asymptotically normal as $N \to \infty$, with asymptotic covariance matrix given by (9).*

(b) Under Condition GFMLN, the GFM estimator $b_N^{\texttt{GFM}}$ is consistent and asymptotically normal as $N \to \infty$, with asymptotic covariance matrix given by (11).

2.4 Efficiency comparison

Without detailed knowledge of the covariance structure of the disturbances u^i, one cannot rank the various estimators in terms of asymptotic efficiency. As already mentioned, it is not even clear how to optimally select the weighting matrices used by the GLS and GFM procedures. We can, however, shed some light on the issue of efficiency under a special assumption of interest. For the rest of this discussion, we assume that $E[P^i] = E[P^i_{\mathtt{Q}}] = 0_{K\times T}$. This assumption holds if Z^i is symmetric around its zero mean. When the homoscedasticity condition $E\left[u^i u^{i\prime}|Z^i\right] = \Sigma^{\mathtt{u}}$ holds, the optimal choice for both GLS and GFM is given by the inverse of the disturbance variance, $(\Sigma^{\mathtt{u}})^{-1}$. Therefore, by construction, GLS is more efficient than LS, and GFM is more efficient than FM. Furthermore, it follows from fact (F1) stated in the Appendix that

$$Avar[b_N^{\mathtt{GFM}}] = E\left[(Z^i(\Sigma^{\mathtt{u}})^{-1}Z^{i\prime})^{-1}\right] \geq (E[Z^i(\Sigma^{\mathtt{u}})^{-1}Z^{i\prime}])^{-1} = Avar[b_N^{\mathtt{GLS}}]$$

and hence the GLS estimator is more efficient than the GFM estimator. A remaining question is whether LS is more efficient than FM. The answer is not clear even under the assumption of conditional homoscedasticity. However, under the stronger assumption $E\left[u^i u^{i\prime}|Z^i\right] = \sigma^2 I_T$, we can conclude

$$Avar[b_N^{\mathtt{FM}}] = \sigma^2 E\left[(Z^i Z^{i\prime})^{-1}\right] \geq \sigma^2 (E[Z^i Z^{i\prime}])^{-1} = Avar[b_N^{\mathtt{LS}}]$$

as it follows from fact (F1). Therefore, in this case, the LS estimator is more efficient than the FM estimator.

2.5 Simulation evidence

We examine the performance of the various estimators under a number of different scenarios. We consider several cases including time-invariant firm effects, firm-invariant time effects, and time-varying firm effects. We examine the effect of asymmetry allowing for different degrees of skewness in both the disturbance and the regressor. We further examine the sensitivity to changes in the cross-section size N and the time-series length T, and we demonstrate the importance of demeaning when firm-invariant time effects are present.

Naturally, the extensive simulations in Petersen (2005) serve as a suitable benchmark for our experiments. We follow his assumptions about the data generating process, although our simulation design differs in some aspects. To reduce the simulation error, we use 50,000 repetitions for each simulation. In addition, we use a cross section of size $N = 300$ instead of 500 since we are interested in the small sample properties of the estimators. Furthermore, we deviate from other panel data simulation studies, such as Kezdi (2003) and Petersen (2005), and do not present average estimated standard errors. This type of evidence is useful for determining whether the estimated standard errors are unbiased estimates of true population standard deviations. Our point of view, however, is that the primary object of interest in empirical research is the empirical distribution of the corresponding t-statistics. Even if we estimate the standard errors very precisely, the finite sample distribution of the t-statistics might deviate from the standard normal distribution. For this reason, we opt for what we hope is a more informative representation of the performance of the various estimators and we present the evidence in terms of the t-statistic empirical distribution.

We summarize the simulation evidence in terms of the empirical p-values of two-sided t-statistic tests for given asymptotic nominal p-values 1%,5%, and 10%. In addition, we report the cut-off values that result in the predetermined rates of rejection and compare them with the asymptotic nominal cut-off values 2.58, 1.96, and 1.64 that correspond to p-values 1%, 5%, and 10%, respectively. Following standard notation, we denote by p the p-values, while we denote by x_p the corresponding cut-off values. To compare the various estimators in terms of efficiency, we also report the root mean square errors.

Following Petersen (2005), we define the fractions of the disturbance variance due to a firm-specific component (ω) and a time-specific component (ϕ) by

$$\rho_\omega = \frac{\sigma_\omega^2}{\sigma_\varepsilon^2} \text{ and } \rho_\phi = \frac{\sigma_\phi^2}{\sigma_\varepsilon^2},$$

respectively. Similarly, we define the fractions of the regressor variance due to a firm-specific component (μ) and a time-specific component (ξ) by

$$\rho_\mu = \frac{\sigma_\mu^2}{\sigma_x^2} \text{ and } \rho_\xi = \frac{\sigma_\xi^2}{\sigma_x^2},$$

respectively. Across different simulations we consider several combinations of the fractions ρ_ω, ρ_ϕ, ρ_μ, and ρ_ξ. Throughout, as in Petersen (2005), the standard deviations of the regressors and the residuals are $\sigma_x = 1$ and $\sigma_\varepsilon = 2$, respectively. To enhance the exposition, we make the following notation change: the regressors x_{it} and the disturbances ε_{it} are standardized so that they have zero mean and unit variance. We guarantee that the right variances are used by multiplying the variables by the appropriate standard deviation. This convention makes the design of the simulation experiments easier to follow. Therefore, the data generating process used in our simulations is given by

$$y_{it} = (\sigma_x x_{it})b + \sigma_\varepsilon \varepsilon_{it}.$$

A number of points are in order regarding the difference between our results and the results of Petersen (2005). First, while we use the modified version of the FM approach, which is more suitable when N is large, Petersen (2005) uses the traditional FM procedure, whose use is justified for panels with large T. Second, we obtain all estimators and t-statistics after a cross-sectional demeaning step (see equation (2)). Third, to compute t-statistics we use consistent estimates of the asymptotically valid covariance matrices as stated in Theorems 1 and 2. In particular, we should stress that we do not use OLS standard errors when we compute the LS t-statistics. In addition to LS and FM estimators, we also consider their generalized versions, GLS and GFM, respectively. Moreover, in the next section, we provide simulation evidence for the large T case typically encountered in asset pricing applications.

We break our simulations into five sets in order to focus on the various aspects that might affect the finite sample behavior of the various estimators. The first set (LN1) examines the case in which time-invariant firm effects and firm-invariant time effects are present in both the disturbance ε_{it} and the regressor x_{it}. In the second set (LN2), we introduce a time-varying firm effect in both ε_{it} and x_{it} to gauge how within-firm serial correlation affects the various estimators. In the first two sets, all simulated variables are normally distributed. The third set (LN3) examines the effect of skewness. We examine the sensitivity to changes in N and T in the fourth set (LN4), where we repeat the simulations from set LN2 for nine different combinations of N and T. The fifth and final set (LN5) illustrates the importance of demeaning in the presence of firm-invariant time effects.

Simulation LN1: Time-invariant firm effects and firm-invariant time effects

Our first simulation experiment assumes that both the disturbance and the regressor contain a time-invariant firm effect as well as a firm-invariant time effect. Specifically, we assume the following structure

$$\begin{aligned}
\varepsilon_{it} &= \sqrt{\rho_\omega}\omega_i + \sqrt{\rho_\phi}\phi_t + \sqrt{1-\rho_\omega-\rho_\phi}v_{it}, \\
x_{it} &= \sqrt{\rho_\mu}\mu_i + \sqrt{\rho_\xi}\xi_t + \sqrt{1-\rho_\mu-\rho_\xi}w_{it}.
\end{aligned}$$

The random variables ω_i, μ_i, ϕ_t, ξ_t, v_{it}, and w_{it} for $i = 1, \ldots, N$, $t = 1, \ldots, T$ are mutually independent and normally distributed. Note that by construction, both ε_{it} and x_{it} have zero mean and unit variance. We consider three choices for (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) : $(0, 0.25)$, $(0.25, 0)$, and $(0.25, 0.25)$. This produces nine possible combinations and covers a number of cases of particular interest. For example, when $(\rho_\omega, \rho_\phi) = (0, 0.25)$, the disturbance contains a firm-invariant time effect but no time-invariant firm effect and, vice versa, when $(\rho_\omega, \rho_\phi) = (0.25, 0)$. The empirical p-values of the two-sided t-statistic tests, and the empirical cut-off values of all four estimators are presented in Table 1. The root mean square errors are presented in Table 6. Examination of Table 1 reveals that all four estimators produce very reliable t-statistics across all nine scenarios. Specifically, for a nominal p-value of 1%, all empirical p-values are between 1% and 1.3%; for a nominal p-value of 5%, all empirical p-values are between 5% and 6.1%; and for a nominal p-value of 10%, all empirical p-values are between 9.9% and 11.7%. The LS and FM t-statistics are consistently more reliable (although by an insignificant amount) than their generalized counterparts. This can be attributed to the fact that the weighting matrices cannot be estimated precisely. On the other hand, Table 6 shows that the GLS and GFM are more efficient than LS and FM, respectively, as predicted by our theoretical arguments. In addition, also as expected, LS is more efficient than FM, and GLS is more efficient than GFM.

Simulation LN2: Time-varying firm effects

In this set we extend the setup used in LN1 and introduce a time-varying and serially correlated firm effect in addition to the time-invariant firm and firm-invariant time effects. As in LN1, we have

$$\begin{aligned} \varepsilon_{it} &= \sqrt{\rho_\omega}\omega_i + \sqrt{\rho_\phi}\phi_t + \sqrt{1-\rho_\omega-\rho_\phi}v_{it}, \\ x_{it} &= \sqrt{\rho_\mu}\mu_i + \sqrt{\rho_\xi}\xi_t + \sqrt{1-\rho_\mu-\rho_\xi}w_{it}. \end{aligned}$$

However, the disturbances v_{it} and w_{it} now follow first-order autoregressive processes instead of being serially independent. Specifically, for each i,

$$v_{it} = \delta_\varepsilon v_{it-1} + \sqrt{1-\delta_\varepsilon^2}\zeta_{it},\ t = 2,\ldots,T,\ v_{i1} = \zeta_{i1}$$

and

$$w_{it} = \delta_x w_{it-1} + \sqrt{1-\delta_x^2}\eta_{it},\ t = 2,\ldots,T,\ w_{i1} = \eta_{i1},$$

where the parameters δ_ε and δ_x lie between -1 and 1. The random variables ω_i, μ_i, ζ_{it}, η_{it} for $i = 1,\ldots,N$, $t = 1,\ldots,T$ are mutually independent and normally distributed. To examine the effect of autocorrelation in isolation, we set (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) equal to $(0.25, 0.25)$ and let δ_ε and δ_x take values -0.5, -0.25, 0.25, and 0.5. This produces 16 possible combinations. The empirical p-values of the two-sided t-statistic tests and the empirical cut-off values of all four estimators are presented in Table 2. The root mean square errors are presented in Table 7. The overall evidence is identical to that from Simulation LN1. Namely, all four estimators yield reliable t-statistics with the LS and FM being slightly better than GLS and GFM, respectively. The generalized versions are again more efficient, as expected. The LS estimator is still more efficient than FM in most cases, but the difference does not seem to be significant.

Simulation LN3: The effect of skewness

To examine the effect of skewness on the performance of the four estimators, especially FM and GFM, we repeat the experiment described in Simulation LN2 above using nonsymmetric random variables. We set the variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) equal to $(0.25, 0.25)$ and the autoregressive coefficients δ_ε and δ_x equal to 0.5. Instead of normal random variables, we generate gamma random variables, with the skewness parameters s_ε and s_x taking the values -1, -0.5, 0.5 and 1. The gamma variables are standardized to have zero mean and unit variance. This produces 16 different combinations. The empirical p-values of the two-sided t-statistic tests and the empirical cut-off values of all four estimators are presented in Table 3. The root mean square errors are presented in Table 8. Again we see no significant change in the performance of all four estimators. We conclude that all four estimators are robust with respect to the presence of skewness in both the disturbances and the regressors.

Simulation LN4: Sensitivity to changes in N and T

Given the robust finite-sample performance of all estimators in the previously examined settings, one might ask at which point the small-sample distribution of the t-statistics starts deviating considerably from the asymptotic distribution. We investigate this aspect by repeating the experiment described in Simulation LN2 above using different choices for the cross-section size N and the time-series length T. Specifically, we set (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) equal to $(0.25, 0.25)$ and $\delta_\varepsilon = \delta_x = 0.5$. Moreover, we let N take values 100, 200, and 300, while T takes values 10, 30, and 50. This produces nine different combinations. The empirical p-values of the two-sided t-statistic tests and the empirical cut-off values of all four estimators are presented in Table 4. The root mean square errors are presented in Table 9. The LS and FM estimators still provide reliable t-statistics in all nine scenarios. Let us focus on the LS estimator first. For a nominal p-value of 1%, the LS empirical p-values are between 1.1% and 1.4%; for a nominal p-value of 5%, the LS empirical p-values are between 5.2% and 5.9%; and for a nominal p-value of 10%, the LS empirical p-values are between 10.3% and 11.1%. For the FM estimator, the corresponding empirical p-values are between 1% and 1.2%, between 5.1% and 5.5%, and between 10.1% and 10.7%, respectively. Recall that the estimators are justified by large N asymptotics. Therefore, we expect the performance to improve as N increases. This is exactly what we see in Table 4. As N increases, the empirical p-values converge to their nominal counterparts. On the other hand, as T increases we expect the estimation of the weighting matrix used by the generalized estimators to become problematic. This is indeed reflected in Table 4. When $N = 100$ and $T = 30$ and for a nominal p-value of 5%, the GLS t-statistic produces an empirical p-value of 20% while the empirical p-value of the GFM t-statistic is 18.4%. To achieve a level of significance of 5% in this scenario, one has to use cut-off values of 3.00 and 2.91 when using the GLS and GFM estimators, respectively. As expected, the size distortion is reduced as N becomes larger and/or T becomes smaller. Again, according to Table 9, the generalized estimators are more efficient than the standard counterparts.

Simulation LN5: The importance of demeaning in the presence of firm-invariant time effects

In our last simulation set in this section, we address the issue of cross-sectional demeaning. This turns out to be crucial when the regressors and/or the disturbances contain firm-invariant time effects. To illustrate this point, we repeat the experiment described in Simulation LN1 above with the exception that we obtain the estimators without first subtracting the cross-sectional mean. The empirical p-values of the two-sided t-statistic tests and the empirical cut-off values of all four estimators are presented in Table 5. The root mean square errors are presented in Table 10. An interesting pattern emerges from Table 5. When there is a firm-invariant time effect in the disturbance ($\rho_\phi = 0.25$) as well as a firm-invariant time effect in the regressor ($\rho_\xi = 0.25$), the performance of all four estimators deteriorates considerably. As an indication, note that for a nominal p-value of 5%, all empirical p-values are above 40%. The correct cut-off values for achieving a level of significance of 5% are between 4.97 and 9.04, as opposed to the nominal cut-off value of 1.96. On the other hand, if either effect is absent, i.e., if either $\rho_\phi = 0$ or $\rho_\xi = 0$, then the performance of all four t-statistics improves dramatically. To understand why this is the case, consider the data generating process stated in LN1. It follows that for $i \neq j$, $E[(x_{it}\varepsilon_{it})(x_{jt}\varepsilon_{jt})] = \rho_\phi\rho_\xi$ and, therefore, we have cross-sectional correlation for every pair of firms as long as $\rho_\phi\rho_\xi \neq 0$. Unless we remove these cross-sectional correlations by means of cross-sectional demeaning, the t-statistics are unreliable as illustrated in Table 5.

3 Long Time Series

The treatment of the large T case will be brief due to obvious similarities to the large N case. We will point out the necessary modifications.

3.1 Model description

The regression model is identical to the one employed in the large N case:

$$y_{it} = x_{it}'b + \varepsilon_{it},$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$. As in the large N case, we assume that the regressors x_{it} and the disturbances ε_{it} contain time and firm effects:

$$\begin{aligned} x_{it} &= \xi_t + \mu_i + w_{it}, \\ \varepsilon_{it} &= \phi_t + \omega_i + v_{it}. \end{aligned}$$

We allow the terms w_{it} and v_{it} to be cross-sectionally correlated in an arbitrary fashion. Therefore, we can, without loss of generality, ignore the terms ξ_t and ϕ_t in the above specification. In other words, we can work with the variables

$$\begin{aligned} z_{it} &= \xi_t + w_{it}, \\ u_{it} &= \phi_t + v_{it}. \end{aligned}$$

We group observations by time period and for each t, we let

$$y_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{Nt} \end{bmatrix} \ (N \times 1), \ X_t' = \begin{bmatrix} x_{1t}' \\ \vdots \\ x_{Nt}' \end{bmatrix} \ (N \times K), \ \varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Nt} \end{bmatrix} \ (N \times 1)$$

so that we can succinctly write the regression model as

$$y_t = X_t' b + \varepsilon_t, \ t = 1, \ldots, T. \tag{13}$$

Further denoting

$$M' = \begin{bmatrix} \mu_1' \\ \vdots \\ \mu_N' \end{bmatrix}, \ Z_t' = \begin{bmatrix} z_{1t}' \\ \vdots \\ z_{Nt}' \end{bmatrix}, \ \omega = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_N \end{bmatrix}, \ u_t = \begin{bmatrix} u_{1t} \\ \vdots \\ u_{Nt} \end{bmatrix},$$

we can write, for $t = 1, \ldots, T$

$$X_t' = M' + Z_t', \ \varepsilon_t = \omega + u_t.$$

As before, without loss of generality, we can assume that

$$E[Z_t'] = 0_{N \times K}, \ E[u_t] = 0_N.$$

Next, we first eliminate the firm effects M and ω by demeaning. In the present context, we subtract the time-series average from each observation. Therefore, after demeaning, regression (13) delivers

$$y_t - \bar{y}_T = (X_t - \bar{X}_T)' b + (\varepsilon_t - \bar{\varepsilon}_T) = (Z_t - \bar{Z}_T)' b + (u_t - \bar{u}_T),$$

where we use the generic notation

$$\bar{r}_T = \frac{1}{T} \sum\nolimits_{t=1}^{T} r_t.$$

We can further simplify the notation by letting

$$\tilde{r}_t = r_t - \bar{r}_T$$

to obtain

$$\tilde{y}_t = \tilde{X}_t' b + \tilde{\varepsilon}_t = \tilde{Z}_t' b + \tilde{u}_t, \ t = 1, 2, \ldots, T. \tag{14}$$

As before, we obtain estimators of the parameter b by applying the least-squares or the Fama-MacBeth method to equation (14).

3.2 Least-squares estimators

Applying least-squares to equation (14) we obtain the LS estimator

$$b_T^{\mathtt{LS}} = \left(\frac{1}{T}\sum\nolimits_{t=1}^{T} \tilde{X}_t \tilde{X}_t'\right)^{-1} \left(\frac{1}{T}\sum\nolimits_{t=1}^{T} \tilde{X}_t \tilde{y}_t\right).$$

The asymptotic properties, as $T \to \infty$, of the LS estimator $b_T^{\mathtt{LS}}$ can be established under appropriate regularity conditions. The main difference between the large N and the present large T case is that we do not have to restrict the observations (Z_t, u_t) to be i.i.d. over time; see Condition LSLN(i) for comparison. Instead, we can allow for serial correlation while we require that the central limit theorem applies. Formally, we make the following assumption:

Condition LSLT. (i) The time series $\{s_t = (vec(Z_t)', u_t')',\ t = 1, 2, \ldots\}$ is stationary and mixing, and the central limit theorem applies to functions of s_t. (ii) The matrix $E[Z_t Z_t']$ has full rank equal to K. (iii) The orthogonality condition $E[Z_t u_t] = 0_K$ holds.

Under Condition LSLT, $b_T^{\mathtt{LS}}$ is consistent and asymptotically normal, as $T \to \infty$. In particular, we can show that

$$\sqrt{T}(b_T^{\mathtt{LS}} - b) = \Sigma_{\mathtt{Z}}^{-1} \frac{1}{\sqrt{T}} \sum\nolimits_{t=1}^{T} Z_t u_t + o_p(1),$$

where

$$\Sigma_{\mathtt{Z}} = E\left[Z_t Z_t'\right]. \tag{15}$$

It follows that the asymptotic covariance of $b_T^{\mathtt{LS}}$ is given by

$$Avar\left[b_T^{\mathtt{LS}}\right] = \Sigma_{\mathtt{Z}}^{-1} \Sigma_{\mathtt{Zu}} \Sigma_{\mathtt{Z}}^{-1}, \tag{16}$$

where

$$\Sigma_{\mathtt{Zu}} = \sum\nolimits_{j=-\infty}^{\infty} E[Z_t u_t (Z_{t-j} u_{t-j})']. \tag{17}$$

Since $\tilde{X}_t = \tilde{Z}_t$ and $\tilde{\varepsilon}_t = \tilde{u}_t$, a consistent estimator of the asymptotic covariance matrix $Avar[b_T^{\mathtt{LS}}]$ can be obtained by

$$V_{\mathtt{b},T}^{\mathtt{LS}} = \Sigma_{\mathtt{X},T}^{-1} \Sigma_{\mathtt{X}\varepsilon,T} \Sigma_{\mathtt{X},T}^{-1},$$

where

$$\Sigma_{\mathtt{X},T} = \frac{1}{T} \sum\nolimits_{t=1}^{T} \tilde{X}_t \tilde{X}_t'$$

and $\Sigma_{\mathtt{X}\varepsilon,T}$ is a heteroscedasticity and autocorrelation consistent (HAC) estimator of the asymptotic variance of the time series $\tilde{X}_t\hat{\varepsilon}_t$, where the estimated residuals $\hat{\varepsilon}_t = y_t - X_t' b_T$ are obtained using b_T, a consistent estimate of b. HAC estimators can be obtained using the methodology of Newey-West (1987, 1994).[3]

Employing a weighting matrix Q other than the identity matrix, we obtain the GLS estimator

$$b_T^{\mathtt{GLS}} = \left(\frac{1}{T}\sum\nolimits_{t=1}^{T} \tilde{X}_t Q_T \tilde{X}_t'\right)^{-1} \left(\frac{1}{T}\sum\nolimits_{t=1}^{T} \tilde{X}_t Q_T \tilde{y}_t\right),$$

where Q_T is a consistent estimator of the symmetric and positive definite matrix Q. It follows that the asymptotic covariance of $b_T^{\mathtt{GLS}}$ is given by

$$Avar\left[b_T^{\mathtt{GLS}}\right] = (\Sigma_{\mathtt{Z}}^{\mathtt{Q}})^{-1}\Sigma_{\mathtt{Zu}}^{\mathtt{Q}}(\Sigma_{\mathtt{Z}}^{\mathtt{Q}})^{-1}, \tag{18}$$

where

$$\Sigma_{\mathtt{Z}}^{\mathtt{Q}} = E\left[Z_t Q Z_t'\right] \tag{19}$$

and $\Sigma_{\mathtt{Zu}}^{\mathtt{Q}}$ is determined by

$$\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T} Z_t Q u_t \xrightarrow{d} N(0_K, \Sigma_{\mathtt{Zu}}^{\mathtt{Q}}). \tag{20}$$

Consistency and asymptotic normality of the GLS estimator $b_T^{\mathtt{GLS}}$ are guaranteed under the following assumption:

Condition GLSLT. (i) Condition LSLT(i) holds. (ii) The matrix $E\left[Z_t Q Z_t'\right]$ has full rank equal to K. (iii) The orthogonality condition $E\left[Z_t \otimes u_t\right] = 0_{NK\times N}$ holds.

The following theorem summarizes the asymptotic behavior of the LS and GLS estimators. Its proof is similar to the proof of Theorem 1 and is omitted.

Theorem 3 *(a) Under Condition LSLT, the LS estimator $b_T^{\mathtt{LS}}$ is consistent and asymptotically normal as $T\to\infty$, with asymptotic covariance matrix given by expression (16).*

(b) Under Condition GLSLT, the GLS estimator $b_T^{\mathtt{GLS}}$ is consistent and asymptotically normal as $T\to\infty$, with asymptotic covariance matrix given by expression (18).

[3] den Haan and Levin (1997, 2000) provide useful guidance for the implementation of HAC estimators.

3.3 Fama-MacBeth estimators

Recall that after demeaning we have

$$\tilde{y}_t = \tilde{X}_t' b + \tilde{\varepsilon}_t = \tilde{Z}_t' b + \tilde{u}_t,\ t = 1, \ldots, T.$$

Therefore, following the original idea of FM, an estimate of b based only on observations from time period t can be obtained as follows

$$\hat{b}_t = (\tilde{X}_t \tilde{X}_t')^{-1} \tilde{X}_t \tilde{y}_t$$

so that $\hat{b}_t = b + (\tilde{Z}_t \tilde{Z}_t')^{-1} \tilde{Z}_t \tilde{u}_t$. The FM estimator is then obtained by averaging over t the individual estimates $\hat{b}_t$:

$$b_T^{\mathtt{FM}} = \frac{1}{T} \sum\nolimits_{t=1}^{T} \hat{b}_t = \frac{1}{T} \sum\nolimits_{t=1}^{T} (\tilde{X}_t \tilde{X}_t')^{-1} \tilde{X}_t \tilde{y}_t.$$

The analogue to Condition FMLN in the large T context is given by

Condition FMLT. (i) Condition LSLT(i) holds. (ii) There exists a $\delta > 0$ arbitrarily small, such that, for all $t = 1, 2, \ldots$ the smallest eigenvalue of the random matrix $Z_t Z_t'$, denoted by $m(Z_t Z_t')$, is greater than δ. (iii) The orthogonality condition $E\left[u_t | Z_t\right] = 0_N$ holds.

Condition FMLT guarantees consistency and asymptotic normality for the FM estimator. In particular, under Condition FMLT, we can adapt the argument in the proof of Theorem 2 and show that

$$\sqrt{T}(b_T^{\mathtt{FM}} - b) = \frac{1}{\sqrt{T}} \sum\nolimits_{t=1}^{T} (P_t - E[P_t]) u_t + o_p(1),$$

where

$$P_t = (Z_t Z_t')^{-1} Z_t.$$

Therefore, the asymptotic covariance matrix of $b_T^{\mathtt{FM}}$ is given by

$$Avar[b_T^{\mathtt{FM}}] = \sum\nolimits_{j=-\infty}^{\infty} E\left[\left[(P_t - E[P_t]) u_t\right]\left[(P_{t-j} - E[P_t]) u_{t-j}\right]'\right]. \quad (21)$$

Under the assumption $E[P_t] = 0_{K \times N}$, which is satisfied when Z_t is symmetric around its zero mean, we have

$$Avar[b_T^{\mathtt{FM}}] = \sum\nolimits_{j=-\infty}^{\infty} E\left[(P_t u_t)(P_{t-j} u_{t-j})'\right].$$

Adapting the original idea of FM to a time-series context, an estimate of the covariance matrix $Avar[b_T^{\mathtt{FM}}]$ is obtained by treating the individual estimates $\hat{b}_t$ as a sample from a stationary time series. This gives rise to the HAC covariance estimator using a Bartlett kernel

$$V_{\mathtt{b},T}^{\mathtt{FM}} = \Omega_{0,T}^{\mathtt{FM}} + \sum\nolimits_{j=1}^{m_T} \left(1 - \frac{j}{m_T + 1}\right) \left[\Omega_{j,T}^{\mathtt{FM}} + (\Omega_{j,T}^{\mathtt{FM}})'\right],$$

where

$$\Omega_{j,T}^{\mathtt{FM}} = \frac{1}{T} \sum\nolimits_{t=j+1}^{T} (\hat{b}_t - b_T^{\mathtt{FM}})(\hat{b}_{t-j} - b_T^{\mathtt{FM}})'$$

and the bandwidth m_T increases with T at an appropriate rate. Arguing as in the large N case, we can show that, for $j = 0, 1, \ldots$

$$\Omega_{j,T}^{\mathtt{FM}} = \frac{1}{T} \sum\nolimits_{t=j+1}^{T} (P_t u_t)(P_{t-j} u_{t-j})' + o_p(1).$$

Therefore, under appropriate regularity conditions (see Newey and West (1987)), $V_{\mathtt{b},T}^{\mathtt{FM}}$ provides a consistent estimator of $Avar[b_T^{\mathtt{FM}}]$ under the assumption that $E[P_t] = 0_{K \times N}$.

Given a symmetric and positive definite $N \times N$ weighting matrix Q, we can also obtain the following generalized FM estimator for the large T case

$$b_T^{\mathtt{GFM}} = \frac{1}{T} \sum\nolimits_{t=1}^{T} (\tilde{X}_t Q_T \tilde{X}_t')^{-1} \tilde{X}_t Q_T \tilde{y}_t,$$

where the weighting matrix Q_T is assumed to converge in probability to Q as $T \to \infty$. The asymptotic properties of the GFM estimator can be established under the following assumption:

Condition GFMLT. (i) Condition LSLT(i) holds. (ii) There exists a $\delta > 0$ arbitrarily small, such that, for all $t = 1, 2, \ldots$ the smallest eigenvalue of the random matrix $Z_t Q Z_t'$, denoted by $m(Z_t Q Z_t')$, is greater than δ. (iii) Condition FMLT(iii) holds.

Given Condition GFMLT, we can show that $b_T^{\mathtt{GFM}}$ is consistent and asymptotically normal. Specifically, it can be shown that

$$\sqrt{T}(b_T^{\mathtt{GFM}} - b) = \frac{1}{\sqrt{T}} \sum\nolimits_{t=1}^{T} (P_t^{\mathtt{Q}} - E[P_t^{\mathtt{Q}}]) u_t + o_p(1),$$

where

$$P_t^{\mathtt{Q}} = (Z_t Q Z_t')^{-1} Z_t Q.$$

Therefore, the asymptotic covariance matrix of $b_T^{\mathtt{GFM}}$ is given by

$$Avar[b_T^{\mathtt{GFM}}] = \sum\nolimits_{j=-\infty}^{\infty} E\left[\left[(P_t^{\mathtt{Q}} - E[P_t^{\mathtt{Q}}]) u_t\right] \left[(P_{t-j}^{\mathtt{Q}} - E[P_t^{\mathtt{Q}}]) u_{t-j}\right]'\right]. \tag{22}$$

Under the assumption that $E[P_t^{\mathtt{Q}}] = 0_{K\times N}$, we can proceed as in the case of the FM estimator and obtain a HAC estimator of $Avar[b_T^{\mathtt{GFM}}]$.

The following theorem summarizes the asymptotic properties of the FM and GFM estimators. The proof is similar to the proof of Theorem 2 and is omitted.

Theorem 4 *(a) Under Condition FMLT, the FM estimator $b_T^{\mathtt{FM}}$ is consistent and asymptotically normal as $T \to \infty$, with asymptotic covariance matrix given by (21).*

(b) Under Condition GFMLT, the GFM estimator $b_T^{\mathtt{GFM}}$ is consistent and asymptotically normal as $T \to \infty$, with asymptotic covariance matrix given by (22).

3.4 Simulation evidence

We adopt the notation and presentation given in subsection 2.5. The difference is that we now focus on the large T case and use the estimators developed in the previous two subsections. The heteroscedasticity and autocorrelation consistent (HAC) estimators of the various covariance matrices are obtained using the procedure developed in Newey-West (1994). Our implementation uses prewhitened residuals and a Bartlett kernel with data-dependent bandwidth parameter. For the exact implementation details see Section 2.2 in den Haan and Levin (1997). The simulation design draws on the simulation study conducted by Driscoll and Kraay (1998). The data generating process is described as follows:

$$y_{it} = (\sigma_x x_{it})b + \sigma_\varepsilon \varepsilon_{it},$$

where

$$\varepsilon_{it} = \sqrt{\rho_\phi}\phi_t + \sqrt{\rho_\omega}\omega_i + \sqrt{1-\rho_\phi-\rho_\omega}v_{it}$$

and

$$x_{it} = \sqrt{\rho_\xi}\xi_t + \sqrt{\rho_\mu}\mu_i + \sqrt{1-\rho_\xi-\rho_\mu}w_{it}.$$

Throughout this subsection, the firm-invariant time effects ϕ_t and ξ_t and the time-invariant firm effects ω_i and μ_i are independent across t and i and normally distributed with zero mean and unit variance. As in subsection 2.5, ρ_ξ and ρ_μ represent the fraction of variability of the regressor due to a time- and firm-specific component, respectively. Similarly, ρ_ϕ and ρ_ω represent the fraction of variability of the disturbance due to a time- and firm-specific component, respectively. It follows from the above specification that the standardized variables ε_{it} and x_{it} will have unit variance as long as the terms v_{it} and w_{it} have unit variance. To capture cross-sectional as well as serial correlations, we model the dynamics of the idiosyncratic terms v_{it} and w_{it} using a factor structure. Specifically, we define

$$v_{it} = \lambda_i f_t + \sqrt{1-\lambda_i^2} s_{it},$$

with the factor f_t following an autoregressive process of order 1 defined by

$$f_t = \delta_f f_{t-1} + \sqrt{1-\delta_f^2}\theta_t,\ t = 2, \ldots, T,\ f_1 = \theta_1.$$

The variables s_{it} and θ_t are independent across i and t and normally distributed with zero mean and unit variance. We restrict the coefficients λ_i and δ_f to be less than one in absolute value so that v_{it} is well-defined. Observe that the above specification guarantees that the factors f_t and the disturbances v_{it} have unit variance. Note that, conditional on the coefficients λ_i and λ_j, the correlation between the terms $v_{i,t}$ and $v_{j,t-\tau}$ is given by

$$Corr[v_{i,t}, v_{j,t-\tau}] = \lambda_i \lambda_j \delta_f^\tau,\ \tau = 0, 1, \ldots$$

To ensure mixing, we restrict the coefficients λ_i to be uniformly bounded. Specifically, we assume that λ_i's are uniformly distributed over the interval $[c_\lambda, 1]$ and independent across i. Similarly, the dynamics of the term w_{it} is described by

$$w_{it} = \kappa_i g_t + \sqrt{1-\kappa_i^2} q_{it},$$

where the factor g_t follows an autoregressive process of order 1 defined by

$$g_t = \delta_g g_{t-1} + \sqrt{1-\delta_g^2}\gamma_t,\ t = 2, \ldots, T,\ g_1 = \gamma_1.$$

The variables q_{it} and γ_t are assumed to be independent across i and t and normally distributed with zero mean and unit variance. The coefficients κ_i are uniformly distributed over the interval $[c_\kappa, 1]$ and independent across i.

To assess the sensitivity of the small-sample performance of the various estimators to changes in the parameters determining the data generating process, we consider several choices for the simulation input. As in subsection 2.5, we break our simulations into five sets. In the first set (LT1), we fix the parameters $\delta_f, \delta_g, c_\lambda$, and c_κ that determine the correlation structure and let the variance fractions $\rho_\phi, \rho_\omega, \rho_\xi$, and ρ_μ vary. In the second set (LT2), we focus on the sensitivity to factor autoregressive coefficients δ_f and δ_g. The third set (LT3) examines the sensitivity to lower bounds c_λ and c_κ. We address the sensitivity to changes in T and N in the fourth set (LT4). The final set (LT5) examines the sensitivity to changes in T and N in the absence of serial correlation. In simulations LT1, LT2, and LT3, we use $T = 240$ (which corresponds to 20 years of monthly data) and $N = 25$ (which is a typical number of portfolios used in empirical asset pricing research). In simulations LT4 and LT5, we consider several choices for T and N.

Simulation LT1: Sensitivity to variance fractions (ρ_ϕ, ρ_ω) and (ρ_ξ, ρ_μ)

We set $\delta_f = \delta_g = 0.5$ and $c_\lambda = c_\kappa = 0.5$ and consider three different values for each pair (ρ_ϕ, ρ_ω) and (ρ_ξ, ρ_μ) : $(0, 0)$, $(0, 0.25)$, and $(0.25, 0)$. This produces a total number of nine scenarios. The empirical p-values of the two-sided t-statistic tests and the empirical cut-off values of all four estimators are presented in Table 11. The root mean square errors are presented in Table 16. From Table 11, it follows that, despite the presence of both serial and cross-sectional correlation, the distribution of the LS, GLS, and FM-based t-statistics adequately approximate the standard normal distribution. For instance, if the nominal p-value equals 5%, the LS empirical p-values are between 6.6% and 7.0%, the GLS empirical p-values are between 5.6% and 6.6%, and the FM empirical p-values are between 6.2% and 6.5%. On the contrary, the GFM empirical p-values are between 11.4% and 13.1%. The GFM t-statistic does not appear to be robust to cross-sectional and serial correlations. Table 16 shows that the generalized estimators are significantly more efficient than their standard counterparts.

Simulation LT2: Sensitivity to factor autoregressive coefficients δ_f and δ_g

To examine the effect of the factor serial correlations in isolation, we remove all firm-invariant time effects and time-invariant firm effects by setting $\rho_\phi = \rho_\omega = \rho_\xi = \rho_\mu = 0$. The parameters c_λ and c_κ are set equal to 0.5. We consider four values for the autoregressive coefficients δ_f and δ_g : -0.75, -0.5, 0.5, and 0.75. This produces 16 possible combinations. The empirical p-values of the two-sided t-statistic tests and the empirical cut-off values of all four estimators are presented in Table 12. The root mean square errors are presented in Table 17. As in the LT1 simulation, we observe that LS, GLS, and FM produce reliable t-statistics, while the distribution of the GFM-based t statistic can considerably deviate from the standard normal. The GLS and GFM estimators are more efficient than LS and FM, respectively, with the difference becoming more pronounced for more extreme values of the autoregressive coefficients.

Simulation LT3: Sensitivity to lower bounds c_λ and c_κ

In this simulation we examine the sensitivity to the variance fractions due to the factors, λ_i and κ_i. Recall, that the λ_i's are assumed to be uniformly distributed over $[c_\lambda, 1]$. Therefore, to achieve this, we let the parameters c_λ and c_κ vary. As in Simulation LT2, we set $\rho_\phi = \rho_\omega = \rho_\xi = \rho_\mu = 0$. The autoregressive coefficients δ_f and δ_g are set equal to 0.5. We consider three values for the lower bounds c_λ and c_κ : 0.25, 0.5, and 0.75. We thus have nine possible scenarios. The empirical p-values of the two-sided t-statistic tests and the empirical cut-off values of all four estimators are presented in Table 13. The root mean square errors are presented in Table 18. As in Simulation

LT3, we observe that LS, GLS, and FM-based t-statistics are reliable, while GFM-based t-statistics can be misleading. The difference in efficiency between GLS (GFM) and LS (FM) increases as c_κ and/or c_λ increases.

Simulation LT4: Sensitivity to T and N

Although the estimators considered in this section are justified as T tends to infinity, we observe that the LS and FM estimators are feasible and can be implemented even if N is much larger than T. It is therefore worth examining their finite-sample behavior when $N \gg T$. We consider four possibilities for the time-series length, $T = 30$, 60, 90, and 120, and four possibilities for the size of the cross section, $N = 250, 500, 750$, and 1000. The rest of the parameters are specified as $(\rho_\phi, \rho_\omega) = (\rho_\xi, \rho_\mu) = (0, 0)$, $\delta_f = \delta_g = 0.5$, and $c_\lambda = c_\kappa = 0.5$. The empirical p-values of the two-sided t-statistic tests and the empirical cut-off values of all four estimators are presented in Table 14. The root mean square errors are presented in Table 19. The LS and FM estimators appear to be equivalent in terms of efficiency. Furthermore, LS-based and FM-based t-statistics behave almost identically. Since both estimators are based on T asymptotics, we expect the distribution of the corresponding t-statistics to converge to the standard normal distribution as T increases. This is exactly what we see in Table 14. For $T = 120$, the empirical p-values are quite close to the nominal p-values. More importantly, the results are the same across all values of the cross-section size N. This suggests that the large-T LS and FM estimators can provide reliable inference even for the large N case, as long as we have enough time series observations. However, if T is small, t-statistics based on these large-T estimators can be rather misleading.

Simulation LT5: Sensitivity to T and N without serial correlation

Fama and French (2002) argue that standard errors based on the (traditional) FM method are typically underestimated by a factor up to 2.5 and they attribute the problem to serial correlation in the individual FM estimates. However, there is evidence that the problem is not caused solely by serial correlation. Rather, it appears to be a short time-series problem that, interestingly enough, does not depend on the size of the cross section. In other words, the t-statistics have a distribution that approximates well the target standard normal distribution when T is large enough, irrespective of the number of cross-sectional units. The other point that deserves emphasis is that, contrary to the claim made by Fama and French (2002), the FM method does not provide a reliable solution to the problem of underestimating standard errors when T is small and N is large. While the LS and FM methods produce t-statistics with very similar small-sample behavior, the distributions of both t-statistics are far from the asymptotic normal distribution. The correction suggested by Fama and French (2002) can then be viewed as a small-sample correction, but it is not clear how this correction should be determined in general. In any case, a similar adjustment could be used for the LS estimator to correct for the small-sample problem. The point that needs to be emphasized is that when we apply LS and FM in the same fashion, we obtain inference of comparable quality. To illustrate this point, we repeat the experiment from Simulation LT4 with one important modification. We set $\delta_f = \delta_g = 0$ so that there is no serial correlation in v_{it} and w_{it}. This implies that the individual FM estimates are serially uncorrelated. As before, we consider four possibilities for the time-series length, $T = 30, 60, 90$, and 120, and four possibilities for the size of the cross section, $N = 250, 500, 750$, and 1000. The rest of the parameters are specified as $(\rho_\phi, \rho_\omega) = (\rho_\xi, \rho_\mu) = (0, 0)$, and $c_\lambda = c_\kappa = 0.5$. The empirical p-values of the two-sided t-statistic tests and the empirical cut-off values of all four estimators are presented in Table 15. The root mean square errors are presented in Table 20. Although the results in Table 15 suggest that the FM t-statistics perform better than the LS t-statistics, both t-statistics are unreliable for small T, such as 30 or 60. As T increases, the performance of both estimators improves. As before, the size N of the cross section does not seem to affect the behavior of the t-statistics.

4 Conclusion

The econometric analysis of panel data depends in a crucial way on the cross-sectional and time-series correlation of the regression residuals. In this paper, we consider four estimators that can be used, with appropriate modifications, for both large N and large T cases. Two estimators (LS and GLS) are based on the least-squares approach, while the other two (FM and GFM) are based on the approach of Fama and MacBeth (1973). All estimators are consistent and asymptotically normal under certain regularity conditions. We provide ample simulation evidence on the small-sample behavior of all estimators for both cases. The main point to emphasize is that both approaches (LS and FM) provide inference of comparable quality, as long as appropriate versions of each estimator are used. Therefore, we suggest that empirical researchers use both approaches to guarantee the validity of their results.

In the case of a large cross sections, the estimators under examination perform remarkably well in the sense that the finite sample distribution of the corresponding t-statistics is very close to the limiting standard normal distribution. This feature is robust across all different settings under consideration, including firm-invariant time effects, time-invariant firm effects, and time-varying firm effects. However, as one might expect, we observe some deterioration in the performance of the GLS and GFM estimators when the time-series length is large relative to the size of the cross section. It appears that this is due to the poor estimation of the disturbance covariance matrix. On the contrary, LS and FM are very reliable even in this challenging situation. Another point that should be emphasized is the importance of cross-sectional demeaning before obtaining the various estimators in the presence of firm-invariant time effects. We present simulation evidence showing that, when the regressors and the disturbances contain firm-invariant time effects, omitting the demeaning step can lead to a deterioration in the performance of the estimators, and inference based on t-statistics can be misleading. The various estimators differ in terms of efficiency. As predicted by theoretical arguments and demonstrated by simulations, GLS is more efficient than LS, and GFM is more efficient than FM. Although the FM estimator appears to be the least efficient, it consistently produces t-statistics as reliable as the LS-based t-statistics. The higher efficiency of GLS and GFM comes at a cost. The optimal weighting matrix can be determined only under conditional homoscedasticity and, even in that case, its estimation can be problematic. In addition, if the disturbances are conditionally heteroscedastic, the GLS and GFM estimators are not guaranteed to be more efficient than their standard counterparts.

A similar picture arises in the case of long time series. Even in the presence of both cross-sectional and serial correlation, all four estimators produce reliable t-statistics in small samples. Although formally justified by T asymptotics, the LS and FM estimators for the large T case are feasible and can be used even if N is much larger than T. Most importantly, they perform remarkably well, regardless of the cross-section size N, as long as the time-series length T is not too small. The generalized estimators are more efficient than the standard versions, and, in our experiments, the FM estimator is more efficient than the LS estimator.

One important remaining issue should be pointed out with regard to the large N case. The regression-based procedures examined in this paper can only handle cross-sectional dependence in the form of a common shock to all firms at a given time period. Generic forms of cross-sectional dependence are far more challenging to address and are beyond the scope of this paper. The spatial methodologies of Conley (1999) and Kapoor, Kelejian and Prucha (2002) appear to be a promising approach to addressing this aspect of the data. This and related issues can be the subject of future research.

5 Appendix: Auxiliary facts and Proofs

The following facts are used in the main text or in the subsequent proofs. We state them here explicitly for the convenience of the reader.

(**F1**) For any symmetric and positive definite $m \times m$ random matrix X, we have

$$E[X^{-1}] \succeq (E[X])^{-1}$$

in the sense that the matrix $E[X^{-1}] - (E[X])^{-1}$ is positive semidefinite.

Fact (F1) follows from combining Theorem 2.1 and Lemma 2.2 in Cacoullos and Olkin (1965). It is an application of Jensen's inequality for convex matrix functions and uses the fact the inverse matrix operator is convex in a matrix-algebraic sense.

(**F2**) Let W and V be positive definite $M \times M$ matrices and C be an $M \times L$ matrix, $M \geq L$ with C having full rank L. Then, the difference

$$(C'WC)^{-1}C'WVWC(C'WC)^{-1} - (C'V^{-1}C)^{-1}$$

is a positive semidefinite matrix.

For a discussion of property (F2) see the Appendix A.7 in Arellano (2003) or subsection 8.3.3 in Wooldridge (2002).

The concepts defined next are used in the following lemma.

Definition. Let $|||\cdot|||$ denote a generic matrix norm and G denote a function from $K \subseteq \mathbb{R}^{p\times q}$ to $\mathbb{R}^{r\times s}$ where K is assumed to be open with respect to $|||\cdot|||$. The matrix function G is called uniformly continuous in $|||\cdot|||$ if, for all $\epsilon > 0$ there exists a $\delta > 0$ such that $A, B \in K$ and $|||A - B||| < \delta$ implies $|||G(A) - G(B)||| < \epsilon$. The matrix function G is called Lipschitz continuous in $|||\cdot|||$ if there exists a constant $c > 0$ such that $|||G(A) - G(B)||| \leq c|||A - B|||$ for all $A, B \in K$.

Clearly, if G is Lipschitz continuous in $|||\cdot|||$ then it is also uniformly continuous in $|||\cdot|||$.

Lemma 5 *Let $|||\cdot|||_2$ denote the matrix norm on $\mathbb{R}^{N\times L}$ $(N > L)$ induced by the standard Euclidean norm $||\cdot||_2$, that is $|||B|||_2 = \sup\{||Bu||_2 : u \in \mathbb{R}^L, ||u||_2 = 1\}$, for all $B \in \mathbb{R}^{N\times L}$. For any square matrix A, denote by $m(A)$ and $M(A)$ its smallest and largest eigenvalues respectively. Fix an arbitrarily small $\delta > 0$ and let $\mathbb{K}_\delta$ be the set of the $N \times L$ matrices X such that $m(X'X) > \delta$. Then, the matrix functions G and H, defined on $\mathbb{K}_\delta$ by $G(X) = (X'X)^{-1}X'$ and $H(X) = (X'X)^{-1}$ respectively, are Lipschitz, and therefore uniformly continuous, with respect to $|||\cdot|||_2$ on $\mathbb{K}_\delta$. Moreover, $|||G(X)|||_2 \leq 1/\sqrt{\delta}$ for all $X \in \mathbb{K}_\delta$.*

Proof: The proof is based on the following six matrix-algebraic properties: (P1) For conformable matrices A and B, the following inequality holds $|||AB|||_2 \leq |||A|||_2|||B|||_2$ (see Section 2, Chapter 10 in Campbell and Meyer (1979)), (P2) $|||B|||_2 = \sqrt{M(B'B)}$ for all $B \in \mathbb{R}^{N\times L}$ (see Proposition 10.2.4 in Campbell and Meyer (1979)), (P3) For any $m \times n$ matrix A, the positive eigenvalues of $A'A$ are the same as the positive eigenvalues of AA' (see Theorem 3.23 in Schott (1997)), (P4) If the $p\times p$ matrix A has eigenvalues $\lambda_1, \ldots, \lambda_p$ then the matrix $A + \alpha I_p$ has eigenvalues $\lambda_1 + \alpha, \ldots, \lambda_p + \alpha$ (it follows directly from the definition of eigenvalues), (P5) If the $p\times p$ matrix A has positive eigenvalues $\lambda_1, \ldots, \lambda_p$ then A is invertible and the eigenvalues of A^{-1} are $\lambda_1^{-1}, \ldots, \lambda_p^{-1}$ (see Theorems 3.4(b) and 3.5(b) in Schott (1997)) and (P6) If λ is an eigenvalue of the matrix A then λ^2 is an eigenvalue of A^2 (see Theorem 3.4(a) in Schott (1997)) . It follows from (P2), (P3) and (P5) that for all $X \in \mathbb{K}_\delta$

$$|||(X'X)^{-1}X'|||_2 = \sqrt{M((X'X)^{-1})} = \frac{1}{\sqrt{m(X'X)}} \leq \frac{1}{\sqrt{\delta}}.$$

Then, for all $X, Y \in \mathbb{K}_\delta$, we have

$$\begin{aligned}
& (X'X)^{-1}X' - (Y'Y)^{-1}Y' \\
= \;& (X'X)^{-1}X'(Y - X)(Y'Y)^{-1}Y' + (X'X)^{-1}X'(I_N - Y(Y'Y)^{-1}Y') \\
= \;& (X'X)^{-1}X'(Y - X)(Y'Y)^{-1}Y' + (X'X)^{-1}(X - Y)'(I_N - Y(Y'Y)^{-1}Y') \\
& +(X'X)^{-1}Y'(I_N - Y(Y'Y)^{-1}Y').
\end{aligned}$$

Now, the last term vanishes to zero, and (P2) combined with (P3) imply that $|||(X-Y)'|||_2 = |||X-Y|||_2$ and (P2) together with (P6) yield $|||(X'X)^{-1}|||_2 = \sqrt{M((X'X)^{-1}(X'X)^{-1})} = M((X'X)^{-1}) \leq \frac{1}{\delta}$. Furthermore, the matrix $I_N - Y(Y'Y)^{-1}Y'$ is symmetric and idempotent and therefore (P2) together with (P4) imply $|||I_N - Y(Y'Y)^{-1}Y'|||_2^2 = M(I_N - Y(Y'Y)^{-1}Y') = 1 - m(Y(Y'Y)^{-1}Y') = 1 - 0 = 1$ where $m(Y(Y'Y)^{-1}Y') = 0$ follows from Theorem 1.3.20 in Horn and Johnson (1990) since $N > L$. Combining the above observations and using the triangular inequality along with the inequality in (P1) yields that

$$|||(X'X)^{-1}X' - (Y'Y)^{-1}Y'|||_2 \leq \frac{2}{\delta}|||X - Y|||_2$$

showing that the matrix function G is Lipschitz with respect to $||| \cdot |||_2$ on $\mathbb{K}_\delta$. Finally, note that

$$(X'X)^{-1} - (Y'Y)^{-1} = -(X'X)^{-1}[X'(X - Y) + (X - Y)'Y](Y'Y)^{-1}.$$

Using property (P1) and the triangular inequality we obtain that for all $X, Y \in \mathbb{K}_\delta$

$$\begin{aligned}
& |||(X'X)^{-1} - (Y'Y)^{-1}|||_2 \\
\leq\; & |||(X'X)^{-1}X'|||_2|||X-Y|||_2|||Y'Y|||_2 + |||X'X|||_2|||(X-Y)|||_2|||(Y'Y)^{-1}Y|||_2 \\
\leq\; & \frac{1}{\sqrt{\delta}}|||X-Y|||_2\frac{1}{\delta} + \frac{1}{\delta}|||X-Y|||_2\frac{1}{\sqrt{\delta}} \\
=\; & \frac{2}{\delta\sqrt{\delta}}|||X-Y|||_2
\end{aligned}$$

and thus the matrix function H is with respect to $|||\cdot|||_2$ on $\mathbb{K}_\delta$. Q.E.D.

Proof of Theorem 2:

We only present the proof for the GFM estimator (part (b)). The proof for the FM estimator is analogous. The starting point is the identity

$$\begin{aligned}
& \sqrt{N}(b_N^{\mathtt{GFM}} - b) \\
=\; & \frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}(\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i Q_N u^i - \left[\frac{1}{N}\sum\nolimits_{i=1}^{N}(\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i Q_N\right]\left[\sqrt{N}\bar{u}_N\right].
\end{aligned}$$

Observe that in order to prove the result it suffices to show that the conditions

$$\frac{1}{N}\sum\nolimits_{i=1}^{N}(\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i Q_N \xrightarrow{p} E[P_{\mathtt{Q}}^i] \tag{23}$$

and

$$\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N}(\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i Q_N u^i = \frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N} P_{\mathtt{Q}}^i u^i + o_p(1) \tag{24}$$

hold where $P_{\mathtt{Q}}^i = (Z^i Q Z^{i\prime})^{-1} Z^i Q$ (see definition (12)). Since Q is positive definite there exists a unique symmetric positive definite matrix S such that $Q = S'S$ as it follows from Theorem 7.2.6 in Horn and Johnson (1990). The matrix S is called the matrix square root of Q. The same theorem yields that there exists a unique symmetric positive definite matrix S_N such that $Q_N = S_N' S_N$. By the continuity of the matrix square root (see problem 18, Section 7.2 in Horn and Johnson (1990)) we have $S_N \xrightarrow{p} S$. According to Condition GFMLN(ii), there exists $\delta > 0$ such that $m(Z^i Q Z^{i\prime}) = m((SZ^{i\prime})'(SZ^{i\prime})) > \delta$ for all i. Note that also $\tilde{Z}^i Q_N \tilde{Z}^{i\prime} = (S_N \tilde{Z}^{i\prime})'(S_N \tilde{Z}^{i\prime})$. It follows from the identity

$$S_N \tilde{Z}^{i\prime} - SZ^{i\prime} = (S_N S^{-1} - I_T)(SZ^{i\prime}) - S_N \bar{Z}_N'$$

and Condition GFMLN(ii) that for all i,

$$|||SZ^{i\prime}|||_2 = \sqrt{M(Z^i Q Z^{i\prime})} \leq \frac{1}{\sqrt{\delta}}$$

where $M(A)$ denotes the largest eigenvalue of a square matrix A (see Lemma 5 and its proof). Hence, since $\bar{Z}_N \xrightarrow{p} 0_{K\times T}$, and $S_N \xrightarrow{p} S$ we obtain

$$\sup_{1\le i\le N} |||S_N\tilde{Z}^{i\prime} - SZ^{i\prime}|||_2 \le |||S_N S^{-1} - I_T|||_2 \sup_{1\le i\le N} |||(SZ^{i\prime})|||_2 + |||S_N|||_2|||\bar{Z}_N'|||_2 \xrightarrow{p} 0.$$

Since m is a continuous operator it follows that, for all $i = 1,\ldots,N$, $m(\tilde{Z}^i Q_N \tilde{Z}^{i\prime}) > \delta/2$ with probability arbitrarily close to 1 for sufficiently large N. We can now employ Lemma 5 to obtain

$$\sup_{1\le i\le N} |||H(S_N\tilde{Z}^{i\prime}) - H(SZ^{i\prime})|||_2 = \sup_{1\le i\le N} |||(\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1} - (Z^i Q Z^{i\prime})^{-1}|||_2 \xrightarrow{p} 0 \tag{25}$$

and

$$\sup_{1\le i\le N} |||G(S_N\tilde{Z}^{i\prime}) - G(SZ^{i\prime})|||_2 = \sup_{1\le i\le N} |||(\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i S_N - (Z^i Q Z^{i\prime})^{-1} Z^i S|||_2 \xrightarrow{p} 0.$$

Next we observe that

$$\begin{aligned}
& (\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i Q_N - (Z^i Q Z^{i\prime})^{-1} Z^i Q \\
= \; & (\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i S_N' S_N - (Z^i Q Z^{i\prime})^{-1} Z^i S' S \\
= \; & \left[(\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i S_N' - (Z^i Q Z^{i\prime})^{-1} Z^i S'\right] S_N \\
& + ((SZ^{i\prime})'(SZ^{i\prime}))^{-1}(SZ^{i\prime})'[S_N - S].
\end{aligned}$$

Since $m((SZ^{i\prime})'(SZ^{i\prime})) > \delta$, Lemma 5 implies

$$|||((SZ^{i\prime})'(SZ^{i\prime}))^{-1}(SZ^{i\prime})'|||_2 \le \frac{1}{\sqrt{\delta}}$$

and therefore using the fact $S_N \xrightarrow{p} S$ once again, we obtain

$$\sup_{1\le i\le N} |||(\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i Q_N - (Z^i Q Z^{i\prime})^{-1} Z^i Q|||_2 \xrightarrow{p} 0. \tag{26}$$

Hence, it follows from (26) that condition (23) is satisfied. We proceed to show that condition (24) also holds. A few steps of algebra show that

$$(\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i Q_N - (Z^i Q Z^{i\prime})^{-1} Z^i Q = R^1_{i,N} - R^2_{i,N} + R^3_{i,N} - R^4_{i,N}$$

where

$$\begin{aligned}
R^1_{i,N} &= (\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i Q_N \bar{Z}_N'(Z^i Q Z^{i\prime})^{-1} Z^i Q \\
R^2_{i,N} &= (\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\bar{Z}_N[Q - QZ^{i\prime}(Z^i Q Z^{i\prime})^{-1} Z^i Q], \\
R^3_{i,N} &= (\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i Q_N (Q_N^{-1} - Q^{-1}) Q \bar{Z}_N'(Z^i Q Z^{i\prime})^{-1} Z^i Q \\
R^4_{i,N} &= (\tilde{Z}^i Q_N \tilde{Z}^{i\prime})^{-1}\tilde{Z}^i Q_N (Q_N^{-1} - Q^{-1}) Q.
\end{aligned}$$

Since $\sqrt{N}\bar{Z}_N$ is asymptotically normal we have $\sqrt{N}\bar{Z}_N = O_p(1)$. Thus, using (26), the property $vec(AXB) = (B' \otimes A)vec(X)$ for conformable matrices A, B, and X, the orthogonality condition $E[u^i|Z^i] = 0_T$, and the fact $\sqrt{N}\bar{Z}_N = O_p(1)$ we obtain

$$\begin{aligned}
& \frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N} R_{i,N}^1 u^i \\
= \quad & \left[\frac{1}{N}\sum\nolimits_{i=1}^{N}[(Z^iQZ^{i\prime})^{-1}Z^iQu^i]' \otimes [(\tilde{Z}^iQ_N\tilde{Z}^{i\prime})^{-1}\tilde{Z}^iQ_N]\right] vec\left[\sqrt{N}\bar{Z}_N'\right] \\
= \quad & \left[\frac{1}{N}\sum\nolimits_{i=1}^{N}[(Z^iQZ^{i\prime})^{-1}Z^iQu^i]' \otimes [(Z^iQZ^{i\prime})^{-1}Z^iQ] + o_p(1)\right] vec\left[\sqrt{N}\bar{Z}_N'\right] \\
= \quad & [o_p(1) + o_p(1)]\, O_p(1) \\
= \quad & o_p(1).
\end{aligned}$$

Following the same line of reasoning we can obtain

$$\frac{1}{\sqrt{N}}\sum\nolimits_{i=1}^{N} R_{i,N}^k u^i = o_p(1), \;\; k = 2, 3, 4.$$

To obtain the result for $k = 2$ we use (25), the fact $\sqrt{N}\bar{Z}_N = O_p(1)$, and the condition $E[u^i|Z^i] = 0_T$. Note that the assumption $\sqrt{N}(Q_N - Q) = O_p(1)$ implies that $\sqrt{N}(Q_N^{-1} - Q^{-1})Q\bar{Z}_N' = O_p(1)$ and $\sqrt{N}(Q_N^{-1} - Q^{-1}) = O_p(1)$. The result for $k = 3$ is obtained using (26), the fact $\sqrt{N}(Q_N^{-1} - Q^{-1})Q\bar{Z}_N' = O_p(1)$ and the condition $E[u^i|Z^i] = 0_T$. Using (26) once again, the fact $\sqrt{N}(Q_N^{-1} - Q^{-1}) = O_p(1)$, and the condition $E[u^i|Z^i] = 0_T$, we can show the result for $k = 4$. Combining the above result for $k = 1, 2, 3$, and 4, we conclude that condition (24) holds and hence the proof is complete. Q.E.D.

References

[1] Anselin, L., 1988, *Spatial Econometrics: Methods and Models*, Kluwer.

[2] Arellano, M., 1987, Computing Robust Standard Errors for Within-Groups Estimators, *Oxford Bulletin of Economics and Statistics* 49, 431-434.

[3] Arellano, M., 2003, *Panel Data Econometrics*, Oxford University Press.

[4] Baltagi, H. B., 2001, *Econometric Analysis of Panel Data*, Wiley.

[5] Brennan, M. J., T. Chordia, and A. Subrahmanyam, 1998, Alternative Factor Specifications, Security Characteristics, and the Cross Section of Expected Returns, *Journal of Financial Economics* 49, 345-373.

[6] Cacoullos, T. and I. Olkin, 1965, On the Bias of Functions of Characteristic Roots of a Random Matrix, *Biometrika* 52, 87-94.

[7] Campbell, S. L. and C. D. Meyer, Jr. 1979, *Generalized Inverses of Linear Transformations*, Dover.

[8] Cliff, A., and J. K. Ord, 1981, *Spatial Processes: Models and Applications*, Pion.

[9] Cochrane, J., 2001, *Asset Pricing*, Princeton University Press.

[10] Conley, T., 1999, GMM Estimation with Cross Sectional Dependence, *Journal of Econometrics* 92, 1-45.

[11] Coval, J. D., and T. Shumway, 2005, Do Behavioral Biases Affect Prices?, *Journal of Finance* 60, 1-34.

[12] den Haan, W. J., and A. T. Levin, 1997, A Practitioner's Guide to Robust Covariance Matrix Estimation, G.S. Maddala and C.R. Rao, Eds.: Handbook of Statistics, Vol. 15, 299-342.

[13] den Haan, W. J., and A. T. Levin, 2000, Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order, NBER Technical Working Paper No. 225.

[14] Driscoll, J. C. and A. C. Kraay, 1998, Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data, *Review of Economics and Statistics* 80, 549-560.

[15] Fama, E. F. and K. French, 2002, Testing Tradeoff and Pecking Order Predictions about Dividends and Debt, *Review of Financial Studies* 15, 1-33.

[16] Fama, E. F. and J. D. MacBeth, 1973, Risk, Return and Equilibrium: Empirical Tests, *Journal of Political Economy* 81, 607-636.

[17] Hsiao, C., 2003, *Analysis of Panel Data,* Cambridge University Press.

[18] Horn, R. A., and Johnson C. R., 1990, *Matrix Analysis*, Cambridge University Press.

[19] Jagannathan, R. and Z. Wang, 1998, Asymptotic Theory for Estimating Beta Pricing Models Using Cross-Sectional Regressions, *Journal of Finance* 53, 1285-1309.

[20] Kapoor, M., H.H. Kelejian, and I.R. Prucha, 2002, Panel Data Models with Spatially Correlated Error Components, Working Paper, University of Maryland.

[21] Kezdi, G., 2003, Robust Standard Error Estimation in Fixed-Effects Panel Models, Working paper, Budapest University of Economics.

[22] Moulton, B. R., 1990, An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units, *Review of Economics and Statistics* 72, 334-338.

[23] Newey, W. K., and K. D. West, 1987, A Simple Positive Semi-Definite Hetescedasticity and Autocorrelation Consistent Matrix Estimation, *Econometrica* 55, 703-708.

[24] Newey, W. K., and K. D. West, 1994, Automatic Lag Selection in Covariance Matrix Estimation, *Review of Economic Studies* 61, 631-653.

[25] Petersen, M., 2005, Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches, Working paper, Northwestern University.

[26] Schott, J. R., 1997, *Matrix Analysis for Statistics*, Wiley.

[27] Shanken, J., 1985, Multivariate Tests of the Zero-Beta CAPM, *Journal of Financial Economics* 14, 327-348.

[28] Shanken, J., 1992, On the Estimation of Beta-Pricing Models, *Review of Financial Studies* 5, 1-34.

[29] Skoulakis, G., 2005, Assessment of Asset-Pricing Models using Cross-Sectional Regressions, Working paper, Northwestern University.

[30] White, H., 1980, A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity, *Econometrica* 48, 817-838.

[31] White, H., 2001, *Asymptotic Theory for Econometricians*, Academic Press.

[32] Wooldridge, J. M., 2002, *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

[33] Wooldridge, J. M., 2003, Cluster-Sample Methods in Applied Econometrics, *American Economic Review* 93, 133-138.

Table 1: Empirical p-values and cut-off values for Simulation LN1

This table presents the empirical p-values and the cut-off values of the t-statistics of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LN1 in subsection 2.5. The nominal p-values, denoted by p, are included in the first column, while the nominal cut-off values, denoted by x_p, are included in the second column. The empirical p-values are included in the upper block, while the empirical cut-off values are included in the lower block. All p-values are stated in percentages. The cross-section size is $N = 300$, and the time-series length is $T = 10$. The variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) take the values $(0, 0.25)$, $(0.25, 0)$, and $(0.25, 0.25)$. The results are obtained using a simulation with 50,000 repetitions.

p	x_p		LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
		(ρ_ω, ρ_ϕ) / (ρ_μ, ρ_ξ)	(0,.25)				(.25,0)				(.25,.25)			
1		(0,.25)	1.1	1.3	1.0	1.3	1.0	1.3	1.0	1.3	1.1	1.3	1.0	1.3
5			5.0	5.7	5.0	5.6	5.2	6.1	5.1	5.7	5.1	5.8	5.3	5.8
10			9.9	11.0	10.1	11.0	10.1	11.3	10.1	11.0	10.1	11.1	10.3	11.1
1		(.25,0)	1.0	1.3	1.0	1.3	1.1	1.3	1.1	1.3	1.1	1.2	1.0	1.3
5			5.1	5.8	5.1	5.7	5.2	6.1	5.1	5.9	5.2	5.8	5.2	5.7
10			10.1	11.2	10.2	11.1	10.4	11.7	10.1	11.1	10.4	11.2	10.2	11.1
1		(.25,.25)	1.1	1.3	1.0	1.3	1.1	1.3	1.0	1.2	1.1	1.2	1.0	1.2
5			5.2	5.9	5.1	5.8	5.1	5.9	5.0	5.6	5.4	5.8	5.1	5.8
10			10.2	11.4	10.3	11.2	10.3	11.3	10.1	10.9	10.3	11.2	10.1	11.1
	2.58	(0,.25)	2.60	2.66	2.59	2.67	2.59	2.66	2.59	2.66	2.62	2.69	2.59	2.66
	1.96		1.96	2.02	1.96	2.01	1.98	2.05	1.97	2.01	1.97	2.03	1.99	2.02
	1.64		1.64	1.69	1.65	1.69	1.65	1.70	1.65	1.69	1.65	1.70	1.66	1.70
	2.58	(.25,0)	2.59	2.67	2.58	2.67	2.61	2.67	2.61	2.68	2.59	2.65	2.59	2.64
	1.96		1.97	2.03	1.97	2.02	1.98	2.05	1.97	2.03	1.98	2.03	1.97	2.02
	1.64		1.65	1.70	1.65	1.70	1.67	1.72	1.65	1.70	1.66	1.71	1.66	1.70
	2.58	(.25,.25)	2.61	2.67	2.56	2.66	2.59	2.70	2.58	2.65	2.62	2.65	2.59	2.64
	1.96		1.97	2.03	1.97	2.03	1.97	2.03	1.96	2.01	1.99	2.03	1.96	2.02
	1.64		1.66	1.71	1.66	1.70	1.66	1.70	1.65	1.69	1.66	1.70	1.65	1.70

Table 2: Empirical p-values and cut-off values for Simulation LN2

This table presents the empirical p-values and the cut-off values of the t-statistics of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LN2 in subsection 2.5. The nominal p-values, denoted by p, are included in the first column, while the nominal cut-off values, denoted by x_p, are included in the second column. The empirical p-values are included in the upper block, while the empirical cut-off values are included in the lower block. All p-values are stated in percentages. The cross-section size is $N = 300$, and the time-series length is $T = 10$. The variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) are both equal to $(0.25, 0.25)$. The autoregressive coefficients δ_ε and δ_x take the values -0.5, -0.25, 0.25, and 0.5. The results are obtained using a simulation with 50,000 repetitions.

p	x_p		LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
		δ_x \ δ_ε	-0.5				-0.25				0.25				0.5			
1		-0.5	1.1	1.2	1.1	1.2	1.2	1.4	1.1	1.3	1.1	1.3	1.1	1.3	1.2	1.3	1.1	1.2
5			5.2	5.8	5.1	5.7	5.2	5.8	5.1	5.7	5.2	5.9	5.2	5.8	5.3	5.8	5.2	5.7
10			10.2	11.0	10.2	11.0	10.3	11.1	10.3	11.1	10.4	11.3	10.3	11.0	10.3	11.3	10.1	11.1
1		-0.25	1.1	1.4	1.1	1.3	1.1	1.3	1.1	1.3	1.0	1.2	1.0	1.2	1.1	1.3	1.1	1.2
5			5.3	6.0	5.2	5.8	5.3	5.8	5.1	5.7	5.3	5.9	5.1	5.7	5.2	5.9	5.1	5.6
10			10.4	11.5	10.1	11.1	10.5	11.2	10.3	11.1	10.4	11.2	10.3	11.0	10.2	11.2	10.1	10.8
1		0.25	1.1	1.3	1.1	1.3	1.2	1.4	1.1	1.3	1.1	1.3	1.0	1.3	1.1	1.3	1.1	1.2
5			5.4	5.9	5.1	5.7	5.4	6.1	5.3	5.8	5.2	5.8	5.0	5.7	5.2	5.8	5.2	5.5
10			10.5	11.3	10.1	11.0	10.5	11.4	10.4	11.2	10.2	11.1	10.3	11.0	10.4	11.2	10.2	10.9
1		0.5	1.1	1.3	1.0	1.2	1.1	1.3	1.0	1.1	1.1	1.3	1.1	1.2	1.1	1.3	1.1	1.3
5			5.2	5.9	5.2	5.8	5.2	5.9	5.0	5.6	5.3	5.9	5.1	5.6	5.3	5.9	5.2	5.7
10			10.4	11.3	10.1	11.0	10.1	11.1	10.0	10.9	10.4	11.2	9.9	10.9	10.4	11.2	10.3	10.9
	2.58	-0.5	2.62	2.65	2.61	2.66	2.63	2.69	2.61	2.67	2.62	2.65	2.59	2.67	2.64	2.67	2.61	2.65
	1.96		1.98	2.02	1.97	2.01	1.98	2.03	1.96	2.02	1.98	2.04	1.98	2.03	1.99	2.03	1.98	2.02
	1.64		1.65	1.70	1.65	1.70	1.66	1.70	1.66	1.70	1.66	1.71	1.66	1.70	1.66	1.71	1.65	1.70
	2.58	-0.25	2.61	2.68	2.61	2.69	2.60	2.65	2.60	2.65	2.58	2.67	2.59	2.65	2.61	2.68	2.61	2.64
	1.96		1.99	2.05	1.98	2.03	1.99	2.03	1.97	2.02	1.99	2.04	1.97	2.01	1.98	2.03	1.97	2.01
	1.64		1.67	1.72	1.65	1.70	1.67	1.70	1.66	1.70	1.66	1.70	1.66	1.70	1.66	1.71	1.65	1.68
	2.58	0.25	2.62	2.66	2.60	2.67	2.65	2.71	2.61	2.68	2.61	2.70	2.58	2.67	2.60	2.66	2.59	2.67
	1.96		1.99	2.03	1.97	2.02	2.00	2.05	1.99	2.04	1.98	2.02	1.96	2.01	1.98	2.02	1.98	2.01
	1.64		1.67	1.71	1.65	1.69	1.67	1.72	1.66	1.70	1.65	1.70	1.66	1.69	1.66	1.70	1.65	1.69
	2.58	0.5	2.60	2.68	2.57	2.65	2.61	2.67	2.57	2.63	2.61	2.67	2.61	2.66	2.59	2.68	2.60	2.67
	1.96		1.97	2.03	1.98	2.02	1.98	2.04	1.96	2.01	1.99	2.04	1.97	2.01	1.99	2.03	1.98	2.01
	1.64		1.66	1.71	1.65	1.69	1.65	1.70	1.64	1.69	1.66	1.70	1.64	1.69	1.66	1.71	1.66	1.69

Table 3: Empirical p-values and cut-off values for Simulation LN3

This table presents the empirical p-values and the cut-off values of the t-statistics of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LN3 in subsection 2.5. The nominal p-values, denoted by p, are included in the first column, while the nominal cut-off values, denoted by x_p, are included in the second column. The empirical p-values are included in the upper block, while the empirical cut-off values are included in the lower block. All p-values are stated in percentages. The cross-section size is $N = 300$, and the time-series length is $T = 10$. The variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) are set equal to $(0.25, 0.25)$. The autoregressive coefficients δ_ε and δ_x are set equal to 0.5. All random numbers are generated from a standardized gamma distribution. The skewness parameters s_ε and s_x take the values -1, -0.5, 0.5, and 1. The results are obtained using a simulation with 50,000 repetitions.

p	x_p		LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
		s_ε																
		s_x	-1				-0.5				0.5				1			
1		-1	1.1	1.4	1.0	1.2	1.2	1.3	1.0	1.3	1.2	1.4	1.0	1.3	1.1	1.3	0.9	1.2
5			5.4	6.1	5.0	5.7	5.2	6.0	5.1	5.8	5.4	6.0	5.0	5.8	5.2	5.9	4.9	5.5
10			10.7	11.5	10.1	11.0	10.3	11.4	10.1	10.9	10.4	11.4	10.1	11.2	10.2	11.3	9.7	10.9
1		-0.5	1.1	1.3	1.0	1.3	1.2	1.4	1.1	1.3	1.1	1.3	1.0	1.2	1.1	1.3	1.0	1.3
5			5.4	5.8	5.0	5.6	5.5	6.2	5.2	5.9	5.4	5.9	5.1	5.9	5.3	5.9	4.9	5.6
10			10.6	11.2	9.9	10.8	10.5	11.6	10.4	11.4	10.5	11.4	10.1	11.3	10.4	11.6	9.8	10.8
1		0.5	1.2	1.3	1.0	1.3	1.1	1.3	1.1	1.3	1.1	1.4	1.1	1.3	1.2	1.3	1.1	1.3
5			5.3	6.1	5.0	5.9	5.3	6.0	5.0	5.7	5.3	5.9	5.0	5.8	5.3	6.0	5.0	5.6
10			10.4	11.5	9.9	11.2	10.4	11.3	10.1	11.1	10.4	11.3	10.1	11.1	10.4	11.3	9.9	10.8
1		1	1.1	1.4	1.0	1.3	1.2	1.5	1.1	1.4	1.1	1.4	1.1	1.3	1.1	1.3	1.0	1.3
5			5.5	6.1	5.0	5.8	5.4	6.3	5.2	6.1	5.3	6.0	5.2	5.9	5.3	5.9	4.9	5.7
10			10.8	11.5	9.9	11.1	10.5	11.8	10.2	11.4	10.5	11.4	10.3	11.1	10.3	11.5	9.9	11.0
	2.58	-1	2.61	2.71	2.58	2.64	2.63	2.68	2.58	2.67	2.62	2.70	2.58	2.67	2.62	2.68	2.56	2.65
	1.96		2.00	2.05	1.96	2.02	1.98	2.04	1.97	2.04	1.99	2.05	1.96	2.02	1.98	2.03	1.95	2.00
	1.64		1.68	1.72	1.65	1.69	1.66	1.71	1.65	1.69	1.67	1.71	1.65	1.70	1.66	1.70	1.63	1.69
	2.58	-0.5	2.63	2.68	2.57	2.67	2.64	2.71	2.61	2.67	2.60	2.67	2.58	2.64	2.62	2.68	2.57	2.67
	1.96		1.99	2.02	1.96	2.01	2.00	2.06	1.98	2.03	1.99	2.04	1.97	2.03	1.99	2.04	1.95	2.01
	1.64		1.67	1.70	1.64	1.68	1.67	1.72	1.66	1.71	1.67	1.71	1.65	1.71	1.66	1.72	1.64	1.69
	2.58	0.5	2.64	2.69	2.58	2.67	2.62	2.67	2.60	2.66	2.62	2.71	2.60	2.67	2.64	2.65	2.59	2.65
	1.96		1.98	2.05	1.96	2.03	1.99	2.04	1.96	2.02	1.98	2.03	1.96	2.02	1.99	2.04	1.96	2.01
	1.64		1.66	1.72	1.64	1.70	1.66	1.71	1.65	1.70	1.67	1.71	1.65	1.70	1.67	1.71	1.64	1.69
	2.58	1	2.60	2.70	2.57	2.67	2.64	2.71	2.62	2.70	2.62	2.71	2.59	2.68	2.62	2.67	2.57	2.66
	1.96		2.00	2.05	1.96	2.02	2.00	2.07	1.98	2.05	1.98	2.05	1.98	2.04	1.98	2.03	1.95	2.02
	1.64		1.68	1.71	1.64	1.69	1.67	1.73	1.66	1.71	1.67	1.71	1.66	1.71	1.66	1.71	1.64	1.69

Table 4: Empirical p-values and cut-off values for Simulation LN4

This table presents the empirical p-values and the cut-off values of the t-statistics of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LN4 in subsection 2.5. The nominal p-values, denoted by p, are included in the first column, while the nominal cut-off values, denoted by x_p, are included in the second column. The empirical p-values are included in the upper block, while the empirical cut-off values are included in the lower block. All p-values are stated in percentages. The variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) are set equal to $(0.25, 0.25)$. The autoregressive coefficients δ_ε and δ_x are set equal to 0.5. The cross-section size N takes values 100, 200, and 300, and the time-series length T takes values 10, 20, and 30. The results are obtained using a simulation with 50,000 repetitions.

p	x_p		LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
		N												
		T	10				20				30			
1		100	1.4	2.1	1.2	1.8	1.4	4.7	1.2	4.2	1.3	9.4	1.2	8.2
5			5.8	8.2	5.5	7.5	5.9	13.0	5.5	11.9	5.8	20.0	5.3	18.4
10			11.1	14.4	10.6	13.4	11.0	20.3	10.7	18.9	11.0	28.1	10.4	26.5
1		200	1.3	1.6	1.1	1.5	1.1	2.2	1.0	2.1	1.2	3.4	1.1	3.1
5			5.5	6.6	5.2	6.2	5.3	8.2	5.1	7.8	5.4	10.5	5.1	9.9
10			10.9	12.4	10.3	11.9	10.3	14.3	10.2	13.9	10.3	17.2	10.3	16.5
1		300	1.1	1.3	1.1	1.4	1.1	1.7	1.2	1.7	1.1	2.2	1.1	2.2
5			5.2	5.9	5.3	5.8	5.3	7.0	5.1	6.8	5.5	8.3	5.4	8.0
10			10.3	11.4	10.1	11.1	10.5	12.8	10.3	12.5	10.8	14.4	10.5	14.1
	2.58	100	2.71	2.90	2.64	2.83	2.72	3.35	2.67	3.29	2.68	3.99	2.64	3.87
	1.96		2.03	2.20	2.00	2.15	2.03	2.55	2.00	2.47	2.03	3.00	1.99	2.91
	1.64		1.70	1.85	1.67	1.81	1.70	2.13	1.68	2.07	1.69	2.53	1.66	2.44
	2.58	200	2.66	2.73	2.61	2.71	2.62	2.90	2.60	2.87	2.64	3.09	2.60	3.07
	1.96		2.00	2.10	1.97	2.07	1.99	2.22	1.97	2.18	2.00	2.37	1.97	2.34
	1.64		1.68	1.76	1.66	1.73	1.66	1.85	1.65	1.82	1.66	1.99	1.66	1.96
	2.58	300	2.61	2.66	2.61	2.70	2.61	2.77	2.63	2.78	2.62	2.90	2.62	2.89
	1.96		1.98	2.04	1.98	2.03	1.99	2.11	1.97	2.11	2.01	2.21	1.99	2.19
	1.64		1.66	1.71	1.65	1.70	1.67	1.78	1.66	1.77	1.68	1.86	1.67	1.84

Table 5: Empirical p-values and cut-off values for Simulation LN5

This table presents the empirical p-values and the cut-off values of the t-statistics of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LN5 in subsection 2.5. The nominal p-values, denoted by p, are included in the first column, while the nominal cut-off values, denoted by x_p, are included in the second column. The empirical p-values are included in the upper block, while the empirical cut-off values are included in the lower block. All p-values are stated in percentages. The cross-section size is $N = 300$, and the time-series length is $T = 10$. The variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) take the values $(0, 0.25)$, $(0.25, 0)$, and $(0.25, 0.25)$. The results are obtained using a simulation with 50,000 repetitions.

p	x_p		LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
		(ρ_ω, ρ_ϕ) / (ρ_μ, ρ_ξ)	(0,.25)				(.25,0)				(.25,.25)			
1		(0,.25)	55.8	30.2	52.6	29.2	1.1	1.3	1.0	1.2	55.8	30.3	54.8	30.6
5			65.6	43.0	63.0	41.9	5.1	5.9	5.0	5.7	65.5	43.0	64.6	43.4
10			70.6	50.6	68.5	49.6	10.2	11.3	10.1	10.9	70.8	50.7	70.1	51.1
1		(.25,0)	1.0	1.3	1.0	1.1	1.1	1.3	1.1	1.3	1.0	1.3	1.1	1.3
5			5.1	5.5	5.2	5.6	5.2	6.1	5.1	5.8	5.1	6.0	5.1	5.7
10			10.0	11.0	10.2	10.8	10.3	11.5	10.1	11.0	10.4	11.4	10.3	11.0
1		(.25,.25)	55.7	33.0	52.7	31.6	1.0	1.2	1.0	1.2	46.0	36.6	49.4	35.7
5			65.4	45.3	62.8	44.2	5.1	5.8	5.0	5.5	57.4	49.0	60.1	48.2
10			70.5	52.7	68.5	51.8	10.1	11.2	10.1	10.8	63.6	56.2	66.0	55.4
	2.58	(0,.25)	12.13	6.84	11.34	6.88	2.60	2.67	2.58	2.66	11.92	6.92	11.76	7.10
	1.96		8.91	5.02	8.31	4.97	1.97	2.04	1.96	2.02	8.88	5.11	8.72	5.17
	1.64		7.44	4.18	6.91	4.11	1.66	1.71	1.65	1.69	7.40	4.20	7.21	4.25
	2.58	(.25,0)	2.58	2.66	2.57	2.61	2.61	2.67	2.62	2.68	2.59	2.66	2.61	2.67
	1.96		1.97	2.01	1.97	2.01	1.98	2.04	1.97	2.03	1.97	2.03	1.97	2.02
	1.64		1.64	1.69	1.65	1.68	1.66	1.72	1.65	1.69	1.66	1.71	1.66	1.70
	2.58	(.25,.25)	12.14	7.42	11.33	7.32	2.57	2.66	2.57	2.63	9.81	8.06	10.58	8.09
	1.96		9.04	5.41	8.36	5.34	1.97	2.02	1.96	2.00	7.23	5.86	7.77	5.83
	1.64		7.47	4.46	6.95	4.39	1.65	1.70	1.65	1.68	5.97	4.84	6.45	4.77

Table 6: Root mean square errors for Simulation LN1

This table presents the root mean square errors of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LN1 in subsection 2.5. The cross-section size is $N = 300$, and the time-series length is $T = 10$. The variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) take the values $(0, 0.25)$, $(0.25, 0)$, and $(0.25, 0.25)$. The results are obtained using a simulation with 50,000 repetitions. All numbers are expressed in multiples of 10^{-2}.

	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
(ρ_ω, ρ_ϕ) / (ρ_μ, ρ_ξ)	(0,.25)				(.25,0)				(.25,.25)			
(0,.25)	3.6	3.7	4.1	4.1	3.2	3.2	3.7	3.7	3.7	3.7	4.3	4.4
(.25,0)	4.2	3.9	4.7	4.3	4.6	3.7	4.8	4.1	5.6	4.5	5.8	5.0
(.25,.25)	3.7	3.2	4.1	3.5	4.2	3.1	4.2	3.4	5.2	3.7	5.2	4.1

Table 7: Root mean square errors for Simulation LN2

This table presents the root mean square errors of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LN2 in subsection 2.5. The cross-section size is $N = 300$, and the time-series length is $T = 10$. The variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) are both equal to $(0.25, 0.25)$. The autocorrelations coefficients δ_ε and δ_x take the values -0.5, -0.25, 0.25, and 0.5. The results are obtained using a simulation with 50,000 repetitions. All numbers are expressed in multiples of 10^{-2}.

	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
δ_ε / δ_x	-0.5				-0.25				0.25				0.5			
-0.5	5.1	3.7	5.5	4.1	5.0	3.4	5.2	3.7	5.0	3.0	5.1	3.5	5.2	3.0	5.4	3.6
-0.25	5.0	3.9	5.3	4.4	5.0	3.7	5.2	4.1	5.2	3.5	5.2	4.0	5.4	3.6	5.6	4.2
0.25	5.1	3.1	5.2	3.7	5.2	3.3	5.2	3.8	5.5	3.7	5.5	4.1	5.9	4.1	6.0	4.5
0.5	5.2	2.5	5.3	3.0	5.4	2.7	5.3	3.1	5.9	3.2	5.8	3.6	6.3	3.7	6.4	4.1

Table 8: Root mean square errors for Simulation LN3

This table presents the root mean square errors of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LN3 in subsection 2.5. The cross-section size is $N = 300$, and the time-series length is $T = 10$. The variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) are set equal to $(0.25, 0.25)$. The autocorrelations coefficients δ_ε and δ_x are set equal to 0.5. All random numbers are generated from a standardized gamma distribution. The skewness parameters s_ε and s_x take values -1, -0.5, 0.5, and 1. The results are obtained using a simulation with 50,000 repetitions. All numbers are expressed in multiples of 10^{-2}.

	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
s_ε / s_x	-1				-0.5				0.5				1			
-1	6.3	3.7	6.5	4.3	6.3	3.7	6.4	4.2	6.3	3.7	6.4	4.2	6.3	3.7	6.5	4.2
-0.5	6.3	3.7	6.5	4.2	6.3	3.8	6.4	4.2	6.3	3.7	6.4	4.2	6.3	3.7	6.5	4.3
0.5	6.3	3.7	6.5	4.3	6.3	3.7	6.4	4.2	6.3	3.7	6.4	4.2	6.3	3.7	6.6	4.2
1	6.3	3.7	6.5	4.3	6.3	3.8	6.4	4.2	6.3	3.7	6.4	4.2	6.3	3.7	6.5	4.3

Table 9: Root mean square errors for Simulation LN4

This table presents the root mean square errors of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LN4 in subsection 2.5. The variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) are set equal to $(0.25, 0.25)$. The autocorrelations coefficients δ_ε and δ_x are set equal to 0.5. The cross-section size N takes values 100, 200, and 300, and the time-series length T takes values 10, 20, and 30. The results are obtained using a simulation with 50,000 repetitions. All numbers are expressed in multiples of 10^{-2}.

	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
N / T	10				20				30			
100	10.9	6.7	11.1	7.3	9.2	5.2	8.5	5.3	8.4	4.6	7.5	4.6
200	7.8	4.7	7.8	5.1	6.5	3.4	6.0	3.5	6.0	2.9	5.3	2.9
300	6.3	3.7	6.4	4.1	5.3	2.7	4.9	2.8	4.9	2.2	4.4	2.3

Table 10: Root mean square errors for Simulation LN5

This table presents the root mean square errors of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LN5 in subsection 2.5. The cross-section size is $N = 300$, and the time-series length is $T = 10$. The variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) take the values $(0, 0.25)$, $(0.25, 0)$, and $(0.25, 0.25)$. The results are obtained using a simulation with 50,000 repetitions. All numbers are expressed in multiples of 10^{-2}.

	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
(ρ_ω, ρ_ϕ) / (ρ_μ, ρ_ξ)	(0,.25)				(.25,0)				(.25,.25)			
(0,.25)	15.8	8.1	16.4	8.8	3.6	3.3	4.2	3.8	15.8	8.2	17.6	9.5
(.25,0)	3.7	3.3	4.1	3.7	4.6	3.7	4.7	4.1	4.6	3.8	4.8	4.1
(.25,.25)	15.8	7.5	16.4	8.2	4.5	3.2	4.7	3.6	16.0	9.0	17.7	9.8

Table 11: Empirical p-values and cut-off values for Simulation LT1

This table presents the empirical p-values and the cut-off values of the t-statistics of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LT1 in subsection 3.4. The nominal p-values, denoted by p, are included in the first column, while the nominal cut-off values, denoted by x_p, are included in the second column. The empirical p-values are included in the upper block, while the empirical cut-off values are included in the lower block. All p-values are stated in percentages. The time-series length is $T = 240$ and the cross-section size is $N = 25$. The autoregressive coefficients δ_f and δ_g are set equal to 0.5. The parameters c_λ and c_κ are set equal to 0.5. The variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) take the values $(0, 0)$, $(0, 0.25)$, and $(0.25, 0)$. The results are obtained using a simulation with 50,000 repetitions.

p	x_p		LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
		(ρ_ω, ρ_ϕ) / (ρ_μ, ρ_ξ)	(0,0)				(0,.25)				(.25,0)			
1		(0,0)	1.9	1.7	1.7	4.4	1.9	1.7	1.6	4.0	1.9	1.5	1.7	4.5
5			6.7	6.1	6.2	12.2	7.0	6.6	6.4	11.6	6.9	6.1	6.4	12.4
10			12.2	11.2	11.7	19.4	12.5	11.8	11.7	18.3	12.4	11.5	11.8	19.3
1		(0,.25)	1.9	1.5	1.6	4.9	1.8	1.6	1.6	4.4	1.8	1.5	1.6	4.9
5			6.9	5.6	6.4	13.1	6.7	5.8	6.4	11.9	6.8	5.6	6.4	12.8
10			12.4	10.2	11.8	20.1	12.3	10.6	11.8	19.1	12.5	10.3	11.7	19.9
1		(.25,0)	1.8	1.7	1.6	4.4	1.9	1.7	1.6	4.0	1.8	1.6	1.5	4.5
5			6.6	6.2	6.5	12.2	6.9	6.5	6.4	11.4	6.7	6.1	6.2	12.3
10			12.2	11.2	11.8	19.3	12.3	11.8	11.7	18.2	12.2	11.1	11.7	19.3
	2.58	(0,0)	2.88	2.79	2.79	3.37	2.85	2.83	2.77	3.28	2.87	2.76	2.78	3.38
	1.96		2.11	2.06	2.07	2.51	2.13	2.10	2.08	2.46	2.13	2.06	2.09	2.51
	1.64		1.75	1.70	1.73	2.09	1.77	1.73	1.73	2.06	1.77	1.72	1.73	2.10
	2.58	(0,.25)	2.87	2.74	2.78	3.40	2.81	2.78	2.75	3.33	2.84	2.74	2.76	3.47
	1.96		2.13	2.02	2.08	2.57	2.11	2.03	2.08	2.50	2.11	2.01	2.07	2.56
	1.64		1.77	1.66	1.73	2.13	1.76	1.68	1.73	2.07	1.77	1.66	1.73	2.12
	2.58	(.25,0)	2.83	2.80	2.78	3.38	2.86	2.80	2.79	3.32	2.83	2.78	2.75	3.34
	1.96		2.11	2.07	2.09	2.51	2.13	2.09	2.08	2.44	2.11	2.06	2.05	2.52
	1.64		1.75	1.71	1.73	2.09	1.76	1.74	1.73	2.04	1.76	1.70	1.72	2.09

Table 12: Empirical p-values and cut-off values for Simulation LT2

This table presents the empirical p-values and the cut-off values of the t-statistics of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LT2 in subsection 3.4. The nominal p-values, denoted by p, are included in the first column, while the nominal cut-off values, denoted by x_p, are included in the second column. The empirical p-values are included in the upper block, while the empirical cut-off values are included in the lower block. All p-values are stated in percentages. The time-series length is $T = 240$ and the cross-section size is $N = 25$. The variance fractions ρ_ϕ, ρ_ω, ρ_ξ, and ρ_μ are all set equal to 0. The parameters c_λ and c_κ are set equal to 0.5. The autoregressive coefficients δ_f and δ_g take the values -0.75, -0.5, 0.5, and 0.75. The results are obtained using a simulation with 50,000 repetitions.

p	x_p		LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
		δ_f / δ_g	-0.75				-0.5				0.5				0.75			
1		-0.75	2.3	1.4	2.0	7.5	1.7	1.6	1.6	5.4	1.5	1.8	1.6	2.6	1.7	1.9	1.5	2.6
5			7.7	5.1	7.3	17.2	6.8	5.9	6.4	13.9	6.3	6.9	6.3	8.9	6.6	7.0	6.3	8.4
10			13.3	9.8	13.3	25.0	12.2	10.6	12.0	21.2	11.7	12.5	11.8	15.5	12.1	12.6	12.1	14.4
1		-0.5	2.0	1.5	1.7	5.2	1.7	1.6	1.5	4.4	1.6	1.9	1.5	2.9	1.8	1.8	1.5	2.6
5			7.2	5.8	6.5	13.8	6.7	6.1	6.1	12.3	6.2	6.9	6.1	9.3	6.7	7.0	6.2	8.8
10			12.6	10.8	12.0	21.1	12.1	11.3	11.4	19.4	11.7	12.5	11.6	15.6	12.1	12.6	11.6	14.9
1		0.5	1.7	1.9	1.6	2.6	1.7	1.9	1.6	3.0	1.9	1.6	1.6	4.5	2.2	1.6	1.7	5.4
5			6.5	6.9	6.3	8.9	6.3	6.7	6.2	9.4	6.8	6.1	6.1	12.2	7.2	5.9	6.6	13.6
10			12.1	12.4	11.8	15.1	11.8	12.3	11.4	15.8	12.3	11.3	11.6	19.4	12.8	10.8	12.0	20.8
1		0.75	1.8	2.0	1.6	2.5	1.6	2.0	1.5	2.7	1.9	1.7	1.7	5.2	2.5	1.3	2.2	7.4
5			6.8	7.3	6.2	8.3	6.5	7.1	6.2	9.2	7.2	6.0	6.7	13.7	8.3	5.2	7.8	16.9
10			12.2	12.7	11.8	14.5	11.9	12.6	11.6	15.3	12.9	11.0	12.6	21.2	14.1	9.7	13.6	24.7
	2.58	-0.75	2.96	2.70	2.86	3.83	2.81	2.77	2.75	3.50	2.75	2.84	2.77	3.00	2.78	2.89	2.74	2.98
	1.96		2.18	1.97	2.14	2.85	2.11	2.04	2.07	2.62	2.07	2.12	2.08	2.27	2.09	2.13	2.06	2.24
	1.64		1.81	1.63	1.81	2.37	1.76	1.68	1.74	2.19	1.73	1.77	1.73	1.90	1.75	1.77	1.74	1.86
	2.58	-0.5	2.88	2.74	2.78	3.51	2.82	2.79	2.75	3.38	2.77	2.85	2.75	3.06	2.83	2.84	2.75	3.04
	1.96		2.15	2.04	2.09	2.61	2.10	2.06	2.06	2.50	2.07	2.13	2.06	2.30	2.11	2.11	2.06	2.26
	1.64		1.78	1.68	1.74	2.17	1.76	1.71	1.72	2.10	1.73	1.77	1.72	1.92	1.75	1.78	1.73	1.89
	2.58	0.5	2.77	2.85	2.76	3.04	2.78	2.85	2.76	3.09	2.84	2.80	2.78	3.39	2.95	2.76	2.80	3.50
	1.96		2.09	2.13	2.06	2.25	2.08	2.12	2.06	2.31	2.11	2.06	2.06	2.50	2.16	2.04	2.10	2.62
	1.64		1.75	1.76	1.73	1.90	1.74	1.76	1.71	1.92	1.76	1.71	1.72	2.09	1.79	1.68	1.74	2.18
	2.58	0.75	2.82	2.86	2.77	3.01	2.75	2.87	2.72	3.06	2.86	2.80	2.79	3.47	3.02	2.70	2.88	3.86
	1.96		2.11	2.14	2.07	2.23	2.08	2.14	2.07	2.29	2.14	2.05	2.10	2.60	2.23	1.98	2.19	2.85
	1.64		1.75	1.78	1.73	1.86	1.74	1.78	1.73	1.91	1.79	1.70	1.77	2.17	1.85	1.63	1.82	2.36

Table 13: Empirical p-values and cut-off values for Simulation LT3

This table presents the empirical p-values and the cut-off values of the t-statistics of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LT3 in subsection 3.4. The nominal p-values, denoted by p, are included in the first column, while the nominal cut-off values, denoted by x_p, are included in the second column. The empirical p-values are included in the upper block, while the empirical cut-off values are included in the lower block. All p-values are stated in percentages. The time-series length is $T = 240$ and the cross-section size is $N = 25$. The variance fractions ρ_ϕ, ρ_ω, ρ_ξ, and ρ_μ are all set equal to 0. The autoregressive coefficients δ_f and δ_g are set equal to 0.5. The parameters c_λ and c_κ take values 0.25, 0.5, and 0.75. The results are obtained using a simulation with 50,000 repetitions.

p	x_p		LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
		c_λ												
		c_κ	0.25				0.5				0.75			
1		0.25	1.9	2.0	1.6	3.4	1.8	2.0	1.6	3.5	2.0	2.1	1.6	3.3
5			6.6	7.1	6.3	10.3	6.9	7.3	6.3	10.7	7.0	7.8	6.2	10.0
10			12.1	12.7	11.8	17.2	12.6	13.0	11.9	17.4	12.5	13.7	11.6	16.6
1		0.5	1.8	1.7	1.7	4.4	1.9	1.7	1.7	4.6	1.9	1.9	1.5	4.1
5			6.8	6.3	6.5	12.2	7.0	6.3	6.4	12.3	7.0	6.7	6.3	11.5
10			12.4	11.4	12.0	19.3	12.5	11.3	11.8	19.3	12.4	12.2	11.8	18.4
1		0.75	1.8	1.1	1.5	6.9	1.8	1.1	1.6	7.3	1.9	1.2	1.5	6.2
5			6.6	4.4	6.4	16.3	6.8	4.3	6.2	17.0	6.9	5.0	6.2	15.2
10			12.2	8.2	11.8	23.9	12.3	8.2	11.5	24.6	12.3	9.3	11.7	22.8
	2.58	0.25	2.85	2.89	2.79	3.19	2.84	2.86	2.75	3.21	2.90	2.91	2.77	3.14
	1.96		2.09	2.14	2.07	2.38	2.12	2.15	2.07	2.39	2.14	2.19	2.08	2.35
	1.64		1.75	1.78	1.74	1.98	1.77	1.79	1.73	2.00	1.77	1.83	1.72	1.96
	2.58	0.5	2.83	2.79	2.79	3.32	2.84	2.81	2.80	3.37	2.84	2.85	2.73	3.32
	1.96		2.11	2.07	2.08	2.50	2.11	2.07	2.08	2.53	2.12	2.10	2.06	2.46
	1.64		1.76	1.72	1.74	2.09	1.77	1.72	1.73	2.10	1.77	1.75	1.73	2.05
	2.58	0.75	2.82	2.62	2.76	3.78	2.83	2.62	2.78	3.76	2.86	2.66	2.76	3.64
	1.96		2.09	1.90	2.07	2.79	2.12	1.89	2.06	2.82	2.12	1.95	2.06	2.71
	1.64		1.75	1.54	1.73	2.32	1.76	1.56	1.72	2.35	1.76	1.61	1.73	2.26

Table 14: Empirical p-values and cut-off values for Simulation LT4

This table presents the empirical p-values and the cut-off values of the t-statistics of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LT4 in subsection 3.4. The nominal p-values, denoted by p, are included in the first column, while the nominal cut-off values, denoted by x_p, are included in the second column. The empirical p-values are included in the upper block, while the empirical cut-off values are included in the lower block. All p-values are stated in percentages. The variance fractions ρ_ϕ, ρ_ω, ρ_ξ, and ρ_μ are all set equal to 0. The parameters c_λ and c_κ are set equal to 0.5. The autoregressive coefficients δ_f and δ_g are set equal to 0.5. The time-series length T takes values 30, 60, 90, and 120 and the cross-section size N takes values 250, 500, 750, and 1000. The results are obtained using a simulation with 50,000 repetitions.

p	x_p		LS	FM	LS	FM	LS	FM	LS	FM
		T								
		N	250		500		750		1000	
1		30	8.3	8.3	8.1	8.1	8.2	8.2	8.1	8.1
5			15.2	15.3	15.2	15.1	15.3	15.2	15.1	15.0
10			21.3	21.3	21.3	21.3	21.4	21.4	21.2	21.2
1		60	3.7	3.7	3.6	3.6	3.6	3.6	3.6	3.6
5			9.7	9.6	9.5	9.5	9.6	9.7	9.3	9.4
10			15.6	15.6	15.3	15.4	15.3	15.2	15.1	15.1
1		90	2.5	2.5	2.4	2.5	2.5	2.5	2.4	2.4
5			7.8	7.7	7.7	7.7	7.6	7.6	7.8	7.8
10			13.5	13.5	13.4	13.3	13.2	13.3	13.2	13.3
1		120	2.4	2.3	2.4	2.3	2.3	2.4	2.4	2.4
5			7.6	7.6	7.5	7.5	7.6	7.6	7.5	7.5
10			13.0	13.0	13.0	13.0	13.3	13.3	12.9	12.9
	2.58	30	5.52	5.44	5.42	5.59	5.59	5.65	5.53	5.62
	1.96		3.12	3.14	3.12	3.12	3.14	3.14	3.12	3.13
	1.64		2.38	2.38	2.38	2.37	2.37	2.37	2.36	2.36
	2.58	60	3.51	3.49	3.52	3.53	3.52	3.52	3.47	3.47
	1.96		2.39	2.39	2.36	2.36	2.38	2.38	2.36	2.36
	1.64		1.94	1.94	1.93	1.93	1.93	1.94	1.91	1.92
	2.58	90	3.05	3.04	3.03	3.04	3.05	3.05	3.07	3.04
	1.96		2.22	2.21	2.20	2.20	2.19	2.19	2.20	2.20
	1.64		1.82	1.82	1.81	1.82	1.80	1.80	1.82	1.81
	2.58	120	2.99	3.01	3.01	3.01	3.01	3.02	3.04	3.02
	1.96		2.19	2.19	2.18	2.18	2.18	2.18	2.18	2.18
	1.64		1.80	1.80	1.80	1.80	1.81	1.81	1.79	1.79

Table 15: Empirical p-values and cut-off values for Simulation LT5

This table presents the empirical p-values and the cut-off values of the t-statistics of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LT5 in subsection 3.4. The nominal p-values, denoted by p, are included in the first column, while the nominal cut-off values, denoted by x_p, are included in the second column. The empirical p-values are included in the upper block, while the empirical cut-off values are included in the lower block. All p-values are stated in percentages. The variance fractions ρ_ϕ, ρ_ω, ρ_ξ, and ρ_μ are all set equal to 0. The parameters c_λ and c_κ are set equal to 0.5. The autoregressive coefficients δ_f and δ_g are set equal to 0. The time-series length T takes values 30, 60, 90, and 120 and the cross-section size N takes values 250, 500, 750, and 1000. The results are obtained using a simulation with 50,000 repetitions.

p	x_p		LS	FM	LS	FM	LS	FM	LS	FM
		T								
		N	250		500		750		1000	
1		30	9.4	8.1	9.4	8.0	9.5	7.9	9.3	7.8
5			17.1	15.1	17.1	15.1	17.0	14.9	17.0	14.8
10			23.2	21.1	23.6	21.3	23.3	21.0	23.3	20.9
1		60	4.2	3.4	4.4	3.7	4.4	3.6	4.3	3.6
5			10.3	9.2	10.6	9.5	10.5	9.2	10.5	9.5
10			16.1	15.1	16.4	15.4	16.3	15.0	16.5	15.4
1		90	3.0	2.5	2.8	2.4	2.9	2.5	2.9	2.4
5			8.6	7.7	8.5	7.7	8.4	7.7	8.4	7.5
10			14.3	13.2	14.4	13.2	14.1	13.2	14.1	13.1
1		120	2.6	2.3	2.7	2.4	2.6	2.4	2.7	2.3
5			7.9	7.6	8.2	7.6	8.1	7.6	8.2	7.6
10			13.5	13.1	13.8	12.9	13.7	13.1	13.7	13.0
	2.58	30	5.95	5.58	5.95	5.52	6.04	5.48	5.89	5.46
	1.96		3.32	3.13	3.30	3.07	3.31	3.07	3.30	3.07
	1.64		2.51	2.36	2.51	2.35	2.51	2.34	2.50	2.34
	2.58	60	3.66	3.45	3.69	3.55	3.70	3.48	3.68	3.55
	1.96		2.44	2.34	2.49	2.37	2.48	2.35	2.46	2.38
	1.64		1.98	1.91	2.00	1.93	1.99	1.91	2.00	1.93
	2.58	90	3.14	3.01	3.12	3.04	3.21	3.04	3.15	3.04
	1.96		2.28	2.20	2.27	2.20	2.26	2.20	2.26	2.18
	1.64		1.87	1.81	1.86	1.81	1.86	1.81	1.86	1.80
	2.58	120	3.11	2.98	3.11	3.03	3.11	3.00	3.10	3.03
	1.96		2.22	2.18	2.23	2.18	2.22	2.18	2.23	2.18
	1.64		1.83	1.81	1.84	1.80	1.83	1.81	1.84	1.80

Table 16: Root mean square errors for Simulation LT1

This table presents the root mean square errors of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LT1 in subsection 3.4. The time-series length is $T = 240$ and the cross-section size is $N = 25$. The variance proportion pairs (ρ_ω, ρ_ϕ) and (ρ_μ, ρ_ξ) take the values $(0, 0)$, $(0, 0.25)$, and $(0.25, 0)$. The results are obtained using a simulation with 50,000 repetitions. All numbers are expressed in multiples of 10^{-2}.

	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
(ρ_ω, ρ_ϕ) / (ρ_μ, ρ_ξ)	(0,0)				(0,.25)				(.25,0)			
(0,0)	9.5	2.3	7.9	2.4	9.6	2.5	8.4	2.7	11.1	2.6	9.1	2.8
(0,.25)	9.6	2.1	8.1	2.2	9.9	2.3	8.7	2.4	11.1	2.4	9.3	2.5
(.25,0)	8.3	2.0	6.8	2.1	8.3	2.2	7.2	2.3	9.5	2.3	7.8	2.4

Table 17: Root mean square errors for Simulation LT2

This table presents the root mean square errors of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LT2 in subsection 3.4. The time-series length is $T = 240$ and the cross-section size is $N = 25$. The variance fractions ρ_ϕ, ρ_ω, ρ_ξ, and ρ_μ are all set equal to 0. The parameters c_λ and c_κ are set equal to 0.5. The autoregressive coefficients δ_f and δ_g take the values -0.75, -0.5, 0.5, and 0.75. The results are obtained using a simulation with 50,000 repetitions. All numbers are expressed in multiples of 10^{-2}.

	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
δ_f / δ_g	-.75				-.5				.5				.75			
-.75	13.7	2.7	10.8	2.8	10.9	2.4	8.8	2.5	5.4	2.0	5.5	2.2	4.5	1.9	4.8	2.1
-.5	10.9	2.4	8.9	2.5	9.6	2.3	7.9	2.4	6.0	2.0	5.8	2.2	5.4	2.0	5.3	2.2
.5	5.4	2.0	5.3	2.2	6.0	2.0	5.8	2.2	9.5	2.3	7.8	2.4	10.9	2.4	8.8	2.5
.75	4.4	1.9	4.7	2.1	5.3	2.0	5.4	2.2	10.8	2.4	8.7	2.5	13.6	2.7	10.7	2.7

Table 18: Root mean square errors for Simulation LT3

This table presents the root mean square errors of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LT3 in subsection 3.4. The time-series length is $T = 240$ and the cross-section size is $N = 25$. The variance fractions ρ_ϕ, ρ_ω, ρ_ξ, and ρ_μ are all set equal to 0. The autoregressive coefficients δ_f and δ_g are set equal to 0.5. The parameters c_λ and c_κ take values 0.25, 0.5, and 0.75. The results are obtained using a simulation with 50,000 repetitions. All numbers are expressed in multiples of 10^{-2}.

	LS	GLS	FM	GFM	LS	GLS	FM	GFM	LS	GLS	FM	GFM
c_λ / c_κ	0.25				0.5				0.75			
0.25	6.9	2.2	5.7	2.3	8.1	2.5	6.8	2.7	9.4	3.2	8.6	3.4
0.5	8.1	1.9	6.6	2.1	9.6	2.3	7.9	2.4	11.0	2.9	10.0	3.1
0.75	9.3	1.6	7.5	1.7	11.0	1.9	9.0	2.0	12.8	2.5	11.5	2.6

Table 19: Root mean square errors for Simulation LT4

This table presents the root mean square errors of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LT4 in subsection 3.4. The variance fractions ρ_ϕ, ρ_ω, ρ_ξ, and ρ_μ are all set equal to 0. The parameters c_λ and c_κ are set equal to 0.5. The autoregressive coefficients δ_f and δ_g are set equal to 0.5. The time-series length T takes values 30, 60, 90, and 120 and the cross-section size N takes values 250, 500, 750, and 1000. The results are obtained using a simulation with 50,000 repetitions. All numbers are expressed in multiples of 10^{-3}.

	LS	FM	LS	FM	LS	FM	LS	FM
T								
N	250		500		750		1000	
30	23.5	23.6	16.6	16.6	13.6	13.6	11.8	11.8
60	16.6	16.6	11.7	11.7	9.5	9.5	8.2	8.2
90	13.5	13.5	9.5	9.5	7.7	7.8	6.7	6.7
120	11.6	11.6	8.2	8.2	6.7	6.7	5.8	5.8

Table 20: Root mean square errors for Simulation LT5

This table presents the root mean square errors of the LS, GLS, FM, and GFM estimators. For a description of the data generating process, see Simulation LT5 in subsection 3.4. The variance fractions ρ_ϕ, ρ_ω, ρ_ξ, and ρ_μ are all set equal to 0. The parameters c_λ and c_κ are set equal to 0.5. The autoregressive coefficients δ_f and δ_g are set equal to 0. The time-series length T takes values 30, 60, 90, and 120 and the cross-section size N takes values 250, 500, 750, and 1000. The results are obtained using a simulation with 50,000 repetitions. All numbers are expressed in multiples of 10^{-1}.

	LS	FM	LS	FM	LS	FM	LS	FM
T								
N	250		500		750		1000	
30	2.6	2.3	2.6	2.3	2.6	2.3	2.6	2.3
60	1.8	1.6	1.9	1.6	1.8	1.6	1.8	1.6
90	1.5	1.3	1.5	1.3	1.5	1.3	1.5	1.3
120	1.3	1.1	1.3	1.1	1.3	1.1	1.3	1.1