Two-Pass Cross-Sectional Regression of Factor Pricing Models

Seung C. Ahn
Arizona State University

Christopher Gadarowski
Rowan University

Marcos Fabricio Perez
Wilfrid Laurier University

This Version: May 4, 2009

Abstract

The two-pass (TP) cross-sectional regression method has been widely used to evaluate linear factor pricing models. This paper examines the finite-sample properties of this method when asset returns are conditionally heteroskedastic and/or autocorrelated. Using minimum distance estimation, we derive heteroskedasticity-and/or-autocorrelation-robust model specification test statistics and asymptotic variances of the TP risk premium estimates. We also derive optimal GLS estimators that are asymptotically more efficient than other two-pass estimators. Also, several findings are obtained from our simulation exercises. First, conditional heteroskedasticity and autocorrelation in the idiosyncratic errors alone do not produce large biases in the TP estimators. Second, use of the *t*-tests from TP estimators based on OLS as Shanken (1985) and OLS with our proposed heteroskedasticity correction is recommended. Third, the use of our optimal model specification test robust to heteroskedasticity is recommended for the cases in which the errors are heteroskedastic but not autocorrelated, especially when a large number of portfolios are analyzed. Forth, all model specification tests are generally unreliable when the idiosyncratic errors are autocorrelated.

***JEL classification*:** C12, C13, C3.

Acknowledgments

The first author gratefully acknowledges the financial support of the College of Business and Dean's Council of 100 at Arizona State University, the Economic Club of Phoenix, and the alumni of the College of Business. We gratefully acknowledge seminar participants at Sogang University and Arizona State University, especially Hank Bessembinder, John Griffin and Mike Lemmon. Zhenyu Wang and Guofu Zhou provided constructive comments on the earlier version of the paper and detailed notes on a topic related to this paper. We also thank the editor and two anonymous referees for their constructive comments and suggestions that helped us improve the paper greatly. All errors are of course our own. Corresponding author: Seung C. Ahn, Department of Economics, W.P. Carey School of Business, Arizona State University, Tempe, AZ 85287, USA; tel: 480-965-6574; fax: 480-965-0748; email: miniahn@asu.edu.

1. Introduction

The two-pass cross-sectional regression method, first used by Black, Jensen and Scholes (1972) and Fama and MacBeth (1973), has been widely used to analyze linear factor pricing models, including the capital asset pricing model (CAPM), arbitrage pricing theory (APT) and their variants.[1] In the two-pass (TP) estimation, asset betas are first estimated by time-series OLS regression of the asset's return on a set of common factors, and then, factor risk prices are estimated by cross-sectional OLS or GLS regressions of mean returns on betas. This methodology is simple and provides several convenient ways to test a given asset pricing model. Fama and Macbeth (1973) provide a convenient way to compute the standard errors of the TP estimators.

However, the simple Fama-MacBeth (FM) method suffers from the well-known errors-in-variables (EIV) problem. That is, because estimated betas are used in place of true betas in the second stage cross-sectional regression, the second-stage regression estimates in the Fama-MacBeth method do not have the usual OLS or GLS properties. Shanken (1992) suggests a correction for EIV on the TP estimators under the restrictive assumptions of no conditional heteroskedasticity and no conditional autocorrelation in asset returns (conditional on factors).

Some papers have proposed alternative estimation methods that are robust to conditional heteroskedasticity and/or autocorrelation. For example, Kim (1995) and Ferson and Harvey (1999) consider cases of conditional heteroskedasticity, but only of particular structures that are often disputed in empirical studies. Jagannathan and Wang (1998a) provide a general form for the correct asymptotic variance matrix of the two-pass estimator, allowing for both conditional heteroskedasticity and autocorrelation. Cochrane (2005) also proposes alternative estimators for the asymptotic variance matrices of the TP estimators based on the generalized method of moments (GMM) of Hansen (1982). However, these studies do not provide empirical evidence for the importance of controlling conditional heteroskedasticity or/and autocorrelation in the two-pass regression.

The main goal of this paper is two-fold. Our first contribution is theoretical, since extending Shanken's results (1985, 1992), we identify the optimal TP estimators for the models with conditionally heteroskedastic (GLS-TP2) and/or autocorrelated idiosyncratic errors (GLS-TP1). The optimal estimators have the form of GLS . We also derive the asymptotic

[1] See Campbell, Lo and MacKinlay (1997) for a summary of the major models and research in this area since the original works in the 1970s.

variances of the TP risk premium estimates. In addition, we develop two alternative model specification test. One robust to heteroskedasticity (Q_{TP2}) and other robust to both heteroskedasticity and autocorrelation (Q_{TP1}).

Second, we investigate the finite-sample properties of the TP estimators and the model specification tests. Though some previous work has investigated finite-sample properties of TP estimators, our study is able to isolate the pure effect of heteroskedastic and/or autocorrelated errors on the TP estimation. This is important since several other studies had found that the finite-sample properties of the TP estimation can be distorted even in the case of *i.i.d* errors. For example Kan and Zhang (1999) investigate the finite-sample properties of the TP estimators when the factors are uncorrelated with returns. They refer to these factors as "useless" factors. Also, . Ahn, Gadarowski and Perez (2009) find that even if a model has *i.i.d.* errors, the finite-sample properties of the TP estimators crucially depend on the structure of the beta matrix and the persistency level of the factors. They conclude that TP estimation can lead to biased statistical inferences when betas have small cross-section variations, and they are highly correlated. Also that persistent factors can distort the finite-sample properties of the TP estimators.

Other studies include Grauser and Janmaat (2005) who investigate the finite-sample power property of a TP-based test for CAPM under the three factor model of Fama and French (1993), and Chen and Kan (2004) consider the finite-sample biases in TP point estimators and their asymptotic standard errors. Both of these studies are limited to the cases in which returns and factors are independently and identically distributed over time, whereas our simulation exercises cover more realistic cases.

The main results of our simulation experiments are as follows. First, conditional heteroskedasticity and autocorrelation in the idiosyncratic errors alone do not produce large biases in the TP estimators. Second, the *t*-tests from TP based on OLS estimators (nonoptimal) are in general better sized than those from optimal GLS estimators. Use of the *t*-tests from TP estimators based on OLS as Shanken (1985) and OLS with heteroskedasticity correction (OLS-TP2) is generally recommended. Third, the use of our optimal model specification test that controls for heteroskedasticity (Q_{TP2}) is recommended for the cases in which the errors are heteroskedastic but not autocorrelated, especially when a large number of portfolios are analyzed. Forth, all model specification tests are generally unreliable when the idiosyncratic errors are autocorrelated. The proposed heteroskedasticity and autocorrelation robust test (Q_{TP1}) performs the best when the number of portfolios is small, but the size distortions are not corrected even in this case.

This paper is organized as follows. In section 2, we briefly discuss the basic asset pricing model of our interest and data generating assumptions. Section 3 presents the asymptotic distributions of the TP estimators, and model specification tests. Section 4 explains our simulation setup, while section 5 reports the simulation results. Concluding remarks follow in section 6.

2. Basic Model and Assumptions

The basic model we consider is a multifactor model in which return data are generated by k common factors:

$$r_{\bullet t} = \alpha + \mathrm{B} f_t + \varepsilon_{\bullet t} \equiv \Lambda z_t + \varepsilon_{\bullet t}, \tag{1}$$

where $r_{\bullet t} = (r_{1t}, r_{2t}, ..., r_{Nt})'$, $\alpha = (\alpha_1, \alpha_2, ..., \alpha_N)'$, $\mathrm{B} = (\beta_1, ..., \beta_N)'$, $\Lambda = (\alpha, \mathrm{B})$, $z_t = (1, f_t')'$, $\varepsilon_{\bullet t}$ = $(\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$, r_{it} is the gross return of asset i (= 1,2,..., N) at time t (= 1,...,T), $f_t = (f_{1t}, \ldots, f_{kt})'$ is the vector of k factors at time t, α_i is the asset-specific intercept term, β_i is the vector of k betas of asset i corresponding to f_t, and ε_{it} is the idiosyncratic error for asset i at time t with zero mean.[2] The factors in f_t are strictly exogenous to the error terms; that is, $E(z_{\bullet t} \otimes \varepsilon_{\bullet s}) = 0_{(k+1)N\times 1}$ for all s and t. The number of time-series observations (T) is large, while the number of assets analyzed (N) is relatively small, and then asymptotics apply as T approaches infinity. That is, this paper considers only the $\sqrt{T}$ -consistency of the TP estimator.[3]

The error vector $\varepsilon_{\bullet t}$ could be heteroskedastic or autocorrelated over time. Chan and Chen (1988) and Jagannathan and Wang (1996) have shown that under some assumptions, conditional CAPM can imply unconditional CAPM or unconditional multi-factor models. These studies, however, suggest models with time-varying betas and risk premia. If we view the factor model (1) as a linear approximation of such models, the error term $\varepsilon_{\bullet t}$ is likely to be autocorrelated.

The usual restriction imposed on (1) by linear asset pricing models is given by:

$$H_o : E(r_{\bullet t}) = 1_N \gamma_o + \mathrm{B}\gamma_f \equiv X\gamma, \tag{2}$$

where $X = [1_N, \mathrm{B}]$, $\gamma = (\gamma_0, \gamma_f')'$, 1_N is the $N \times 1$ vectors of ones, γ_o is an unknown constant (*e.g.*,

---

[2]If a risk-free asset yielding return r_{ft} is available, r_{it} may denote excess return ($r_{it} - r_{ft}$).

[3]For the conditions for $\sqrt{N}$ -consistency of the two-pass estimator, see Shanken (1992).

zero-beta return), and γ_f is the $k \times 1$ vector of factor risk prices. The traditional two-pass (TP) approach estimates the vector γ by regressing $\bar{r} = T^{-1}\Sigma_{t=1}^{T} r_{\bullet t}$ and $\hat{X} = [1_N, \hat{\mathrm{B}}]$ with an arbitrary positive-definite (and asymptotically nonstochastic) weighting matrix A, where $\hat{\mathrm{B}}$ is the OLS estimator of B:

$$\hat{\gamma}_{TP} = (\hat{\gamma}_{0,TP}, \hat{\gamma}'_{1,TP})' = (\hat{X}'A\hat{X})^{-1}\hat{X}'A\bar{r} \, . \tag{3}$$

There are many possible choices for A. If we choose $A = I_N$, then the two-pass estimator $\hat{\gamma}_{\mathrm{TP}}$ becomes an OLS estimator. In contrast, with the choice of $A = [Var(\varepsilon_{\bullet t})]^{-1}$, the inversed unconditional variance matrix of the $\varepsilon_{\bullet t}$, the two-pass estimator $\hat{\gamma}_{\mathrm{TP}}$ becomes the GLS estimator considered by Shanken (1992) and Kandel and Stambaugh (1995).

A problem of the TP estimator (3) is that it uses the estimated beta matrix, $\hat{\mathrm{B}}$, because the true beta matrix, B, is not observed. This generates the well-known EIV problem. Shanken (1992) shows that despite this problem, the TP estimator is consistent and asymptotically normal. Further, under the assumption that the $\varepsilon_{\bullet t}$ are independently and identically distributed (*i.i.d.*) over time, he provides the correct asymptotic variance matrix of the TP estimator explicitly incorporating the sampling errors generated by the use of the estimated beta. A more general variance matrix can be found in Jagannathan and Wang (1998a).

An alternative method often used in the literature to avoid the EIV problem is maximum likelihood estimation. Work in this area includes Gibbons (1982), Kandel (1984), Shanken (1986), Gibbons, Ross and Shanken (1989), and Zhou (1998). This method assumes that asset returns are normally distributed and homoskedastic conditional on given factors. Under these assumptions, asset betas and factor risk premiums are jointly estimated. In particular, the maximum likelihood estimation (MLE) approach focuses on an alternative null hypothesis,

$$H_o^{\alpha} : \alpha = \lambda_o 1_N + \mathrm{B}\lambda_f \equiv X\lambda \, , \tag{4}$$

where $\lambda = (\lambda_o, \lambda_f')'$, α is the vector of individual intercept terms in the first-pass model (1), and λ_f is a unknown $k \times 1$ vector. In fact, this hypothesis is equivalent to H_o in (2), because $\lambda_o = \gamma_o$ and $\lambda_f + E(f_t) = \gamma_f$ (Campbell, Lo and MacKinlay, 1997, p. 227). Note that given the specification (4), the vector of risk prices, γ_f, is decomposed into the population mean of the factor vector, $E(f_t)$, and the lambda component, $\lambda_f = \gamma_f - E(f_t)$. This lambda component can be interpreted as the vector of factor-mean adjusted risk prices (Zhou, 1998). The MLE method jointly estimates B, λ_o, and λ_f, and tests the hypothesis H_o^{α} by a standard likelihood ratio (LR) test. Then the vector of risk prices, γ_f, is estimated by the sum of the estimated λ_f and the sample mean of the factor vector, $\overline{f} = T^{-1}\Sigma_{t=1}^{T} f_t$. This MLE procedure is efficient under the assumption that the returns and factors are jointly normal.

We can also estimate λ by regressing the OLS estimate of α ($\hat{\alpha}$) on $\hat{X}$:

$$\hat{\lambda}_{TP} = (\hat{\lambda}_{o,TP}, \hat{\lambda}'_{f,TP})' = (\hat{X}'A\hat{X})^{-1}\hat{X}'A\hat{\alpha}. \tag{5}$$

It is straightforward to show:

$$\hat{\gamma}_{o,TP} = \hat{\lambda}_{o,TP}\,;\ \hat{\gamma}_{f,TP} = \hat{\lambda}_{f,TP} + \overline{f} \tag{6}$$

Thus, we can compute the asymptotic variance of $\hat{\gamma}_{TP}$ using the asymptotic variance matrices of $\hat{\lambda}_{TP}$ and $\overline{f}$.

In order to derive the asymptotic variance matrices of the TP estimators $\hat{\lambda}_{TP}$ and $\hat{\gamma}_{TP}$, we need to make some assumptions on the time-series model (1). Specifically, the following set of conditions is sufficient to obtain the main results of this paper.

Assumption 1: (i) The data $r_{\bullet t}$ and f_t are covariance-stationary, ergodic, and have finite moments up to the fourth order. (ii) $E(z_{\bullet t} \otimes \varepsilon_{\bullet s}) = 0_{(1+k)N\times 1}$ for all *t* and *s*: That is, the factors are strictly exogenous to the errors. (iii) $X = [1_N, \mathrm{B}]$ is of full column; that is, all the columns in B are linearly independent and every element in β_i varies over different *i*.

Several comments on Assumption 1 are worth noting.[4] First, Assumption 1 is general enough to subsume most of the assumptions that have been adopted in the literature. Under Assumptions 1(i) and (ii),

$$\mathrm{B} = E[(r_{\bullet t} - E(r_{\bullet t}))(f_t - E(f_t))'][Var(f_t)]^{-1}.$$

Thus, the matrix B reserves the usual beta interpretation. Assumption 1(ii) guarantees the consistency of OLS estimation of Λ. Second, Assumption (i) rules out nonstationary factors and

[4]More technical conditions are required for the consistency and asymptotic normality of the estimators discussed below. For detailed technical conditions, see Hansen (1982) or White (1984).

nonstationary idiosyncratic errors. But it allows the factors f_t and returns $r_{\bullet t}$ to be conditionally heteroskedastic and/or serially correlated over time. Third, Assumption 1(ii) also warrants that $Var(\hat{\gamma}_f) = Var(\hat{\lambda}_f) + Var(\bar{f})$, as discussed later. Thus, the asymptotic distribution of $\hat{\gamma}_f$ can be found from those of $\hat{\lambda}_f$ and $\bar{f}$. Assumption 1 also allows for cross-section dependence among the idiosyncratic errors in $\varepsilon_{\bullet t}$.

Assumption 1(iii) is necessary for the identification of factor prices by cross-section regression. The assumption rules out perfect multicollinearity in the beta matrix B. For the assumption to hold, $Var(f_t)$ should be nonsingular: that is, f_t contains no redundant factors, and all of the columns in $E[(r_{\bullet t} - E(r_{\bullet t}))(f_t - E(f_t))']$ are linearly independent. Even if no perfect multicollinearity exists in B, near multicollinearity could hurt the finite-sample properties of the TP estimators as shown by Ahn, Gadarowski and Perez (2009).

Most of the previously mentioned studies using the TP estimators have imposed stronger assumptions than Assumption 1. We here consider two additional assumptions:

Assumption 2: In addition to Assumption 1, $E(\varepsilon_{\bullet t}\varepsilon_{\bullet s}' \mid f_1,..., f_T) = 0_{N\times N}$, for all $t \neq s$. That is, the errors $\varepsilon_{\bullet t}$ are serially uncorrelated given factors.

Assumption 3: In addition to Assumption 2, $Var(\varepsilon_{\bullet t} \mid f_1,..., f_T) = \Sigma_\varepsilon$, for any t, where Σ_ε is the unconditional variance matrix of $\varepsilon_{\bullet t}$.

Assumption 2 is stronger than Assumption 1 because it rules out autocorrelation in the errors. However, it is still a general and plausible assumption because it allows for heteroskedasticity and the autocorrelations in observed return data are generally weak. Assumption 3 now rules out heteroskedasticity as well as autocorrelation in the errors. Although this assumption is often empirically disputed, it is often assumed in many previous studies. For example, under the assumption, Shanken (1985, 1992) derived the asymptotic distributions of the TP estimators and proposed a specification test statistic for the hypothesis (2). In the next section, we consider the asymptotic distribution of the TP estimator under each of Assumptions 1-3.

3. Asymptotic Distribution of the Two-Pass Estimator

Under Assumptions 1 or 2, Jagannathan and Wang (1998a) and Cochrane (2005) have studied the asymptotic distribution of the TP estimator under Assumptions 1 or 2. In this section, we reexamine the asymptotic distribution of the TP estimator using the minimum distance approach developed by Ferguson (1958), Chamberlain (1984), Amemiya (1978) and Newey (1987). There are three reasons why the minimum-distance approach is attractive to study the asymptotic distribution of the TP estimator. First, using the approach, we can see in a more systematic and clear way how heteroskedasticity and autocorrelation in the errors can influence the asymptotic variance matrices of the TP estimators. Second, the approach enables us to find the optimal weighting matrix (A) which minimizes the asymptotic variance matrix of the TP estimator of λ under general conditions. Third, we can derive a model specification statistic, which resembles that of Shanken (1985), but which is robust to heteroskedastic and/or autocorrelated idiosyncratic errors.

3.1. Asymptotic Distribution of the Time-Series OLS Estimators

The distribution of a TP estimator depends on the distribution of the OLS estimator of $\Lambda = [\alpha, \mathrm{B}]$. So we here briefly review the OLS estimator's asymptotic distribution. The OLS estimator is given by $\hat{\Lambda} = \hat{\Delta}_{rz}\hat{\Delta}_{zz}^{-1}$, where $\hat{\Delta}_{zz} = T^{-1}\Sigma_{t=1}^{T} z_{\bullet t} z_{\bullet t}'$ and $\hat{\Delta}_{rz} = T^{-1}\Sigma_{t=1}^{T} r_{\bullet t} z_{\bullet t}'$. Using usual asymptotic theories (White, 1984, Chapters 3 and 4), we can show that under Assumption 1, as $T \to \infty$,

$$\sqrt{T}[vec(\hat{\Lambda}) - vec(\Lambda)] \Rightarrow N\left(0_{(k+1)N\times 1}, p\lim_{T\to\infty}(\hat{\Delta}_{zz}^{-1} \otimes I_N)\hat{\Xi}_1(\hat{\Delta}_{zz}^{-1} \otimes I_N)\right),$$

where "$\Rightarrow$" means "converges in distribution", $vec(\bullet)$ is a matrix operator stacking all the columns in a matrix into a column vector, and $\hat{\Xi}_1$ is a consistent estimator of

$$\Xi = \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\Sigma_{t=1}^{T}(z_{\bullet t} \otimes \varepsilon_{\bullet t})\right). \tag{7}$$

Under Assumption 1, we can consistently estimate Ξ by using one of the nonparametric methods developed in Newey and West (1987), Andrews (1991) and Andrews and Monahan (1993).

Assumptions 2 and 3 can simplify the estimation of Ξ. Under Assumption 2 (no autocorrelation), $(z_{\bullet t} \otimes \varepsilon_{\bullet t})$ is serially uncorrelated. Thus,

$$\Xi = \lim_{T\to\infty} \frac{1}{T} \Sigma_{t=1}^{T} Var(z_{\bullet t} \otimes \varepsilon_{\bullet t}) = \lim_{T\to\infty} \frac{1}{T} \Sigma_{t=1}^{T} E\left(z_{\bullet t} z_{\bullet t}' \otimes \varepsilon_{\bullet t}\varepsilon_{\bullet t}'\right),$$

which can be consistently estimated by $\hat{\Xi}_2 = T^{-1}\Sigma_{t=1}^{T}(z_{\bullet t}z_{\bullet t}' \otimes \hat{\varepsilon}_{\bullet t}\hat{\varepsilon}_{\bullet t}')$, where $\hat{\varepsilon}_{\bullet t} = r_{\bullet t} - \hat{\Lambda} z_{\bullet t}$ is the OLS residual vector at time *t*. Under Assumption 3 (no autocorrelation and no conditional heteroskedasticity), $E(z_{\bullet t}z_{\bullet t}' \otimes \varepsilon_{\bullet t}\varepsilon_{\bullet t}') = E(z_{\bullet t}z_{\bullet t}') \otimes \Sigma_\varepsilon$. Thus, Ξ can be consistently estimated by $\hat{\Xi}_3 = \hat{\Delta}_{zz} \otimes \hat{\Sigma}_\varepsilon$, where $\hat{\Sigma}_\varepsilon = T^{-1}\Sigma_{t=1}^{T}\hat{\varepsilon}_{\bullet t}\hat{\varepsilon}_{\bullet t}'$.

In view of asymptotic theory, the estimator Ξ_1 is always consistent as long as Assumption 1, which is most general among the three assumptions, holds. However, as a practical matter, it may be useful to consider the simpler estimator, $\hat{\Xi}_2$. We can conjecture that $\hat{\Xi}_2$ would be more accurate estimator when Assumption 2 holds. This is so because $\hat{\Xi}_2$ is computed explicitly utilizing the information that the vector $(z_{\bullet t} \otimes \varepsilon_{\bullet t})$ is serially uncorrelated under Assumption 2.

3.2. Asymptotic Distribution of the Two-Pass Estimator

The asymptotic distribution of the TP estimator of λ can be easily derived using the minimum distance approach of Ferguson (1957), Chamberlain (1984), Amemiya (1978) and Newey (1987). Consider the following minimization problem:

$$\min_\lambda T(\hat{\alpha} - \hat{X}\lambda)' A (\hat{\alpha} - \hat{X}\lambda), \tag{8}$$

where *A* is an arbitrary positive definite and asymptotically nonstochastic weighting matrix. The solution of this problem is called a minimum-distance estimator. However, a straightforward algebra shows that the TP estimator $\hat{\lambda}_{TP}$ coincides with the minimum-distance estimator. Thus, using the general properties of minimum-distance estimators, we can derive the distribution of the TP estimator. Specifically, we can show that

$$Var(\hat{\lambda}_{TP}) = T^{-1}(\hat{X}'A\hat{X})^{-1}\hat{X}'A\hat{\Omega}A\hat{X}(\hat{X}'A\hat{X})^{-1}, \tag{9}$$

where $\hat{\Omega}$ is a consistent estimate of $\Omega = \lim_{T\to\infty} Var[\sqrt{T}(\hat{\alpha} - \hat{X}\lambda)]$. The following theorem guides us on how to estimate Ω.

Theorem 1: Define $\hat{\Omega}_1 = (\hat{\lambda}_*'\hat{\Delta}_{zz}^{-1} \otimes I_N)\hat{\Xi}_1(\hat{\Delta}_{zz}^{-1}\hat{\lambda}_* \otimes I_N)$, where $\hat{\lambda}_* = (1, -\hat{\lambda}_{f,TP}')'$. Then, under Assumptions 1 and (4), $\Omega = p\lim_{T\to\infty}\hat{\Omega}_1$. If Assumption 1 is strengthened by Assumption 2, Ω

can be also consistently estimated by $\hat{\Omega}_2$, which replaces $\hat{\Xi}_1$ in $\hat{\Omega}_1$ by $\hat{\Xi}_2$. We can also show that under Assumption 3, $\hat{\Omega}_3 \equiv (\hat{\lambda}_*'\hat{\Delta}_{ZZ}^{-1} \otimes I_N)\hat{\Xi}_3(\hat{\Delta}_{ZZ}^{-1}\hat{\lambda}_* \otimes I_N)$ is a consistent estimator of Ω. In addition, it can be shown that

$$\hat{\Omega}_3 = (\hat{\lambda}_*'\hat{\Delta}_{zz}^{-1}\hat{\lambda}_*)\hat{\Sigma}_\varepsilon = (1+\hat{c})\hat{\Sigma}_\varepsilon, \tag{10}$$

where $\hat{c} = \hat{\gamma}_{f,TP}'\hat{\Sigma}_f^{-1}\hat{\gamma}_{f,TP}$, $\hat{\gamma}_{f,TP} = \hat{\lambda}_{f,TP} + \bar{f}$, and $\hat{\Sigma}_f = T^{-1}\Sigma_{t=1}^T(f_t - \bar{f})(f_t - \bar{f})'$.[5]

As discussed earlier, the TP estimator, $\hat{\gamma}_{f,TP}$, of the vector of risk prices (γ_f) is simply computed by the sum of $\hat{\lambda}_{f,TP}$ and $\bar{f}$. If factors are strictly exogenous as in Assumption 1(ii), $\hat{\lambda}_{f,TP}$ and $\bar{f}$ are uncorrelated. This is so because, as shown in Appendix A, the asymptotic distribution of $\sqrt{T}(\hat{\lambda}_{TP} - \lambda)$ depends on $T^{-1/2}\Sigma_{t=1}^T z_{\bullet t} \otimes \varepsilon_{\bullet t}$. When factors are strictly exogenous, the law of iterative expectation implies that $(z_{\bullet t} \otimes \varepsilon_{\bullet t})$ and $z_{\bullet s}$ are not correlated for all $t \neq s$. Accordingly, under Assumption 1, we have

$$Var(\hat{\gamma}_{TP}) = Var(\hat{\lambda}_{TP}) + J(\hat{\Sigma}_{\bar{f}}/T)J', \tag{11}$$

where $J = (0_{k\times 1}, I_k)'$, and $\hat{\Sigma}_{\bar{f}}$ is a consistent estimator of $\Sigma_{\bar{f}} = \lim_{T\to\infty} Var\left[\sqrt{T}(\bar{f} - E(f_t))\right]$.

The matrix $\hat{\Sigma}_{\bar{f}}/T$ is the estimate of the variance matrix of $\bar{f}$, and the matrix $J\hat{\Sigma}_{\bar{f}}J'$ is equivalent to the "bordered version" of $\hat{\Sigma}_{\bar{f}}$ in Shanken (1992). This result implies that under Assumption 1(ii) (the assumption of strictly exogenous factors), the variance matrix of $\hat{\gamma}_{f,\mathrm{TP}}$ can be estimated simply by the sum of the variance matrices of $\hat{\lambda}_{f,TP}$ and $\bar{f}$.

Under Assumption 3, Shanken (1992) derives the asymptotic variance matrix of the TP estimator $\hat{\gamma}_{\mathrm{TP}}$. In fact, if we replace $\hat{\Xi}_1$ in $\hat{\Omega}$ by $\hat{\Xi}_3$, we immediately obtain Theorem 1 of Shanken (1992).[6] Thus, the results (10) and (11) are a generalization of his result in the case in which the asset returns are heteroskedastic or autocorrelated conditional on the realized factors.

Since market returns or other macroeconomic factors are likely to be autocorrelated in

[5]All of the proofs of this and other theorems are given in Appendix A.
[6]Strictly speaking, (10) is equivalent to Theorem 1 of Shanken (1992) applied to the case in which portfolio factors are absent. For this case, our result (10) coincides with the notation (1+c)Ω in Theorem 1 of Shanken (1992).

practice, the variance matrix $\hat{\Sigma}_{\bar{f}}$ may have to be estimated nonparametrically. Nonetheless, we below show that the test of model specification (2) or (4) requires only the estimation of the lambda component (λ) of the factor price vector. As long as Assumption 1 holds, the potential autocorrelation in the factor vector f_t is irrelevant for model specification tests.

3.3. Optimal Two-Pass Estimator and Specification Test

Because the choice of A is not restricted for (8), there are many possible TP estimators of λ. Amemiya (1978), however, shows that the optimal choice of A is the inverse of $\hat{\Omega}$. With this choice, the TP estimator has the smallest asymptotic variance matrix among the TP estimators with different choices of A. This optimal TP (OTP) estimator is of generalized least squares (GLS) form:

$$\hat{\lambda}_{OTP} = (\hat{\lambda}_{o,OTP}, \hat{\lambda}'_{f,OTP})' = (\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1}\hat{X}'\hat{\Omega}^{-1}\hat{\alpha}, \tag{11}$$

where, under Assumption 1,

$$Var(\hat{\lambda}_{OTP}) = T^{-1}[\hat{X}'\hat{\Omega}_1^{-1}\hat{X}]^{-1}, \tag{12}$$

while under Assumption 2, we have an identical form as (12) but with $\hat{\Omega}_2$ replacing $\hat{\Omega}_1$.

Following Amemiya (1980), we can also show that an optimal TP estimator of γ equals

$$\hat{\gamma}_{OTP} = [\hat{X}'(\hat{\Omega}^*)^{-1}\hat{X}]^{-1}\hat{X}'(\hat{\Omega}^*)^{-1}\bar{r},$$

where $\hat{\Omega}^*$ is a consistent estimator of $\Omega^* = \lim_{T\to\infty} Var[\sqrt{T}(\bar{r} - \hat{X}\gamma)]$. But, by the same reason for (11), $\Omega^* = \Omega + \mathrm{B}\Sigma_{\bar{f}}\mathrm{B}'$, under Assumption 1. Thus, $\hat{\Omega}^* = \hat{\Omega} + \hat{\mathrm{B}}\hat{\Sigma}_{\bar{f}}\hat{\mathrm{B}}'$ is a consistent estimator of Ω^*. Perhaps surprisingly, using $\hat{\Omega}^*$ for $\hat{\gamma}_{OTP}$, we obtain the following result:

Theorem 2: $\hat{\gamma}_{OPT} = \hat{\lambda}_{OPT} + J\bar{f}$.

An interesting result arises if Assumption 3 holds. Substituting $\hat{\Omega}_3$ into (12), we can show that the optimal TP estimator of λ exactly equals $[\hat{X}'\hat{\Sigma}_\varepsilon^{-1}\hat{X}]\hat{X}'\hat{\Sigma}_\varepsilon^{-1}\hat{\alpha}$. Shanken (1992, Theorems 3 and 4) shows that this GLS estimator is asymptotically equivalent to the maximum likelihood estimator under Assumption 3 and the joint normality of asset returns and factors. His result implies that the OTP estimator of λ computed with $\hat{\Omega}_1$ or $\hat{\Omega}_2$ is also asymptotically equivalent to the maximum likelihood estimator under the same assumptions because the estimates $\hat{\Omega}_1$, $\hat{\Omega}_2$ and $\hat{\Omega}_3$ are consistent estimates of Ω. However, it is important to note that when Assumption 3 is violated, the GLS estimator $[\hat{X}'\hat{\Sigma}_\varepsilon^{-1}\hat{X}]\hat{X}'\hat{\Sigma}_\varepsilon^{-1}\hat{\alpha}$ is no longer efficient although it is still consistent. When Assumption 3 is violated, the weighting matrix $\hat{\Omega}_3^{-1}$ (which results in the GLS estimator) is suboptimal. This is so because $\hat{\Omega}_3^{-1}$ is no longer a consistent estimator of Ω. For this case, a more (asymptotically) efficient TP estimator is obtained using $\hat{\Omega}_1$ or $\hat{\Omega}_2$.

Despite its asymptotic efficiency, OTP should be used in practice only with caution. Altonji and Segal (1996) have shown that optimal minimum-distance estimates can be more biased than the non-optimal minimum-distance estimates in finite samples. Their results indicate that the *t*-tests based on the OLS estimator instead of OTP might result in more reliable statistical inferences. Thus, using OTP as a supplement to the OLS estimation with correct variance matrix would be a prudent empirical practice.

When asset returns and factors are not jointly normal, the OTP estimator is not the most (asymptotically) efficient estimator. However, it remains an efficient estimator among those utilizing the first pass OLS estimator $\hat{\Lambda} = [\hat{\alpha}, \hat{\mathrm{B}}]$. The proof of this claim is given in Appendix B.

One advantage of using the OTP estimator is that it provides a convenient specification test statistic for testing the restrictions (4). Using the minimum-distance principle, we can show that under (4) and Assumption 1,

$$Q_{TP} = \frac{T-N+1}{N-1-k}(\hat{\alpha} - \hat{X}\hat{\lambda}_{OTP})'\hat{\Omega}^{-1}(\hat{\alpha} - \hat{X}\hat{\lambda}_{OTP}) \Rightarrow \frac{\chi^2(N-1-k)}{N-1-k}, \tag{13}$$

where $\hat{\Omega} = \hat{\Omega}_1$ under Assumption 1.[7] We refer to this statistic as Q_{TP} statistic. When Assumptions 2 or 3 hold, $\hat{\Omega}$ can be replaced by $\hat{\Omega}_2$ or $\hat{\Omega}_3$, respectively. There are two ways to use the Q_{TP} statistic. First, $(N-1-k)Q_{TP}$ can be compared with $\chi^2(N-1-k)$ to determine rejection or acceptance of a given model specification. Second, the statistic can be compared with $F(N-1-k, T-N+1)$. Asymptotically speaking, these two test strategies are equivalent. But we prefer to use the *F*-test, because it performs better in our simulation exercises. The *F*-test strategy

[7] The formal proof for (13) is available upon request.

requires some theoretical consideration, as we discuss below.

The alternative statistic, Q_{TP}, can be also computed with $\hat{\gamma}_{OTP} = \hat{\lambda}_{OTP} + J\bar{f}$. Specifically, it can be shown that under Assumption 1,

$$Q_{TP}^{*} = \frac{T-N+1}{N-1-k}(\bar{r} - \hat{X}\hat{\gamma}_{OTP})'(\hat{\Omega}^{*})^{-1}(\bar{r} - \hat{X}\hat{\gamma}_{OTP}) \Rightarrow \frac{\chi^2(N-1-k)}{N-1-k}.$$

However, $Q_{TP}^{*} = Q_{TP}$. Stated formally:

Theorem 3:

$$Q_{TP}^{*} = \frac{T-N+1}{N-1-k}(\bar{r} - \hat{X}\hat{\gamma}_{OTP})'\hat{\Omega}^{-1}(\bar{r} - \hat{X}\hat{\gamma}_{OTP}) = Q_{TP}.$$

Theorem 3 reveals the relationship between the Q_{TP} test and the GLS residual test proposed by Shanken (1985). Define $\hat{\gamma}_{GLS} = (\hat{\gamma}_{o,GLS}, \hat{\gamma}'_{f,GLS})' \equiv (\hat{X}'\hat{\Sigma}_{\varepsilon}^{-1}\hat{X})^{-1}\hat{X}'\hat{\Sigma}_{\varepsilon}^{-1}\bar{r}$, which is the OTP estimator under Assumption 3. Then Shanken's GLS residual test statistic has the form $Q_C \equiv [(T-N+1)/(N-1-k)]Q_S$, where

$$Q_S = \frac{(\bar{r} - \hat{X}\hat{\gamma}_{GLS})'\hat{\Sigma}_{\varepsilon}^{-1}(\bar{r} - \hat{X}\hat{\gamma}_{GLS})}{1 + \hat{\gamma}'_{f,GLS}\hat{\Sigma}_{f}^{-1}\hat{\gamma}_{f,GLS}}. \tag{14}$$

Notice that this statistic can be obtained if we insert $\hat{\Omega}_3 = (1+\hat{c})\hat{\Sigma}_{\varepsilon}$ into Q_{TP}^{*} and use $\hat{\gamma}_{f,GLS}$ to compute the EIV correction term $\hat{c}$. Shanken shows that under Assumption 3 and the normality assumption, this statistic is asymptotically $F(N-1-k, T-N+1)$-distributed. An important advantage of using the Q_C test instead of its χ^2-version, TQ_S, is that the Q_C test can control for the potential size distortions in TQ_S that may occur when the number of assets (*N*) is large relative to the number of time series observations (*T*). It is possible that the TQ_S statistic is severely upward biased when *N* is too large (Shanken, 1992). In contrast, the Q_C statistic mitigates the size distortions by using the degrees-of-freedom adjustment term (*T*-*N*+1)/(*N*-1-*k*) instead of *T*. Amsler and Schmidt (1985) have shown through Monte Carlo simulations the Q_C test is well sized in finite samples when Assumption 3 and the normality assumption hold. Our simulation results indicate that their result is extended to the Q_{TP} test: The *F*-test using Q_{TP} is better than the χ^2-test using $T(N-1-k)/(T-N+1)Q_{TP}$ when the errors are not *i.i.d.*

Notice that the Q_{TP} or Q_C tests, differently from the *t*-tests for risk prices, do not require the use of a nonparametric estimate of the variance-matrix of factor means. This property is particularly appealing for the analysis of a model with persistent factors, because the nonparametric estimators generally have poor finite-sample properties when they are applied to persistent data.

4. Simulation Setup

4.1. Focuses of Simulations

So far we have discussed the asymptotic properties of the two-pass (TP) estimators. In this section, we examine the finite-sample properties of the TP estimators through a series of Monte Carlo simulations. We also investigate the finite-sample properties of the TP-based model specification tests (Q_{TP}). We consider both non-optimal and optimal estimators. The non-optimal estimator we consider is the OLS estimator, TP using $A = I_N$ as the weighting matrix (from now on, OLS-TP). We use the notation TP2 and TP1 to refer to the assumptions under which the variance matrix of the OLS-TP estimator is computed. For example, TP1 indicates that the variance matrix is computed under Assumption 1 (conditionally heteroskedastic and/or autocorrelated errors). However, because the TP estimation under Assumption 3 is equivalent to Shanken's method (1985, 1992), we use SH to refer to what would otherwise be TP3. The Newey-West (1987) method is used to control for autocorrelation in the idiosyncratic errors and factors. The required bandwidths for the method are automatically selected by the method of Newey and West (1994). The variance matrix of a TP estimator of risk price vector depends on the variance matrix of factor means. In our simulations, we always estimate the variance matrix of factor means using the Newey-West estimator, whether or not generated factors are autocorrelated. We do so because actual factors are generally autocorrelated.

We also consider the Fama-MacBeth (FM) estimation. This method computes the standard errors of OLS risk price estimates, ignoring autocorrelations in both factors and idiosyncratic errors. For fair comparisons with other TP estimators, we consider alternative versions of the FM method. One, which we call OLS-FM, controls for the possible autocorrelations in factors, but not for autocorrelations in the idiosyncratic errors. The other, which we refer to as OLS-NW, computes the standard errors controlling for both autocorrelations in factors and idiosyncratic errors. Neither OLS-FM nor OLS-NW corrects the EIV problem.[8]

In addition to the OLS-TP estimators, we consider the three optimal TP estimators which are computed with the optimal weighting matrices under Assumptions 1-3, respectively. Since the estimators are of GLS form, we refer to them as GLS-TP1, GLS-TP2 and GLS-SH (GLS-TP3), respectively. The GLS-TP1 estimator uses $\hat{\Omega}_1^{-1}$ as the weighting matrix, while the GLS-TP2 and GLS-SH estimators are computed with $\hat{\Omega}_2^{-1}$ and $\hat{\Omega}_3^{-1}$, respectively. The model specification test statistics based on GLS-TP1, GLS-TP2 and GLS-SH are referred to as Q_{TP1}, Q_{TP2} and Q_C, respectively.

Our simulation exercises focus in the performance of the estimators under different variance-covariance structure of idiosyncratic errors. We investigate how the time-series properties of idiosyncratic error terms influence the empirical distributions of the TP estimators and the specification test statistics.

4.2. Simulation Design

The main objective of our simulation design is to isolate the pure effect of different variance-covariance structure of idiosyncratic errors on the TP estimation. Ahn, Gadarowski and Perez (2009) show that even if a model is has *i.i.d.* errors, the finite-sample properties of the TP estimators crucially depend on the structure of the beta matrix and the persistency level of the factors. They find TP estimation can lead to biased statistical inferences when betas have small cross-section variations, and they are highly correlated. Also TP are significantly biased when the factors are persistent. In order to isolate the pure effects of the different time-series structures of idiosyncratic error terms, we generate data assuming uncorrelated betas with relatively high cross-sectional variation. Also, we generate factors that are not serially correlated.

The foundation of our experiments are the three factor model by Fama and French (1993) hereafter, FF . The FF model is based on the (excess) market return (VW), SBM, and HML factors. For our

---

[8]The variance matrix used by FM is equivalent to $T^{-1}\left[(\hat{X}'\hat{X})^{-1}\hat{X}'\hat{\Sigma}_\varepsilon\hat{X}(\hat{X}'\hat{X})^{-1}+J\hat{\Sigma}_f J'\right]$.
In our simulations, OLS-FM and OLS-NW use $T^{-1}\left[(\hat{X}'\hat{X})^{-1}\hat{X}'\hat{\Sigma}_\varepsilon\hat{X}(\hat{X}'\hat{X})+J\hat{\Sigma}_{\bar{f}}J'\right]$, while OLS-NW uses $T^{-1}(\hat{X}'\hat{X})^{-1}\hat{X}'\hat{\Sigma}^*\hat{X}(\hat{X}'\hat{X})$, where $\hat{\Sigma}^*$ is the Newey-West estimator of the asymptotic variance matrix of $T^{-1/2}\Sigma_{t=1}^T r_{\bullet t}$.

simulation exercises, we generate data mimicking the actual returns and factors as much as we can. Specifically, for the three factor model (FF), we estimate betas and risk prices by OLS using actual data. We compute the means, standard errors and correlations of the estimated betas, and use them to generate the betas used for simulations. We use the estimated betas and risk prices from actual data to calculate the expected asset returns. Simulated return data are obtained by adding generated errors (ε_{it}) to these expected returns. Our simulation results are based on 1,000 trials.[9]

We use actual returns, not excess returns, in our simulation exercises. In unreported simulations, we also examined excess returns, but the results are not materially different from those reported below. The OLS-TP estimates of factor prices from actual data are used as the true factor prices in simulations. Factor prices used are -0.256, 0.142 and 0.520 for the VW, SBM, and HML factors. The FF factors are excess returns. So, when excess, not raw, returns are used to estimate the FF model, the sample means of the FF factors are the efficient estimators of the risk prices (see Shanken, 1992).

We simulate both factors and returns data for each trial. Factors are simulated *i.i.d.* replicating the mean and standard deviations of the true FF factors. Betas are generated only once and are used for 1,000 simulations by:

$$\mathrm{B} = 1_N m_{\mathrm{B}}' + V(s_{\mathrm{B}} C_{\mathrm{B}} s_{\mathrm{B}}')^{1/2},$$

where $m_{\mathrm{B}} = E[(\beta_{i1}, \beta_{i2}, \beta_{i3})']$, V is a $N \times 3$ orthogonal matrix such that $1_N' V = 0_{1\times 3}$, $V'V / N = I_3$, s_{B} is the 3×1 vector of standard errors of factor betas, and C_{B} is the 3×3 correlation matrix of betas. We set C_{B} equal to an 3×3 identity matrix in order to have uncorrelated betas. The same mean and standard deviation values are used for all betas. This warranties that all betas have relatively high cross sectional variation[10]. For the cases with 25 portfolios, the mean and standard deviation values of each beta are fixed at 0.317 and 1, respectively[11]. For the cases with 100 portfolios, the mean and standard deviation values are set at 0.350 and 1.08, respectively.

---

[9]We also tried 5,000 trials for selected simulations and found qualitatively similar results.

[10] Ahn, Gadarowski and Perez (2009) propose a diagnostic measure degree of cross sectional variation of betas *VB* . Their simulations show that when *VB* is larger than 0.50 the TP coefficients can be biased and the *t*-statistic has low power. Our simulation design generates a *VB* = 0.09.

[11] These values are simply the averages of the mean and standard deviations of the three estimated betas from FF model. The use of this values is not critical for our results. Any other values give substantially the same results as long as VB<0.50.

In terms of the idiosyncratic errors, we consider three possible cases. In the first case, the errors are *i.i.d.* over time. For this case, we consider two sub-cases: one case in which the idiosyncratic errors are cross-sectionally independent; the other case in which the errors are cross-sectionally correlated. We consider these two sub-cases to investigate how cross-sectional correlation could influence finite-sample properties of the TP estimators. In both the second and third cases, the idiosyncratic errors are cross-sectionally independent (with different variances for different assets). But the errors are conditionally heteroskedastic over time in the second case, while they are autocorrelated in the third cases. For all cases, TP estimators are computed to be asymptotically robust to cross-sectional correlations.

For the first case (*i.i.d.* errors), we estimate the variance matrix, Π, of the residuals from the first-pass time-series regressions of actual data and compute a lower triangular matrix Γ such that $\Gamma\Gamma' = \Pi$. Then, we generate error vectors $\varepsilon_t = \Gamma v_t$, where the v_t are random vectors drawn from $N(0_{N\times 1}, I_N)$. This procedure generates cross-sectionally correlated errors. For the simulated data with cross-sectionally uncorrelated errors, we replace all off-diagonal entries of Π with zeros.

For the second case (heteroskedastic errors), we use $\varepsilon_{it} = (s_i v_{it}) F_{1t} / \sqrt{T^{-1}\Sigma_{t=1}^{T} v_{it}^2 F_{1t}^2}$, where s_i is the estimated standard error of the errors in asset *i*, and v_{it} is a random number from $N(0,1)$, and F_{1t} is the VW factor. Notice that for this data generating process, the unconditional variance of ε_{it}, $\mathrm{var}(\varepsilon_{it})$, is set at s_i^2. For the third case (autocorrelated errors), we use $\varepsilon_{i1} = s_i v_{i1}$ and ε_{it} = $\rho\varepsilon_{i,t-1} + \sqrt{1-\rho^2}\, s_i v_{it}$, where $t = 2, \ldots, T$ and $\rho = 0.1$.[12] We choose this small value for ρ, because asset returns are only weakly autocorrelated over time. Again, for this process, we use $\mathrm{var}(\varepsilon_{it})$ = s_i^2, where the error terms are drawn to be cross-sectionally independent.

5. Simulation Results

5.1. Preliminary Results from Actual Data

The reliability and appropriateness of the OLS-TP and GLS-TP estimators would depend on

[12] In unreported experiments, we also have considered the cases with $\rho = 0.5$. The results were qualitatively similar results to those reported here.

whether or not error terms ($\varepsilon_{\bullet t}$) are autocorrelated or heteroskedastic over time. Thus, as a preliminary measure, we test the presence of heteroskedasticity and autocorrelation in the error terms in the FF model using the actual data. For the purpose of comparison, we also include test results for the Jagannathan and Wang (1996) factor model (hereafter, JW-96). The factors used for the JW-96 model are debt premium (PRE) and labor return (LAB) and the market return.

Table 1 reports our test results for heteroskedasticity and autocorrelation in the error terms of individual asset returns. To save space, we only report the results from 25 assets. Table 1 reveals that the residuals from the time-series regressions of returns on the FF factors, as well as those from the regressions with the JW-96 factors, are conditionally heteroskedastic. Using White's (1980) test for heteroskedasticity, we found that both the factor models result in a large percentage of assets having heteroskedastic errors. For the FF model, 21 of 25 assets failed this test at both the 5% and 10% levels, while 24 and 25 of 25 did so at these levels, respectively, for the JW-96 model. A smaller portion of assets failed the Breusch-Godfrey Lagrangean-Multiplier (LM) test for autocorrelation for both models, but the number of rejections appears higher for the JW-96 model (11 and 6 at the 10% and 5% levels, respectively) than for the FF model (7 and 3, similarly).

The number of rejections, alone, does not provide a sufficient indication of heteroskedasticity or autocorrelation. Even if no return is conditionally heteroskedastic, we can expect $\alpha \times 100\,\%$ of rejections by the White test when it is performed at the α level. To check whether the frequency of the White (or Breusch-Pagan) tests rejected is significantly different from the number of rejections expected from a chosen α level, we conduct a proportion test based on a normal approximation of the binomial distribution. Suppose that there is no conditional heteroskedasticity in any portfolio return. Then the hypothesis of conditional homoskedasticity in returns could be falsely rejected with the probability equal to α. Thus, the number of rejections would follow a binomial distribution, with the variance of the rejection number equal to $N\alpha(1-\alpha)$. If the number of portfolios, *N*, is large enough, then the binomial distribution can be approximated by normal distribution. Using this information, we can test for statistical significance of the number of rejections. Significance by this test may be indicative of the presence of heteroskedasticity in the given model. Not surprisingly, the numbers of rejections by the White test appear statistically significant for each of the FF and JW-96 models. For both models, the *p*-values for the test of proportions are close to zero regardless of whatever significance level we choose.

While the test of proportions strongly suggests that autocorrelation matters in the JW-96 model, the test result is less strong for the FF model. The rejection number of seven at the 10% significance level is statistically significant (*p*-value equals 0.38%), but the rejection number of three at the 5% significance level is not (*p*-value equals 12.57%). Admittedly, the proportion test we use is asymptotically valid only if the Breusch-Pagan LM test statistics are independently distributed across different assets. This assumption is likely to be violated in practice because the residuals from time-series regressions could be correlated across assets. Nevertheless, it might be fair to say that autocorrelation would be more of a concern for the case in which the JW-96 model is a correctly specified model than for the alternative case in which the FF model is the correct one. This does not mean that both of the two models are correctly specified models. If one is a correctly specified model, the other is likely to be a misspecified model. Determining which model is correctly specified is beyond the scope of this paper. We just report the results of this two models in order to show the relevance of our simulation study.

5.5. Heteroskedasticity and Autocorrelations in Idiosyncratic Errors

We examine the effects on the two pass estimators of the covariance structure of idiosyncratic errors. We consider OLS-NW, OLS-FM, OLS-SH, GLS-SH, GLS-TP1 and GLS-TP2, estimators. Panel A of Table 2 shows the relative biases in the OLS and GLS estimators. There is no strong evidence that the error structure affects the biases in the risk price estimates. The biases are negligible for all cases we consider: cross-correlation, conditional heteroskedasticity and autocorrelation.

Panel B shows the size properties of the *t*-tests. Asymptotically, the *t*-tests based on OLS-SH and GLS-SH would not be correctly sized unless the idiosyncratic errors are *i.i.d.* over time. The tests based on TP2 would not be asymptotically correctly sized unless the errors are not autocorrelated. The tests based on TP1 are the ones that are asymptotically valid even if the errors are heteroskedastic and/or autocorrelated. Despite these asymptotic results, the empirical sizes of the *t*-tests do not appear to be sensitive to what TP estimator is used. Even if the errors are autocorrelated, OLS-TP1 does not necessarily have better finite-sample properties than OLS-SH and OLS-TP2. When the errors are heteroskedastic, there is no strong evidence that OLS-TP2 produces better sized *t*-tests than OLS-SH. Somewhat surprisingly, OLS-SH produces the *t*-tests comparable to the tests from OLS-TP1 and OLS-TP2 even if the errors are not *i.i.d.*. Optimal GLS estimators produce large size distortions in the *t*-tests for intercept, but not for risk prices. For the cases with *i.i.d.* errors, cross-section correlation does not seem to influence the finite-sample sizes of the *t*-tests.

Panel C reports the empirical (size-adjusted) power of the *t*-tests. As expected, the *t*-tests have low power to reject the incorrect hypotheses of zero factor prices, when the true risk prices are small. For example for the case of SMB factor, where the true risk price 0.142, the power of the test goes from 14.6 to 22.6 depending of the estimator used. However for a higher risk price, like the HML factor, with a risk price of 0.52 the power to the test is higher than 92.5.

Except those from OLS-NW, all of the *t*-tests have similar power properties. The *t*-tests from OLS-NW have greater power for some risk prices but lower power for the others. Overall, there is no evidence that OLS-NW produces *t*-tests with lower power.

Panel D reports the empirical sizes of the Q_{TP} tests. Asymptotic theory indicates that the Q_{TP1} test is correctly sized even if the idiosyncratic errors are heteroskedastic and/or autocorrelated. However, Panel D shows that for the cases with autocorrelated errors, the test has finite-sample size distortions at least as large as the Q_C and Q_{TP2} tests do. The Q_C test is reasonably sized even if the errors are conditionally heteroskedastic. In terms of size, the Q_C test is as good as the Q_{TP2} test, which is robust to heteroskedasticity unless the errors are autocorrelated.

Panel E reports the size-adjusted power of the specification tests. The Q_{TP1} test has the lowest power even if the errors are autocorrelated. Overall, the Q_{TP2} test has the greatest power, but only marginally so compared to the Q_C test.

Table 3 shows the empirical sizes and powers of the *t*-tests and the Q_{TP} tests obtained from the simulations of 100 portfolio data. The reported biases in the TP estimators and the empirical size and power properties of the *t*-tests are similar to those reported in Table 2. Panel D shows that the Q_C test is reasonably well sized when the errors are *i.i.d.* over time. However, the test tends to under-reject correct models when the errors are heteroskedastic and to over-reject when the errors are autocorrelated. The Q_{TP2} test, which is robust to heteroskedasticity, tends to over-reject correct models when the errors are *i.i.d.* However, the test is better sized than the Q_C test when the errors are heteroskedastic. When the errors are autocorrelated, the Q_C, Q_{TP2}, and Q_{TP1} all over-reject correct models too often. In particular, the Q_{TP1} test always rejects correct hypotheses no matter how the errors are generated. It appears that the Q_{TP1} test is inappropriate for the analysis of a large number of portfolio returns, whether or not the errors are autocorrelated.

Panel E reports the empirical power of the Q_{TP} tests. Overall, the tests appear to have only limited power to reject BCAPM.

The major findings from Tables 2 and 3 are summarized as follows. First, conditional heteroskedasticity and autocorrelation in the idiosyncratic errors alone do not produce large biases in the TP estimators we consider. Second, the *t*-tests from nonoptimal OLS-TP estimators are in general better sized than those from optimal GLS-TP estimators. Use of the *t*-tests from OLS-SH and OLS-TP2 is generally recommended. Third, the model specification tests are generally unreliable when the idiosyncratic errors are autocorrelated. However, use of the Q_{TP2} test is recommended for the cases in which the errors are heteroskedastic but not autocorrelated, especially when a large number of portfolios are analyzed.

6. Conclusion

In this paper, we consider the asymptotic and finite-sample properties of the two-pass estimators under different variance-covariance structures of idiosyncratic errors. By virtue of the minimum distance principle, we have derived the general asymptotic distributions of the two-pass estimators and optimal GLS estimators when the idiosyncratic errors are not *i.i.d.* over time. As a byproduct, we also have derived alternative model specification tests that are extensions of Shanken's Q_C test.

Our simulation results indicate it is important to check whether the residuals from the time series regressions are autocorrelated or not. In our simulations, only mildly autocorrelated idiosyncratic errors are used (AR(1) coefficient equals 0.1). Nonetheless, all of the specification tests we consider suffer from large size distortions. If significant evidence for autocorrelations is detected from the regressions with actual data, little credence should be given to the model specification tests.

Finally, we find little evidence that optimal GLS estimators have better finite-sample size and power properties than nonoptimal OLS estimators. The simulation results reported in this paper are obtained using data with 330 time series observations. In unreported simulations, we also have used 1,000 time series observations. Even in these simulations, the power gained by using optimal GLS was marginal. However, the optimal GLS estimators are useful to compute model specification tests. Use of the heteroskedasticity-robust GLS estimator and model specification test is recommendable unless the idiosyncratic errors are autocorrelated.

Appendix A

Proof of Theorem 1:

By the definition of $\hat{\lambda}_{TP}$, we can show that

$$\sqrt{T}(\hat{\lambda}_{TP}-\lambda)=\sqrt{T}[(\hat{X}'A\hat{X})^{-1}\hat{X}'A\hat{\alpha}-\lambda]=(\hat{X}'A\hat{X})^{-1}\hat{X}'A\sqrt{T}(\hat{\alpha}-\hat{X}\lambda), \tag{A.1}$$

which implies (9). Under (4), $\alpha-\lambda_o 1_N-\mathrm{B}\lambda_f=0_{N\times 1}$. Using this restriction and standard matrix theories, we can show

$$\begin{aligned}\sqrt{T}(\hat{\alpha}-\hat{X}\lambda)&=\sqrt{T}[(\hat{\alpha}-\lambda_o 1_N-\hat{\mathrm{B}}\lambda_f)-(\alpha-\lambda_o 1_N-\mathrm{B}\lambda_f)]=\sqrt{T}[(\hat{\alpha}-\alpha)-(\hat{\mathrm{B}}-\mathrm{B})\lambda_f]\\&=\sqrt{T}(\hat{\Lambda}-\Lambda)\lambda_*=\sqrt{T}vec[(\hat{\Lambda}-\Lambda)\lambda_*]=(\lambda_*\otimes I_N)vec(\hat{\Lambda}-\Lambda)\\&=(\lambda_*'\Delta_{zz}^{-1}\otimes I_N)\frac{1}{\sqrt{T}}\Sigma_{t=1}^{T}(z_{\bullet t}\otimes\varepsilon_{\bullet t}),\end{aligned} \tag{A.2}$$

where $\lambda_*=(1,-\lambda_f')'$. Then, it follows that

$$\sqrt{T}(\hat{\alpha}-\hat{X}\lambda)\Rightarrow N(0,\Omega), \tag{A.3}$$

where $\Omega=(\lambda_*'\Delta_{ff}^{-1}\otimes I_N)\Xi(\Delta_{ff}^{-1}\lambda_*\otimes I_N)$ and Ξ is defined in (7). Thus, under Assumptions 1, 2 and 3, Ξ can be estimated by $\hat{\Xi}_1$, $\hat{\Xi}_2$ and $\hat{\Xi}_3$, respectively. Finally, the equality (10) results from the fact that

$$\hat{\lambda}_*'\hat{\Delta}_{zz}^{-1}\hat{\lambda}_*=\begin{pmatrix}1 & -\hat{\lambda}_{f,TP}'\end{pmatrix}\begin{pmatrix}1+\bar{F}'\hat{\Sigma}_F^{-1}\bar{F} & -\bar{F}'\hat{\Sigma}_F^{-1}\\ -\hat{\Sigma}_F^{-1}\bar{F} & \hat{\Sigma}_F^{-1}\end{pmatrix}\begin{pmatrix}1\\ -\hat{\lambda}_{f,TP}\end{pmatrix}=1+(\hat{\lambda}_{f,TP}+\bar{f})'\hat{\Sigma}_f^{-1}(\hat{\lambda}_{f,TP}+\bar{f}).$$

Proof of Theorem 2:

Using the two equalities $(\Omega+\mathrm{B}\Sigma_{\bar{f}}\mathrm{B}')^{-1}=\Omega^{-1}-\Omega^{-1}\mathrm{B}(\Sigma_{\bar{f}}^{-1}+\mathrm{B}'\Omega^{-1}\mathrm{B})^{-1}\mathrm{B}'\Omega^{-1}$ and $\bar{r}=\hat{\alpha}+\hat{X}J\bar{f}$, we can show that

$$[\hat{X}'(\hat{\Omega}^*)^{-1}\hat{X}]^{-1}=\begin{pmatrix}0 & 0_{1\times k}\\ 0_{k\times 1} & \hat{\Sigma}_{\bar{f}}\end{pmatrix}+(\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1}.$$

We also have

$$(\hat{\Omega}^*)^{-1}\hat{X}=\hat{\Omega}^{-1}\hat{X}-\hat{\Omega}^{-1}\hat{\mathrm{B}}[\hat{\mathrm{B}}'\hat{\Omega}^{-1}\hat{\mathrm{B}}+\hat{\Sigma}_{\bar{f}}^{-1}]^{-1}\hat{\mathrm{B}}'\hat{\Omega}^{-1}\hat{X};$$

$$[\hat{X}'\hat{\Omega}^{-1}\hat{X}]^{-1}\hat{X}'\hat{\Omega}^{-1}\hat{\mathrm{B}}=J$$

Using these results, we can now show

$$
\begin{aligned}
\hat{\gamma}_{OTP} &= [\hat{X}'(\hat{\Omega}^*)^{-1}\hat{X}]^{-1}\hat{X}'(\hat{\Omega}^*)^{-1}\overline{r} = J\overline{f} + [\hat{X}'(\hat{\Omega}^*)^{-1}\hat{X}]^{-1}\hat{X}'(\hat{\Omega}^*)^{-1}\hat{\alpha} \\
&= J\overline{f} + [\hat{X}'\hat{\Omega}^{-1}\hat{X}]^{-1}\hat{X}'(\hat{\Omega}^*)^{-1}\hat{\alpha} + \begin{pmatrix} 0 & 0_{1\times k} \\ 0_{k\times 1} & \hat{\Sigma}_{\overline{f}} \end{pmatrix}\hat{X}'(\hat{\Omega}^*)^{-1}\hat{\alpha} \\
&= J\overline{f} + [\hat{X}'\hat{\Omega}^{-1}\hat{X}]^{-1}\hat{X}'\hat{\Omega}^{-1}\hat{\alpha} + J\hat{\Sigma}_{\overline{f}}\hat{B}'(\hat{\Omega}^*)^{-1}\hat{\alpha} - J[\hat{B}'\hat{\Omega}^{-1}\hat{B} + \hat{\Sigma}_{\overline{f}}^{-1}]^{-1}\hat{B}'\hat{\Omega}^{-1}\hat{\alpha} = J\overline{f} + \hat{\lambda}_{OTP},
\end{aligned}
$$

because,

$$
\begin{aligned}
&\hat{B}'(\hat{\Omega}^*)^{-1} - \hat{\Sigma}_{\overline{f}}^{-1}[\hat{B}'\hat{\Omega}^{-1}\hat{B} + \hat{\Sigma}_{\overline{f}}^{-1}]^{-1}\hat{B}'\hat{\Omega}^{-1} \\
&\quad = \hat{B}'\hat{\Omega}^{-1} - \hat{B}'\hat{\Omega}^{-1}\hat{B}[\hat{B}'\hat{\Omega}^{-1}\hat{B} + \hat{\Sigma}_{\overline{f}}^{-1}]^{-1}\hat{B}'\hat{\Omega}^{-1} - \hat{\Sigma}_{\overline{f}}^{-1}[\hat{B}'\hat{\Omega}^{-1}\hat{B} + \hat{\Sigma}_{\overline{f}}^{-1}]^{-1}\hat{B}'\hat{\Omega}^{-1} \\
&\quad = (\hat{B}'\hat{\Omega}^{-1}\hat{B} + \hat{\Sigma}_{\overline{f}}^{-1} - \hat{B}'\hat{\Omega}^{-1}\hat{B} - \hat{\Sigma}_{\overline{f}}^{-1})[\hat{B}'\hat{\Omega}^{-1}\hat{B} + \hat{\Sigma}_{\overline{f}}^{-1}]^{-1}\hat{B}'\hat{\Omega}^{-1} = 0.
\end{aligned}
$$

Proof of Theorem 3:

Using the fact that $0 = \hat{X}'\hat{\Omega}^{-1}(\hat{\alpha} - \hat{X}\hat{\lambda}_{OTP}) = \hat{X}'\hat{\Omega}^{-1}(\overline{r} - \hat{X}J\overline{f} - \hat{X}\hat{\lambda}_{OTP}) = \hat{X}'\hat{\Omega}^{-1}(\overline{r} - \hat{X}\hat{\gamma}_{OTP})$, we can show

$$
\begin{aligned}
(\hat{\Omega}^*)^{-1}(\overline{r} - \hat{X}\hat{\gamma}_{OTP}) &= \hat{\Omega}^{-1}(\overline{r} - \hat{X}\hat{\gamma}_{OTP}) - \hat{\Omega}^{-1}\hat{B}(\hat{B}'\hat{\Omega}^{-1}\hat{B} + \Sigma_{\overline{f}}^{-1})^{-1}\hat{B}'\hat{\Omega}^{-1}(\overline{r} - \hat{X}\hat{\gamma}_{OTP}) \\
&= \hat{\Omega}^{-1}(\overline{r} - \hat{X}\hat{\gamma}_{OTP}) = \hat{\Omega}^{-1}(\hat{\alpha} - \hat{X}\hat{\lambda}_{OTP}).
\end{aligned}
$$

Appendix B: Asymptotic Efficiency of OTP

In this appendix, we address the asymptotic efficiency (minimum variance) of OTP among a certain class of estimators. We have shown above that the OTP estimator of λ is asymptotically equivalent to maximum likelihood under Assumption 3 and the normality assumption. This section considers the efficiency of the estimator under weaker assumptions. In particular, we examine the efficiency properties of the OTP estimator $\hat{\lambda}_{OTP}$ among a class of estimators utilizing the first-pass OLS estimator $\hat{\Lambda} = [\hat{\alpha}, \hat{B}]$. This class of estimators is of interest because any TP estimator of the form (5) belongs to the class. We here restrict our discussion only to cases in which Assumption 3 holds to save space.

Defining $b = vec(B)$ and $\Lambda(\lambda, b) = [1_N \lambda_o + B\lambda_f, B]$, we consider the following minimization problem:

$$
\min_{\lambda,b} Q_{MCS}(\lambda, b) \equiv vec[\hat{\Lambda} - \Lambda(\lambda,b)]'[Var(vec(\hat{\Lambda}))]^{-1} vec(\hat{\Lambda} - \Lambda(\lambda,b)), \tag{A.4}
$$

where $Var(vec(\hat{\Lambda})) = T^{-1}(\Delta_{zz}^{-1} \otimes \hat{\Sigma}_{\varepsilon})$. Solutions for this type of problems are called "minimum chi-square" (MCS) estimators (Ferguson, 1958; and Newey, 1987). We use notation $(\hat{\lambda}'_{MCS}, \hat{b}'_{MCS})' = (\hat{\lambda}_{o,MCS}, \hat{\lambda}'_{f,MCS}, \hat{b}'_{MCS})'$ to denote the solution for (A.4). Newey (1987) shows that this MCS estimator is asymptotically efficient among estimators based on the OLS estimator $\hat{\Lambda}$ = $(\hat{\alpha}, \hat{\mathrm{B}})$. Further, by Chamberlain (1982, Proposition 8), under Assumptions 3 and (4),

$$Q_{MCS} \equiv Q_{MCS}(\hat{\lambda}_{MCS}, \hat{b}_{MCS}) \Rightarrow \chi^2(N-1-k).$$

Thus, using this MCS method, researchers can test for the model specification (2) or (4). We also obtain the following result:

Theorem 4: Under Assumptions 3 and (4) ,

$$\sqrt{T}\left(\hat{\lambda}_{MCS} - \lambda\right) \Rightarrow N\left[0_{(1+k)\times 1}, (1+c)(X'\Sigma_\varepsilon^{-1}X)^{-1}\right],$$

where $c = (\lambda_f + \mu_f)'\Sigma_f^{-1}(\lambda_f + \mu_f)$, $\mu_f = E(f_t)$, and $\Sigma_f = Var(f_t)$. The c and X can be estimated by using any consistent estimates of λ_f and B.

Proof of Theorem 4: Note that

$$\frac{\partial \begin{pmatrix} \hat{\alpha} - 1_N \lambda_o - \mathrm{B}\lambda_f \\ \hat{b} - b \end{pmatrix}}{\partial(\lambda_o, \lambda'_f, b')} = -\begin{pmatrix} X & \lambda'_f \otimes I_N \\ 0_{Nk\times(1+k)} & I_{Nk} \end{pmatrix}.$$

Then, Chamberlain (1982, Proposition 7) implies that $(\hat{\lambda}'_{MCS}, \hat{b}'_{MCS})'$ is asymptotically normal and,

$$\lim_{T\to\infty} Var\begin{pmatrix} \sqrt{T}(\hat{\lambda}_{MCS} - \lambda) \\ \sqrt{T}(\hat{b}_{MCS} - b) \end{pmatrix} = \left[\begin{pmatrix} X & \lambda'_f \otimes I_k \\ 0_{k\times(1+k)} & I_{Nk} \end{pmatrix}' (\Delta_{zz} \otimes \Sigma_\varepsilon^{-1}) \begin{pmatrix} X & \lambda'_f \otimes I_k \\ 0_{k\times(1+k)} & I_{Nk} \end{pmatrix}\right]^{-1},$$

where $b = vec(\mathrm{B})$. But, using usual partitioned matrix theories and a little algebra, we can show:

$$\begin{aligned} &\lim_{T\to\infty} Var\left[\sqrt{T}\left(\hat{\lambda}_{MCS} - \lambda\right)\right] \\ &= \left[1 - (\lambda_f + \mu_f)'\left\{(\lambda_f + \mu_f)(\lambda_f + \mu_f)' + \Sigma_f\right\}^{-1}(\lambda_f + \mu_f)\right]^{-1}(X'\Sigma_\varepsilon^{-1}X)^{-1} \\ &= \left[1 + (\lambda_f + \mu_f)'\Sigma_f^{-1}(\lambda_f + \mu_f)\right](X'\Sigma_\varepsilon^{-1}X)^{-1} = (1+c)(X'\Sigma_\varepsilon^{-1}X)^{-1}. \end{aligned}$$

Theorem 4 has an important implication. If we choose $\hat{\lambda}_{f,TP}$ and $\hat{\mathrm{B}}$ to compute

$\hat{\gamma}_{f,TP} = \hat{\lambda}_{f,TP} + \bar{f}$ and $\hat{X}$, we obtain $Var(\hat{\lambda}_{MCS}) = T^{-1}(1+\hat{c})(\hat{X}'\hat{\Sigma}_\varepsilon^{-1}\hat{X})^{-1}$. However, this variance matrix is exactly identical to the variance matrix of the OTP estimator given in (12), if we replace $\hat{\Omega}_1$ by $\hat{\Omega}_3$. This result implies that $\hat{\lambda}_{OTP}$ and $\hat{\lambda}_{MCS}$ have the same asymptotic distribution. Accordingly, we can conclude that $\hat{\lambda}_{OTP}$ (as well as $\hat{\lambda}_{MCS}$) is asymptotically efficient among the estimators utilizing the OLS estimator $\hat{\Lambda}$. That is, there is no estimator which utilizes $\hat{\Lambda}$ and is more (asymptotically) efficient than the OTP estimator.

Although the MCS estimator is not of direct interest, it is useful to clarify the relationship between our OTP and MLE. In spite of the fact that MCS does not require the normality assumption, the MCS estimator can be shown to be MLE derived under the normality assumption. The criterion function $Q_{MCS}(\lambda,b)$ in (A.4) is highly nonlinear in λ and $b = vec(\mathrm{B})$. However, perhaps surprisingly, the solution for the problem (A.4), $(\hat{\lambda}'_{MCS},\hat{b}'_{MCS})' = (\hat{\lambda}_{o,MCS},\hat{\lambda}'_{f,MCS},\hat{b}'_{MCS})'$ has a closed form. Thus, when Assumption 3 holds, $\hat{\lambda}_{MCS}$ could be used as an alternative to $\hat{\lambda}_{OTP}$. Furthermore, the specification test statistic (A.5) can be dramatically simplified. We summarize these results in the following theorem:

Theorem 5: Define $\hat{\lambda}_{*,MCS} = (1,-\hat{\lambda}'_{f,MCS})'$. Then, the followings are true: (i) Let $\hat{\rho}$ be the smallest eigenvalue of the matrix $\hat{\Delta}_{zz}\hat{\Lambda}(\hat{\Sigma}_\varepsilon^{-1} - \hat{\Sigma}_\varepsilon^{-1}1_N(1'_N\hat{\Sigma}_\varepsilon^{-1}1_N)^{-1}1'_N\hat{\Sigma}_\varepsilon^{-1})\hat{\Lambda}$. Then, $\hat{\lambda}_{*,MCS}$ is an eigenvector corresponding to $\hat{\rho}$, which is normalized such that the first element equals one. (ii) $\hat{\lambda}_{o,MCS} = (1'_N\hat{\Sigma}_\varepsilon^{-1}1_N)^{-1}1'_N\hat{\Sigma}_\varepsilon^{-1}\hat{\Lambda}\hat{\lambda}_{*,MCS}$. (iii) $Q_{MCS} = T\hat{\rho}$.

Proof of Theorem 5: Define

$$W(\lambda_f) = \begin{pmatrix} 1 & -\lambda'_f \\ 0_{k\times 1} & I_k \end{pmatrix} \otimes I_N .$$

Note that $W(\lambda_f)$ is nonsingular for any λ_f. Thus, we can have

$$Q_{MCS}(\lambda,b) = [W(\lambda_f)d(\lambda,b)]'[W(\lambda_f)(\hat{\Delta}_{zz}^{-1}\otimes\hat{\Sigma}_\varepsilon)W(\lambda_f)']^{-1}[W(\lambda_f)d(\lambda,b)], \quad \text{(A.6)}$$

where $d(\lambda,b) = [(\hat{\alpha} - X\lambda)',(\hat{b}-b)']'$. It is straightforward to show

$$W(\lambda_f)d(\lambda,b) = \begin{pmatrix} \hat{\alpha} - \hat{X}\lambda \\ \hat{b} - b \end{pmatrix} = \begin{pmatrix} \hat{\Lambda}\lambda_* - 1_N\lambda_o \\ \hat{b} - b \end{pmatrix}. \quad \text{(A.7)}$$

Note that

$$\hat{\Delta}_{zz}^{-1} = \begin{pmatrix} 1 & \bar{f}' \\ \bar{f} & \hat{\Delta}_{ff} \end{pmatrix}^{-1} = \begin{pmatrix} 1+\bar{f}'\hat{\Sigma}_f^{-1}\bar{f} & -\bar{f}'\hat{\Sigma}_f^{-1} \\ -\hat{\Sigma}_f^{-1}\bar{f} & \hat{\Sigma}_f^{-1} \end{pmatrix},$$

where $\hat{\Delta}_{ff} = T^{-1}\Sigma_{t=1}^{T} f_t f_t'$. Using this fact, we can show:

$$\begin{aligned} W(\lambda_f)(\hat{\Delta}_{zz}^{-1}\otimes I_N)W(\lambda_f)' &= \begin{pmatrix} 1 & -\lambda_f' \\ 0_{k\times 1} & I_k \end{pmatrix}\begin{pmatrix} 1+\bar{f}'\hat{\Sigma}_f^{-1}\bar{f} & -\bar{f}'\hat{\Sigma}_f^{-1} \\ -\hat{\Sigma}_f^{-1}\bar{f} & \hat{\Sigma}_f^{-1} \end{pmatrix}\begin{pmatrix} 1 & 0_{1\times k} \\ -\lambda_f & \hat{\Sigma}_f^{-1} \end{pmatrix}\otimes\hat{\Sigma} \\ &= \begin{pmatrix} 1+(\lambda_f+\bar{f})'\hat{\Sigma}_f^{-1}(\lambda_f+\bar{f}) & -(\lambda_f+\bar{f})'\hat{\Sigma}_f^{-1} \\ -\hat{\Sigma}_f^{-1}(\lambda_f+\bar{f}) & \hat{\Sigma}_f^{-1} \end{pmatrix}\otimes\hat{\Sigma}_\varepsilon. \end{aligned} \tag{A.8}$$

Substitute (A.7) and (A.8) into (A.6) and let $\hat{K} = \hat{\Sigma}_f + (\lambda_f+\bar{f})(\lambda_f+\bar{f})'$. Then, tedious but straightforward algebra yields

$$Q_{MCS}(\lambda,b) = Q_M(\lambda) + Ts(\lambda,b)'[\hat{K}\otimes\hat{\Sigma}^{-1}]s(\lambda,b),$$

where $s(\lambda,b) = \hat{b} - b + [(\lambda_*'\hat{\Delta}_{zz}^{-1}\lambda_*)^{-1}\hat{\Sigma}_f^{-1}(\lambda_f+\bar{f})\otimes I_N](\hat{\Lambda}\lambda_* - 1_N\lambda_o)$, and

$$Q_M(\lambda) = T\frac{(\hat{\Lambda}\lambda_* - 1_N\lambda_o)'\hat{\Sigma}_\varepsilon^{-1}(\hat{\Lambda}\lambda_* - 1_N\lambda_o)}{\lambda_*'\hat{\Delta}_{zz}^{-1}\lambda_*}.$$

We now consider the minimization solutions for b and λ_o, given λ_f, which we denote by $\bar{b}$, $\bar{\lambda}_o$, respectively. From the first-order conditions, $\partial Q_{MCS}/\partial b = 0$ and $\partial Q_{MCS}/\partial\lambda_o = 0$, we can easily show

$$\bar{b} = \hat{b} + [(\lambda_*'\hat{\Delta}_{zz}^{-1}\lambda_*)^{-1}\hat{\Sigma}_f^{-1}(\lambda_f+\bar{f})\otimes I_N](\hat{\Lambda}\lambda_* - 1_N\lambda_o); \tag{A.9}$$

$$\bar{\lambda}_o = (1_N'\hat{\Sigma}_\varepsilon^{-1}1_N)^{-1}1_N'\hat{\Sigma}_\varepsilon^{-1}\hat{\Lambda}\lambda_*. \tag{A.10}$$

Substituting (A.9) and (A.10) into $Q_{MCS}(\lambda,b) = Q_{MCS}(\lambda_o,\lambda_f,b)$, we can obtain a concentrated minimand:

$$Q_{CM}(\lambda_f) = Q_{MCS}(\bar{\lambda}_o,\lambda_f,\bar{b}) = T\frac{\lambda_*'\hat{\Lambda}'[\hat{\Sigma}_\varepsilon^{-1} - \hat{\Sigma}_\varepsilon^{-1}1_N(1_N'\hat{\Sigma}_\varepsilon^{-1}1_N)^{-1}1_N'\hat{\Sigma}_\varepsilon^{-1}]\hat{\Lambda}\lambda_*}{\lambda_*'\hat{\Delta}_{zz}^{-1}\lambda_*}. \tag{A.11}$$

Thus, minimizing (A.11) with respect to λ_f results in the MCS estimator of λ_f. However, the Rayleigh-Ritz theorem implies that the eigenvector corresponding to the smallest eigenvalue of the matrix $\hat{\Delta}_{ff}\hat{\Lambda}'[\hat{\Sigma}_\varepsilon^{-1} - \hat{\Sigma}_\varepsilon^{-1}1_N(1_N'\hat{\Sigma}_\varepsilon^{-1}1_N)^{-1}1_N'\hat{\Sigma}_\varepsilon^{-1}]\hat{\Lambda}$ is a solution for the minimization of (A.11). Thus, we have proven the result (i). The result (ii) comes from (A.10). Finally, since $\hat{\lambda}_{*,MCS}$ is an eigenvector of $\hat{\Delta}_{zz}\hat{\Lambda}'[\hat{\Sigma}_{\varepsilon}^{-1} - \hat{\Sigma}_{\varepsilon}^{-1}1_N(1_N'\hat{\Sigma}_{\varepsilon}^{-1}1_N)^{-1}1_N'\hat{\Sigma}_{\varepsilon}^{-1}]\hat{\Lambda}$ corresponding to the eigenvalue $\hat{\rho}$, we have $\hat{\Lambda}'[\hat{\Sigma}_{\varepsilon}^{-1} - \hat{\Sigma}_{\varepsilon}^{-1}1_N(1_N'\hat{\Sigma}_{\varepsilon}^{-1}1_N)^{-1}1_N'\hat{\Sigma}_{\varepsilon}^{-1}]\hat{\Lambda}\hat{\lambda}_{*,MCS} = \hat{\rho}\hat{\Delta}_{zz}^{-1}\hat{\lambda}_{*,MCS}$. Substituting this result into (A.11) yields $Q_{MCS}(\hat{\lambda}_{MCS}, \hat{b}_{MCS}) = Q_{CM}(\hat{\lambda}_{f,MCS}) = T\hat{\rho}$.

A notable result from Theorem 5 is that $\hat{\lambda}_{f,MCS}$ is exactly identical to the closed-form solution of the maximum likelihood estimator derived by Zhou (1998). Since $\hat{\lambda}_{OTP}$ is asymptotically equivalent to $\hat{\lambda}_{MCS}$, it is also asymptotically equivalent to the maximum likelihood estimator. That is, if Assumption 3 holds and the errors $\varepsilon_{\bullet t}$ are normal, $\hat{\lambda}_{MCS}$ is the efficient estimator. However, when Assumption 3 is violated, $\hat{\lambda}_{OTP}$ is strictly more efficient than the MCS or MLE estimator of λ. This is so because, when Assumption 3 is violated, the weighting matrix, $(\hat{\Delta}_{zz} \otimes \hat{\Sigma}_{\varepsilon}^{-1})$, which is used for the MCS estimator, becomes suboptimal. An asymptotically more efficient MCS estimator can be obtained by minimizing (A.4) with the optimal weight, $[(\hat{\Delta}_{zz}^{-1} \otimes I_N)\hat{\Xi}_1(\hat{\Delta}_{zz}^{-1} \otimes I_N)]^{-1}$ (or $[(\hat{\Delta}_{zz}^{-1} \otimes I_N)\hat{\Xi}_2(\hat{\Delta}_{zz}^{-1} \otimes I_N)]^{-1}$). It can be shown that this alternative MCS estimator of λ is asymptotically equivalent to our OTP estimator of λ when Assumption 1 (or 2) holds. Another interesting point of Theorem 4 is (iii). The test statistic $T\hat{\rho}$ is comparable to the likelihood ratio (LR) test statistic $T \times \ln(1+\hat{\rho})$, which is also developed by Zhou (1998). An important difference between these two statistics is that the latter requires the normality assumption while the former does not.

References

Ahn, Seung C., Christopher Gadarowski and M. Fabricio Perez, 2009, Effects of Beta Distribution and Persistent Factors on the Two-Pass Cross-Sectional Regression, Working paper, Arizona State University.

Ahn, Seung C. and Peter Schmidt, 1995, Efficient estimation of a model with dynamic panel data, Journal of Econometrics 68, 5-27.

Altonji, Joseph G. and Lewis M. Segal, 1996, Small-sample bias in GMM estimation of covariance structures, Journal of Business and Economic Statistics 14, 353-366.

Amemiya, Takeshi, 1978, The estimation of a simultaneous equation generalized probit model, Econometrica 46, 1193-1205.

Andersen, T. G. and R. E. Sørensen, 1996, GMM estimation of a stochastic volatility model: A Monte Carlo Study, Journal of Business & Economic Statistics 14, 328-352.

Amsler, Christine E., and Peter Schmidt, 1985, A Monte Carlo investigation of the accuracy of multivariate CAPM tests, Journal of Financial Economics 14, 359-375

Andrews, Donald W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, 817-858.

Andrews, Donald W.K., and J. Christopher Monahan, 1993, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, Econometrica 60, 953-966.

Berk, Jonathan B., 1995, A critique of size-related anomalies, Review of Financial Studies 8, 275-286.

Black, Fisher, 1972, Capital market equilibrium with restricted borrowing, Journal of Business, 45, 444-454.

Black, Fisher, Michael C. Jensen, and Myron S. Scholes, 1972, The Capital asset pricing model: Some empirical tests, in Michael C. Jensen, ed.: Studies in the Theory of Capital Markets (Praeger, New York, New York).

Breusch, T., 1978, Testing for autocorrelation in dynamic linear models, Australian Economic Papers 17, 1978, 334-355.

Campbell, John Y., Andrew W. Lo and A. Craig MacKinlay, 1997, The Econometrics of Financial Markets (Princeton University Press, Princeton, New Jersey).

Chamberlain, Gary, 1982, Multivariate regression models for panel data, Journal of Econometrics 18, 5-46.

Chamberlain, Gary, 1984, Panel data, in Zvi Griliches, and Michael D. Intriligator, eds.: Handbook of Econometrics, Volume 2 (North-Holland, New York, New York).

Chen, Robert, and Raymond Kan, 2004, Finite sample analysis of two-pass cross-sectional regressions, mimeo, University of Toronto.

Chochrane, John H., 2005, Asset Pricing, Revised Edition (Princeton University Press, Princeton, New Jersey).

Copeland, Thomas E., and J. Fred Weston, 1992, Financial theory and corporate policy (Addison-Wesley Publishing Company, Reading, Massachusetts).

Dickey, D.A., and W.A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74, 427B431.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, Journal of Financial Economics 33, 3-56.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, Return, and Equilibrium: empirical tests, Journal of Political Economy 71, 607-636.

Ferguson, T.S., 1953, A method of generating best asymptotically normal estimates with an application to the estimation of bacterial densities, Annals of Mathematical Statistics 29, 1046-1062.

Ferson, Wayne E., and Campbell R. Harvey, 1999, Conditioning variables and the cross-section of stock returns, Journal of Finance (forthcoming).

Gibbons, Michael R., 1982, Multivariate tests of financial models: a new approach, Journal of Financial Economics 10, 3-27.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, Econometrica 57, 1121-1152.

Godfrey, L., 1978, Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables, Econometrica 46, 1293-1302.

Grauser, Robert R., and Johannus A. Janmaat, 2005, On the power of cross-sectional and multivariate tests of the CAPM, mimeo, Simon Fraser University.

Greene, William H., 2003, Econometric Analysis, 5th eds., (Prentice Hall, Upper Saddle River, New Jersey).

Hamilton, James D., 1994, Time Series Analysis, (Princeton University Press, Princeton, New Jersey).

Hansen, Lars P., 1982, Large sample properties of generalized method of moments estimators, Econometrica 50, 1029-1054.

Hansen, L.P., J. Heaton and A. Yaron, 1996, Finite-sample properties of some alternative GMM estimators, Journal of Business & Economic Statistics, 14, 262-280.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the Cross-Section of Expected Return, Journal of Finance 51, 3 - 53.

Jagannathan, Ravi, and Zhenyu Wang,1998a, An asymptotic theory for estimating beta-pricing models using cross-sectional regression, Journal of Finance 53, 1285-1309.

Jagannathan, Ravi, and Zhenyu Wang,1998b, A note on the asymptotic covariance in Fama-MacBeth regressions, Journal of Finance 53, 799-801.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, Journal of Finance 45, 881-898.

Kan, Raymond, and Chu Zhang, 1999, Two-pass tests of asset pricing models with useless factors, Journal of Finance 54, 204-235.

Kandel, Samuel, 1984, The likelihood ratio test of mean-variance efficiency without a riskless asset, Journal of Financial Economics 13, 575-592

Kandel, Samuel, and Robert F. Stambaugh, 1995, Portfolio efficiency and the cross section of expected returns, Journal of Finance 50, 185-224.

Kim, Dongcheol, 1995, The errors in the variables problem in the cross-section of expected returns, Journal of Finance 50, 1605-1634.

Lintner, John, 1965a, Security prices, risk and maximal gains from diversification, Journal of Finance 20, 587-615.

Lintner, John, 1965b, The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, Review of Economics and Statistics 47, 13-47.

MacKinlay, A. Craig, and Matthew P. Richardson, 1991, Using generalized methods of moments to test mean-variance efficiency, Journal of Finance 46, 511-527.

MacKinnon, J.G., 1991, Critical Values for Cointegration Tests, Chapter 13 in Long-run Economic Relationships: Readings in Cointegration, E.d. R.F. Engle and C.W.J. Granger, Oxford University Press.

Mossin, Jan, 1966, Equilibrium in a capital asset market, Econometrica 35, 768-783.

Newey, Whitney K., 1987, Efficient estimation of limited dependent variable models with endogenous explanatory variables, Journal of Econometrics 36, 231-250.

Newey, Whitney K. and Kenneth D. West, 1987, Hypothesis testing with efficient method of moments estimation, International Economic Review, 28, 777-787.

Newey, Whitney K. and Kenneth D. West, 1994, Automatic lag selection in covariance matrix estimation, Review of Economic Studies, 61, 631-654.

Phillips, P.C.B., and P. Perron, 1988, Testing for a unit root in time series regression, Biometrika 75, 335-346.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing model, Journal of Economic Theory 13, 341-360.

Schmidt, Peter, 1976, Econometrics, Marcel Dekker, Inc., New York, New York.

Shanken, Jay, 1985, Multivariate tests of the zero-beta CAPM, Journal of Financial Economics 14, 327-348.

Shanken, Jay, 1986, Testing portfolio efficiency when the zero-beta rate is unknown, Journal of Finance 41, 269-276.

Shanken, Jay, 1992, On the estimation of beta-pricing models, Review of Financial Studies 5, 1-34.

Sharpe, William, F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, Journal of Finance 19, 425-442.

White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity, Econometrica 48, 817B838.

White, Halbert, 1984, Asymptotic Theory for Econometricians (Academic Press, INC, San Diego, California).

Zhou, Guofu, 1994, Analytical GMM tests: Asset pricing with time-varying risk premiums, Review of Financial Studies 7, 687-709.

Zhou, Guofu, 1998, On cross-sectional stock returns: Maximum likelihood approach, mimeo, Washington University.

Table 1
Diagnostic Test Results from Time-Series Regressions of Fama-French and Jagannathan-Wang Models

Reported below are diagnostic test results for conditional heteroskedasticity and autocorrelation on idiosyncratic errors in time-series regressions of asset returns for 25 size/pre-beta portfolios on the Fama-French (1993) and the Jagannathan-Wang (1996) factors. For each portfolio, The information matrix test statistic of White (1980) are used to test for conditional heteroskedasticity and the Breusch-Godfrey LM statistic (Breusch, 1978; Godfrey, 1978) using up to six lags is used to test for autocorrelation. For each model and test, the number of rejections at various test sizes is indicated, along with the 1-tailed p-value based on the mid-point Normal approximation to a binomial distribution that the true count is greater than or equal to the number of rejections shown given the test size.

Portfolio Quintiles by Size & Pre-beta		Fama-French (1993)				Jagannathan-Wang (1996)			
		White Test for Conditional Heteroskedasticity		Breusch-Godfrey LM-test for Autocorrelation		White Test for Conditional Heteroskedasticity		Breusch-Godfrey LM-test for Autocorrelation	
		Obs*R^2	p-value	Obs*R^2	p-value	Obs*R^2	p-value	Obs*R^2	p-value
1	1	72.99	0.00%	9.78	13.41%	33.22	0.01%	9.15	16.52%
1	2	40.38	0.00%	4.60	59.55%	54.02	0.00%	11.03	8.75%
1	3	39.39	0.00%	11.67	6.97%	49.05	0.00%	17.36	0.81%
1	4	33.33	0.01%	6.52	36.71%	60.36	0.00%	15.79	1.49%
1	5	43.40	0.00%	7.38	28.71%	49.22	0.00%	11.36	7.79%
2	1	13.50	14.11%	8.64	19.50%	24.56	0.35%	9.80	13.31%
2	2	39.12	0.00%	7.28	29.60%	97.78	0.00%	15.54	1.64%
2	3	10.76	29.22%	6.65	35.40%	27.42	0.12%	12.90	4.47%
2	4	31.34	0.03%	20.01	0.28%	64.87	0.00%	10.55	10.32%
2	5	41.91	0.00%	18.46	0.52%	34.43	0.01%	21.43	0.15%
3	1	73.94	0.00%	5.11	52.99%	45.58	0.00%	7.93	24.29%
3	2	5.29	80.85%	1.93	92.60%	41.37	0.00%	5.90	43.47%
3	3	41.22	0.00%	6.88	33.24%	64.33	0.00%	10.07	12.18%
3	4	17.72	3.85%	10.28	11.32%	43.04	0.00%	*7.94	24.28%
3	5	30.64	0.03%	11.42	7.63%	31.37	0.03%	10.66	9.95%
4	1	23.75	0.47%	12.57	5.05%	38.37	0.00%	10.20	11.65%
4	2	30.00	0.04%	7.98	23.98%	54.88	0.00%	11.08	8.60%
4	3	49.07	0.00%	4.38	62.53%	32.53	0.02%	10.97	8.93%
4	4	24.90	0.31%	10.85	9.32%	21.83	0.94%	7.87	24.75%
4	5	30.45	0.04%	9.14	16.60%	35.94	0.00%	9.09	16.85%
5	1	102.21	0.00%	3.50	74.38%	45.83	0.00%	0.84	99.09%
5	2	70.18	0.00%	6.17	40.48%	67.33	0.00%	3.54	73.80%
5	3	13.95	12.40%	13.57	3.49%	20.39	1.56%	15.76	1.51%
5	4	37.85	0.00%	5.28	50.90%	24.94	0.30%	3.62	72.81%
5	5	22.69	0.69%	6.24	39.72%	16.62	5.49%	4.29	63.72%

Number of rejections at the indicated test size across assets

Test Size	#	p-value	#	p-value	#	p-value	#	p-value
10%	21	0.00%	7	0.38%	25	0.00%	11	0.00%
5%	21	0.00%	3	12.57%	24	0.00%	6	0.00%

Table 2
Simulations with Heteroskedastic or Autocorrelated Idiosyncratic Errors: 25 Portfolios

1,000 simulations are used. Data are generated using the risk price and error-variance estimates from the estimation of the Fama-French model with actual data. Betas are generated are generated only once and they are orthogonal to each other. Factors are *i.i.d.* over time. The errors of the first two types are *i.i.d.* over time: those of the first type are cross-sectionally independent, and those of the second type are cross-sectionally correlated. The errors of the third and fourth types are conditional heteroskedastic, and autocorrelated over time, respectively, while they are cross-sectionally independent. Panel A reports the relative biases of estimated risk prices by OLS, GLS-SH, GLS-TP2 and GLS-TP1. Panels B and C show the empirical sizes and power (%) of the *t*-tests based on eight TP estimators, respectively. Panels D and E reports the empirical size and power (%) of the three TP model specification tests. The empirical size and power of a *t*-test are the rejection rate for the correct hypothesis of risk price being equal to its true value and the incorrect hypothesis of risk price being equal to zero, respectively. The empirical size and power of the Q_C test is the rejection rates for the model used for simulations and Black's CAPM respectively. All hypotheses are tested at 5% significance level.

Panel A: Relative Bias

	I.I.D. Errors															
	No Cross-correlated Errors				Cross-Correlated Errors				Conditional Het. Errors				Autocorrelated Errors			
	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML
OLS	0.2	1.3	-1.9	-0.2	0.2	1.5	-1.5	-0.1	0.2	1.5	-1.2	-0.2	0.2	1.4	-2.0	-0.2
GLS-SH	0.2	1.8	-1.9	-0.1	0.1	1.7	-1.1	-0.1	0.1	1.7	-1.4	-0.1	0.2	2.0	-2.0	-0.1
GLS-TP2	0.2	1.9	-1.9	-0.1	0.2	1.7	-1.1	-0.1	0.1	1.7	-1.3	0.0	0.2	2.0	-1.9	-0.1
GLS-TP1	0.2	2.3	-1.7	-0.2	0.0	2.3	-0.8	-0.2	0.1	1.9	-1.2	-0.1	0.2	2.4	-1.8	-0.2

Panel B: Rejection rate (size) of the *t*-test for H_o: price = true value

	I.I.D. Errors															
	No Cross-Correlated Errors				Cross-Correlated Errors				Conditionally Het. Errors				Autocorrelated Errors			
	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML
OLS																
OLS-NW	5.0	3.4	5.2	2.4	5.2	3.2	4.8	2.3	4.8	3.6	4.5	2.5	6.1	3.4	5.2	2.5
OLS-FM	6.0	4.7	5.5	6.7	6.0	4.7	5.1	6.0	5.5	4.8	4.8	6.1	8.3	4.3	5.1	5.6
OLS-SH	5.7	4.4	5.3	5.9	5.0	5.4	5.8	6.8	5.0	5.3	6.3	6.7	7.9	4.7	5.4	6.8
OLS-TP2	5.5	4.8	5.5	6.7	5.1	5.4	5.9	6.8	4.7	5.3	6.3	6.6	7.4	4.7	5.4	6.8
OLS-TP1	6.5	5.0	5.6	7.1	6.0	5.0	5.5	6.6	4.7	5.5	6.0	6.8	6.7	4.9	5.4	6.9
GLS																
GLS-SH	7.8	4.5	5.1	6.4	7.0	4.7	5.3	7.0	6.5	4.7	6.3	6.8	11.3	4.5	5.6	6.9
GLS-TP2	8.3	4.6	5.6	7.0	7.7	4.7	5.3	7.2	6.5	4.6	6.2	6.9	12.1	4.6	5.8	6.9
GLS-TP1	12.8	4.8	5.9	7.0	14.5	4.6	5.6	6.7	16.6	5.4	6.0	7.4	14.3	4.8	6.0	7.0

Table 2 continues.

Panel C: Rejection rate (power) of the t-test for H_o: price = 0

	I.I.D. Errors															
	No Cross-Correlated Errors				Cross-Correlated Errors				Conditionally Het. Errors				Autocorrelated Errors			
	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML
OLS																
OLS-NW	100.0	11.9	22.6	99.0	100.0	11.7	22.7	99.0	100.0	12.7	22.3	98.9	100.0	11.8	22.5	99.0
OLS-FM	100.0	20.0	18.3	94.7	100.0	20.8	17.4	93.9	100.0	19.5	16.4	94.5	100.0	20.1	18.1	94.6
OLS-SH	100.0	19.3	18.2	93.3	100.0	18.7	16.6	91.8	100.0	18.3	15.9	92.7	100.0	18.2	17.5	92.1
OLS-TP2	100.0	19.3	18.2	93.2	100.0	18.5	16.6	91.8	100.0	18.3	15.8	92.8	100.0	18.2	17.5	92.1
OLS-TP1	100.0	18.9	17.9	93.3	100.0	18.3	17.3	92.3	100.0	17.9	16.3	92.5	100.0	18.2	17.7	92.3
GLS																
GLS-SH	100.0	20.8	17.4	93.1	100.0	20.2	18.6	92.1	100.0	19.0	15.7	91.3	100.0	19.4	16.2	91.9
GLS-TP2	100.0	20.2	16.4	93.4	100.0	20.1	18.7	92.3	100.0	19.1	17.3	91.3	100.0	19.5	15.9	91.9
GLS-TP1	100.0	20.1	14.6	92.5	100.0	19.6	17.1	92.1	100.0	18.2	16.7	90.4	100.0	19.9	15.1	91.6

Panel D: Rejection rate (size) of two-pass specification test

	I.I.D. Errors			
	No Cross-Correlated Errors	Cross-Correlated Errors	Conditionally Het. Errors	Autocorrelated Errors
Q_C	4.5	5.2	4.3	16.5
Q_{TP2}	5.1	5.9	3.8	19.3
Q_{TP1}	9	39.7	23	16.1

Panel E: Rejection rate (power) of two-pass specification test

	I.I.D. Errors			
	No Cross-Correlated Errors	Cross-Correlated Errors	Conditionally Het. Errors	Autocorrelated Errors
Q_C	43.0	45.5	45.8	31.2
Q_{TP2}	44.6	47.5	48.7	33.7
Q_{TP1}	25.9	26.8	41.1	16

Table 3
Simulations with Heteroskedastic or Autocorrelated Idiosyncratic Errors: 100 Portfolios

1,000 simulations are used. The data generating process used is the same as the process used for Table 2, data is generated for 100 portfolios. The errors of the first two types are *i.i.d.* over time: those of the first type are cross-sectionally independent, and those of the second type are cross-sectionally correlated. The errors of the third and fourth types are conditional heteroskedastic, and autocorrelated over time, respectively, while they are cross-sectionally independent. Panel A reports the relative biases of estimated risk prices by OLS, GLS-SH, GLS-TP2 and GLS-TP1. Panels B and C show the empirical sizes and power (%) of the *t*-tests based on eight TP estimators, respectively. Panels D and E reports the empirical size and power (%) of the three TP model specification tests. The empirical sizes and power of the *t* and Q_C tests are computed by the same ways that are used for Table 2.

Panel A: Relative Bias

	I.I.D. Errors															
	No Cross-Correlated Errors				Cross-Correlated Errors				Conditionally Het. Errors				Autocorrelated Errors			
	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML
OLS	0.0	0.4	0.2	0.2	0.1	0.4	0.2	0.2	0.0	0.2	0.3	0.2	0.0	0.4	0.2	0.2
GLS-SH	0.0	0.4	0.4	0.1	0.1	0.4	0.5	0.2	0.1	0.6	0.3	0.2	0.0	0.4	0.4	0.1
GLS-TP2	0.0	0.4	0.4	0.1	0.1	0.4	0.4	0.2	0.1	0.7	0.4	0.2	0.0	0.3	0.4	0.1
GLS-TP1	0.0	0.3	0.2	0.2	0.1	0.5	0.3	0.2	0.0	0.6	0.4	0.4	0.1	0.3	0.2	0.2

Panel B: Rejection rate (size) of the t-test for H_o: price = true value

	I.I.D. Errors															
	No Cross-Correlated Errors				Cross-Correlated Errors				Conditionally Het. Errors				Autocorrelated Errors			
	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML
OLS																
OLS-NW	4.0	3.7	3.8	3.3	4.1	3.8	4.3	3.3	5.2	3.8	4.1	3.1	6.5	3.9	3.9	3.3
OLS-FM	4.8	6.2	5.3	6.7	5.1	5.7	4.1	5.6	6.2	5.8	4.0	5.7	7.8	5.7	4.3	5.8
OLS-SH	4.3	6.0	4.8	6.4	4.9	6.2	4.7	6.6	5.5	6.5	4.8	6.4	7.5	6.1	4.9	6.4
OLS-TP2	4.1	6.0	4.9	6.4	4.6	6.3	4.8	6.7	5.4	6.6	4.8	6.5	7.4	6.1	4.9	6.5
OLS-TP1	5.2	6.1	4.8	6.4	4.9	6.3	4.8	6.3	6.8	6.6	5.4	6.6	5.7	6.0	4.8	6.3
GLS																
GLS-SH	15.7	6.2	5.2	6.4	15.4	6.2	5.0	6.5	15.2	6.5	5.1	6.7	21.0	6.1	5.3	6.4
GLS-TP2	17.2	6.0	5.3	6.4	16.7	6.3	5.0	6.7	16.8	6.5	5.1	6.5	23.3	5.9	5.3	6.3
GLS-TP1	50.4	6.2	4.8	6.6	64.5	6.6	5.5	6.6	58.1	6.7	5.1	6.4	53.9	6.3	4.8	6.5

Table 3 continues.

Panel C: Rejection rate (power) of the t-test for H_o: price = 0

	I.I.D. Errors															
	No Cross-Correlated Errors				Cross-Correlated Errors				Conditionally Het. Errors				Autocorrelated Errors			
	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML	Cst.	VW	SMB	HML
OLS																
OLS-NW	100.0	32.1	33.1	75.4	100.0	31.3	33.8	75.2	100.0	30.3	33.7	74.6	100.0	31.5	33.1	75.3
OLS-FM	100.0	39.5	28.3	59.9	100.0	40.1	28.8	58.6	100.0	39.0	26.8	59.4	100.0	39.4	28.1	60.8
OLS-SH	100.0	39.4	24.5	57.3	100.0	38.8	24.6	57.1	100.0	38.2	23.9	54.3	100.0	38.6	24.5	57.0
OLS-TP2	100.0	39.4	24.4	57.3	100.0	38.8	24.7	57.1	100.0	38.2	23.9	54.3	100.0	38.6	24.4	57.0
OLS-TP1	100.0	39.8	23.9	55.9	100.0	39.3	25.4	57.5	100.0	37.6	24.5	54.2	100.0	38.8	24.0	55.6
GLS																
GLS-SH	100.0	37.4	24.1	56.1	100.0	37.7	24.4	55.4	100.0	38.4	25.4	54.8	100.0	37.3	24.3	56.0
GLS-TP2	100.0	36.8	23.5	56.6	100.0	37.9	24.5	55.0	100.0	38.0	24.9	55.8	100.0	37.5	23.0	56.7
GLS-TP1	100.0	38.5	25.9	56.9	100.0	39.6	24.9	54.9	100.0	38.3	24.9	53.9	100.0	38.9	25.2	57.7

Panel D: Rejection rate (size) of two-pass specification test

	I.I.D. Errors			
	No Cross-Correlated Errors	Cross-Correlated Errors	Conditionally Het. Errors	Autocorrelated Errors
Q_C	4.8	5	1.9	32.4
Q_{TP2}	8.3	7.6	3.1	41.3
Q_{TP1}	100	100	100	100

Panel E: Rejection rate (power) of two-pass specification test

	I.I.D. Errors			
	No Cross-Correlated Errors	Cross-Correlated Errors	Conditionally Het. Errors	Autocorrelated Errors
Q_C	10.9	10.7	15.6	8.8
Q_{TP2}	10.2	10.9	17.7	9.0
Q_{TP1}	4.1	0.8	59.8	4.2