Two-Pass Cross-Sectional Regression of Factor Pricing Models

Seung C. Ahn
Arizona State University

Christopher Gadarowski
Rowan University

This Version: December 2003

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 and factors are conditionally heteroskedastic and/or autocorrelated. Using minimum distance (MD) method, we derive heteroskedasticity-and/or-autocorrelation-robust model specification test statistics and asymptotic variances of the TP risk premium estimates. We also derive optimal MD estimators that are asymptotically more efficient than other two-pass estimators. Several findings are obtained from our simulation exercises. First, the model specification tests and t-tests based on Shanken (1985, 1992) and the heteroskedasticity-robust MD method perform reasonably well in finite samples, unless asset returns are autocorrelated conditionally on factors and the number of assets analyzed is not too large. The model specification tests and t-tests based on optimal MD method tend to over-reject correct hypotheses, especially when autocorrelation is present in data and/or too many assets are analyzed. Second, the t-tests based on non-optimal two-pass estimation often perform better than those based on optimal MD estimation. Third, the finite-sample properties of TP or MD estimation are sensitive to what factors generate returns. Our results suggest that the TP or MD methods may be inappropriate for the analysis of the models with highly persistent factors.

Acknowledgment

The title of earlier versions of this paper was “Two-Pass Cross-Sectional Regression of Factor Pricing Models Model: Minimum Distance Approach.” 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, Mike Lemmon and Byung-Sam Yoo. We also thank Zhenyu Wang for comments on the earlier version of the paper, and Guofu Zhou for generously providing us detailed notes on a topic related with this paper. All remaining errors are of course our own. The first version of this paper was completed while the first author was visiting the Korea Economic Research Institute and University College London. Corresponding author: Christopher Gadarowski, College of Business, Rowan University, Glassboro, NJ, 08028; email: gadarowski@rowan.edu.
Two-Pass Cross-Sectional Regression of Factor Pricing Models

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 and factors are conditionally heteroskedastic and/or autocorrelated. Using minimum distance (MD) method, we derive heteroskedasticity-and/or-autocorrelation-robust model specification test statistics and asymptotic variances of the TP risk premium estimates. We also derive optimal MD estimators that are asymptotically more efficient than other two-pass estimators. Several findings are obtained from our simulation exercises. First, the model specification tests and t-tests based on Shanken (1985, 1992) and the heteroskedasticity-robust MD method perform reasonably well in finite samples, unless asset returns are autocorrelated conditionally on factors and the number of assets analyzed is not too large. The model specification tests and t-tests based on optimal MD method tend to over-reject correct hypotheses, especially when autocorrelation is present in data and/or too many assets are analyzed. Second, the t-tests based on non-optimal two-pass estimation often perform better than those based on optimal MD estimation. Third, the finite-sample properties of TP or MD estimation are sensitive to what factors generate returns. Our results suggest that the TP or MD methods may be inappropriate for the analysis of the models with highly persistent factors.
1. Introduction

The two-pass (TP) cross-sectional regression method, first used by Black, Jensen and Scholes (1972) and Fama and MacBeth (1973), has been widely used to evaluate linear factor pricing models, including the capital asset pricing model (CAPM), arbitrage pricing theory (APT) and their variants.\(^1\) TP estimation, where the asset betas are first estimated by time-series linear regression of the asset’s return on a set of common factors, then, factor risk prices are estimated by cross-sectional regressions of mean returns on betas, is both simple and provides several convenient ways to test a given asset pricing model. These include evaluating the significance of an asset (firm)-specific regressor in the second-stage ordinary least squares (OLS) regression of returns on factor betas and the regressor (Fama and MacBeth (1973)), and the specification of factor model based on the residuals from a GLS two-pass regression (Shanken (1985)).

Despite its simplicity, the Fama-MacBeth 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 for multivariate OLS estimation following Fama-MacBeth, while Shanken (1985) provides it for GLS estimation.\(^2\) Unfortunately, Shanken’s EIV-corrected standard errors are consistent only under the restrictive assumptions of no conditional heteroskedasticity and no conditional autocorrelation in asset returns (given factors).\(^3\) Because these assumptions are often disputed in empirical studies, Shanken’s EIV adjustments could produce potentially biased statistical inferences. Recently, Jagannathan and Wang (1998a) provide a general form for the correct asymptotic variance matrix of the two-pass estimator, allowing for both conditional heteroskedasticity as well as autocorrelation in asset returns. They, however, do not detail the

---

\(^1\)Sharpe (1964), Lintner (1965a,b) and Mossin (1966) pioneered CAPM while Ross (1976) developed the original theory behind APT. See Copeland and Weston (1992) and Campbell, Lo and MacKinlay (1997) for a summary of the major models and research in this area since then.


\(^3\)Kim (1995) and Ferson and Harvey (1999) also consider cases of conditional heteroskedasticity, but only of particular structures. See equation (18) of Kim’s paper and the appendix in Ferson and Harvey.
estimation procedure for the variance matrix, nor provide empirical evidence for the importance of controlling conditional heteroskedasticity or/and autocorrelation in the two-pass regression. In this paper, we present a detailed TP estimation procedure that is EIV-corrected, robust to autocorrelation and conditional heteroskedasticity, and general to both OLS and GLS estimation. Through some limited Monte Carlo experiments, we also examine the finite sample properties of TP regression when asset returns are conditionally heteroskedastic or autocorrelated.

To derive the EIV-correction method robust to conditional heteroskedasticity and/or autocorrelation, we reexamine the asymptotic properties of two-pass estimators and generalize the estimation and model test methods developed by Shanken (1985, 1992). A novelty of this paper is that we use the method of minimum distance (MD) which has been developed by Ferguson (1958), Amemiya (1977), Chamberlain (1982, 1984) and Newey (1987). Use of this method makes three contributions to the literature of linear factor pricing models. First, the method provides a systematic method to derive EIV-corrected standard errors of the traditional OLS or GLS two-pass estimators, under both general and special distributional assumptions on asset returns. We also show that the MD approach is general enough to subsume the methods proposed by previous studies.4

Second, we derive an optimal MD estimator in the sense that it is asymptotically efficient (minimum-variance) among a class of two-pass regression estimators. In fact, this estimator is a generalization of the GLS two-pass estimator by Shanken (1992). He shows that his GLS estimator is asymptotically equivalent to MLE, if the asset returns and factors are Gaussian, serially uncorrelated, and homoskedastic conditional on realized factors. Under the same conditions, we show that the optimal MD (OMD) estimator reduces to the GLS estimator. However, if there exists conditional heteroskedasticity or autocorrelation, the OMD estimator can be easily formulated to be more (asymptotically) efficient than the GLS estimator.5

Third, using the optimal MD estimator, we construct a simple F-statistic for testing a given factor pricing model, which has properties similar to the generalized method of moments test

4We note that we are not trying to claim that we develop a new estimation technique that can control heteroskedasticity and autocorrelation in the estimation of linear asset pricing models. The generalized method of moments can be also used to do so.

5Use of more efficient estimation is desirable in practice because the power of a test statistic usually increases with the efficiency of the estimator used to compute the statistic.
(Hansen, 1982). This statistic can be also viewed as a heteroskedasticity-and/or-autocorrelation-robust version of Shanken’s (1985) GLS residual test. This is so because, despite its robustness, the statistic is asymptotically equivalent to the GLS residual test under the conditions justifying GLS.

Our Monte Carlo experiments are based on the three factor model of Fama and French (1993), and the Premium-Labor (PL) model of Jagannathan and Wang (1996). The main results are as follows. First, although the model specification test and t-tests proposed by Shanken (1985, 1992) are designed for models with no heteroskedasticity and autocorrelation, they perform reasonably well, if returns are not conditionally autocorrelated, and if the number of assets analyzed is not too large. Second, the model specification tests and t-tests based on optimal MD method generally over-reject correct hypotheses, especially when autocorrelation presents in data and/or too many assets are analyzed. The t-tests based on non-optimal two-pass estimation often perform better than those based on optimal MD estimation. Third, our simulation results indicate that the TP or MD methods are inappropriate for the analysis of the PL-model. The specification tests and t-tests based on the TP or MD are severely biased toward rejecting correct hypotheses when returns are generated with the factors of the PL-model. This unexpected result is mainly due to the fact that one factor of the model is, if not nonstationary, near nonstationary.

The remainder of this paper is organized as follows. In section 2, we discuss the basic asset pricing model of our interest and assumptions. In section 3, we present the minimum-distance (MD) approach to estimate and test for the model. Section 4 examines our simulation results obtained based on the Fama-French and the PL-models. Finally, section 5 summarizes our findings and suggests a practical guidance on the TP or MD estimation of linear factor models.

2. Basic Model and Assumptions

In this section, we introduce the basic asset pricing model of our interest and assumptions. As with most work in this area, we assume returns are linearly related to some common factors. Specifically, we consider the following population projection of asset returns on k common factors:

\[ R_t = \alpha_t + \beta_t F_t + \epsilon_t, \quad \epsilon_t = \sigma_t Z_t, \quad (1) \]
where \( R_t = [R_{1t}, \ldots, R_{Nt}] \) and \( \epsilon_{t} = [\epsilon_{1t}, \ldots, \epsilon_{Nt}] \) are matrices with \( N \) rows and \( T \) columns, \( R_{it} \) is the gross return of asset \( i \) at time \( t \), \( \epsilon_{it} \) is the idiosyncratic error for asset \( i \) at time \( t \) with zero mean. \( F_t = [F_{1t}, \ldots, F_{kt}] \) is the vector of \( k \) factors at time \( t \), \( \beta_{it} \) is the vector of \( k \) betas of asset \( i \) corresponding to \( F_{it} \), and \( \alpha_{i} \) is the asset-specific intercept term.

Viewing (1) as a projection model, we assume \( E(Z_t \bar{\epsilon}_t) = 0 \) for \( t = 1, \ldots, T \), where \( Z_t \bar{\epsilon}_t \) is the Kronecker product of the two matrices obtained by multiplying each entry of \( Z_t \) by \( \epsilon_t \). We assume that \( T \) is large and \( N \) is relatively small, so that asymptotics apply as \( T \) approaches infinity. That is, this paper considers only the \( \sqrt{T} \)-consistency of two-pass estimators.

Since model (1) is a linear projection model, not a data generating process of asset returns, we can allow the \( \epsilon_t \) to be autocorrelated. 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 projection model (1) as a linear approximation of such models, the error term \( \epsilon_t \) is likely to be autocorrelated.

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

\[
H_0: \quad E(R_t) = (\alpha_0, \beta_0)\epsilon_{N},
\]

where \( e_N \) is the \( N \times 1 \) vector of ones, \( \alpha_0 \) is a unknown constant (e.g., zero-beta return), \( \beta_0 \) is the \( k \times 1 \) vector of factor risk prices. However, tests of asset pricing models using asset-specific regressors have arisen with mounting evidence inconsistent with the basic factor-structure (2). The two-pass regression approach often uses a generalized model by which the hypothesis \( H_0 \) can be tested. Specifically, many previous studies consider the following auxiliary model

\[
E(R_t) = (\alpha_0, \beta_0)\epsilon_{N} + S \cdot \beta_0 + X, \quad (3)
\]

where \( S \) is a \( N \times q \) matrix of asset-specific variables, \( \beta_0 \) is a \( q \times 1 \) vector of unknown parameters, \( X = [e_N, \%S] \), and \( \%S = \frac{1}{N} \sum_{i=1}^{N} S \cdot \beta_0 \). The restriction \( (\beta_0 = 0) \) on (3) implies \( H_0 \) in (2). Thus, the test

---

6If a risk-free asset yielding return \( R_{ft} \) is available, \( R_{ft} \) may denote excess return \( (R_{ft} - R_{ft}) \).

7For the conditions for \( \sqrt{N} \)-consistency of two-pass estimators, see Shanken (1992).

8See Copeland and Weston (1992) for a summary discussion of earlier works and Campbell, Lo and MacKinlay (1997) for more recent studies.
of $H_0$, against (3) can be conducted based on a two-pass regression method applied to (3). The traditional two-pass (TP) approach estimates the vector $\hat{\beta}$ by regressing $\tilde{R} \tilde{\Phi} \tilde{\chi} \tilde{\Phi} \tilde{\chi} \tilde{\Phi} \tilde{\chi} \tilde{\phi} \tilde{A} \tilde{R}$ on $\hat{X}' (e_N, \hat{\beta}, S)$ with an arbitrary positive-definite (and asymptotically nonstochastic) weighting matrix, where $\hat{\beta}$ is the ordinary least squares (OLS) estimate of $\beta$.

There are many possible choices for $A$. If we choose $A = I_n$, then the two-pass estimator $\hat{\beta}_{TP}$ becomes an OLS estimator. In contrast, with the choice of $A = [\text{Var}(\hat{\beta})]^{-1}$, it becomes a GLS estimator (Shanken, 1992; and Kandel and Stambaugh, 1995).

A problem of the TP estimator (4) is that it uses the estimated beta, $\hat{\beta}$, because the true beta, $\beta$, is not observed. It 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 $\epsilon_i$ are independently and identically distributed (i.i.d.) over time, he provides the correct asymptotic variance matrix of the TP estimator explicitly incorporating estimation errors generated by the use of the estimated beta. A more general variance matrix can be found in Jagannathan and Wang (1998a).

Once the TP estimator (4) is computed, the asset-pricing restriction (2) can be examined by testing the restriction $\hat{\beta}_2 = 0$ by a Wald test statistic, $\hat{\chi}^T [\text{Var}(\hat{\beta}_2)]^{-1} \hat{\beta}_2$. This statistic is $F^2$-distributed with degrees of freedom equal to $q$. Alternately, one can use individual $t$-statistics corresponding to each of the elements in $\hat{\beta}_2$. Jagannathan and Wang (1996) use this approach to test their Premium-Labor model. In particular, they test their model against the residual size effects suggested by Berk (1995). In their study, the matrix $S$ includes only the logarithm of the firm’s market value. Alternately, the matrix $S$ could include asset-specific variables which capture the so-called anomalies effects, such as those attributed to proxy variables for past winners and losers (Jegadeesh and Titman, 1993).

An alternative method often used in the literature to avoid the EIV problem existing in the TP estimation is maximum likelihood estimation. Work in this area includes Gibbons (1982), Kandel (1984), Shanken (1986), Gibbons, Ross and Shanken (1989), and Zhou (1998). This

---

9For any number $g$, we hereafter use $I_g$ to denote an $g \times g$ identity matrix.

10$\text{Var}(\epsilon_i)$ means the unconditional variance matrix of the $\epsilon_i$. 

---
method assumes 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,

\[
H_0'' : " ^ { ' \beta_0 e_N % % ^ \beta_1 , } \tag{5}
\]

where " is the vector of individual intercept terms in the first-pass model (1), and \( \beta_i \) is a unknown \( k \times 1 \) vector. In fact, this hypothesis is equivalent to \( H_o \) in (2). To see this, note that model (1) implies \( \text{E}(R_i) = " ^ { + \% \text{E}(F_i) } \). Let \( \beta_0 = (\_0 \) and \( \beta_1 + \text{E}(F_i) = (\_1 \). Then, (5) implies \( \text{E}(R_i) = \beta_0 e_N + \% ^ \beta_1 + \% \text{E}(F_i) = (\_0 e_N + \% ^ \_1 \) (Campbell, Lo and MacKinlay, 1997, p. 227). Note that given the specification (5), the vector of risk prices, \( (\_1 \), is decomposed into the population mean of the factor vector, \( \text{E}(F_i) \), and the lambda component, \( \beta_i = (\_1 - \text{E}(F_i) \). This lambda component can be interpreted as the vector of factor-mean adjusted risk prices (Zhou, 1998).

The MLE approach estimates \( \% \beta_0 \) and \( \beta_1 \) jointly, and test the hypothesis \( H_0'' \) by a standard likelihood ratio (LR) test. Then, the vector of risk prices, \( (\_1 \), is estimated by the sum of the estimated \( \beta_1 \) and the sample mean of the factor vector, \( \text{E}(F_i) \), and the lambda component, \( \beta_i = (\_1 - \text{E}(F_i) \). This MLE procedure is efficient under the assumption that the returns and factors are jointly normal.

Although the MLE approach focuses on the LR test for the hypothesis \( H_0'' \), we can think of an alternative test procedure. As we have extended (2) to (3), we can extend the restriction (5) into the model

\[
" ^ { ' \beta_0 e_N % ^ \beta_1 % ^ \beta_2 / X \beta , } \tag{6}
\]

where \( \theta = [\theta_0, \theta_1, \theta_2] \) \( N \) \( \theta_0 = (\_0, \beta_1 = (\_1 - \text{E}(F_i) \), and \( \theta_2 = (\_2 \). Thus, we can test the null hypothesis \( H_0'' \) by testing the restriction \( \theta_2 = 0 \).\(^{11}\) A way to estimate \( \theta = [\theta_0, \theta_1, \theta_2, \theta_3, \theta_4] \) which has not been considered in the literature, is to apply the two-pass method to (6) with the OLS estimator " replacing \( R \) in (4). That is, we estimate \( \beta \) by regressing the OLS estimate of " (" ) on \( \hat{X} \):

\[
\hat{\beta}_{TP} / \left[ \hat{\beta}_{0,TP}, \hat{\beta}_{1,TP}, \hat{\beta}_{2,TP} \right] \ \ (\hat{X}^\dagger A \hat{X}^\dagger \hat{X}) \ A'' . \tag{7}
\]

\(^{11}\)We note that this restriction is not a sufficient, but a necessary condition for \( H_0'' \).
To see the relationship of $\hat{\zeta}_{TP}$ and $\hat{\Sigma}_{TP}$, we substitute the equality $\widetilde{R} = \bar{A}^\prime \bar{\Sigma} \bar{F}^\prime$ (Campbell, Lo and MacKinlay, 1997, p. 223) into (7). Then, we have

$$
\begin{bmatrix}
\zeta_{0,TP} \\
\zeta_{1,TP} \\
\zeta_{2,TP}
\end{bmatrix} = (\bar{X}^\prime A \bar{X})\hat{\Sigma}_{TP} \hat{\Sigma}_{TP} \hat{\Sigma}_{TP} \hat{\Sigma}_{TP}
$$

for $J = [0, I, 0]$ and any choice of the weighting matrix $A$. This result implies that the vector of risk prices, $\zeta$, always can be estimated by the sum of the mean factor vector, $\bar{F}$, and the two-pass estimator $\hat{\Sigma}_{1,TP}$.

In order to derive the asymptotic variance matrix for the TP estimator $\hat{\Sigma}_{TP}$, we need to make some general assumptions on the time-series model (1). Specifically, the following set of conditions are sufficient to obtain the main results of this paper.

**Assumption 1**

(I) The data $R_t$ and $F_t$ are covariance stationary, ergodic, and have finite moments up to fourth order. (ii) $E(Z_t U_{t+1}) = 0$ for all $t$: That is, the errors are uncorrelated with the contemporaneous factors. (iii) $\beta = [\beta, \gamma]$ is of full column: That is, all the columns in $\beta$ are linearly independent.

Several comments on Assumption 1 are worth noting. First, Assumption 1 is general enough to subsume most of the assumptions frequently adopted in the literature. Under both Assumptions 1(I) and (ii),

$\beta = E[(R_t \bar{\Sigma} E(R_t))(F_t \bar{\Sigma} E(F_t))]\bar{\Sigma}$,

where $\text{Var}(F_t) = E[(F_t - E(F_t))(F_t - E(F_t))]\bar{\Sigma}$ is the variance matrix of $F_t$. Thus, the parameter matrix $\beta$ reserves the usual beta interpretation. Assumption 1(ii) also guarantees the consistency of OLS estimation of $\beta$. Note that Assumption 1(ii) is much weaker than the assumption of stochastic independence between the factor $F_t$ and the error $\epsilon_t$. Furthermore, Assumption 1(ii) does not rule out conditional heteroskedasticity in the errors, $\epsilon_t$, and allows the factors, $F_t$,

---

12Here and throughout our discussion, $E(\bullet)$ means expectation defined over time.

13More conditions are required for the consistency and asymptotic normality of the estimators discussed below. For such additional conditions, see Hansen (1982).
returns, $R$, and errors to be serially correlated.

Finally, Assumption 1(iii) implies that $\text{Var}(F)$ is nonsingular\(^{14}\) and all of the columns in $E[(R_t - E(R_t))(R_t - E(R_t))]N$ are linearly independent. This assumption rules out the case in which some columns of $E[(R_t - E(R_t))(F_t - E(F_t))]N$ equal zero vectors. That is, under Assumption 1(iii), returns, $R$, and each factor in $F_t$ should be contemporaneously correlated. It also implies that there is no factor in $F_t$ that is ‘useless’ in the sense of Kan and Zhang (1997); that is, all of the factors in $F_t$ can explain $R_t$. Assumption 1(iii), which is also adopted by Jagannathan and Wang (1998a), is essential for the identification of parameters in the restricted models which we discuss below.

The OLS estimator of $\theta = [\theta^\prime, \theta^\prime']$ is given by

$$\hat{\theta} \leftarrow (\theta^\prime, \theta^\prime')' \hat{\theta}_{\theta_{\theta\theta}}^{RZ} \theta_{\theta\theta}$$

(9)

where $\hat{\theta}_{\theta_{\theta\theta}}^{RZ} = T^{\theta_{\theta_{\theta\theta}}} \theta_{\theta_{\theta\theta}}^{T} \theta_{\theta_{\theta\theta}}^{RZ}$ and $\theta_{\theta_{\theta\theta}}^{RZ} = T^{\theta_{\theta_{\theta\theta}}} \theta_{\theta_{\theta\theta}}^{T} \theta_{\theta_{\theta\theta}}^{RZ}$. We assume $\lim_{t \to T} \theta_{\theta_{\theta\theta}}^{RZ} = \theta_{\theta_{\theta\theta}}$ and $\lim_{t \to T} \theta_{\theta_{\theta\theta}}^{RZ} = \theta_{\theta_{\theta\theta}}$. The asymptotic distribution of the OLS estimator $\hat{\theta}$ plays an important role in finding the correct asymptotic distribution of two-pass estimators. Thus, we here briefly review the distribution of the OLS estimator under several different sets of assumptions.

Usual asymptotic theories (White, 1984, Chapters 3 and 4) imply that under Assumption 1, the OLS estimator $\hat{\theta}$ is consistent and asymptotically normal: That is, as $T \to T_0$,

$$\sqrt{T}(\text{vec}(\hat{\theta}) \theta_{\theta\theta}(\theta)) \to N(0, \theta_{\theta\theta}(\theta))$$

(10)

where “$\to$” means “converges in distribution”, vec$(\cdot)$ is a matrix operator stacking all the columns in a matrix into a column vector, and

$$= \lim_{t \to T_0} \text{Var}\left( \frac{1}{\sqrt{T}} \theta_{\theta_{\theta\theta}}^{T} \theta_{\theta_{\theta\theta}}^{\theta_{\theta_{\theta\theta}}}(\theta) \right)$$

(11)

Note that with Assumption 1(ii), we allow autocorrelation in the errors, $e_t$. Under these relatively general conditions, we can consistently estimate $\theta$ by using a nonparametric method developed by Newey and West (1987), Andrews (1991) or Andrews and Monahan (1993).

\(^{14}\) That is, there is no redundant factor in $F_t$. 

8
Hereafter, we denote this nonparametric estimate of \( \gamma \) by \( \hat{\gamma}_1 \).\(^{15}\) Alternately, Assumption 2 as follows simplifies estimation of \( \gamma \) while retaining the possibility of conditional heteroskedasticity and serially correlated factors.

**Assumption 2** (I) In addition to Assumption 1, \( E(\epsilon, F_1, \ldots, F_T) = 0 \) for all \( t \): The factors are strictly exogenous with respect to the errors \( \epsilon \). (ii) \( E(\epsilon, \epsilon', N^F_1, \ldots, N^F_T) = 0 \) for all \( t \neq s \): The errors are serially uncorrelated given the factors.

Assumption 2(I), which we call the assumption of strictly exogenous factors, has been implicitly adopted by many empirical studies of unconditional capital asset pricing models, which treat \( \epsilon \) as the modeling error of \( R_t \) and \( \epsilon + \% F_t \) as a conditional mean of \( R_t \) given the entire history of the factors, \( F_t \). Note that Assumption 2(I) is still weaker than the assumption of stochastic independence between the errors and factors. For example, Assumption 2(I) does allow conditional heteroskedasticity in the errors, i.e. \( \text{Var}(\epsilon, F_t) \neq \text{Var}(\epsilon, F_s) \), for \( s \neq t \). Assumption 2(ii) does not allow the errors to be serially uncorrelated. However, it allows the factors, \( F_t \), to be serially correlated. Thus, under Assumptions 2, the return vector, \( R_t \), can be (unconditionally) serially correlated because it is a function of the factors.

Estimation of \( \gamma \) can be quite simplified under Assumption 2. Using the law of iterative expectation, we can show that under Assumptions 2, \( E[(Z_t, \hat{U}_t)(Z_s, \hat{U}_s)] = 0 \), for any \( t \neq s \). Thus, the vector \( (Z_t, \hat{U}_t) \) is serially uncorrelated. Furthermore, Assumption 1(I) implies that \( \text{Var}(Z_t, \hat{U}_t) = \text{Var}(Z_s, \hat{U}_s) \) for any \( t \) and \( s \). Using these results, we can show that

\[
\begin{align*}
\gamma &= (Z_t, \hat{U}_t) = E(Z_t, \hat{U}_t) = \text{Var}(Z_t, \hat{U}_t) = \text{Var}(Z_s, \hat{U}_s) \text{ for any } t \text{ and } s.
\end{align*}
\]

Accordingly, the variance matrix \( \gamma \) can be consistently estimated by

\[
\hat{\gamma} = \left( T^{\frac{1}{2}} \right) \left( (Z_t, \hat{U}_t) \right)^T, \tag{12}
\]

where the \( \hat{\epsilon}_t \) are OLS residuals, i.e., \( \hat{\epsilon}_t \equiv R_t \& \hat{Z}_t \).\(^{16}\)

\(^{15}\)This matrix estimate is a weighted sum of autocovariance matrices of \( (Z_t, \hat{U}_t) \). In practice, OLS residuals, say \( \hat{\epsilon}_t \), can be used to compute this matrix.

\(^{16}\)In fact, the consistency of \( \hat{\gamma} \) only requires that the \( Z_t, \hat{U}_t \) be serially uncorrelated, which is much weaker than Assumption 2.
In terms of asymptotics, there really is no need to distinguish between $\hat{\Xi}_1$ and $\hat{\Xi}_2$, because both of them are consistent estimators of $\Xi$. However, as a practical matter, it may be useful to consider the simpler estimate $\hat{\Xi}_2$. We can conjecture that it would have better finite sample properties if Assumption 2 holds. This is so because $\hat{\Xi}_2$ is computed explicitly utilizing the information that the vector $(Z_t, \tilde{U}_t)$ is serially uncorrelated under Assumption 2.

Estimation of $\Xi$ can be further simplified under the following assumption:

**Assumption 3** In addition to Assumption 2, $\text{Var}(\hat{\gamma}, \hat{\gamma}^*, F, ..., F_t) = E$, for any $t$, where $E$ is the unconditional variance matrix of $\gamma$. Var($\gamma$, $\gamma^*$).

Because Assumption 3 implies that the variance matrix of $\gamma$, does not depend on time or realized factors, we call it the assumption of no conditional heteroskedasticity and autocorrelation. This assumption has been adopted by Shanken (1992), and Jagannathan and Wang (1996).\(^{17}\) Note that if returns and factors are jointly normal and i.i.d. over time, returns are warranted to be homoskedastic conditional on factors. Under Assumption 3, an alternative simple consistent estimator of $\Xi$ is

$$\hat{\Xi}_3 = \tilde{U}_t \bar{Z}_t \hat{E}, \tag{13}$$

where $\hat{E} = \begin{bmatrix} \hat{E}_{1\tilde{t}} & \hat{E}_{1\hat{t}} \\ \hat{E}_{\hat{t}1} & \hat{E}_{\hat{t}\hat{t}} \end{bmatrix}$. Also, under Assumption 3, we obtain

$$\sqrt{T}(\text{vec}(\hat{\gamma}) & \text{vec}(\hat{\gamma})) \sim N(0, \text{Var}(\hat{\gamma})) \tag{14}$$

as $T \to 4$. Assumption 3 may be quite restrictive for real data analysis. For example, when returns and factors are jointly t-distributed, returns should be conditionally heteroskedastic (MacKinlay and Richardson, 1991).\(^{18}\) Nonetheless, Assumption 3 has been frequently assumed in empirical studies. Thus, this assumption can be used to compare our estimation procedures with other methods.

\(^{17}\)Jagannathan and Wang (1998b) provide a correction to the asymptotic results of Jagannathan and Wang (1996).

\(^{18}\)Note that Assumption 2 allows returns and factors to be jointly t-distributed.
3. Minimum Distance Approach

This section introduces a minimum distance (MD) approach to estimation and tests of the restrictions (3) or (6). Using this approach, we derive the asymptotic distributions of the two-pass estimator under general assumptions, identify the asymptotically most efficient two-pass estimator, and obtain a simple specification test statistic.

3.1. Basic Results

Following Chamberlain (1984), Amemiya (1978) and Newey (1987), we can obtain a MD estimator of the lambda ($\lambda$) vector in (6) by solving the following minimization problem:

$$ \min_{\lambda} J(\lambda' \hat{X}) A(\lambda' \hat{X}), \quad (15) $$

where $A$ is an arbitrary positive definite and asymptotically nonstochastic weighting matrix. However, a straightforward algebra shows that the two-pass estimator $\hat{\lambda}_{tp}$ coincides with the solution of the problem (15). Thus, $\hat{\lambda}_{tp}$ is a MD estimator. In addition, using the MD principle, we can easily show that

$$ \text{Var}(\hat{\lambda}_{tp}) = T(\hat{X}^\top A \hat{X}) \text{det}(A) \text{det}(\hat{X}^\top A \hat{X}), \quad (16) $$

where $\hat{S}$ is a consistent estimate of $S/lim_{T \to \infty} \text{Var}[\sqrt{T}(\lambda' \hat{X})]$. The following theorem guides us on how to estimate $S$.

**Theorem 1** Under Assumption 1 and (6), $S = (\lambda)' = \lambda_{N} \lambda_{N}'$, where $\lambda_{N} = [1, I, \lambda_{N}]'$. Thus, $S$ can be consistently estimated by

$$ \hat{S} \sim (\hat{\lambda}_I)' = (\hat{\lambda}_I)' \hat{\lambda} \hat{I}_N, \quad (17) $$

where $\hat{\lambda}_I \sim (1, \lambda_{N})'$, and $\hat{\lambda}_I$ is any consistent estimator of the factor-mean adjusted risk price vector, and $\hat{\lambda}_I$ can be obtained by using a nonparametric estimation method. If Assumption 1 is strengthened by Assumption 2, $S$ can alternately be consistently estimated by

$$ \hat{S}_2 \sim (\hat{\lambda}_I)' \hat{\lambda} \hat{I}_N, \quad (18) $$

If Assumption 3 also holds, $S$ can be consistently estimated by

$$ \hat{S}_3 \sim (\hat{\lambda}_I)' \hat{\lambda} \hat{I}_N, \quad (19) $$
where $\hat{c}' \left( \hat{\epsilon}'_F \hat{\epsilon}_F + T \hat{\epsilon}'_T \right)$, $\hat{F}' \% \hat{\epsilon}_1$, and $\hat{E}_F' \left( \hat{T}'_i (F_i \& \hat{F}) (F_i \& \hat{F}) \right)$.

All proofs are given in Appendix A. Theorem 1 suggests some tractable estimation procedures when we allow some structure in the errors as provided in Assumptions 2 or 3. The estimated variance matrix $\hat{S}_1$ is consistent for $S$ under quite general assumptions. Even if we impose stronger assumptions about the error structure, such as Assumptions 2 or 3, $\hat{S}_1$ remains consistent, although, under Assumptions 2 or 3, $\hat{S}_2$ or $\hat{S}_3$ would be better estimates in finite samples. Note that using $\hat{S}_2$ for $\text{Var}(\hat{S}_{TP})$, we can still control for potential conditional heteroskedasticity in the error, $\epsilon_i$. Further, Assumption 2 allows autocorrelation in returns (if the factor vector $F_i$ is autocorrelated). However, as long as the errors $\epsilon_i$, $i = 1, \ldots, T$, are serially uncorrelated, the autocorrelation in returns does not affect the asymptotic distribution of $\hat{S}_{TP} = [\hat{S}_{0,TP}, \hat{S}_{1,TP}, \hat{S}_{2,TP}]$.

Theorem 1 is useful to compare our results with those in Shanken (1992). He shows that under Assumption 3, the traditional TP estimator $(\hat{\epsilon}'_T \hat{\epsilon}_T)$ [$(\hat{\epsilon}'_0, \hat{\epsilon}'_1, \hat{\epsilon}'_2)$] given in (4) has the asymptotic variance matrix of the form (16) with $\hat{S}$ replaced by $\hat{S}_3$. Shanken interprets the $\hat{E}$ component in (19) as the error component of the TP estimator caused by the residual errors, $\epsilon_i$; and the component $\hat{c}$ as an adjustment for the EIV problem caused by the use of estimated beta in the two-pass regression.

In order to see whether factors are priced or not, researchers need to estimate the vector of risk prices, $\hat{\beta}_i$. The traditional two-pass estimator, $(\hat{\beta}'_{1,TP})$, of $\beta_i$ can be simply computed by the sum of $\hat{S}_{1,TP}$ and $\hat{F}$. Unfortunately, however, as Jagannathan and Wang (1998a) have shown, the asymptotic variance matrix of $\hat{\beta}'_{1,TP}$ (and the variance matrix of $\hat{\beta}'_{TP}$) is somewhat complicated under Assumption 1. We here state essentially the same asymptotic result as Theorem 1 of Jagannathan and Wang, but in a different representation:

**Theorem 2** Define $\hat{c}' \left( [(Z_i, \hat{U}_i, \beta_i)' \cdot (F_i \& \hat{E}(F_i))]' \right)$ and

$$ Q \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{c} \right) $$

Under Assumption 1 and (6),

$$ \sqrt{T} (\hat{S}_{TP} \& \hat{S}) \mathcal{Y} N(0, \text{Var}(\hat{S}_{TP} \& \hat{S})) $$
where $\mathbf{\hat{\varphi}}_{TP}$, $\mathbf{\hat{\varphi}}_{TP} \% J \mathbf{F}$, $J = (0_{k \times 1}, I_1, 0_{k \times q}) N \mathbf{M}$, $[(X^2)^{\mathcal{S}} A (\mathcal{S})]^{\mathcal{S}} A (\mathcal{S}) z_2 \mathbf{M}^\prime$, and $\mathbf{Q}$ is a consistent estimator of $\mathbf{Q}$. That is, $\mathbf{T} \& \hat{M} Q \hat{M}$ is a consistent estimator of $\text{Var}(\mathbf{\hat{\varphi}}_{TP})$.

Although this theorem may be merely a rehearse of Jagannathan and Wang (1998a), it provides some additional insights into the traditional two-pass estimation. First, estimation of $\mathbf{Q}$ requires the nonparametric methods of Newey and West (1987), or Andrews (1991). The reason for this complexity is that Assumption 1 does not rule out the possibility that the model errors, $\iota$, and the factors, $F_t$, are autocorrelated. Then, $\mathbf{Q}$ becomes the sum of all of autocovariance matrices of the time series $\mathcal{S}$.

Some stronger assumptions can simplify the estimation of $\text{Var}(\mathbf{\hat{\varphi}}_{TP})$. Many studies of asset pricing models assume that factors are i.i.d. over time and stochastically independent of the errors, $\iota$. This assumption can simplify the structure of $\mathbf{Q}$ considerably. However, in fact, Assumption 2(I), the assumption of strictly exogenous factors, is sufficient to obtain similar results. Under Assumption 2(I), we can easily show that $\mathbf{T}^{\mathcal{S} \& \hat{F}} T_{\mathcal{S} \& \hat{F}} T_{\mathcal{S} \& \hat{F}} (Z_{\mathcal{S} \& \hat{F}})$ and $\mathbf{T}^{\mathcal{S} \& \hat{F}} T_{\mathcal{S} \& \hat{F}} (F_{\mathcal{S} \& \hat{F}})$ are uncorrelated (by the law of iterated expectation). This is so because under the assumption, the model errors, $\iota$, cannot be correlated with any function of the factors.

Thus, under this assumption, the variance matrix $\mathbf{Q}$ is a diagonal matrix whose diagonal blocks are equal to the variance matrices of $\mathbf{T}^{\mathcal{S} \& \hat{F}} T_{\mathcal{S} \& \hat{F}} (Z_{\mathcal{S} \& \hat{F}})$ and $\mathbf{T}^{\mathcal{S} \& \hat{F}} T_{\mathcal{S} \& \hat{F}} (F_{\mathcal{S} \& \hat{F}})$, respectively. Thus, a consistent estimator of $\mathbf{Q}$ can be obtained by

$$\mathbf{\hat{Q}} = \begin{bmatrix} \hat{\Sigma}_1 & 0 \\ 0 & \hat{\Sigma}_F \end{bmatrix} \quad (22)$$

---

19 Aside from the notational differences between our approach and that of Jagannathan and Wang, readers may find that our asymptotic variance of the two-pass estimator $\mathbf{\hat{\varphi}}_{TP}$ is quite different from that of Jagannathan and Wang. This difference, however, is due to the fact that their asymptotics apply to $[(F_{\mathcal{S} \& \hat{F}}) \hat{U}_t, (R_{\mathcal{S} \& \hat{F}}) \hat{E}(R_{\mathcal{S} \& \hat{F}})]^1$, while our asymptotics apply to $[(Z_{\mathcal{S} \& \hat{F}}) \hat{U}_t, (F_{\mathcal{S} \& \hat{F}}) \hat{E}(F_{\mathcal{S} \& \hat{F}})]^1$. Nonetheless, the variance matrix given (21) is asymptotically equivalent to that given in Theorem 1 of Jagannathan and Wang. A supplemental note on this equivalence is available from the authors on request.

20 Note that Assumption 1(ii) rules out non-zero correlation between $\iota$ and $F_t$. But it does not rule out non-zero correlation between the errors, $\iota$, and squared factors. Thus, in principle, $(Z_{\mathcal{S} \& \hat{F}}) \hat{U}_t$ and $(F_{\mathcal{S} \& \hat{F}})$ could be correlated under Assumption 1.
where \( \hat{E}_F \) is a consistent estimator of \( E_F = \lim \text{Var}\left[ \sqrt{T}(F \delta E(F)) \right] = \lim \text{Var}[T^{d/2t} \hat{F}_t (F \delta \hat{F})] \).

If the factor vectors are serially uncorrelated, then we can choose \( \hat{E}_F = \hat{F}_F \). Otherwise, we need to use nonparametric methods to estimate \( E_F \) (Shanken, 1992). Note that the diagonal form (22) does not require Assumption 2(ii), the assumption of no autocorrelation.

Substituting (22) into (21), we obtain

\[
\text{Var}(\hat{\lambda}_{1,TP}) \quad \text{Var}(\hat{\Theta}_{1,TP}) \quad \% J \hat{E}_F J / T, \tag{23}
\]

where \( \hat{E}_F / T \) is the variance matrix of \( \hat{F} \).\(^{21}\) This result implies that under Assumption 2(I) (the assumption of strictly exogenous factors), the variance matrix of \( \lambda_{1,TP} \) can be estimated simply by the sum of the variance matrices of \( \hat{\Theta}_{1,TP} \) and \( \hat{F} \). Note that Assumption 2(I) still allows conditional heteroskedasticity in the error, \( \epsilon \), and autocorrelation in the factors, \( F \). The formula (23) is relevant even for the case in which \( F_i \) and \( R_i \) are jointly t-distributed (MacKinlay and Richardson, 1991).

Under Assumption 3, Shanken (1992) derives the asymptotic variance matrix of the TP estimator, \( \hat{\lambda}_{TP} \). In fact, if we replace \( \hat{\lambda}_1 \) in (22) by \( \hat{\lambda}_3 \), we immediately obtain Theorem 1 of Shanken (1992).\(^{22}\) Thus, the result (23) can be regarded as a generalization of his result to 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 practice, the variance matrix \( \hat{E}_F \) may have to be estimated nonparametrically. Nonetheless, we below show that the test of model specification (3) or (6) requires only the estimation of the lambda component (\( \Theta \)) of the factor price vector. As long as Assumption 1 holds, the potential autocorrelation in the factor vector, \( F \), is irrelevant for model specification tests.

3.2. Optimal Minimum-Distance Estimation and Specification Tests

Because the choice of \( A \) is not restricted for (15), there are many possible MD (TP) estimators. Amemiya (1978), however, shows that the optimal choice of \( A \) is the inverse of \( \hat{S} \). With this

\(^{21}\)The matrix \( J \hat{E}_F J \) is equivalent to the “bordered version” of \( \hat{E}_F \) in Shanken (1992).

\(^{22}\)Strictly speaking, (23) is equivalent to Theorem 1 of Shanken (1992) applied to the case in which portfolio factors are absent. For this case, our result (23) coincides with the notation \((1+c)S \) in Theorem 1 of Shanken (1992).
choice, the MD estimator has the smallest asymptotic variance matrix among the MD estimators with different choices of A. This optimal MD estimator (OMD) is of a generalized least squares (GLS) form:

\[
\hat{\beta}_{OMD} = [\hat{\beta}_{0,OMD}, \hat{\beta}_{1,OMD}, \hat{\beta}_{2,OMD}]', \quad [\hat{X}^{\prime} \hat{S}_1 \hat{X}]^{-1} \hat{X}^{\prime} \hat{S}_2 \hat{X}^{\prime}, 
\]

(24)

where, under Assumption 1,

\[
\text{Var}(\hat{\beta}_{OMD}) = \left[ X^{\prime} \hat{S}_1 X \right]^{-1}, \quad (25)
\]

while under Assumption 2, we have an identical form as (25) but with \( \hat{S}_2 \) replacing \( \hat{S}_1 \).

An interesting result arises if Assumption 3 holds. Substituting (19) into (25), we can show that the OMD estimator of \( \beta \) exactly equals the GLS estimator applied to (6): That is, \( [\hat{X}^{\prime} \hat{E}_d \hat{X}]_{\hat{S}_1} \hat{X}^{\prime} \hat{E}_d \hat{X}^{\prime} \)

Shanken (1992, Theorems 3 and 4) shows that this GLS estimator is asymptotically equivalent to maximum likelihood under Assumption 3 and the joint normality of asset returns and factors. His result implies that the OMD estimator of \( \beta \) computed with \( \hat{S}_1 \) or \( \hat{S}_2 \) are also asymptotically equivalent to maximum likelihood under the same assumptions, because all of the estimates \( \hat{S}_1 \), \( \hat{S}_2 \) and \( \hat{S}_3 \) are consistent estimates of \( S \). However, it is important to note that when Assumption 3 is violated, the GLS estimator is no longer efficient, although it is still consistent. When Assumption 3 is violated, the weighting matrix \( \hat{S}_3 \) (which results in the GLS estimator) is suboptimal. This is so because \( \hat{S}_3 \) is no longer a consistent estimator of \( S \). For this case, more (asymptotically) efficient MD estimator is obtained using \( \hat{S}_1 \) or \( \hat{S}_2 \).

Despite its asymptotic efficiency, OMD should be used in practice with caution. A recent study by Altonji and Segal (1996) shows that optimal MD estimates could be more biased than the non-optimal MD estimates in finite samples. Their results indicate that the t-tests based on the traditional two-pass estimation (MD using identity matrix as a weighting matrix) instead of OMD could result in more reliable statistical inferences. Thus, using OMD as supplementary to the traditional two-pass estimation with correct variance matrix, not as a substitute, would be a prudent empirical practice.

When asset returns and factors are not jointly normal, the OMD estimator is not the most (asymptotically) efficient estimator. However, it remains an efficient estimator among a class of estimators utilizing the first pass OLS estimator \( \hat{\gamma} = [\gamma, \#] \). The proof of this claim is given in
Appendix B.

One advantage of using the OMD estimator is that it provides a convenient specification test statistic for testing the restrictions (5) or (6). That is, it can be shown that under (6),

$$Q_{OMD} = \frac{T-N-k}{N-k} (\mathbf{S}^\dagger (\mathbf{S}^\dagger \hat{\mathbf{S}}_{OMD} \mathbf{S}^\dagger )^{-1} S^\dagger (\mathbf{S}^\dagger \hat{\mathbf{S}}_{OMD} \mathbf{S}^\dagger ) Y F(\mathbf{S}^\dagger \hat{\mathbf{S}}_{OMD} \mathbf{S}^\dagger )}{N-k},$$

where \( \hat{\mathbf{S}} = \hat{\mathbf{S}}_1 \) under Assumption 1. If Assumptions 2 or 3 hold, \( \hat{\mathbf{S}} \) can be replaced by \( \hat{\mathbf{S}}_2 \) or \( \hat{\mathbf{S}}_3 \), respectively. There are two ways to use this OMD test statistic. First, \((N-1-k-q)Q_{OMD}\) can be compared with \(P^2(N-1-k-q)\) to determine rejection or acceptance of a given model specification. Second, the OMD statistic \(Q_{OMD}\) is directly compared with \(F(N-1-k-q,T-N+1)\). Asymptotically speaking, these two test strategies are equivalent. But we prefer the latter to the former because we find the latter performs better in our simulation exercises. In addition, the latter strategy is not without any theoretical consideration as we discuss below.

The OMD test is related with a test considered by Shanken (1985). To see this, consider the restriction (3). Define a GLS two-pass estimator of \( (\mathbf{t}^\dagger \mathbf{X}) \) by \( \mathbf{t}^\dagger = [\mathbf{t}^\dagger_{GLS}, \mathbf{t}^\dagger_{1,GLS}, \mathbf{t}^\dagger_{2,GLS}] / (\mathbf{X}^\dagger \mathbf{S}^\dagger \mathbf{X})^\dagger \mathbf{X}^\dagger \mathbf{R} \). Then, Shanken’s GLS residual test statistic has the form \(Q_C / [(T!N+1)/(N-1-k-q)]Q_S\), where

$$Q_S / \frac{(\mathbf{R} \& \mathbf{S}) \mathbf{t}_{GLS}^\dagger (\mathbf{R} \& \mathbf{S})}{1% \mathbf{t}_{GLS}} \mathbf{E}_{F} \mathbf{F}^\dagger_{1,GLS}$$

and, as above, \( \mathbf{t} \) is the risk price vector. Shanken shows that under Assumption 3 and the normality assumption, this statistic is asymptotically \(F(N-1-k-q,T-N+1)\)-distributed. We now suppose that Assumption 3 holds, so we use \( \hat{\mathbf{S}}_3 \) for the OMD estimator and the OMD statistic. Then, as mentioned before, the OMD estimator of \( \mathbf{8} = [\mathbf{g}_0, \mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3] \) becomes the GLS estimator \( \hat{\mathbf{8}}_{GLS} = (\mathbf{X}^\dagger \mathbf{E}^\dagger \mathbf{X})^\dagger (\mathbf{X}^\dagger \mathbf{E}^\dagger \mathbf{X})^\dagger \mathbf{E}^\dagger \mathbf{R} \). If we substitute \( \hat{\mathbf{S}}_3 \) and this GLS estimator into (26), and if we use the GLS estimator \( \hat{\mathbf{8}}_{1,GLS} \) and the factor mean vector \( \mathbf{F} \) to estimate the risk factor price, we can obtain \(Q_{OMD} = Q_C\), using the fact that \( \mathbf{t}^\dagger_{GLS} = \hat{\mathbf{8}}_{GLS} + \mathbf{F} \) and \( \mathbf{S}^\dagger \hat{\mathbf{S}}_{OMD} \mathbf{S}^\dagger \mathbf{t} \). Thus, the proof for (26) is available upon request.

If we could replace the estimated \( \mathbf{t} \) in the denominator by the true value of \( \mathbf{t} \), the statistic would become exactly \( F \)-distributed if the error terms \( \mathbf{e} \) are normally distributed. Shanken suggests that the \( Q_C \) statistic be compared to the critical values from the \( F(N-1-k,T-N+1) \) distribution.
Q_{OMD} statistic computed with \( \hat{S}_1 \) or \( \hat{S}_2 \) can be viewed as a heteroskedasticity-and/or-autocorrelation-robust version of the Q_{c} statistic.

An important advantage of using the Q_{c} test instead of its P²-version, TQ_{S}, is that it 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 penalizes itself through the coefficient (T-N+1) whenever N is too large. Thus, we can conjecture that the Q_{c} test would have better finite sample-properties than the test based on TQ_{S}. Indeed, Amsler and Schmidt (1985) confirm this conjecture through Monte Carlo simulations, although their simulations are confined to the cases in which Assumption 3 and the normality assumption hold. This discussion also motivates our use of our OMD statistic Q_{OMD}.

The Q_{OMD} statistic computed without asset-specific variables can serve as a test for the asset-pricing restriction (5). One advantage of this test (as well as the Q_{c} test) may be that it does not require any particular alternative hypothesis. That is, the OMD test without asset-specific variables can test the restriction (5) against a broader range of possible deviations from (5).

4. Simulation Results

So far we have discussed the asymptotic properties of the TP or MD estimators. In this section, we examine the performances of the TP or optimal MD (OMD) estimators in finite samples through a series of Monte Carlo simulations. In particular, our experiments are designed to investigate the size and power properties of the specification tests based on OMD, and the size properties of the t-tests for factor prices. We consider three different types of MD (TP) or OMD. From now on, we use the terms MD3 and OMD3, MD2 and OMD2, and MD1 and OMD1, to denote the non-optimal MD (TP) and OMD under Assumption 3 (i.i.d. model errors), Assumption 2 (conditionally heteroskedastic errors), and Assumption 1 (conditionally heteroskedastic and/or autocorrelated errors), respectively. The Newey-West (1987) method is used to control for autocorrelation in model errors and factors. We fix the value of bandwidth at 3, although many different values were also examined in unreported experiments.

4.1. Simulation Design

The foundation of our experiments are the two competing three factor models by Fama and
French (1993) and Jagannathan and Wang (1996) -- hereafter, FF and JW-96, respectively. The FF model is based on the market return (VW), SBM, and HML factors, while the Premium-Labor (PL) model of JW-96 uses the debt premium (PREM) and labor return (LABOR) factors, as well as the market return.  

For our simulation exercises, we generate data mimicking the actual returns and factors. Specifically, for a given three factor model (FF or JW-96), we estimate factor betas and risk prices by OMD using actual data. The estimated betas and risk prices are then used to calculate the expected returns for simulated assets. Simulated return data are obtained by adding model error terms \( (\epsilon_{it}) \) to these expected returns. For simplicity, the error terms are generated so that they are cross-sectionally independent. Our simulation results are based on 1,000 trials.

As mentioned above, we consider three possible cases in which the error terms are (a) i.i.d., or (b) heteroskedastic conditionally on factors, or (c) serially correlated over time. For (a), \( \epsilon_{it} = s_{i}v_{it} \), where \( s_{i} \) is the estimated standard error of asset \( i \), and \( v_{it} \) is a random number from i.i.d. \( N(0,1) \). For (b), we use \( \epsilon_{it} = (s_{i}v_{it})F_{it}/\sqrt{T_{it}D_{it}}v_{it}^{2}F_{it}^{2} \), where \( F_{it} \) is a factor. We use the market return factor (VW) for \( F_{it} \). Notice that with this data generating process, the unconditional variance of the \( \epsilon_{it} \), \( \text{var}(\epsilon_{it}) \), remains unchanged at \( s_{i}^{2} \). Finally, for (c), we use \( \epsilon_{it} = s_{i}v_{it} \) and \( \epsilon_{it} = D_{i0} + \sqrt{T_{it}D_{it}}s_{i}v_{it} \) where \( t = 2, \ldots, T \) and \( D = 0.1 \). We choose this small value for \( D \) because asset returns are generally only weakly autocorrelated over time. Again, with this process, \( \text{var}(\epsilon_{it}) = s_{i}^{2} \).

For the analysis of our \( Q_{\text{OMD}} \) and Shanken’s \( Q_{C} \) tests, we use the same actual factors while different simulated return data are used for each trial. However, for the analysis of the t-tests for factor prices, we use simulated factors for each trial. We do so because the factor price estimates depend on sample factor means (while the specification tests do not). Factors are generated based on a simple vector autoregression (VAR) applied to actual factors. Specifically, we

\[ 25 \text{Similar to JW-96, we perform analysis on raw returns and use the raw market returns as the first factor in each model.} \]

\[ 26 \text{In unreported experiments, we also have considered the cases with } D = 0.5, \text{ but the results from such experiments were not materially different from those reported here.} \]

\[ 27 \text{In all of our reported simulations, the error terms are assumed to be cross-sectionally independent. In unreported experiments, we have examined the cases with cross-sectionally correlated errors. But the results obtained from these experiments were materially similar to those reported in this paper.} \]
estimate the following VAR model of order one (VAR(1)) for each set of factors corresponding to the FF and JW-96 models:

\[ F_t = \gamma_0 + \sum_{i=1}^{1} \gamma_i F_{t-i} + \eta_t, \]  

(28)

where \( \eta_t \sim N(0, \Sigma) \).

The actual returns and factors we use for simulations are the data on raw returns for FF portfolios, which JW-96 have created and used. JW-96 replicate the FF method of constructing 100 size/pre-beta decile portfolios for NYSE/AMEX firms from July 1963 to December 1990. To check that our data set matches JW-96, we replicate their OLS and Fama-MacBeth analysis with univariate betas for the models common to our analysis. We are able to replicate JW-96's univariate-beta FM estimation of point estimates and standard errors for their three factor model to within three significant digits for most variables. However, because their data set does not contain FF factors, we use data for these series as currently available from FF. For the FF model, our estimates and t-statistics do not deviate more than 8% (in relative terms) from those reported by JW-96, but the OLS R^2 are identical to three significant digits. We suspect that these deviations are due to slightly different values for FF factors in our respective data sets. Our results using FF factors, however, appear close enough to theirs as to render any differences in inference immaterial. To save space, we make these results available upon request.

In order to examine the sensitivity of the TP (MD) and OMD estimation to the sample size, we repeat the analysis of each model using 25 value-weighted size/pre-beta quintile portfolios.

---

28Our simulation experiments are based on two separate VAR regressions, not on one VAR regression using all of the factors in the FF and PL-models. However, our simulation results do not depend on how the factors are generated. For the simulation exercises designed to investigate size and power properties of the OMD specification tests, we use actual factors. Our simulation results for the finite-sample size and power properties of the OLD t-tests also do not depend on whether factors are simulated based on two separate VAR’s or one VAR using all of the FF and JW-96 factors.

29We also examined excess returns, but the results are not materially different from those shown here.

30We obtained this data set through the FTP server at the University of Minnesota. We gratefully thank Jagannathan and Wang for access to their data.
We construct the 25 value-weighted portfolios from JW-96's 100 portfolios as follows. First, we identify groups of 4 original portfolios to form 25 portfolios that roughly relate to the 5-by-5 size/pre-beta quintiles used by FF. Second, while the 100 portfolios constructed by JW-96 are reported to be based on equally-weighted returns, it is common practice to evaluate 25 portfolios using value-weighted returns to avoid creating portfolios that are not representative of what an actual investor can realistically construct (see FF). To achieve value-weighting, we use the average firm size values reported for each 100 portfolio.

### 4.2. Preliminary Heteroskedasticity and Autocorrelation Test Results

The reliability and appropriateness of the TP or MD estimation may crucially depend on whether or not error terms \( (\epsilon_t) \) are autocorrelated. Thus, as a preliminary measure, we test the presence of heteroskedasticity and autocorrelation in the error terms in the FF model and the Premium-Labor (PL) model, using the actual data. Table 1 reports our test results for the cases of 25 assets. The table reveals that both the Fama-French and Premium-Labor models should ideally be robust to conditional heteroskedasticity. Using White’s (1980) test for heteroskedasticity, we found that both 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% level, while 24 and 25 of 25 did so at these levels, respectively, for the PL-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 PL-model (11 and 6 at the 10% and 5% levels, respectively) than for the FF model (7 and 3, similarly).

Of course, the non-zero number of rejections, alone, would not be a sufficient indication of heteroskedasticity or autocorrelation. Even if no return is conditionally heteroskedastic, we can expect \( n \times 100 \) percent of rejections by the White test when it is performed at the \( n \) level. To

---

\(^{31}\)We do so using neighboring size and pre-beta portfolios. Because FF first sort firms by size, combining neighboring size-decile portfolios into size-decile portfolios should exactly replicate true size quintile sorting. In contrast, because sub-sorting by pre-beta is performed over firms in each size quintile, combining neighboring pre-beta deciles that were constructed in different size deciles may result in a different grouping of firms across pre-beta versus true pre-beta quintile sub-sorting. However, because the average pre-betas in neighboring size deciles in JW-98's original 100 portfolio are similar, it is not likely that this difference in pre-beta sub-sorting results in materially different portfolios.
check whether the frequency of the White (or Breusch-Pagan) tests rejected is significantly
different from the number of rejections expected from a chosen \( \eta \) level, we conduct a proportion
test based on a normal approximation of the binomial distribution. To motivate our test, suppose
that in the FF model, 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 \( \eta \). Thus, the number of rejections would follow a binomial distribution. If
the number of trials (in our case, the number of portfolios, \( N \)) is large enough, then the binomial
distribution can be approximated by normal distribution.\(^{32}\) 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 model
and the PL-model. For both models, the p-values for the test of proportions are close to zeros
regardless of whatever significant level we choose.

While the test of proportions strongly suggests that autocorrelation matters in the PL-model
(the p-value is far smaller than 5 or 10% of significance level), the test result is less strong for
the FF model. The rejection number of seven at the 10% of significance level is statistically
significant (p-value equals 0.38%), but the rejection number of three at the 5% of 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 results 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 through unspecified common factors. Nevertheless,
it would be fair to say that the PL-model is more likely to suffer from autocorrelation than the FF
model does.

4.2. Size and Power Properties of the Specification Tests
Table 1 examines the size and power properties of the \( Q_C \) and \( Q_{OMD} \) specification tests when
asset returns are generated by the three FF factors (VW, SMB and HML). We begin by
considering the finite-sample size properties of the \( Q_C \) and \( Q_{OMD} \) tests for the cases with 25
assets. Panel A indicates that the \( Q_C \) test by Shanken (1992) has a good size property when

\(^{32}\)The variance of the rejection number equals \( N^\eta (1-\eta) \).
model error terms are i.i.d. over time. Amsler and Schmidt (1985) obtained similar results. At the 1% asymptotic significance level, the $Q_c$ test rejects the FF model exactly 10 times out of 1,000 trials. The test tends to under-reject correct models at the 10% significance level, but only slightly so. Not surprisingly, the test significantly over-rejects correct models when error terms are not i.i.d. For example, for the cases with autocorrelated errors, the actual rejection rates of the $Q_c$ test are 4.4% and 24.7% at the 1% and 10% significance levels, respectively.

The $Q_{OMD2}$ test, which is designed to be robust to conditional heteroskedasticity, performs better than the $Q_c$ and the $Q_{OMD1}$ test in terms of size when model error terms are heteroskedastic, although they tend to over-reject correct models especially at the 1% of significance level. When the errors are i.i.d., the $Q_{OMD2}$ test tends to over-reject correct models more than the $Q_c$ does, but only marginally so. For the cases with autocorrelated errors, the $Q_{OMD2}$ test rejects correct models much more often than asymptotic significance levels indicate. Somewhat surprisingly, for the cases with autocorrelated errors, the over-rejection rate of the $Q_{OMD1}$ test (which is robust to heteroskedasticity and/or autocorrelation) is worse than those of the $Q_c$ and $Q_{OMD2}$ tests. These results are consistent with previous findings that the Newey and West (1987) method has poor finite sample properties (see, for example, Andrews, 1991). In response to this problem, in some unreported exercises, we have tried an alternative method by Andrews (1991). The performances of the $Q_{OMD1}$ test computed by the Andrews method, however, were similar to those reported in Table 2. These results indicate that caution is required when statistical inferences are made based on OMD1. Testing conditional heteroskedasticity and autocorrelation would be useful. When only conditional heteroskedasticity in errors is detected, the $Q_{OMD2}$ test would be a desirable choice. In contrast, when autocorrelation is also detected, the statistical inferences made based on the $Q_{OMD1}$ test may be as unreliable as those based on the $Q_c$, $Q_{OMD3}$, or $Q_{OMD2}$ tests. Clearly, developing improved versions of the $Q_{OMD1}$ test would be a promising future research agenda.

[INSERT TABLE 2 HERE]

Panel A of Table 2 also shows that the degrees of over-rejection by the $Q_c$ and other $Q_{OMD}$ tests are quite severe when 100 assets are used. In particular, the $Q_{OMD1}$ test rejects correct models for most of the cases considered (at least 90% of all cases) whatever significance level is used. The $Q_c$ test has better size properties than any other $Q_{OMD}$ test, but it also rejects correct models too often except for the cases with i.i.d. errors. These results indicate that if too many
assets are analyzed, neither the $Q_c$ nor $Q_{OMD}$ tests produce reliable specification test results, especially when model errors are conditionally heteroskedastic or/and autocorrelated.

Panels B and C of Table 1 demonstrate the power properties of the $Q_c$ and $Q_{OMD}$ tests. Panel B reports the rejection rate of each test when Black’s (1972) CAPM is tested. Since asset returns are generated using the FF factors, the CAPM is a misspecified model to be rejected. In general, the $Q_c$ and other $Q_{OMD}$ tests have higher power to reject CAPM correctly, when model errors are heteroskedastic or autocorrelated, or when more assets are analyzed. This result is not surprising given that the tests tend to be more oversized for such cases as Panel A indicates. The ability to reject the CAPM of the $Q_c$ and $Q_{OMD}$ tests are somewhat low when 25 assets are analyzed, especially when the 1% asymptotic significance level is used. It appears that in terms of power, the 10% level is a better choice for the $Q_c$ and $Q_{OMD}$ tests. Panel C reports the rejection rates of the specification tests for the PL-model of JW-96. When 25 assets are analyzed, the $Q_c$ and $Q_{OMD}$ tests are less successful in rejecting correctly the PL-model than the CAPM. This result may imply that the PL-model is quite compatible to the FF model in the analysis of 25 assets. The power of the $Q_{OMD}$ tests to correctly reject the PL-model is much higher when 100 assets are analyzed. However, these tests are seriously biased to rejecting correct models in the analysis of 100 assets.

For a sensitivity analysis, Table 3 considers the cases in which asset returns are generated by the PL-model of JW-96: That is, Table 3 reports the finite-sample size and power properties of the $Q_c$ and $Q_{OMD}$ tests when the PL model is a correct specification. Tables 2 and 3 indicate that the sizes and power of the $Q_c$ and $Q_{OMD}$ tests are sensitive to the true returns-generating factors. Unexpectedly, Panel A’s of Tables 2 and 3 demonstrate that the specification tests incorrectly reject the PL-model too often, more often than they reject the FF model incorrectly. Exceptions are the $Q_c$ test when model errors are conditionally heteroskedastic. Although the test is (asymptotically) valid only for models with i.i.d. errors, they are sized better when the errors are conditionally heteroskedastic than when the errors are i.i.d. We are unable to provide a reasonable conjecture for this abnormal finding. However, it may be fair to say that the sizes of the $Q_c$ is much less distorted by conditional heteroskedasticity than by autocorrelation.

A puzzling question arising from Table 3 is why the sizes of the specification tests become more distorted when asset returns are generated with the JW-96 instead of FF factors. As

[INSERT TABLE 3 HERE]
discussed in the previous sections, the asymptotic properties of the traditional two-pass (TP) and OMD estimators hinge on the assumption of stationary factors (see Assumption 1). The asymptotic results derived in the previous sections do not depend on how close to a unit root process each factor time series would be, as long as each is stationary. However, the finite-sample properties of the TP or OMD estimates and test statistics may depend on whether a factor is near-unit root. To explore this possibility, we conducted some additional tests and simulation using the factors.

Panel A of Table 4 reports the VAR(1) estimation results for the factors used in the FF or JW-96 models. Panel A clearly indicates that of all the factors, PREM has the highest level of autocorrelation even with the effect of the other factors (0.961). This level of persistence suggests that PREM in this sample has a unit root or near-unit root. Furthermore, as Panel B of Table 4 indicates, Dickey-Fuller (1979) and Phillips-Perron (1988) tests decisively reject unit roots in all of the three FF factors and the LABOR factor, but do not so with PREM using either of the tests at the 5% significance level. Admittedly, unit root tests are known to have low power to reject unit root. Additionally, it is intuitive that PREM, the difference between the interest rates on Baa and Aaa corporate bond rates, is likely to be stationary over a long time because the two rates are likely to be cointegrated. However, because our sample is only finite, it is quite possible that the inference of the TP estimation is materially perturbed by the presence of a near-unit root factor in a finite sample. As a further test, we generated a factor assuming a simple AR(1) structure with increasing AR(1) coefficients. To save space, we do not report the detailed results from these simulations, but this exercise confirmed our conjecture: The over-rejection rates of the $Q_{OMD}$ tests increase with the value of the AR(1) coefficient. In particular, our simulation results obtained setting the AR(1) coefficient at 0.95 (a near nonstationary case) were quite comparable to those reported in Table 3. From this we conclude that the high

33 The asymptotic normality of the TP and OMD estimators of factor prices crucially hinges on the asymptotic normality of the beta estimates (#). However, the presence of near unit-root factors can distort the finite-sample distributions of the beta estimates, and therefore, the TP and OMD estimators. In addition, given that the $Q_C$ and $Q_{OMD}$ test statistics are functions of the beta and factor price estimates, their finite-sample distributions can be also affected by the presence of near-unit root factors.

34 A Gauss program for this exercise is available from the authors upon request.
autocorrelation of PREM likely biases upward the rejection rate of the $Q_C$ and $Q_{OMD}$ tests, rendering them unreliable in the presence of such a factor.

Returning to Table 3, Panels B and C of Table 3 show that the power of the $Q_C$ and $Q_{OMD}$ tests is higher when asset returns are generated with the PL-model. With few exceptions, the rejection rates of the CAPM by the specification tests are above 94%. Comparing Panel C’s of Tables 2 and 3, we can see that the specification tests are much more successful in rejecting the FF model correctly when asset returns are generated by the PL-model, than rejecting the PL-model correctly when asset returns are generated by the FF factors. However, these results are hardly intriguing given that all of the specification tests are severely over-sized when returns are generated by the PL model.

The main results from Tables 2 and 3 can be summarized as follows. First, the $Q_C$ and $Q_{OMD}$ tests tend to over-reject correct model specifications, especially when too many assets are analyzed and/or model errors are autocorrelated. Second, the $Q_C$ test is often better sized than the $Q_{OMD2}$ and $Q_{OMD1}$ tests, while the power of the former is quite similar to that of the latter. The $Q_{OMD1}$ test, which is designed to be robust to both conditional heteroskedasticity and autocorrelation, does not necessarily perform better than the $Q_C$ and $Q_{OMD2}$ tests, even if autocorrelation truly is present in the data. These results indicate that testing autocorrelation in model error terms is important to determine the reliability of the specification results. Third, the finite-sample size and power properties of the $Q_C$ and $Q_{OMD}$ tests are sensitive to factors generating actual asset returns. The $Q_C$ and $Q_{OMD}$ tests seem to reject true models too often when true factors are highly autocorrelated or near nonstationary. The TP (or MD) or OMD estimation is not a useful tool to examine models with highly persistent factors.

### 4.3. Size and Power Properties of the T-Test

We now consider the finite-sample properties of the t-test statistics computed by Fama and MacBeth’s (1973), Shanken’s (1992), and our MD methods. For our experiments, we simulate both the FF factors and asset returns as we have discussed above. For each set of simulated data, the factor prices are estimated by the traditional two-pass (TP) and our OMD methods. The asymptotic standard errors for the TP estimators are estimated by the Fama-MacBeth’s (1973)
method based on multivariate betas (FM),\textsuperscript{35} Shanken’s (1992) EIV correction method (SH, which is equivalent to our MD3), MD2 and MD1. We also examine the t-statistics based on Shanken’s (1992) GLS (SH-GLS, which is equivalent to OMD3), and two other OMD estimates, OMD2 and OMD1, that are robust to conditional heteroskedasticity and heteroskedasticity-and/or-autocorrelation, respectively. For each set of simulated data, we compute the t-statistics for testing the correct hypotheses that the true factor prices equal the values used to simulate data. Table 5.1 shows the rejection rates in these simulations at 5% of the asymptotic significance level. Thus, the rejection rates of properly sized t-tests should be near 5%.

| INSERT TABLE 5.1 HERE |

We begin by considering the tests based on nonoptimal MD estimators. Shanken (1992) has shown that the FM method underestimates correct asymptotic standard errors of the two-pass MD estimator when error terms are i.i.d. over time. His theoretical result indicates that the t-tests based on FM may over-reject correct hypotheses. Table 5.1 confirms his result with only a few exceptions (i.e., the t-tests for the HML factor price with 100 assets). Jagannathan and Wang (1998) have shown that the FM method would not necessarily over-reject correct hypotheses when errors are not i.i.d.\textsuperscript{36} In our simulation exercises, the FM method tends to over-reject correct hypotheses even for the cases with conditionally heteroskedastic or autocorrelated errors.

The EIV correction method by Shanken (1992) provides better sized t-statistics than the FM method does, even if model errors are conditionally heteroskedastic or autocorrelated. Since the Shanken method is valid only for models with i.i.d. errors, there is no theoretical reason why the t-tests based on it should perform better than the test based on the FM method when errors are

\textsuperscript{35}In unreported exercises, we also have considered the Fama-MacBeth estimator based on univariate betas. However, the results were materially similar to the results obtained using multivariate betas.

\textsuperscript{36}An intuition follows. To begin with, assume that the error term \( \epsilon \) is i.i.d. Then, the true variance of the TP estimate of a factor price consists of two parts: the variance when the true betas are used in TP (say, part a), and the other variance due to the use of estimated betas in TP (say, part b). Part a equals the variance computed by the FM method. Accordingly, the FM variance understates the true variance. However, when the error term \( \epsilon \) is conditionally heteroskedastic, the FM variance need not equal part a. Instead, it is possible that the FM variance is greater than the sum of part a and b. For such cases, the FM variance can overstate the true variance.
not i.i.d.. However, our simulation results indicate that the Shanken method generally leads to better statistical inferences even if model errors are conditionally heteroskedastic or autocorrelated. As a matter of fact, the performances of the t-test based on it are quite compatible to those of the tests based on MD2 when model errors are conditionally heteroskedastic, regardless of the number of assets analyzed. The t-tests based on MD1 also perform well when model errors are not autocorrelated. When the errors are autocorrelated, the tests based on MD1 dominate the tests based on other methods.

Overall, the t-tests based on nonoptimal MD estimators tend to over-reject the correct hypotheses related to the individual factors (except for the HML factor with 100 assets). Except for the HML factor, the degree of over-rejections increases with the number of assets. However, differently from the analyses of the model specification tests, the size properties of the t-tests are not too sensitive to the number of assets analyzed.

As we have shown in the previous section, Shanken’s (1992) GLS, which is an OMD under the assumption of i.i.d. model errors (OMD3), and the other two OMD estimators (OMD2 and OMD1), which are robust under Assumptions 1 and 2, respectively, are asymptotically more efficient than the traditional TP estimator. However, as we have discussed in the previous section, the Monte Carlo results of Altonji and Segal (1996) imply that the sizes of the t-tests based on OMD might be distorted. Our results reported in Table 5.1 are generally consistent with their findings. The t-tests for intercept term based on optimal estimators are over-sized especially when a large number of assets are analyzed. The t-tests for other factor prices based on Shanken’s GLS or OMD also tend to over-reject correct hypotheses more often than the tests based on non-optimal MD, although the difference is not substantial. The over-rejection rates of the t-tests for factor prices based on optimal estimators are higher when error terms are autocorrelated, or when a larger number of assets are analyzed. The t-tests based on Shanken's GLS generally perform well, when model errors are not autocorrelated, and/or when a smaller number of assets are analyzed. Shanken's GLS appears to dominate OMD2, even for the cases with conditional heteroskedasticity. But, the GLS method is also dominated by Shanken's EIV correction (SH) to the traditional TP estimation (FM). The t-tests based on OMD1 perform worst even if model error terms are autocorrelated. It appears that the t-tests based on MD1 are ideal for the analyses of models with autocorrelation.

[INSERT TABLE 5.2 HERE]
Table 5.2 analyzes the finite-sample power properties of the t-tests. The table reports the rejection rates for the incorrect hypotheses that the intercept term or the factor premia are equal to zero. The power of the t-tests based on nonoptimal MD estimators is low (except the t-tests for the hypotheses regarding intercept terms). The FM t-tests have slightly greater power, but it is not surprising given that they are oversized tests. When nonoptimal MD estimators are used, the rejection rates never exceed 40%. The power of the nonoptimal t-tests increases with the number of assets analyzed, except for the cases with the HML factor. The t-test based on optimal MD estimators generally greater power than the nonoptimal counterparts, but only in a small margin (except for the case with the VW factor, i.i.d. model errors and 25 assets). Given the size distortions reported in Table 5.1, the power gains by the tests based on optimal MD do not seem substantial.

As a sensitivity analysis, Table 6.1 reports the results obtained by simulating returns and JW-96 (instead of FF) factors. From both Tables 5.1 and 6.1, it is clear that the sizes of the t-tests are much more distorted when returns are generated with simulated JW-96 factors. The degree of over-rejection by t-tests are particularly severe for PREM (as well as LABOR), regardless of what estimation method is used. This implies that all of the TP and OMD methods severely underestimate the standard errors for the estimates of the PREM (as well as LABOR) factor price. That is, the t-tests based on TP, MD or OMD are likely to overstate the significance of the PREM factor price. As with analysis of the MD specification tests, these poor finite-sample properties of the t-tests are due to the fact that PREM a highly persistent factor. Our unreported simulation results obtained based on a simple artificial factor model revealed that when a factor is near nonstationary (the AR(1) coefficient is greater than 0.9), the standard errors for the two-pass risk-price estimators are severely downward biased.37

[INSERT TABLE 6.2 HERE]

Table 6.2 reports our results for the power analysis of the t-test statistics when the JW-96 factors generate returns. We report the results without further comments, because the power analysis of the t-tests is not very informative given their poor size properties in Table 6.1.

37The over-rejection rates of the t-tests can be reduced if higher values of bandwidth (around 40) are used for the Newey-West method. However, this improvement is only marginal. The over-rejection degrees of the t-tests remain severe regardless of the bandwidth value when factors are persistent.
The main results from Tables 5 and 6 can be summarized as follows. First, the t-test based on the Fama-MacBeth method generally over-rejects correct hypotheses. The Shanken's EIV correction method dominates the Fama-MacBeth even for the cases with conditional heteroskedasticity or autocorrelation. Second, the t-tests based on nonoptimal MD are better sized than those based on optimal MD, regardless of the number of assets analyzed. In particular, the t-test based on MD1 performs much better than those based on OMD1. Both the t-tests based on Shanken's GLS (OMD3) or OMD2 perform reasonably well for the analysis of 25 assets with no autocorrelation. Third, the t-tests based on MD or OMD appear to be inappropriate to make meaningful statistical inferences regarding persistent factors.

5. Conclusion

The two-pass cross-sectional regression method is widely used to evaluate numerous linear factor pricing models. However, using estimated betas in the second-stage cross-section regression causes a well-known errors-in-variable (EIV) problem. Shanken (1985, 1992) provides a simple specification test based on the two-pass regression, as well as a correct formula for two-pass regression standard errors at the presence of the EIV problem. However, his work is based on the assumption of no conditional heteroskedasticity and autocorrelation in returns. Generalizing his results, we derive alternative minimum-distance (MD) estimation and testing procedures that are robust to conditional heteroskedasticity and/or autocorrelation.

Our simulation results suggest the following practical strategies for the minimum-distance (MD) estimation of linear factor models. First, before researchers estimate their factor models by the MD methods, they must check persistency of their factors (Tables 3, 4, 6.1 and 6.2). When factors follow unit root or near-unit root processes, the statistical inferences produced by MD methods are not reliable. Second, it is not desirable to analyze too many assets (Tables 2 and 5.1). The MD analyses of one hundred assets tend to over-reject correct hypotheses. Analyses of twenty five or fewer assets would be much more appropriate. Third, it is important to test whether asset returns are heteroskedastic and/or autocorrelated conditionally on factors. When returns are conditionally autocorrelated, all of the specification tests based on Shanken and optimal OMD tend to over-reject correct models (Table 2). When returns are heteroskedastic but not autocorrelated, use of the heteroskedasticity-robust specification tests is recommended. When returns are neither heteroskedastic nor autocorrelated, the test of Shanken
is recommended. However, it is important to note that these model specification tests would have not have great power. Use of higher significance levels (e.g., 10 percent) is recommended to improve the power of the tests. Fourth, the t-tests for factor prices based on nonoptimal MD estimators are better sized than those based on optimal MD estimators (Tables 5.1 and 5.2). This result is consistent with Altonji and Segal (1996). The appropriate choice among homoskedastic, heteroskedasticity-robust, and heteroskedastic-and-autocorrelation-robust MD methods should be made based on the test results for heteroskedasticity and autocorrelation.
Appendix A

Proof of Theorem 1:

By the definition of $\hat{8}_{i\ell}$, we can show that
\[ \sqrt{T}(\hat{8}_{i\ell} \& \hat{8})' \sqrt{T}[(\hat{X}'AX)' \hat{X}'A'AX] (\hat{X}'AX)' \hat{X}'A'AX(\hat{X}'AX)' (\hat{X}'AX)' \hat{X}'A'AX(\hat{X}'AX)' (\hat{X}'AX)' \hat{X}'A'AX. \] \[ \text{(A1)} \]

Under (6), $-8_{0}e_{N} - %8_{1} - S8_{2} = 0$. Using this restriction and standard matrix theories, we can show
\[ \sqrt{T}(\hat{8} \& \hat{8})' \sqrt{T}[(\hat{8} \& \hat{8})' \hat{8} I_{N} \& \hat{8} I_{N}] \sqrt{T}(\hat{8} \& \hat{8}) \hat{8}. \]
\[ \text{(A2)} \]

where $\hat{8}_{i} = (1, -8NN)$. Then, it follows that
\[ \sqrt{T}(\hat{8} \& \hat{8}) \overset{Y}{\sim} N(0, S), \]
\[ \text{(A3)} \]

where $S = (8_{1})^{\frac{\&\&}{\&\&} \hat{8} I_{N}) = (8_{1})^{\frac{\&\&}{\&\&} \hat{8} I_{N})$ and $=\hat{8}$ is defined in (11). Under Assumption 1, $=\hat{8}$ can be estimated by $\hat{8}_{1}$. Thus, (A1) and (A2) imply (17). (18) results from the fact that under Assumption 2, $\hat{8}_{2}$, which is defined in (12), is a consistent estimator of $=\hat{8}$. Finally, we obtain (19) if we replace $\hat{8}_{1}$ in $\hat{8}_{1}$ by (13). The equality in (19) results from the fact that
\[ \text{(A4)} \]

Proof of Theorem 2:

Using (A1) and (A2), we can show
\[ \sqrt{T} (\hat{\delta}_{\text{TP}} \& \hat{\delta}) \] 
\[ \times (\hat{X}^1 A \hat{X}) \delta \hat{X}^1 A \sqrt{T} (\hat{\delta} \& \hat{\delta}) \] 
\[ \times (\hat{X}^1 A \hat{X}) \delta \hat{X}^1 A (\hat{Z}) \frac{1}{\sqrt{T}} \frac{T}{r'} (Z_1 \hat{U},) \cdot \] 
\[ (A5) \]

Using (8), and the equalities \( \delta = \delta_0 \), \( \delta_1 = \delta_1 + E(F) \), and \( \delta_2 = \delta_2 \), we can also show that

\[ \sqrt{T} (\hat{\delta}_{\text{TP}} \& \hat{\delta}) \] 
\[ \times \sqrt{T} (\hat{\delta}_{\text{TP}} \& \hat{\delta}) \] 
\[ \% J \sqrt{T} (\hat{F} \& \delta E(F)) \] 
\[ \times (\hat{X}^1 A \hat{X}) \delta \hat{X}^1 A (\hat{Z}) \frac{1}{\sqrt{T}} \frac{T}{r'} (Z_1 \hat{U},) \] 
\[ \% J \frac{1}{\sqrt{T}} \frac{T}{r'} (F, \delta E(F)) \] 
\[ (A6) \]

\[ \hat{M} \frac{1}{\sqrt{T}} \frac{T}{r'} < \gamma, \]

which gives us the desired result.
Appendix B: Asymptotic Efficiency of OMD

In this appendix, we address the asymptotic efficiency (minimum variance) of OMD among a certain class of estimators. We have shown above that the OMD estimator 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 OMD estimator \( \hat{\theta}_{\text{OMD}} \) among a class of estimators utilizing the first-pass OLS estimator \( \hat{7} = [*, \#] \). We here concern only with the efficiency of \( \hat{\theta}_{\text{OMD}} \), not of \( \theta_{\text{OMD}} = \hat{\theta}_{\text{OMD}} \{ \ldots \} \). This class of estimators is of our interest, because any TP estimator of the form (7) belongs to the class. We here restrict our discussion only to cases in which Assumption 3 holds, to save space.

Defining \( b = \text{vec}(\%) \) and \( 7(8, b) = [\theta_{\text{OMD}} + \% \theta_{1} + S_{2}, \%] \), we consider the following minimization problem:

\[
\min_{b, \theta} \quad O_{\text{MCS}}(b, \theta) / \text{vec}(\hat{7} \& \theta(8, b))) \{ \text{Var(vec(\hat{7}))) \} \text{vec}(\hat{7} \& \theta(8, b)), \tag{A7}
\]

where \( \text{Var(vec(\hat{7}))) \sim T \&(\ldots) \{ \ldots \} \). Solutions for this type of problems are called “minimum chi-square” (MCS) estimators (Ferguson, 1958; and Newey, 1987). We use notation \((\hat{b}_{\text{MCS}}, \hat{\theta}_{\text{MCS}})_{N} = (\hat{\theta}_{\text{OMD}, \text{MCS}}, \hat{\theta}_{1, \text{MCS}}, \hat{\theta}_{2, \text{MCS}}, \hat{\theta}_{\text{MCS}})_{N} \) to denote the solution for (A7). Newey (1987) shows that this MCS estimator is asymptotically efficient among estimators based on the OLS estimator \( \hat{7} = (\ast, \#) \). Further, by Chamberlain (1982, Proposition 8), under Assumption 3 and (6),

\[
O_{\text{MCS}} / O_{\text{MCS}}(\hat{\theta}_{\text{MCS}}, \hat{\theta}_{\text{MCS}}) \quad \sim \quad \mathcal{N}(\mathcal{M}, \mathcal{M}) \tag{A8}
\]

Thus, using this MCS method, researchers can test for the model specification (3) or (6). We also obtain the following result:

**Theorem 3** Under Assumption 3 and (6),

\[
\sqrt{T}(\hat{\theta}_{\text{MCS}} \& \theta) \quad \sim \quad \mathcal{N}[0(1\% \times 1), (1\%)(X^{\prime}E \& X)^{\times 1}],
\]

where \( c^\prime (\theta_{1} \& E(F_{j}))^{\times 1}E_{F}^{\times 1}(\theta_{1} \& E(F_{j})) \) and \( E_{F} = \text{Var}(F_{j}) \). The \( c \) and \( X \) can be estimated by using any consistent estimates of \( \theta_{1} \) and \( \%

33
Proof of Theorem 3: Note that
\[
\begin{bmatrix}
\mathbb{M} & \mathbb{S}_0, \mathbb{S}_1, \mathbb{S}_2 \\
\hat{b} \otimes \hat{b} & X \otimes \hat{b} \hat{I}_N
\end{bmatrix}
\]
Then, Chamberlain (1982, Proposition 7) implies that \((\hat{\theta}_{\text{MCS}}, \hat{b}_{\text{MCS}})\)'s asymptotically normal and,
\[
\lim_{T \to \infty} \text{Var}\left(\sqrt{T}(\hat{\theta}_{\text{MCS}} \otimes \hat{b})\right) = \left[\begin{bmatrix}
X & \hat{b} \hat{I}_N
\end{bmatrix} \otimes \left(\sum \hat{U} \hat{E} \hat{U} \hat{E} \hat{X} \hat{I}_N\right) \right]
\]
where \(\$ = \text{vec}(\#)\). But, using usual partitioned matrix theories and a little algebra, we can show:
\[
\lim_{T \to \infty} \text{Var}[\sqrt{T}(\hat{\theta}_{\text{MCS}} \otimes \hat{b})]' = \left[1 \otimes (\mathbb{S}_1 \% \mathbb{I}_F) \right]' \left(1 \otimes (\mathbb{S}_1 \% \mathbb{I}_F) \right) \text{Var}(\hat{\theta}_{\text{OMD}}) \left(1 \otimes (\mathbb{S}_1 \% \mathbb{I}_F) \right)
\]
where \(\mathbb{S}_1 = \text{vec}(\#)\). However, this variance matrix is exactly identical to the variance matrix of the OMD estimator given in (25), if we replace \(\hat{S}_1\) by \(\hat{S}_3\). This result implies that \(\hat{\theta}_{\text{OMD}}\) and \(\hat{\theta}_{\text{MCS}}\) have the same asymptotic distribution. Accordingly, we can conclude that \(\hat{\theta}_{\text{OMD}}\) (as well as \(\hat{\theta}_{\text{MCS}}\)) is asymptotically efficient among the estimators utilizing the OLS estimator \(\hat{\theta}\). That is, there is no estimator which utilizes \(\hat{\theta}\) and is more (asymptotically) efficient than the OMD estimator.

Although the MCS estimator is not of our direct interest, it is useful to clarify the relation between our OMD 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 \(Q_{\text{MCS}}(\theta, b)\) in (A7) is highly nonlinear in \(\theta\) and \(b = \text{vec}(\#)\). However, perhaps surprisingly, the solution for the problem (A7), \((\hat{\theta}_{\text{MCS}}, \hat{b}_{\text{MCS}})\)'s has a closed form. Thus, when Assumption 3 holds, \(\hat{\theta}_{\text{MCS}}\) could be used as an alternative to \(\hat{\theta}_{\text{OMD}}\). Furthermore, the specification test statistic (A8) can be dramatically simplified. We summarize these results in the following theorem:
Theorem 4: Define \( \hat{\mathcal{G}}_{\text{MCS}} / (1, \mathcal{G}^1_{\text{MCS}}) \). Then, the following are true: (I) Let \( \mathbb{D}_s \) be the smallest eigenvalue of \( \hat{\mathcal{G}}_{\text{MCS}} / (S_e, \mathcal{G}) \), where \( S_e = [e_N, S] \). Then, \( \hat{\mathcal{G}}_{\text{MCS}} \) is an eigenvector corresponding to \( \mathbb{D}_s \) which is normalized such that the first element equals one. (ii) \( \hat{\mathcal{G}}_{\text{MCS}} / (S_0, S_2) = (S_e, \mathcal{G}) \). (iii) \( Q_{\text{MCS}} = \mathbb{T} \mathbb{D}_s \).

Proof of Theorem 4: Define

\[
W(\mathcal{G}_i) = \begin{bmatrix}
1 & \mathcal{G}_{\text{MCS}} \\
0_{k \times 1} & I_k
\end{bmatrix} \mathbb{I} \mathbb{I}_N. 
\tag{A9}
\]

Note that \( W(\mathcal{G}_i) \) is nonsingular for any \( \mathcal{G}_i \). Thus, we can have

\[
Q_{\text{MCS}}(\mathcal{G}, \mathbb{B}) = [W(\mathcal{G}_i) d(T, \mathbb{B})] \mathbb{I} W(\mathcal{G}_i) (\hat{\mathcal{G}}_{\text{MCS}} \hat{\mathcal{U}} \hat{\mathcal{E}}) W(\mathcal{G}_i)] \mathbb{I} [W(\mathcal{G}_i) d(\mathcal{G}, \mathbb{B})], 
\tag{A10}
\]

where \( d(\mathcal{G}, \mathbb{B}) = [(\mathcal{G} X \mathcal{G}) \mathbb{B}] \mathbb{B} \mathbb{B} \). It is straightforward to show

\[
W(\mathcal{G}_i) d(\mathcal{G}, \mathbb{B}) = \begin{bmatrix}
1 & \mathcal{G} \\
\hat{\mathcal{B}}_{\mathcal{G}} & \hat{\mathcal{B}}_{\mathcal{G}}
\end{bmatrix}, \quad \begin{bmatrix}
\mathcal{G} E_{\mathcal{G}} \\
\hat{\mathcal{B}}_{\mathcal{G}}
\end{bmatrix} = \begin{bmatrix}
\hat{\mathcal{B}}_{\mathcal{G}} E_{\mathcal{G}} \\
\hat{\mathcal{B}}_{\mathcal{G}}
\end{bmatrix}. 
\tag{A11}
\]

Note that

\[
\hat{\mathcal{G}}_{\text{MCS}} = \begin{bmatrix}
1 & \mathcal{G} \\
\hat{\mathcal{B}}_{\mathcal{G}} & \hat{\mathcal{B}}_{\mathcal{G}}
\end{bmatrix}, \quad \begin{bmatrix}
1 & \mathcal{G} E_{\mathcal{G}} \\
\hat{\mathcal{B}}_{\mathcal{G}} & \hat{\mathcal{B}}_{\mathcal{G}}
\end{bmatrix} = \begin{bmatrix}
\hat{\mathcal{B}}_{\mathcal{G}} E_{\mathcal{G}} \\
\hat{\mathcal{B}}_{\mathcal{G}}
\end{bmatrix}. 
\tag{A12}
\]

Using this fact, we can show:

\[
W(\mathcal{G}_i) (\mathcal{G}_{\text{MCS}} \hat{\mathcal{U}} \hat{\mathcal{E}}) W(\mathcal{G}_i) 
\]

Substitute (A11) and (A13) into (A10) and let \( K = \hat{\mathcal{E}}_{\mathcal{G}} \mathcal{G} (\mathcal{G}_{\text{MCS}} \hat{\mathcal{E}}_{\mathcal{G}}) \). Then, a tedious but
straightforward algebra yields
\[
Q_{MCS}(\tilde{b}, \mathbf{b})^{'} = Q_M(\mathbf{b})^{'} \% T_{s}(\mathbf{b}, \mathbf{b}) \% K_{\hat{\mathbf{U}}_{\psi}} \hat{\mathbf{E}}_{\psi}^{\psi}{\delta}^{\prime} I_{s}(\mathbf{b}, \mathbf{b}),
\]  
where \( s(\mathbf{b}, \mathbf{b}) = \tilde{b} \% \hat{\mathbf{q}}_{\psi}(\mathbf{b})^{'} \% \hat{\mathbf{E}}_{\psi}^{\psi}{\delta}^{\prime} \hat{\mathbf{U}}_{\psi} I_{s}(\mathbf{b}, \mathbf{b}) \), and

\[
Q_M(\mathbf{b})^{'} / T_{s}(\mathbf{b}, \mathbf{b}) \% K_{\hat{\mathbf{U}}_{\psi}} \hat{\mathbf{E}}_{\psi}^{\psi}{\delta}^{\prime} I_{s}(\mathbf{b}, \mathbf{b}) \% (A15)
\]

We now consider the minimization solutions for \( \mathbf{b}, \mathbf{b}_0, \) and \( \mathbf{b}_2 \) given \( \mathbf{b}_1 \), which we denote by \( \tilde{\mathbf{b}}, \mathbf{b}_0, \) and \( \mathbf{b}_2 \), respectively. From the first-order conditions \( M_{Q_{MCS}}/\partial \mathbf{b} = 0 \) and \( M_{Q_{MCS}}/\partial \mathbf{b}_2 \mathbf{b}_2 = 0 \), we can easily show

\[
\tilde{\mathbf{b}}^{'} = \hat{\mathbf{q}}_{\psi}(\mathbf{b}_1)^{\psi}{\delta}^{\prime} \hat{\mathbf{E}}_{\psi}^{\psi}{\delta}^{\prime} \hat{\mathbf{U}}_{\psi} I_{s}(\mathbf{b}, \mathbf{b})^{'}(A16)
\]

and

\[
\begin{bmatrix}
\mathbf{b}_0 \\
\mathbf{b}_2
\end{bmatrix} = \begin{bmatrix}
\mathbf{S}_e \mathbf{E}_{\psi}^{\psi}{\delta}^{\prime} \mathbf{S}_e \\
\mathbf{S}_e \mathbf{E}_{\psi}^{\psi}{\delta}^{\prime} \mathbf{S}_e
\end{bmatrix} \mathbf{\tilde{b}} (A17)
\]

Substituting (A16) and (A17) into \( Q_{MCS}(\mathbf{b}, \mathbf{b}) = Q_{MCS}(\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}) \), we can obtain a concentrated mininhand:

\[
Q_{CM}(\mathbf{b}_1)^{'} / Q_{MCS}(\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b})^{'} \% T_{s}(\mathbf{b}, \mathbf{b}) \% K_{\hat{\mathbf{U}}_{\psi}} \hat{\mathbf{E}}_{\psi}^{\psi}{\delta}^{\prime} I_{s}(\mathbf{b}, \mathbf{b})^{'} (A18)
\]

Thus, minimizing (A18) with respect to \( \mathbf{b}_1 \) results in the MCS estimator of \( \mathbf{b}_1 \). However, the Rayleigh-Ritz theorem implies that the eigenvector corresponding to the smallest eigenvalue of the matrix \( \mathbf{\tilde{\mathbf{b}}^{'}} \mathbf{E}_{\psi}^{\psi}{\delta}^{\prime} \mathbf{S}_e (\mathbf{S}_e \mathbf{E}_{\psi}^{\psi}{\delta}^{\prime} \mathbf{S}_e) \mathbf{E}_{\psi}^{\psi}{\delta}^{\prime} \mathbf{\tilde{\mathbf{b}}} \) is a solution for the minimization of (A18). Thus, we have proven the result (I). The result (ii) comes from (A17). Finally, since \( \mathbf{b}_1(MCS) \) is an eigenvector of \( \mathbf{\tilde{\mathbf{b}}^{'}} \mathbf{E}_{\psi}^{\psi}{\delta}^{\prime} \mathbf{S}_e (\mathbf{S}_e \mathbf{E}_{\psi}^{\psi}{\delta}^{\prime} \mathbf{S}_e) \mathbf{E}_{\psi}^{\psi}{\delta}^{\prime} \) corresponding to the eigenvalue \( \mathbf{D}_1 \), we have \( \mathbf{\tilde{\mathbf{b}}^{'}} \mathbf{E}_{\psi}^{\psi}{\delta}^{\prime} \mathbf{S}_e (\mathbf{S}_e \mathbf{E}_{\psi}^{\psi}{\delta}^{\prime} \mathbf{S}_e) \mathbf{E}_{\psi}^{\psi}{\delta}^{\prime} \mathbf{\tilde{\mathbf{b}}} = \mathbf{D}_1^{'} \mathbf{\tilde{\mathbf{b}}} (A18) \)

Substituting this result into (A18) yields \( Q_{MCS}(\mathbf{b}_1(MCS), \mathbf{b}_1(MCS) = Q_{CM}(\mathbf{b}_1(MCS)) = \mathbf{T}\mathbf{D}_1 \).
A technical note related to Theorem 4 follows. If the first entry of $\hat{\theta}_{1,MCS}$ equals zero, this method suggested by Theorem 4 would not work out. For such cases, computation of $\hat{\theta}_{1,MCS}$ would be complicated. If we define $A = \mathbf{E}^k \mathbf{E}^k$, let $O$ be the vector of the last $k$ entries of $\hat{\theta}_{1,MCS}$, a necessary condition for such cases is that $O$ is an eigenvector of $\mathbf{E}_F \hat{A} \# \mathbf{R} \hat{A} \# O = 0$. But this condition is unlikely to hold in practice.

A notable result from Theorem 4 is that for models without firm-specific variables $S$, $\hat{\theta}_{MCS}$ is exactly identical to the closed-form solution of the maximum likelihood estimator derived by Zhou (1998). Since $\hat{\theta}_{OMD}$ is asymptotically equivalent to $\hat{\theta}_{MCS}$, it is also asymptotically equivalent to the maximum likelihood estimator. That is, if Assumption 3 holds and the errors $\epsilon_t$ are normal, $\hat{\theta}_{OMD}$ is the efficient estimator. However, when Assumption 3 is violated, $\hat{\theta}_{OMD}$ is strictly more efficient than the MCS or MLE estimator of $\theta$. This is so because, when Assumption 3 is violated, the weighting matrix, $(\hat{\mathbf{J}}_{\mathbf{ZZ}} \hat{\mathbf{U}} \mathbf{E}^k)$, which is used for the MCS estimator, becomes suboptimal. An asymptotically more efficient MCS estimator can be obtained by minimizing (A7) with the optimal weight, $[(\hat{\mathbf{J}}_{\mathbf{ZZ}} \hat{\mathbf{U}} \mathbf{I}_N)]^{\mathbf{E}^k}$ (or $[(\hat{\mathbf{J}}_{\mathbf{ZZ}} \hat{\mathbf{U}} \mathbf{I}_N)]^{\mathbf{E}^k}$). It can be shown that this alternative MCS estimator of $\theta$ is asymptotically equivalent to our OMD estimator of $\theta$ when Assumption 1 (or 2) holds.

Another interesting point of Theorem 4 is (iii). The test statistic $\mathbf{T}_{\mathbf{D}_s}$ is comparable to the likelihood ratio (LR) test statistic $T \mathbf{H}(1\%\mathbf{D}_s)$, 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. and Peter Schmidt, 1995, Efficient estimation of a model with dynamic panel
Altonji, Joseph G. and Lewis M. Segal, 1996, Small-sample bias in GMM estimation of
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
Amsler, Christine E., and Peter Schmidt, 1985, A Monte Carlo investigation of the accuracy of
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
Campbell, John Y., Andrew W. Lo and A. Craig MacKinlay, 1997, The Econometrics of
Chamberlain, Gary, 1982, Multivariate regression models for panel data, Journal of
Econometrics 18, 5-46.


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


