Preliminary. Do not circulate.

The Analysis of the Cross Section of Security Returns*

Ravi Jagannathan†

Georgios Skoulakis‡

Zhenyu Wang§

*This paper will appear as a chapter in the forthcoming *Handbook of Financial Econometrics* edited by Yacine Aït-Sahalia and Lars P. Hansen. The authors wish to thank Bob Korajczyk, Ernst Schaumburg and Jay Shanken for comments and Aiyesha Dey for editorial assistance.

†Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60201, USA. E-mail: rjaganna@kellogg.northwestern.edu

‡Department of Finance, Kellogg School of Management,Northwestern University, Evanston, IL 60201, USA. E-mail: g-skoulakis@kellogg.northwestern.edu

§Department of Finance and Economics, Graduate School of Business, Columbia University, New York, NY 10027, USA. E-mail: zwang@columbia.edu

Contents

1 Introduction

Financial assets exhibit wide variation in their historical average returns. For example, during the period from 1926 to 1999 large stocks earned an annualized average return of 13.0%, whereas long-term bonds earned only 5.6%. Small stocks earned 18.9% - substantially higher than large stocks. These differences are statistically and economically significant (Jagannathan and McGrattan (1996), Ferson and Jagannathan (1996)). Furthermore, such significant differences in average returns are also observed among other classes of stocks. If investors were rational, they would have anticipated such differences. Nevertheless, they still preferred to hold financial assets with such widely different expected returns. A natural question that arises is why this is the case. A variety of asset pricing models have been proposed in the literature for understanding why different assets earn different expected rates of return. According to these models difference assets earn different expected returns only because they differ in their systematic risk. The models differ based on the stand they take regarding what constitutes systematic risk. Among them, the linear beta pricing models form an important class.

According to linear beta pricing models, a few economy-wide pervasive factors are sufficient to represent systematic risk, and the expected return on an asset is a linear function of its factor betas (Ross (1976), Connor (1984)). Some beta pricing models specify what the risk factor should be based on theoretical arguments. According to the standard Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965), the return on the market portfolio of all assets that are in positive net supply in the economy is the relevant risk factor. Other models specify factors based on economic intuition and introspection. For example, Chen, Roll and Ross (1986) specify unanticipated changes in the term premium, default premium, the growth rate of industrial production and inflation as the factors, whereas Fama and French (1993) construct factors that capture the size and book-to-market effects documented in the literature and examine if these are sufficient to capture all economy-wide pervasive sources of risk. Campbell (1996) and Jagannathan and Wang (1996) use innovations to labor income as an aggregate risk factor. Another approach is to identify the pervasive risk factors based on systematic statistical analysis of historical return data as in Connor and Korajczyk (1988) and Lehmann and Modest (1988).

In this chapter we discuss econometric methods that have been used to evaluate linear beta pricing models using historical return data on a large cross section of stocks. Three approaches have been suggested in the literature for examining linear beta pricing models: (a) Cross sectional regressions method; (b) Maximum Likelihood (ML) methods; and (c) Generalized method of moments (GMM). Shanken (1992) and MacKinlay and Richardson (1991) show that the cross-sectional method is asymptotically equivalent to ML and GMM when returns are conditionally homoscedastic. In view of this, we focus our attention primarily on the cross-sectional regression method since it is more robust and easier to implement in large cross sections, and provide only a brief overview of the use of ML and GMM.

Fama and MacBeth (1973) developed the two pass cross sectional regression method to examine whether the relation between expected return and factor betas are linear. Betas are estimated using time series regression in the first pass and the relation between returns and betas are estimated using a second pass cross sectional regression. The use of estimated betas in the second pass introduces the classical errors-in-variables problem. The standard method for handling errors in variables problem is to group stocks into portfolios following Black, Jensen and Scholes (1972). Since each portfolio has a large number of individual stocks, portfolio betas are estimated with sufficient precision and this fact allows one to ignore the errors-in-variables problem as being of second order in importance. One, however, has to be careful to ensure that the portfolio formation method does not highlight or mask characteristics in the data that have valuable information about the validity of the asset pricing model under examination. Put in other words, one has to avoid data snooping biases discussed in Lo and MacKinlay (1990).

Shanken (1992) provided the first comprehensive analysis of the statistical properties of the classical two-pass estimator under the assumption that returns and factors exhibit conditional homoscedasticity. He demonstrated ed how to take into account the sampling errors in the betas estimated in the first pass and generalized-least-squares in the second stage cross-sectional regressions. Given these adjustments, Shanken (1992) conjectured that it may not be necessary to group securities into portfolios in order to address the errors in variables problem. Brennan, Chordia and Subrahmanyam (1998) make the interesting observation that the errors in variables problem can be avoided without grouping securities into portfolios by using risk-adjusted returns as dependent variables in tests of linear beta pricing models, provided all the factors are excess returns on traded assets. However, the relative merits of this approach as compared to portfolio grouping procedures has not been examined in the literature.

Jagannathan and Wang (1998) extended Shanken's analysis to allow for conditional heteroscedasticity and consider the case where the model is misspecified. This may happen even when the model holds in the population, if the econometrician uses the wrong factors or misses factors in computing factor betas. When the linear factor pricing model is correctly specified, firm characteristics such as firm size should not be able to explain expected return variations in the cross section of stocks. In the case of misspecified factor models, Jagannathan and Wang (1998) showed that the t-values associated with firm characteristics will typically be large. Hence, model misspecification can be detected using firm characteristics in cross-sectional regression. Such a test does not require that the number of assets be small relative to the length of the time series of observations on asset returns, as is the case with standard multivariate tests of linearity.

Gibbons (1982) showed that the classical maximum likelihood method can be used to estimate and test linear beta pricing models when stock returns are i.i.d and jointly normal. Kandel (1984) developed a straight forward computational procedure for implementing the maximum likelihood method. Shanken (1992) extended it further and showed that the cross sectional regression approach can be made asymptotically as efficient as the maximum likelihood method. Kim (1985) developed a maximum likelihood procedure that allows for the use of betas estimated using past data. Jobson and Korkie (1982) and MacKinlay (1987) developed exact multivariate tests for the CAPM and Gibbons, Ross and Shanken (1989) exact multivariate tests for linear beta pricing models when there is a risk free asset.

MacKinlay and Richardson (1991) show how to estimate the parameters of the CAPM by applying the GMM to its beta representation. They illustrate the bias in the tests based on standard maximum likelihood methods when stock returns exhibit contemporaneous conditional heteroscedasticity and show that the GMM estimator and the maximum likelihood method are equivalent under conditional homoscedasticity. An advantage of using the GMM is that it allows estimation of model parameters in a single pass thereby avoiding the error-in-variables problem. Linear factor pricing models can also be estimated by applying the GMM to their stochastic discount factor (SDF) representation. Jagannathan and Wang (2001) show that parameters estimated by applying the GMM to the SDF representation and the beta representation of linear beta pricing models are asymptotically equivalent.

The rest of the chapter is organized as follows. In Section 2 we set up the necessary notation and describe the general linear beta pricing model. We discuss in detail the two pass cross sectional regression method in Section 3 and provide an overview of the maximum likelihood methods in Section 4 and the GMM in Section 5. We summarize in Section 6.

2 Linear Beta Pricing Models, Factors and Characteristics

In this section we first describe the linear beta pricing model and then provide a brief history of its development. We need the following notation. Consider an economy with a large collection of assets. The econometrician has observations on the returns on N of the assets in the economy. Denote by $\mathbf{R}_t = [R_t^1 \;\; R_t^2 \;\; \cdots \;\; R_t^N]'$ the vector of gross returns on the N securities at time t, by $\mathbf{\Sigma}_{\mathrm{R}} = E\left[(\mathbf{R}_t - E[\mathbf{R}_t])(\mathbf{R}_t - E[\mathbf{R}_t])'\right]$ the covariance matrix of the return vector and by $\mathbf{f}_t = [f_t^1 \;\; f_t^2 \;\; \cdots \;\; f_t^K]'$ the vector of time-t values taken by the K factors.

2.1 Linear beta pricing models

Suppose the intertemporal marginal rate of substitution of the marginal investor is a time invariant function of only K economy-wide factors. Further assume that the returns on the N securities are generated according to the following linear factor model with the same K factors:

$$R_t^i = \alpha_i + \mathbf{f}_t'\boldsymbol{\beta}_i + u_{it} \qquad E[u_{it}|\mathbf{f}_t] = 0, \; i = 1, \ldots, N \tag{1}$$

where $\boldsymbol{\beta}_i$ is the vector of betas for security i given by

$$\boldsymbol{\beta}_i = \boldsymbol{\Sigma}_{\mathrm{F}}^{-1} E[(R_t^i - E[R_t^i])(\mathbf{f}_t - E[\mathbf{f}_t])], \tag{2}$$

and $\boldsymbol{\Sigma}_{\mathrm{F}}$ is the variance matrix of $\mathbf{f}_t$ given by

$$\boldsymbol{\Sigma}_{\mathrm{F}} = E\left[(\mathbf{f}_t - E[\mathbf{f}_t])\,(\mathbf{f}_t - E[\mathbf{f}_t])'\right]. \tag{3}$$

Under these assumptions, following Connor (1984), it can be shown that the expected return on any asset i, $i = 1, 2, \ldots, N$, is given by the linear beta pricing model as follows

$$E[R_t^i] = a_0 + \boldsymbol{\lambda}'\boldsymbol{\beta}_i, \; i = 1, \ldots, N \tag{4}$$

where $\boldsymbol{\lambda}$ is the $K \times 1$ vector of constants. The jth element of $\boldsymbol{\lambda}$, λ_j, corresponds to the jth factor risk premium - it is the expected return on a portfolio, p, of the N assets which has the property $\boldsymbol{\beta}_{p,k} = 1$ when $k = j$ and $\boldsymbol{\beta}_{p,k} = 0$ when $k \neq j$.

Sharpe (1964), Lintner (1965) and Mossin (1966) developed the first linear beta pricing model, the standard capital asset pricing model (CAPM). Merton (1973) derived the first linear multi beta pricing model by examining the intertemporal portfolio choice problem of a representative investor in continuous time. Long (1974) proposed a related multibeta pricing model in discrete time. Ross (1976) showed that an approximate version of the linear multi beta pricing model based on the assumptions that returns had a factor structure, the economy was large and there were no arbitrage opportunities. Chamberlain and Rothschild (1983) extended Ross's result to the case where returns had only an approximate factor structure, i.e., the covariance matrix of asset returns had only K unbounded eigenvalues. Dybvig (1985) and Grinblatt and Titman (1985) provided theoretical arguments supporting the view that deviations from exact linear beta pricing may not be economically important.

2.2 Factor selection

Three approaches have been followed in the literature for choosing the right factors. The first approach makes use of theory and economic intuition to identify the factors. For example, according to the standard CAPM there is only one factor and it is the return on the market portfolio of all assets in positive net supply. The Intertemporal Capital Asset Pricing Model (ICAPM) of Merton (1973) identifies one of the factors as the return on the market portfolio of all assets in positive net supply as in the CAPM. The other factors are those that help predict future changes in investment opportunities. As Campbell (1993) and Jagannathan and Wang (1996) point out, factors that help predict future return on the market portfolio would be particularly suitable as additional factors in an ICAPM setting. The common practice is to use the return on a large portfolio of stocks as a proxy for the return on the market portfolio, and innovations to macroeconomic variables as proxies for the other factors, as in Chen, Roll and Ross (1986).

The second approach uses statistical analysis of return data for extracting the factors. Factor analysis is one of the statistical approaches used. For expositional purposes it is convenient to rewrite the linear factor model in matrix notation as follows.

$$\mathbf{R}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \mathbf{u}_t \tag{5}$$

where $\mathbf{R}_t$ is the $N \times 1$ vector of date t returns on the N assets, $\boldsymbol{\alpha}$ is the N vector of α_i's, $\mathbf{B}$ is the $N \times K$ dimensional matrix of factor betas of the N assets, $\mathbf{f}_t$ is the $K \times 1$ vector of date t factor realizations, and $\mathbf{u}_t$ is the date t linear factor model innovations to the N returns. Let $\mathbf{\Sigma}_{\mathrm{U}}$ denote the diagonal covariance matrix of $\mathbf{u}_t$, and $\mathbf{\Sigma}_{\mathrm{F}}$ denote the covariance matrix of the factors. The covariance matrix of asset returns $\mathbf{R}_t$, $\mathbf{\Sigma}_{\mathrm{R}}$, can be decomposed as follows:

$$\mathbf{\Sigma}_{\mathrm{R}} = \mathbf{B}\mathbf{\Sigma}_{\mathrm{F}}\mathbf{B}' + \mathbf{\Sigma}_{\mathrm{U}}. \tag{6}$$

Note that the matrix of factor betas, $\mathbf{B}$, is identified only up to a linear transformation. For example consider transforming the factors to get $\mathbf{P}^{-1}\mathbf{f}_t$ as a new set of factors and transforming $\boldsymbol{\beta}$ to get $\mathbf{BP}$. Then

$$\mathbf{\Sigma}_{\mathrm{R}} = \mathbf{BP}(\mathbf{P}^{-1})\Sigma_{\mathrm{F}}(\mathbf{P}^{-1})'\mathbf{P}'\mathbf{B}' + \mathbf{\Sigma}_{\mathrm{U}} = \mathbf{B}\mathbf{\Sigma}_{\mathrm{F}}\mathbf{B}' + \mathbf{\Sigma}_{\mathrm{U}}. \tag{7}$$

The indeterminacy is eliminated by specifying that the factors are orthogonal along with other restrictions. For a discussion the reader is referred to Anderson (1984). The parameters $\mathbf{B}$ and $\mathbf{\Sigma}_{\mathrm{U}}$ are typically estimated using the maximum likelihood method under the assumption that stock returns are jointly normal and i.i.d over time. The estimates of $\mathbf{B}$ obtained in this way are then used in econometric evaluation of the linear beta pricing model given in equation (4) by applying the cross-sectional regression method described in the next section. When returns on the assets in excess the risk free rate are used, the multivariate test proposed by Gibbons, Ross and Shanken (1989) is more convenient after grouping securities into portfolios to reduce the cross sectional dimension. For many interesting applications, such as portfolio performance evaluation and risk management, is necessary to have estimates of the factors in addition to the factor betas. There are several approaches to get estimates of the realized value of the factors. The most common is the Fama-MacBeth style GLS cross-sectional regression where the returns are regressed on factor betas obtained through factor analysis. The estimated factors correspond to returns on specific portfolios of the primitive assets used to estimate the factor betas in the first stage using factor analysis. For a detailed discussion of factor selection using factor analysis the reader is referred to Lehmann and Modest (1985).

Chamberlain and Rothschild (1983) showed that, when the covariance matrix of asset returns has an approximate factor structure with K factors, the eigenvectors corresponding to the K exploding eigenvalues converge to the factor loadings. Let $\mathbf{R}$ denote the $N \times T$ matrix of T returns on the N assets, $\widehat{\boldsymbol{\Sigma}}_{\mathrm{R}T}$ denote the sample analogue of $\boldsymbol{\Sigma}_{\mathrm{R}}$, y_j denote the jth largest eigenvalue of the $N \times N$ matrix $\widehat{\boldsymbol{\Sigma}}_{\mathrm{R}T}$, $j = 1, 2, \ldots, K$, $\mathbf{t}_j$ denote the corresponding eigenvector, $\mathbf{F} = [\mathbf{f}_1 \cdots \mathbf{f}_T]$ denote the $K \times T$ matrix of T observations on the K factors, and $\mathbf{B}$ denote the $N \times K$ matrix of the factor betas of the N assets. Consider the matrix, $\tilde{\mathbf{B}}$ with $y_j^{1/2}\mathbf{t}_j$ as its jth column, $j = 1, 2, \ldots, K$. Then it can be shown that $\tilde{\mathbf{B}}\tilde{\mathbf{B}}'$ converges to $\mathbf{B}\mathbf{B}'$ almost surely under suitable regularity conditions. Hence principal component analysis can be used to identify the factor structure. As in the case of factor analysis discussed earlier, the factors can then be estimated through Fama and MacBeth (1973) style cross-sectional regression of the returns on the betas.

Connor and Korajczyk (1986) showed that, under suitable regularity conditions, the $K \times T$ matrix that has the first K eigenvectors of the $T \times T$ matrix $\widehat{\boldsymbol{\Sigma}}_{\mathrm{R}N} = \frac{1}{N}\mathbf{R}'\mathbf{R}$ as rows converges almost surely to $\mathbf{F}$ as $N \to \infty$. Note that a precise estimation of the factors requires the number of securities N to be large. When the number of assets N is much larger than the length of the time series of return observations T, which is usually the case, the Connor and Korajczyk (1986) approach that involves computing the eigenvectors of the $T \times T$ matrix $\widehat{\boldsymbol{\Sigma}}_{\mathrm{R}N}$ is preferable to the other approaches that are equivalent to computing the eigenvectors of the $N \times N$ matrix $\widehat{\boldsymbol{\Sigma}}_{\mathrm{R}T}$. Connor, Korajczyk and Uhlander (2002) showed that an iterated two-pass procedure that starts with an arbitrary set of K factors, estimates the factor betas using these factors in the first stage, and then computes the factors using Fama-MacBeth style cross-sectional regressions in the second stage ultimately converges to the K realized factor values as the number of asset N becomes large if returns have an approximate K factor structure and some additional regularity conditions are satisfied. The basic intuition behind these methods of estimating factors is that, when returns have an approximate K factor structure, any K distinct well diversified portfolios span the space of realized values of the K factors as N becomes very large. When the number of assets involved is very large, the sampling errors associated with the factor estimates, as compared to model misspecification errors, are of secondary importance and can be ignored.

The third approach is based on empirical anomalies. Empirical studies in the asset pricing area have documented several well known anomalies. The size and book to price anomalies are the more prominent ones among them. Firms that are relatively small and firms with relatively large book value to market value ratios have historically earned a higher average return after controlling for risk according to the standard CAPM. Banz (1981), Reinganum (1981) and Keim (1983) document the association between size and average returns in the cross section. Stattman (1980), Rosenberg, Reid, and Lanstein (1985) and Chan, Hamao and Lakonishok (1991) document the relation between book to price ratios and average returns. Schwert (1983) provides a nice discussion of the size and stock return relation and other anomalies. Berk (1995) forcefully argues that relative size and relative book to price ratios should be correlated with future returns on average in the cross section so long as investors have rational expectations. Suppose firms with these characteristics earn a higher return on average to compensate for some pervasive risk factor that is not represented by the standard CAPM. Then the return differential between two portfolios of securities, one with a high score on a characteristic and another that has a low score on the same characteristic, but otherwise similar in all other respects, would mimic the missing risk factor provided the two portfolios have similar exposure to other risk factors. Fama and French (1993) constructed two pervasive risk factors in this way that are now commonly used in empirical studies. One is referred to as the book to market factor (HML, short for high minus low) and the other as the size factor (SMB, short for small minus big). Daniel and Titman (1997) present evidence suggesting that these two risk factors constructed by Fama and French may not fully account for the ability of size and book to price ratios to predict future returns.

2.3 Characteristics

Theoretical models are abstractions from reality and the beta pricing model is no exception. Hence it should come as no surprise that empirical studies in the asset pricing area find that most pricing models are statistically rejected. However, statistical tests that reject model validity by themselves do not convey much information about what is missing in a model. As a result, the common practice in empirical studies is to identify firm specific characteristics and macroeconomic variables that help forecast future returns and create portfolios, based on such predictability, that pose the greatest challenge to an asset pricing model. Such portfolios, in a way, summarize what is missing in an asset pricing model.

Two such characteristics, namely the relative firm size and the book to price ratio of a firm, were discussed earlier. Another important characteristic that has received attention in the empirical literature is relative strength or momentum. It is typically measured by a score that depends on the return on an asset relative to other assets in the comparison group. Jegadeesh and Titman

(1993) show how to construct portfolios that earn apparently superior risk-adjusted returns using clever trading strategies that exploit momentum, i.e., the tendency of past winners to continue to win and past losers to continue to lose. Other characteristics that have received attention include liquidity (Brennan and Subrahmanyam (1996)), earnings to price ratio (Basu (1977)), dividend yield (Fama and French (1988), and leverage (Bhandari (1988)). In order to estimate the factors, the econometrician has to decide on the number of factors to be considered. Theoretical models do not provide much of a guidance on this issue. Hence empirical studies typically examine the sensitivity of the conclusions to the number of factors used. When factor analysis is used and multivariate normality of the factors and the residuals is assumed, maximum likelihood methods can be used to estimate the factor loadings and test whether an exact K-factor structure obtains using the likelihood ratio test. The reader is referred to Andersen (1984) for further details. Connor and Korajczyk (1993) show how to relax the exact factor structure assumption in tests for the number of factors.

3 Cross-Sectional Regression Methods

3.1 Description of the CSR method

In this subsection, we provide a description of the cross-sectional regression estimator originally employed by Fama and MacBeth (1973) in a slightly different form, and present the Fama-MacBeth covariance matrix estimator. We will use succinct vector-matrix notation to ease the exposition. Recall that the $N \times K$ matrix of betas or factor loadings is denoted by $\mathbf{B} = [\boldsymbol{\beta}_1 \ \cdots \ \boldsymbol{\beta}_N]'$. Then (2) can be rewritten compactly as

$$\mathbf{B} = E\left[(\mathbf{R}_t - E[\mathbf{R}_t])(\mathbf{f}_t - E[\mathbf{f}_t])'\right]\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}. \tag{8}$$

Next, we define the vector of risk premia

$$\mathbf{c} = [a_0 \ \ \boldsymbol{\lambda}']' \ \ ((K+1)\times 1 \text{ vector}) \tag{9}$$

and the matrix

$$\mathbf{X} = [\mathbf{1}_N \ \ \mathbf{B}] \ \ (N \times (K+1) \text{ matrix}). \tag{10}$$

We assume that $N > K$, as is typically the case in practice. The rank of matrix of $\mathbf{X}$ is assumed to equal $K+1$, that is, $\mathbf{X}$ is of full rank. The beta representation equation (4) is then concisely expressed as

$$E[\mathbf{R}_t] = a_0\mathbf{1}_N + \mathbf{B}\boldsymbol{\lambda} = \mathbf{X}\mathbf{c} \tag{11}$$

and therefore the unknown parameter $\mathbf{c}$ can be expressed as

$$\mathbf{c} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{R}_t]. \tag{12}$$

Standard time-series regression yields

$$\mathbf{R}_t = E[\mathbf{R}_t] + \mathbf{B}(\mathbf{f}_t - E[\mathbf{f}_t]) + \mathbf{u}_t \quad \text{with} \quad E[\mathbf{u}_t] = \mathbf{0}_N \quad \text{and} \quad E[\mathbf{u}_t\mathbf{f}_t'] = \mathbf{0}_{N\times K} \tag{13}$$

as it follows from the definition of $\mathbf{B}$. Thus, using the beta representation (11) we can rewrite equation (13) as

$$\mathbf{R}_t = a_0\mathbf{1}_N + \mathbf{B}(\mathbf{f}_t - E[\mathbf{f}_t] + \boldsymbol{\lambda}) + \mathbf{u}_t. \tag{14}$$

Equation (14) can be viewed as the model describing the return data generating process.

Suppose that the econometrician observes a time series of length T of security return and factor realizations, denoted as follows

$$\left[\mathbf{R}_t' \;\; \mathbf{f}_t'\right]' = (R_{1t}, \ldots, R_{Nt}, f_t^1, \ldots, f_t^K)', \quad t = 1, \ldots, T. \tag{15}$$

Some standard econometric assumptions about the dynamics of the preceding time series are in order. These assumptions will be necessary for the development of the asymptotic theory that follows. We assume that the vector process $[\mathbf{R}_t' \;\; \mathbf{f}_t']'$ is stationary and ergodic and that the law of large numbers applies so that the sample moments of returns and factors converge to the corresponding population moments.

The CSR testing method involves two steps and for this reason it is also referred to as the two-pass procedure. In the first step we estimate $\boldsymbol{\Sigma}_{\mathrm{F}}$ and $\mathbf{B}$ by the standard sample analogue estimates

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T} = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)' \quad \text{where} \quad \overline{\mathbf{f}}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t \tag{16}$$

and

$$\widehat{\mathbf{B}}_T = \left[\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{R}_t - \overline{\mathbf{R}}_T\right)\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)'\right]\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1} \quad \text{where} \quad \overline{\mathbf{R}}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t. \tag{17}$$

Then, in the second step, for each $t = 1, \ldots, T$ we use the estimate $\widehat{\mathbf{B}}_T$ of the beta matrix $\mathbf{B}$ and simple cross-sectional regression to obtain the following ordinary least-squares (OLS) estimates of $\mathbf{c}$

$$\widehat{\mathbf{c}}_t = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\mathbf{R}_t, \; t = 1, \ldots, T \tag{18}$$

where

$$\widehat{\mathbf{X}}_T = [\mathbf{1}_N \quad \widehat{\mathbf{B}}_T] \tag{19}$$

as suggested by equation (11). The standard Fama-MacBeth estimate of $\mathbf{c}$ then is the time-series average of the T estimates

$$\bar{\widehat{\mathbf{c}}}_T = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{c}}_t = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\overline{\mathbf{R}}_T. \tag{20}$$

To conduct inference regarding the parameter of interest $\mathbf{c}$, one also needs estimates of the asymptotic covariance of the estimator $\bar{\widehat{\mathbf{c}}}_T$. Fama and MacBeth (1973) proposed treating the set of the individual CSR estimates $\{\widehat{\mathbf{c}}_t : t = 1, \ldots, T\}$ as a random sample and therefore estimating the covariance matrix of $\sqrt{T}(\bar{\widehat{\mathbf{c}}}_T - \mathbf{c})$ by

$$\widehat{\mathbf{V}}_T = \frac{1}{T}\sum_{t=1}^{T}(\widehat{\mathbf{c}}_t - \bar{\widehat{\mathbf{c}}}_T)(\widehat{\mathbf{c}}_t - \bar{\widehat{\mathbf{c}}}_T)'. \tag{21}$$

The Fama-MacBeth procedure has an intuitive appeal and is rather easy to implement. However, one has to use caution when using this procedure since it is subject to the well-known errors-in-variables (EIV) problem. As first pointed out in Shanken (1992), some corrections are required to ensure the validity of the method.

A more flexible estimate that has been suggested in the literature is the feasible GLS version of the foregoing estimate. The following notation for the GLS weighting matrix and the corresponding estimator will be used throughout the section 3. Let $\mathbf{Q}$ be a symmetric and positive definite $N \times N$ matrix and $\widehat{\mathbf{Q}}_T$ be a consistent estimator of $\mathbf{Q}$ which is also assumed to be symmetric and positive definite for all T. Then the feasible GLS estimator of $\mathbf{c}$ obtained from the cross-sectional regression at time t is

$$\widehat{\mathbf{c}}_t = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\mathbf{R}_t \tag{22}$$

and therefore the Fama-MacBeth estimator of $\mathbf{c}$ is given by

$$\bar{\widehat{\mathbf{c}}}_T = \frac{1}{T}\sum_{t=1}^{T}(\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\mathbf{R}_t = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\overline{\mathbf{R}}_T. \tag{23}$$

The subsequent analysis will employ the feasible GLS version of the CSR estimator.

3.2 Consistency and asymptotic normality of the CSR estimator

In this subsection, we address the issues of consistency and asymptotic normality of the two-pass cross-sectional regression estimator that was described in the previous section. Using the law of large numbers and Slutsky's theorem, it follows from (17) and (16) that $\widehat{\mathbf{X}}_T \xrightarrow{P} \mathbf{X}$ as $T \to \infty$. Applying the law of large numbers and Slutsky's theorem once again and using (11) we obtain from (23) that $\overline{\widehat{\mathbf{c}}}_T$ converges in probability to $(\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}E[\mathbf{R}_t] = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}\mathbf{X}\mathbf{c} = \mathbf{c}$. Thus we have shown the following

Proposition 3.1 *The time series average $\overline{\widehat{\mathbf{c}}}_T$ of the cross-sectional estimates*

$$\widehat{\mathbf{c}}_t = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\mathbf{R}_t,\ t = 1, \ldots, T$$

where $\widehat{\mathbf{X}}_T = [\mathbf{1}_N \quad \widehat{\mathbf{B}}_T]$ is a consistent estimator of $\mathbf{c} = \begin{bmatrix} a_0 & \boldsymbol{\lambda}' \end{bmatrix}'$, that is

$$\overline{\widehat{\mathbf{c}}}_T \xrightarrow{P} \mathbf{c} \text{ as } T \to \infty. \tag{24}$$

Next we proceed to address the issue of precision of the estimator $\overline{\widehat{\mathbf{c}}}_T$ by deriving its asymptotic distribution. The derivation will require some additional mild assumptions. First we need some notation. Define

$$\mathbf{D} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q} \tag{25}$$

and

$$\mathbf{h}_t^1 = \mathbf{R}_t - E[\mathbf{R}_t],\ \mathbf{h}_t^2 = \left[(\mathbf{f}_t - E[\mathbf{f}_t])'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}\boldsymbol{\lambda}\right]\mathbf{u}_t \text{ and } \mathbf{h}_t = [(\mathbf{h}_t^1)' \quad (\mathbf{h}_t^2)']'. \tag{26}$$

Clearly $E[\mathbf{h}_t^1] = \mathbf{0}_N$ and (13) implies $E[\mathbf{h}_t^2] = \mathbf{0}_N$. The following assumption will be essential in the derivation of the asymptotic distribution of the CSR estimator.

Assumption A. The central limit theorem applies to the random sequence $\mathbf{h}_t$ defined in (26), that is $\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbf{h}_t$ converges in distribution to a multivariate normal with zero mean and covariance matrix given by

$$\boldsymbol{\Sigma}_{\mathbf{h}} = \begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma}' & \boldsymbol{\Pi} \end{bmatrix}$$

where

$$\boldsymbol{\Psi} = \sum_{k=-\infty}^{+\infty} E\left[\mathbf{h}_t^1(\mathbf{h}_{t+k}^1)'\right],\ \boldsymbol{\Gamma} = \sum_{k=-\infty}^{+\infty} E\left[\mathbf{h}_t^1(\mathbf{h}_{t+k}^2)'\right] \text{ and } \boldsymbol{\Pi} = \sum_{k=-\infty}^{+\infty} E\left[\mathbf{h}_t^2(\mathbf{h}_{t+k}^2)'\right]. \tag{27}$$

Assumption A is rather mild and can be obtained under standard stationarity, mixing and moment conditions. Related results can be found, for instance, in Hall and Heyde (1980) and Davidson (1994). Note that when the time series $[\mathbf{R}_t' \quad \mathbf{f}_t']'$ is stationary and serially independent we have $\boldsymbol{\Psi} = E\left[\mathbf{h}_t^1(\mathbf{h}_t^1)'\right] = \boldsymbol{\Sigma}_{\mathrm{R}}$.

We are now in a position to state and prove the theorem that gives the asymptotic distribution of the CSR estimator. A more general version of this theorem, dealing also with security characteristics, appeared as Theorem 1 in Jagannathan and Wang (1998) with slightly different notation. We will denote $\mathbf{A}_T \stackrel{\text{LD}}{=} \mathbf{B}_T$ when the two multivariate time series $\mathbf{A}_T$ and $\mathbf{B}_T$ have the same asymptotic distribution as $T \to \infty$.

Theorem 3.2 *Let* $\mathbf{c} = [a_0 \quad \boldsymbol{\lambda}']'$ *and* $\overline{\widehat{\mathbf{c}}}_T = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\overline{\mathbf{R}}_T$ *where* $\widehat{\mathbf{X}}_T = [\mathbf{1}_N \quad \widehat{\mathbf{B}}_T]$*. Under Assumption A, as* $T \to \infty$*,* $\sqrt{T}(\overline{\widehat{\mathbf{c}}}_T - \mathbf{c})$ *converges in distribution to a multivariate normal with zero mean and covariance*

$$\boldsymbol{\Sigma}_c = \mathbf{D}\boldsymbol{\Psi}\mathbf{D}' + \mathbf{D}\boldsymbol{\Pi}\mathbf{D}' - \mathbf{D}\left(\boldsymbol{\Gamma} + \boldsymbol{\Gamma}'\right)\mathbf{D}' \tag{28}$$

where $\mathbf{D} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}$ *with* $\mathbf{X} = [\mathbf{1}_N \quad \mathbf{B}]$ *and* $\boldsymbol{\Psi}, \boldsymbol{\Gamma}$ *and* $\boldsymbol{\Pi}$ *are defined in (27).*

Proof. First we note that the identity $\mathbf{R}_t = \widehat{\mathbf{X}}_T\mathbf{c} + (\mathbf{B} - \widehat{\mathbf{B}}_T)\boldsymbol{\lambda} + \mathbf{R}_t - \mathbf{X}\mathbf{c}$ holds. Using the pricing equation (11) and the foregoing identity we obtain from (23)

$$\widehat{\mathbf{c}}_t = \mathbf{c} + (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\left[(\mathbf{B} - \widehat{\mathbf{B}}_T)\boldsymbol{\lambda} + (\mathbf{R}_t - E[\mathbf{R}_t])\right]$$

from which it follows, using Slutsky's theorem, that

$$\begin{aligned}\sqrt{T}(\overline{\widehat{\mathbf{c}}}_T - \mathbf{c}) &= (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\left[\sqrt{T}(\mathbf{B} - \widehat{\mathbf{B}}_T)\boldsymbol{\lambda} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\mathbf{R}_t - E[\mathbf{R}_t])\right] \\ &\stackrel{\text{LD}}{=} \mathbf{D}\left[\sqrt{T}(\mathbf{B} - \widehat{\mathbf{B}}_T)\boldsymbol{\lambda} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\mathbf{R}_t - E[\mathbf{R}_t])\right]. \end{aligned} \tag{29}$$

¿From (14) we have

$$\mathbf{R}_t - \overline{\mathbf{R}}_T = \mathbf{B}(\mathbf{f}_t - \overline{\mathbf{f}}_T) + \mathbf{u}_t - \overline{\mathbf{u}}_T \Rightarrow$$

$$\frac{1}{T}\sum_{t=1}^{T}(\mathbf{R}_t - \overline{\mathbf{R}}_T)(\mathbf{f}_t - \overline{\mathbf{f}}_T)' = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{u}_t - \overline{\mathbf{u}}_T)(\mathbf{f}_t - \overline{\mathbf{f}}_T)' + \mathbf{B}\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}$$

and therefore (17) yields

$$\widehat{\mathbf{B}}_T - \mathbf{B} = \left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_t(\mathbf{f}_t - \overline{\mathbf{f}}_T)'\right]\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1}.$$

Using the last equation and Slutsky's theorem again we obtain from (29)

$$\sqrt{T}(\overline{\widehat{\mathbf{c}}}_T - \mathbf{c}) \stackrel{\text{LD}}{=} \mathbf{D}\left[-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{u}_t\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}\boldsymbol{\lambda} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\mathbf{R}_t - E[\mathbf{R}_t])\right]$$

$$\stackrel{\text{LD}}{=} \mathbf{D}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathbf{h}_t^1 - \mathbf{h}_t^2\right)\right] = \mathbf{DH}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{h}_t$$

where $\mathbf{H} = [\mathbf{I}_N \quad -\mathbf{I}_N]$. Using Assumption A, which states that the central limit theorem applies to the random sequence $\mathbf{h}_t$, yields the asymptotic distribution of $\overline{\widehat{\mathbf{c}}}_T$

$$\sqrt{T}(\overline{\widehat{\mathbf{c}}}_T - \mathbf{c}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Sigma}_c)$$

where

$$\boldsymbol{\Sigma}_c = \mathbf{DH}\begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma}' & \boldsymbol{\Pi} \end{bmatrix}\mathbf{H}'\mathbf{D}' = \mathbf{D}\boldsymbol{\Psi}\mathbf{D}' + \mathbf{D}\boldsymbol{\Pi}\mathbf{D}' - \mathbf{D}\left(\boldsymbol{\Gamma} + \boldsymbol{\Gamma}'\right)\mathbf{D}'$$

thus completing the proof. ■

Using the previous theorem, we can compute appropriate standard errors and thus test hypotheses of interest, such as $\boldsymbol{\lambda} = \mathbf{0}_K$. Actual application of the theorem, though, requires knowledge of several matrices. Since the matrices are unknown, we use the values of their consistent estimators instead. The matrix $\mathbf{D}$ can be consistently estimated by $(\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T$ while the spectral density matrix $\boldsymbol{\Sigma}_{\mathbf{h}} = \sum_{k=-\infty}^{k=\infty} E[\mathbf{h}_t\mathbf{h}_t']$ can be estimated by the methods proposed by Newey-West (1987), Andrews (1991) and Andrews and Monahan (1992). In large cross sections it would be necessary to impose a block diagonal structure on $\boldsymbol{\Sigma}_{\mathbf{h}}$ to ensure that the law of large numbers starts kicking in given the length of the time series of observations available to the econometrician.

3.3 Fama-MacBeth variance estimator

As mentioned in subsection 3.1, the Fama-MacBeth estimator of the asymptotic covariance of $\overline{\widehat{\mathbf{c}}}_T$ is $\widehat{\mathbf{V}}_T = \frac{1}{T}\sum_{t=1}^{T}(\widehat{\mathbf{c}}_t - \overline{\widehat{\mathbf{c}}}_T)(\widehat{\mathbf{c}}_t - \overline{\widehat{\mathbf{c}}}_T)'$. In this subsection we examine the limiting behavior of $\widehat{\mathbf{V}}_T$. Substituting $\widehat{\mathbf{c}}_t$ from (22) into (21) we obtain

$$\widehat{\mathbf{V}}_T = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\left[\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{R}_t - \overline{\mathbf{R}}_T\right)\left(\mathbf{R}_t - \overline{\mathbf{R}}_T\right)'\right]\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T(\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}. \tag{30}$$

Applying the law of large numbers and the fact that $\widehat{\mathbf{Q}}_T \xrightarrow{P} \mathbf{Q}$ then yields that $\widehat{\mathbf{V}}_T \xrightarrow{P} \mathbf{D}\boldsymbol{\Sigma}_{\mathrm{R}}\mathbf{D}' = \mathbf{V}$ where $\mathbf{D}$ is defined in (25) and $\boldsymbol{\Sigma}_{\mathrm{R}}$ is the return covariance matrix. Hence we have proved the following

Proposition 3.3 *The Fama-MacBeth covariance estimator $\widehat{\mathbf{V}}_T$ converges in probability to the matrix $\mathbf{V} = \mathbf{D}\boldsymbol{\Sigma}_R\mathbf{D}'$, that is,*

$$\widehat{\mathbf{V}}_T \xrightarrow{P} \mathbf{V} = \mathbf{D}\boldsymbol{\Sigma}_R\mathbf{D}', \ \textit{as } T \to \infty \tag{31}$$

where $\mathbf{D} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}$ with $\mathbf{X} = [\mathbf{1}_N \quad \mathbf{B}]$.

The preceding proposition is the mathematical statement representing the well-known errors-in-variables problem. On comparing the expression (28) for the asymptotic covariance $\mathbf{\Sigma}_c$ of the estimator $\bar{\bar{\mathbf{c}}}_T$ with the expression for $\mathbf{V}$, it follows that, in general, the Fama-MacBeth covariance estimator $\widehat{\mathbf{V}}_T$ is not a consistent estimator of $\mathbf{\Sigma}_c$. The bias introduced is due to the fact that estimated betas are used instead of the true betas in the cross-sectional regression. In the case of serially uncorrelated returns we have $\mathbf{\Psi} = \mathbf{\Sigma}_{\mathrm{R}}$ and thus $\mathbf{V} = \mathbf{D\Psi D}'$ is the first term appearing in the expression (28) for the asymptotic covariance of the estimator $\bar{\bar{\mathbf{c}}}_T$. The second term in (28) is clearly a positive semi-definite matrix. However, the presence of the last term in (28) complicates the situation and makes unclear whether the Fama-MacBeth procedure leads to underestimation or overestimation of the covariance of the CSR estimator, even in the case of serially uncorrelated returns. The situation becomes more straightforward under the simplifying assumption of conditionally homoscedasticity as we illustrate next.

3.4 Conditionally homoscedastic residuals given the factors

In this subsection, we look at the asymptotic behavior of the CSR estimator under the assumption that the time series residuals $\mathbf{u}_t$ are conditionally homoscedastic given the factors $\mathbf{f}_t$. The assumption of conditional homoscedasticity, which has been employed in the analysis by Shanken (1992) (see Assumption 1 in his paper), holds when the returns and the factors are serially independent, identically distributed and have a joint normal distribution. Some extra notation will enable us to state and prove the related results. Denote by $\mathbf{\Sigma}_{\bar{\mathrm{F}}}$ the asymptotic covariance matrix of $\bar{\mathbf{f}}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbf{f}_t$, that is

$$\mathbf{\Sigma}_{\bar{\mathrm{F}}} = \sum_{k=-\infty}^{\infty} E\left[(\mathbf{f}_t - E[\mathbf{f}_t])\,(\mathbf{f}_{t+k} - E[\mathbf{f}_{t+k}])'\right] \tag{32}$$

and further define the so-called bordered version of $\mathbf{\Sigma}_{\bar{\mathrm{F}}}$ by

$$\mathbf{\Sigma}^*_{\bar{\mathrm{F}}} = \begin{bmatrix} 0 & \mathbf{0}'_K \\ \mathbf{0}_K & \mathbf{\Sigma}_{\bar{\mathrm{F}}} \end{bmatrix}. \tag{33}$$

Let $\mathcal{F}$ denote the information set generated by the entire factor sequence $\{\mathbf{f}_t : t = 1, 2, \ldots\}$ and let $\mathbf{\Sigma}_{\mathrm{U}}$ be a constant $N \times N$ symmetric and positive definite matrix. Consider the following

Assumption B. Given the information set $\mathcal{F}$, the time series regression residuals $\mathbf{u}_t$ have zero conditional mean, i.e. $E[\mathbf{u}_t|\mathcal{F}] = \mathbf{0}_N$. Furthermore, given $\mathcal{F}$, the residuals $\mathbf{u}_t$ have constant conditional covariance equal to $\mathbf{\Sigma}_{\mathrm{U}}$ and are conditionally serially uncorrelated, i.e. $E[\mathbf{u}_t\mathbf{u}'_t|\mathcal{F}] = \mathbf{\Sigma}_{\mathrm{U}}$ and $E[\mathbf{u}_t\mathbf{u}'_{t+k}|\mathcal{F}] = \mathbf{0}_{N\times N}$ for all integers k.

The main result under the conditional homoscedasticity assumption is captured in the following theorem. A more general version of this theorem, directly imposing pricing restrictions to factors that are returns to traded portfolios, appeared as Theorem 1 (iii) in Shanken (1992).

Theorem 3.4 *Let* $\mathbf{c} = [a_0 \quad \boldsymbol{\lambda}']'$ *and* $\bar{\widehat{\mathbf{c}}}_T = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\overline{\mathbf{R}}_T$ *with* $\widehat{\mathbf{X}}_T = [\mathbf{1}_N \quad \widehat{\mathbf{B}}_T]$. *Assumption B implies* $\boldsymbol{\Psi} = \mathbf{B}\boldsymbol{\Sigma}_{\bar{F}}\mathbf{B}' + \boldsymbol{\Sigma}_U$, $\boldsymbol{\Pi} = \left(\boldsymbol{\lambda}'\boldsymbol{\Sigma}_F^{-1}\boldsymbol{\lambda}\right)\boldsymbol{\Sigma}_U$ *and* $\boldsymbol{\Gamma} = \mathbf{0}_{N\times N}$ *where* $\boldsymbol{\Psi}, \boldsymbol{\Pi}$ *and* $\boldsymbol{\Gamma}$ *are defined in (27). Therefore, under Assumptions A and B, the result in Theorem 3.2 becomes*

$$\sqrt{T}(\bar{\widehat{\mathbf{c}}}_T - \mathbf{c}) \xrightarrow{\mathcal{D}} N(\mathbf{0}_{K+1}, \boldsymbol{\Sigma}_c) \quad as \quad T \to \infty \tag{34}$$

where

$$\boldsymbol{\Sigma}_c = \boldsymbol{\Sigma}_{\bar{F}}^* + \left(1 + \boldsymbol{\lambda}'\boldsymbol{\Sigma}_F^{-1}\boldsymbol{\lambda}\right)\mathbf{D}\boldsymbol{\Sigma}_U\mathbf{D}' \tag{35}$$

with $\boldsymbol{\Sigma}_{\bar{F}}^*$ *being the bordered version of* $\boldsymbol{\Sigma}_{\bar{F}}$ *defined in (33) and* $\mathbf{D} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}$ *with* $\mathbf{X} = [\mathbf{1}_N \ \ \mathbf{B}]$.

Proof. Let k be any integer. First, we obtain the expression for $\boldsymbol{\Psi}$. Using (26), (13), the law of iterated expectations and Assumption A we obtain

$$E\left[\mathbf{h}_t^1(\mathbf{h}_{t+k}^1)'\right] = E\left[\left[\mathbf{B}(\mathbf{f}_t - E[\mathbf{f}_t]) + \mathbf{u}_t\right]\left[\mathbf{B}(\mathbf{f}_{t+k} - E[\mathbf{f}_{t+k}]) + \mathbf{u}_{t+k}\right]'\right]$$

$$= \mathbf{B}E\left[(\mathbf{f}_t - E[\mathbf{f}_t])(\mathbf{f}_{t+k} - E[\mathbf{f}_{t+k}])'\right]\mathbf{B}' + \mathbb{I}_{[k=0]}E\left[\mathbf{u}_t\mathbf{u}_t'\right]$$

where $\mathbb{I}$ denotes the indicator function. Therefore from (27) and (32) it follows that $\boldsymbol{\Psi} = \mathbf{B}\boldsymbol{\Sigma}_{\bar{\mathrm{F}}}\mathbf{B}' + \boldsymbol{\Sigma}_{\mathrm{U}}$. Next, we obtain the expression for $\boldsymbol{\Pi}$. From (26), the law of iterated expectations and Assumption A, it follows that

$$E[\mathbf{h}_t^2(\mathbf{h}_{t+k}^2)'] = E\left[\left[\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}(\mathbf{f}_t - E[\mathbf{f}_t])\right]\left[\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}(\mathbf{f}_{t+k} - E[\mathbf{f}_{t+k}])\right]\mathbf{u}_t\mathbf{u}_{t+k}'\right]$$

$$= E\left[\left[\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}(\mathbf{f}_t - E[\mathbf{f}_t])\right]\left[\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}(\mathbf{f}_{t+k} - E[\mathbf{f}_{t+k}])\right]E\left[\mathbf{u}_t\mathbf{u}_{t+k}'|\mathcal{F}\right]\right]$$

$$= \mathbb{I}_{[k=0]}E\left[\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}(\mathbf{f}_t - E[\mathbf{f}_t])(\mathbf{f}_t - E[\mathbf{f}_t])'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}\boldsymbol{\lambda}\right]\boldsymbol{\Sigma}_{\mathrm{U}}$$

$$= \mathbb{I}_{[k=0]}\left(\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}\boldsymbol{\Sigma}_{\mathrm{F}}\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}\boldsymbol{\lambda}\right)\boldsymbol{\Sigma}_{\mathrm{U}} = \mathbb{I}_{[k=0]}\left(\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}\boldsymbol{\lambda}\right)\boldsymbol{\Sigma}_{\mathrm{U}}.$$

The expression for $\boldsymbol{\Pi}$ then follows from (27). Finally, we obtain the expression for $\boldsymbol{\Gamma}$. From (26) and (13) we obtain

$$E\left[\mathbf{h}_t^1\mathbf{h}_{t+k}^2)'\right] = E\left[(\mathbf{R}_t - E[\mathbf{R}_t])\left[\mathbf{u}_{t+k}(\mathbf{f}_{t+k} - E[\mathbf{f}_{t+k}])'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}\boldsymbol{\lambda}\right]'\right]$$

$$= E\left[\left[\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}(\mathbf{f}_{t+k} - E[\mathbf{f}_{t+k}])\right]\left(\mathbf{B}(\mathbf{f}_t - E[\mathbf{f}_t]) + \mathbf{u}_t\right)\mathbf{u}_{t+k}'\right].$$

Thus, by the assumptions $E[\mathbf{u}_t|\mathcal{F}] = \mathbf{0}_N$ and $E[\mathbf{u}_t\mathbf{u}'_{t+k}|\mathcal{F}] = \mathbf{I}_{[k=0]}\mathbf{\Sigma}_{\mathrm{U}}$ and the law of iterated expectations, it follows that $E\left[\mathbf{h}_t^1(\mathbf{h}_{t+k}^2)'\right] = \mathbf{0}_{N\times N}$ for every integer k and therefore (27) yields $\mathbf{\Gamma} = \mathbf{0}_{N\times N}$. Using the expressions $\mathbf{\Psi} = \mathbf{B}\mathbf{\Sigma}_{\bar{\mathrm{F}}}\mathbf{B}' + \mathbf{\Sigma}_{\mathrm{U}}$, $\mathbf{\Pi} = \left(\boldsymbol{\lambda}'\mathbf{\Sigma}_{\mathrm{F}}^{-1}\boldsymbol{\lambda}\right)\mathbf{\Sigma}_{\mathrm{U}}$ and $\mathbf{\Gamma} = \mathbf{0}_{N\times N}$ and the fact that $\mathbf{D}\mathbf{B}\mathbf{\Sigma}_{\bar{\mathrm{F}}}\mathbf{B}'\mathbf{D}' = \mathbf{\Sigma}^*_{\bar{\mathrm{F}}}$, which we prove in the following lemma, we obtain from Theorem 3.2 that the asymptotic distribution of $\bar{\bar{\mathbf{c}}}_T$ is the one given in (34) under the assumptions A and B. ■

Lemma 3.5 *The identity* $\mathbf{D}\mathbf{B}\mathbf{\Sigma}_{\bar{F}}\mathbf{B}'\mathbf{D}' = \mathbf{\Sigma}^*_{\bar{F}}$ *holds, where* $\mathbf{\Sigma}^*_{\bar{F}}$ *is defined in (33) and* $\mathbf{D} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}$ *with* $\mathbf{X} = [\mathbf{1}_N \quad \mathbf{B}]$.

Proof. From the definition of $\mathbf{X}$ it follows that

$$\mathbf{X}'\mathbf{Q}\mathbf{X} = \begin{bmatrix} \mathbf{1}'_N \\ \mathbf{B}' \end{bmatrix} \mathbf{Q} \begin{bmatrix} \mathbf{1}_N & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'_N\mathbf{Q}\mathbf{1}_N & \mathbf{1}'_N\mathbf{Q}\mathbf{B} \\ \mathbf{B}'\mathbf{Q}\mathbf{1}_N & \mathbf{B}'\mathbf{Q}\mathbf{B} \end{bmatrix}.$$

Using the formula for the inverse of a partitioned matrix (see Theorem 7.1 in Schott (1997)) we obtain

$$(\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}$$
$$= \begin{bmatrix} d & -d\mathbf{1}'_N\mathbf{Q}\mathbf{B}(\mathbf{B}'\mathbf{Q}\mathbf{B})^{-1} \\ -d(\mathbf{B}'\mathbf{Q}\mathbf{B})^{-1}\mathbf{B}'\mathbf{Q}\mathbf{1}_N & (\mathbf{B}'\mathbf{Q}\mathbf{B})^{-1} + d(\mathbf{B}'\mathbf{Q}\mathbf{B})^{-1}\mathbf{B}'\mathbf{Q}\mathbf{1}_N\mathbf{1}'_N\mathbf{Q}\mathbf{B}(\mathbf{B}'\mathbf{Q}\mathbf{B})^{-1} \end{bmatrix}$$

where d is a scalar defined by $d = (\mathbf{1}'_N\mathbf{Q}\mathbf{1}_N - \mathbf{1}'_N\mathbf{Q}\mathbf{B}(\mathbf{B}'\mathbf{Q}\mathbf{B})^{-1}\mathbf{B}'\mathbf{Q}\mathbf{1}_N)^{-1}$. Using the above expression and (25) we have

$$\mathbf{D}\mathbf{B} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}\mathbf{B} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1} \begin{bmatrix} \mathbf{1}'_N\mathbf{Q}\mathbf{B} \\ \mathbf{B}'\mathbf{Q}\mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{0}'_K \\ \mathbf{I}_K \end{bmatrix}$$

from which, using the definition (33), we obtain

$$\mathbf{D}\mathbf{B}\mathbf{\Sigma}_{\bar{\mathrm{F}}}\mathbf{B}'\mathbf{D}' = \begin{bmatrix} \mathbf{0}'_K \\ \mathbf{I}_K \end{bmatrix} \mathbf{\Sigma}_{\bar{\mathrm{F}}} \begin{bmatrix} \mathbf{0}_K & \mathbf{I}_K \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{0}'_K \\ \mathbf{0}_K & \mathbf{\Sigma}_{\bar{\mathrm{F}}} \end{bmatrix} = \mathbf{\Sigma}^*_{\bar{\mathrm{F}}}$$

which completes the proof. ■

The preceding theorem and its consequences deserve some further discussion. Consider the commonly used case in which the factors are assumed to be serially uncorrelated, so that $\mathbf{\Sigma}_{\bar{\mathrm{F}}} = \mathbf{\Sigma}_{\mathrm{F}}$. If, in addition, Assumption B is satisfied, then it follows that the returns are also serially uncorrelated and following the lines of the preceding proof one can show that $\mathbf{\Sigma}_{\mathrm{R}} = \mathbf{B}\mathbf{\Sigma}_{\mathrm{F}}\mathbf{B}' + \mathbf{\Sigma}_{\mathrm{U}}$. In this case, the probability limit of the Fama-MacBeth covariance estimator is $\mathbf{V} = \mathbf{D}\mathbf{\Sigma}_{\mathrm{R}}\mathbf{D}' =$

$\boldsymbol{\Sigma}_{\tilde{\mathrm{F}}}^{*} + \mathbf{D}\boldsymbol{\Sigma}_{\mathrm{U}}\mathbf{D}'$. On comparing the true asymptotic covariance matrix $\boldsymbol{\Sigma}_c$ to $\mathbf{V}$, we observe that, since the matrix $\left(\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}\boldsymbol{\lambda}\right)\mathbf{D}\boldsymbol{\Sigma}_{\mathrm{U}}\mathbf{D}'$ is positive definite, the standard errors obtained from the Fama-MacBeth method always overstate the precision of the estimates, under the assumption of conditional homoscedasticity. However, this is generally not true when a time series is conditionally heteroscedastic.

As mentioned above, Assumption B is satisfied when the joint time series of returns and factors is i.i.d. and normally distributed. However, there is a large body of literature presenting evidence in favor of non-normality and heteroscedasticity. The early papers by Fama (1965) and Blattberg and Gonedes (1974) document non-normality while the papers by Barone-Adesi and Talwar (1983), Schwert and Sequin (1990) and Bollerslev et al. (1988) document conditional heteroscedasticity. Employing an i.i.d. sequence of returns following a multivariate t-distribution with more than four degrees of freedom, MacKinlay and Richardson (1991) demonstrate that returns are conditionally heteroscedastic and the test of mean-variance efficiency will be biased under the assumption of conditional homoscedasticity. They further demonstrate that stock returns are not homoscedastic based on a bootstrapping experiment. For these reasons, MacKinlay and Richardson (1991) advocate the GMM method developed by Hansen (1982), which does not require conditional homoscedasticity.

On the other hand, Assumption A may be satisfied by many stationary time series that are not conditionally homoscedastic. It follows from Lindeberg-Lèvy central limit theorem that any serially i.i.d. time series of returns and factors with finite fourth moments satisfies Assumption A, while it might not satisfy the assumption of conditional homoscedasticity unless the time series is also normally distributed. Clearly, the i.i.d. time series of returns with t-distribution in MacKinlay and Richardson (1991) is such an example.

3.5 Using security characteristics to test factor pricing models

The methodology developed in the two previous sections can be used to assess the validity of a proposed linear asset pricing factor model. The distribution theory presented in Theorems 3.2 and 3.4 allows one to construct t-statistics that have an asymptotic normal distribution. A significantly large t-value indicates that the corresponding factor is indeed priced while a small t-value suggests that the factor is not priced and should be excluded from the model. An alternative route that has been taken by several researchers is to use firm characteristics to detect misspecification errors. If a linear beta pricing is correctly specified, security characteristics added to the model should not explain the cross-sectional variation of expected returns after the factor betas prescribed by the model have been taken into account. In this case, the t-value of a characteristic reward should be insignificant. On the other hand, a significant t-value of a characteristic reward indicates model misspecification and should lead to rejection of the linear factor model. Common examples of firm characteristics used in the literature include relative firm size, the ratio of book value to market value and price-earnings ratio. Banz (1981) first used the firm size to examine the validity of the CAPM. Chan, Hamao and Lakonishok (1991) and Fama and French (1992) use the book-to-market ratio and provide evidence that this variable explains a larger fraction of the cross-sectional variation in expected returns. This evidence led Fama and French (1993) to propose a three-factor model for stock returns. Daniel and Titman (1997) add firm size and book-to-market ratio to the Fama-French three-factor model and find that the t-values associated with these firm characteristics are still significant.

Jagannathan and Wang (1998) extended the analysis of Shanken (1992) to allow for conditional heteroscedasticity. Further, they were the first to provide a rigorous econometric analysis of the cross-sectional regression method when firm characteristics are employed in addition to factors. Their framework assumed that the firm characteristics are constant over time. Let $\mathbf{Z}^i$ denote the vector of M characteristics associated with the ith security. Then the beta pricing model equation (4) augmented to include firm characteristics becomes

$$E[R_t^i] = a_0 + \mathbf{a}'\mathbf{Z}^i + \boldsymbol{\lambda}'\boldsymbol{\beta}_i, \; i = 1, \ldots, N \tag{36}$$

where $\mathbf{a}$ is an M-dimensional constant vector representing the characteristics rewards. Let $L = 1 + M + K$ and define

$$\mathbf{c} = [a_0 \;\; \mathbf{a}' \;\; \boldsymbol{\lambda}']' \quad (L \times 1 \text{ vector}), \tag{37}$$

$$\mathbf{Z} = \begin{bmatrix}\mathbf{Z}^1 & \mathbf{Z}^2 & \cdots & \mathbf{Z}^N\end{bmatrix}' \quad (N \times M \text{ matrix}) \tag{38}$$

and

$$\mathbf{X} = [\mathbf{1}_N \;\; \mathbf{Z} \;\; \mathbf{B}] \quad (N \times L \text{ matrix}) \tag{39}$$

where $\mathbf{B}$ is defined in (8). Then equation (36) can be compactly written as

$$E[\mathbf{R}_t] = a_0\mathbf{1}_N + \mathbf{Z}\mathbf{a} + \mathbf{B}\boldsymbol{\lambda} = \mathbf{X}\mathbf{c}. \tag{40}$$

As is typically the case in applications, we assume that $N \geq L$ and that $\mathbf{X}$ is of full rank equal to L. As in subsection 3.1, let $\mathbf{Q}$ be a symmetric and positive definite $N \times N$ matrix and $\widehat{\mathbf{Q}}_T$ be a consistent estimator of $\mathbf{Q}$ which is also assumed to be symmetric and positive definite for all T. Following the development in subsection 3.1, a GLS estimator of $\mathbf{c}$ is obtained by a cross-sectional regression at each time t as follows

$$\widehat{\mathbf{c}}_t = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\mathbf{R}_t \tag{41}$$

where

$$\widehat{\mathbf{X}}_T = [\mathbf{1}_N \;\; \mathbf{Z} \;\; \widehat{\mathbf{B}}_T]. \tag{42}$$

Here, $\widehat{\mathbf{B}}_T$ is the sample analogue estimator of $\mathbf{B}$ defined in (17) and the matrix $\widehat{\mathbf{X}}_T$ is assumed to be of full rank equal to L. As before, the Fama-MacBeth estimator is then obtained by time averaging the T cross-sectional estimates

$$\overline{\widehat{\mathbf{c}}}_T = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{c}}_t = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\overline{\mathbf{R}}_T. \tag{43}$$

3.5.1 Consistency and asymptotic normality of the CSR estimator

It turns out that the consistency and asymptotic normality properties of the Fama-MacBeth estimator are maintained when we include security characteristics in the analysis. Following the steps in the proof of Proposition 3.1 and using the relation (40), one can derive the analogous result for the case in which characteristics are used, which we state next.

Proposition 3.6 *The time series average $\overline{\widehat{\mathbf{c}}}_T$ of the cross-sectional estimates*

$$\widehat{\mathbf{c}}_t = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\mathbf{R}_t,\; t = 1,\ldots,T$$

where $\widehat{\mathbf{X}}_T = [\mathbf{1}_N \;\; \mathbf{Z} \;\; \widehat{\mathbf{B}}_T]$ is a consistent estimator of $\mathbf{c} = \left[a_0 \;\; \mathbf{a}' \;\; \boldsymbol{\lambda}'\right]'$, that is

$$\overline{\widehat{\mathbf{c}}}_T \xrightarrow{P} \mathbf{c} \text{ as } T \to \infty. \tag{44}$$

The asymptotic behavior of the CSR estimator, when security characteristics are used in the analysis, is captured in the next theorem. The proof of this theorem, which appeared as Theorem 1 in Jagannathan and Wang (1998) with slightly different notation, closely resembles the proof of Theorem 3.2.

Theorem 3.7 *Let $\mathbf{c} = [a_0 \;\; \mathbf{a}' \;\; \boldsymbol{\lambda}']'$ and $\overline{\widehat{\mathbf{c}}}_T = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\overline{\mathbf{R}}_T$ where $\widehat{\mathbf{X}}_T = [\mathbf{1}_N \;\; \mathbf{Z} \;\; \widehat{\mathbf{B}}_T]$. Under Assumption A, as $T \to \infty$, $\sqrt{T}(\overline{\widehat{\mathbf{c}}}_T - \mathbf{c})$ converges in distribution to a multivariate normal with zero mean and covariance*

$$\boldsymbol{\Sigma}_c = \mathbf{D}\boldsymbol{\Psi}\mathbf{D}' + \mathbf{D}\boldsymbol{\Pi}\mathbf{D}' - \mathbf{D}\left(\boldsymbol{\Gamma} + \boldsymbol{\Gamma}'\right)\mathbf{D}' \tag{45}$$

where $\mathbf{D} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}$ with $\mathbf{X} = [\mathbf{1}_N \;\; \mathbf{Z} \;\; \mathbf{B}]$ and $\boldsymbol{\Psi}, \boldsymbol{\Gamma}$ and $\boldsymbol{\Pi}$ are defined in (27).

A result, similar to Proposition 3.3, can be shown in the present context stating that the Fama-MacBeth covariance matrix estimator is not a consistent estimator of the asymptotic covariance matrix of the cross-sectional regression estimator.

3.5.2 Misspecification bias and protection against spurious factors

The main assumption of the preceding analysis was that the null hypothesis model is correctly specified. When the model is correctly specified, the CSR estimator is consistent under very general conditions that we stated above. However, if the null hypothesis model is misspecified, the estimator in cross-sectional regression will be asymptotically biased. Assume that $\tilde{\mathbf{f}}$ is a different vector of factors than $\mathbf{f}$, that is, at least some of the factors in $\tilde{\mathbf{f}}$ and $\mathbf{f}$ are different. Suppose the true factor vector is $\tilde{\mathbf{f}}$ and thus the true model is

$$E[\mathbf{R}_t] = a_0\mathbf{1}_{\mathrm{N}} + \mathbf{Za} + \widetilde{\mathbf{B}}\boldsymbol{\lambda} \tag{46}$$

where

$$\widetilde{\mathbf{B}} = E[(\mathbf{R}_t - E[\mathbf{R}_t])(\tilde{\mathbf{f}} - E[\tilde{\mathbf{f}}])']\widetilde{\mathbf{\Sigma}}_{\mathrm{F}}^{-1} \quad \text{and} \quad \widetilde{\mathbf{\Sigma}}_{\mathrm{F}} = E[(\tilde{\mathbf{f}}_t - E[\tilde{\mathbf{f}}_t])(\tilde{\mathbf{f}}_t - E[\tilde{\mathbf{f}}_t])']. \tag{47}$$

A researcher who incorrectly specifies the model as

$$E[\mathbf{R}_t] = a_0\mathbf{1}_N + \mathbf{Za} + \mathbf{B}\boldsymbol{\lambda}, \tag{48}$$

where

$$\mathbf{B} = E\left[(\mathbf{R}_t - E[\mathbf{R}_t])(\mathbf{f}_t - E[\mathbf{f}_t])'\right]\mathbf{\Sigma}_{\mathrm{F}}^{-1} \quad \text{and} \quad \mathbf{\Sigma}_{\mathrm{F}} = E\left[(\mathbf{f}_t - E[\mathbf{f}_t])\,(\mathbf{f}_t - E[\mathbf{f}_t])'\right], \tag{49}$$

will estimate betas by regressing returns on the vector of misspecified factors $\mathbf{f}$ and then estimate the risk premia in cross-sectional regression. The resulting bias is presented in the following theorem. We make the standard assumption that the time-series $[\mathbf{R}_t' \;\; \mathbf{f}_t']'$ is stationary and ergodic so that the law of large numbers applies and sample moments converge to population moments.

Theorem 3.8 *Assume that* $\mathbf{X} = [\mathbf{1}_N \;\; \mathbf{Z} \;\; \mathbf{B}]$ *has full rank equal to* L. *If equation (46) holds for the time series* $[\mathbf{R}_t' \;\; \tilde{\mathbf{f}}_t']'$ *but betas are estimated using the time-series* $[\mathbf{R}_t' \;\; \mathbf{f}_t']'$, *then the cross-sectional estimator* $\bar{\hat{\mathbf{c}}}_T = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\overline{\mathbf{R}}_T$ *with* $\widehat{\mathbf{X}}_T = [\mathbf{1}_N \;\; \mathbf{Z} \;\; \widehat{\mathbf{B}}_T]$ *converges to* $\mathbf{c} + (\mathbf{X}'\mathbf{QX})^{-1}\mathbf{X}'\mathbf{Q}(\widetilde{\mathbf{B}} - \mathbf{B})\boldsymbol{\lambda}$ *in probability as* $T \to \infty$.

Proof. Using equation (46) and appropriately rearranging the terms yields

$$\overline{\mathbf{R}}_T = (\overline{\mathbf{R}}_T - E[\mathbf{R}_t]) + \widehat{\mathbf{X}}_T\mathbf{c} - (\widehat{\mathbf{B}}_T - \mathbf{B})\boldsymbol{\lambda} + (\widetilde{\mathbf{B}} - \mathbf{B})\boldsymbol{\lambda}.$$

On multiplying the above equation by $(\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}'\widehat{\mathbf{Q}}_T$, one obtains the following expression for the cross-sectional regression estimator

$$\bar{\hat{\mathbf{c}}}_T = \mathbf{c} + (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}[(\overline{\mathbf{R}}_t - E[\mathbf{R}_t]) - (\widehat{\mathbf{B}}_T - \mathbf{B})\boldsymbol{\lambda}]$$

$$+(\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T(\widetilde{\mathbf{B}}-\mathbf{B})\boldsymbol{\lambda}.$$

Using the law of large numbers, the assumption that $\mathbf{X}$ has full rank and Slutsky's theorem, it follows from the last equation that

$$\bar{\widehat{\mathbf{c}}}_T \xrightarrow{P} \mathbf{c} + (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}(\widetilde{\mathbf{B}}-\mathbf{B})\boldsymbol{\lambda}.$$

This completes the proof. ■

In view of this theorem, the estimator in cross-sectional regression is asymptotically biased if and only if $\mathbf{X}'\mathbf{Q}(\widetilde{\mathbf{B}}-\mathbf{B})\boldsymbol{\lambda} \neq \mathbf{0}_L$. Notice that not only the estimates for the premium on the misspecified betas can be biased, but the estimates for the premium on those correctly specified betas can also be biased when some other factor is misspecified.

3.6 Time-varying security characteristics

Jagannathan, Skoulakis and Wang (2002) extended the analysis of Jagannathan and Wang (1998) to allow for time-varying firm characteristics. They study the case in which no pricing restrictions are imposed on the traded factors - if there are any such factors - and the case in which all factors are traded and pricing restrictions are imposed on all of them. Their main result is that using the observed time-varying characteristics in each cross-sectional regression induces a bias making the cross-sectional regression estimator generally inconsistent. They provide an expression for the bias and derive the asymptotic theory for the CSR estimator in both cases. They also show how one can avoid the bias problem by using time-averages of the firm characteristics. In the next three subsections, we state some of the more important results while we refer the interested reader to Jagannathan, Skoulakis and Wang (2002) for the full analysis including proofs and details.

3.6.1 No pricing restrictions imposed on traded factors

In this subsection, we proceed without directly imposing any pricing restrictions on traded factors, if any such factors are employed in the analysis. In other words, we do not distinguish between traded and nontraded factors. On the other hand, we allow for time-varying firm-specific characteristics. Let $\mathbf{Z}_t^i$ be a vector of M characteristics associated with the ith asset observed at time $t-1$. The factor pricing equation (4) is expanded to include the firm characteristics as follows

$$E[R_t^i] = a_0 + \mathbf{a}'E[\mathbf{Z}_t^i] + \boldsymbol{\lambda}'\boldsymbol{\beta}_i \quad \text{for} \quad i = 1, \ldots, N \tag{50}$$

where $\mathbf{a}$ is the M-dimensional constant vector of characteristics rewards. As in subsection 3.5, we let $L = 1 + M + K$, $\mathbf{c} = [a_0 \quad \mathbf{a}' \quad \boldsymbol{\lambda}']'$ $\quad (L \times 1$ vector) and define the time-varying characteristics matrix

$$\mathbf{Z}_t = \begin{bmatrix}\mathbf{Z}_t^1 & \mathbf{Z}_t^2 & \cdots & \mathbf{Z}_t^N\end{bmatrix}' \quad (N \times M \text{ matrix}) \tag{51}$$

and

$$\mathbf{X}_t = [\mathbf{1}_N \quad \mathbf{Z}_t \quad \mathbf{B}] \quad (N \times L \text{ matrix}) \tag{52}$$

where $\mathbf{B}$ is defined in (8). Then equation (50) can be written as

$$E[\mathbf{R}_t] = a_0\mathbf{1}_N + E[\mathbf{Z}_t]\mathbf{a} + \mathbf{B}\boldsymbol{\lambda} = E[\mathbf{X}_t]\mathbf{c}. \tag{53}$$

Again assume that $N \geq L$ and that $\mathbf{X}_t$ is of full rank equal to L for all t. Following the development in subsection 3.1, a GLS estimate of $\mathbf{c}$ is obtained by a cross-sectional regression at each time t as follows

$$\widehat{\mathbf{c}}_t = (\widehat{\mathbf{X}}'_{T,t}\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_{T,t})^{-1}\widehat{\mathbf{X}}'_{T,t}\widehat{\mathbf{Q}}_T\mathbf{R}_t \tag{54}$$

where

$$\widehat{\mathbf{X}}_{T,t} = [\mathbf{1}_N \quad \mathbf{Z}_t \quad \widehat{\mathbf{B}}_T] \tag{55}$$

and $\widehat{\mathbf{B}}_T$ is the sample analogue estimate of $\mathbf{B}$ defined in (17). The matrix $\widehat{\mathbf{X}}_{T,t}$ is assumed to be of full rank equal to L for all t. As before, the time average of the T cross-sectional estimates provides the Fama-MacBeth CSR estimate

$$\overline{\widehat{\mathbf{c}}}_T = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{c}}_t = \frac{1}{T}\sum_{t=1}^{T}(\widehat{\mathbf{X}}'_{T,t}\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_{T,t})^{-1}\widehat{\mathbf{X}}'_{T,t}\widehat{\mathbf{Q}}_T\mathbf{R}_t. \tag{56}$$

Our goal is to obtain the probability limit and the asymptotic distribution of the CSR estimator in the case of time-varying firm characteristics. We make the standard assumption that the vector process $\left[(\mathbf{R}_t^e)' \;\; (\mathbf{f}_t^e)' \;\; (\mathrm{vec}(\mathbf{Z}_t))'\right]'$ is stationary and ergodic so that the law of large numbers applies. The derivation of the results that follow requires a few mild technical assumptions, the first of which is stated next.

Assumption C. Consider an arbitrarily small $\delta > 0$ and assume that, for all t, the smallest eigenvalue of the matrix $\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t$ is greater than δ, where $\mathbf{X}_t = [\mathbf{1}_N \quad \mathbf{Z}_t \quad \mathbf{B}]$. In addition, assume that all elements of the characteristics matrix $\mathbf{Z}_t$ are bounded uniformly in t.

Under the usual assumption that, for a given t, the matrix $\mathbf{X}_t$ is of full rank, namely L, we have that the $L \times L$ matrix $\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t$ also has rank L and is positive definite since $\mathbf{Q}$ is positive definite. In this case, the smallest eigenvalue of $\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t$ will be strictly positive. Assumption C requires a slightly stronger condition, namely that the smallest eigenvalue of $\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t$ is outside a fixed small neighborhood of 0 for all t. On the other hand, it turns out that if the OLS estimator of $\mathbf{c}$ is used, instead of the feasible GLS estimator, the analysis still goes through without the boundedness assumption on $\mathbf{Z}_t$.

The derivation of the results stated in this subsection relies heavily on the following lemma, the proof of which is based on Assumption C. We state the lemma in order to give the reader an idea of how the proofs proceed in the present setting of time-varying firm characteristics.

Lemma 3.9 *Under Assumption C, the matrix random sequence* $\mathbf{A}_{T,t} = (\widehat{\mathbf{X}}_{T,t}'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_{T,t})^{-1}\widehat{\mathbf{X}}_{T,t}'\widehat{\mathbf{Q}}_T - (\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Q}$ *converges in probability to* $\mathbf{0}_L$ *as* $T \to \infty$ *uniformly in* t.

An important result about the CSR estimator in the present setting is that the use of time-varying firm characteristics in the fashion described by equations (54) and (55) produces an estimator which is not necessarily a consistent estimator of the unknown parameter $\mathbf{c}$. The following proposition describes the limiting behavior of the CSR estimator and its asymptotic bias, as $T \to \infty$.

Proposition 3.10 *Let* $\mathbf{c} = [a_0 \quad \mathbf{a}' \quad \boldsymbol{\lambda}']'$. *Under assumption C, the probability limit of the cross-sectional regression estimator* $\bar{\widehat{\mathbf{c}}}_T = \frac{1}{T}\sum_{t=1}^{T}(\widehat{\mathbf{X}}_{T,t}'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_{T,t})^{-1}\widehat{\mathbf{X}}_{T,t}'\widehat{\mathbf{Q}}_T\mathbf{R}_t$ *is given by*

$$\bar{\widehat{\mathbf{c}}}_T \xrightarrow{P} \mathbf{c} + \boldsymbol{\gamma}, \quad \text{as} \;\; T \to \infty$$

where the asymptotic bias is given by

$$\boldsymbol{\gamma} = E[(\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Q}\,(\mathbf{R}_t - \mathbf{X}_t\mathbf{c})] \tag{57}$$

assuming that the expectation that defines $\boldsymbol{\gamma}$ *exists and is finite. Under the null hypothesis* H_0 : $\mathbf{a} = \mathbf{0}_M$ *the bias is given by* $\boldsymbol{\gamma} = E[(\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Q}\,(\mathbf{R}_t - E[\mathbf{R}_t])]$.

¿From Proposition 3.10 one can see that the cross-sectional estimator $\bar{\widehat{\mathbf{c}}}_T$ is an asymptotically biased estimator of $\mathbf{c}$ unless $\boldsymbol{\gamma} = \mathbf{0}_L$. Since the firm characteristics will presumably be correlated with the returns, it follows from the expression $\boldsymbol{\gamma} = E[(\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Q}\,(\mathbf{R}_t - \mathbf{X}_t\mathbf{c})]$ that, in principle, the bias $\boldsymbol{\gamma}$ will not be equal to the zero vector. Therefore, ignoring this potential bias of the CSR estimator might lead to erroneous inferences. However, when the firm characteristics $\mathbf{Z}_t$ used in the study are constant over time, say equal to $\mathbf{Z}$, then $\mathbf{X}_t = \mathbf{X} \equiv [\mathbf{1}_N \quad \mathbf{Z} \quad \mathbf{B}]$ and so (53) implies that $\boldsymbol{\gamma} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}E[\mathbf{R}_t - \mathbf{X}\mathbf{c}] = \mathbf{0}_L$, which is equivalent to $\bar{\widehat{\mathbf{c}}}_T$ being a consistent estimator of $\mathbf{c}$. Thus we obtain Proposition 3.6 as a corollary to Proposition 3.10.

Next, we study the asymptotic distribution of the CSR estimator. Two additional assumptions will ensure the validity of the theorem that follows. To precisely state the necessary assumptions, we need an additional piece of notation. Define

$$\widetilde{\mathbf{h}}_t^1 = (\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Q}(\mathbf{R}_t - \mathbf{X}_t\mathbf{c}) - \boldsymbol{\gamma},\ \widetilde{\mathbf{h}}_t^2 = \left[(\mathbf{f}_t - E[\mathbf{f}_t])'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}\boldsymbol{\lambda}\right]\mathbf{u}_t,$$
$$\text{and}\ \ \widetilde{\mathbf{h}}_t = [(\widetilde{\mathbf{h}}_t^1)' \ \ (\widetilde{\mathbf{h}}_t^2)']'. \tag{58}$$

Note that $E[\widetilde{\mathbf{h}}_t^1] = \mathbf{0}_N$ and $E[\widetilde{\mathbf{h}}_t^2] = \mathbf{0}_N$ as it follows from (57) and (13) respectively. The first of the following assumptions is the analogue of Assumption A used in Subsection 3.2, appropriately modified in the context of time-varying characteristics.

Assumption D.1. The central limit theorem applies to the random sequence $\widetilde{\mathbf{h}}_t$ defined in (58), that is $\frac{1}{\sqrt{T}}\sum_{t=1}^T \widetilde{\mathbf{h}}_t$ converges in distribution to a multivariate normal with zero mean and covariance matrix given by

$$\begin{bmatrix} \widetilde{\boldsymbol{\Psi}} & \widetilde{\boldsymbol{\Gamma}} \\ \widetilde{\boldsymbol{\Gamma}}' & \widetilde{\boldsymbol{\Pi}} \end{bmatrix}$$

where

$$\widetilde{\boldsymbol{\Psi}} = \sum_{k=-\infty}^{+\infty} E\left[\widetilde{\mathbf{h}}_t^1(\widetilde{\mathbf{h}}_{t+k}^1)'\right],\ \widetilde{\boldsymbol{\Gamma}} = \sum_{k=-\infty}^{+\infty} E\left[\widetilde{\mathbf{h}}_t^1(\widetilde{\mathbf{h}}_{t+k}^2)'\right]\ \text{and}\ \widetilde{\boldsymbol{\Pi}} = \sum_{k=-\infty}^{+\infty} E\left[\widetilde{\mathbf{h}}_t^2(\widetilde{\mathbf{h}}_{t+k}^2)'\right]. \tag{59}$$

We also make the following

Assumption D.2. The random sequence $\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbf{A}_{T,t}\mathbf{Y}_t$ converges to $\mathbf{0}_L$ in probability, where $\mathbf{A}_{T,t} = (\widehat{\mathbf{X}}_{T,t}'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_{T,t})^{-1}\widehat{\mathbf{X}}_{T,t}'\widehat{\mathbf{Q}}_T - (\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Q}$ and $\mathbf{Y}_t = \mathbf{R}_t - \mathbf{X}_t\mathbf{c}$.

Although it might seem redundant given the rest of the assumptions we have made, Assumption D.2 is actually essential for the proof of the next theorem. In the case of constant firm characteristics, it follows from Lemma 3.9 that Assumption D.2 holds since then $\mathbf{A}_{T,t}$ does not depend on t and it factors out. It is also easily verified that Assumption D.2 holds in the case that $\mathbf{Z}_t$ takes only a finite number of values and is stochastically independent of $\mathbf{R}_t - \mathbf{X}_t\mathbf{c}$.

The next theorem provides the asymptotic distribution of the CSR estimator.

Theorem 3.11 *Let $\mathbf{c} = [a_0 \ \ \mathbf{a}' \ \ \boldsymbol{\lambda}']'$ and $\widetilde{\widehat{\mathbf{c}}}_T = \frac{1}{T}\sum_{t=1}^T(\widehat{\mathbf{X}}_{T,t}'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_{T,t})^{-1}\widehat{\mathbf{X}}_{T,t}'\widehat{\mathbf{Q}}_T\mathbf{R}_t$ where $\widehat{\mathbf{X}}_{T,t} = [\mathbf{1}_N \ \ \mathbf{Z}_t \ \ \widehat{\mathbf{B}}_T]$. Under Assumptions C, D.1 and D.2, as $T \to \infty$, $\sqrt{T}(\widetilde{\widehat{\mathbf{c}}}_T - [\mathbf{c} + \boldsymbol{\gamma}])$ converges in distribution to a multivariate normal with zero mean and covariance matrix*

$$\widetilde{\boldsymbol{\Sigma}} = \widetilde{\boldsymbol{\Psi}} + \widetilde{\mathbf{D}}\widetilde{\boldsymbol{\Pi}}\widetilde{\mathbf{D}}' - \left(\widetilde{\boldsymbol{\Gamma}}\widetilde{\mathbf{D}}' + \widetilde{\mathbf{D}}\widetilde{\boldsymbol{\Gamma}}'\right)$$

where $\widetilde{\mathbf{D}} = E\left[(\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Q}\right]$ with $\mathbf{X}_t = [\mathbf{1}_N \ \ \mathbf{Z}_t \ \ \mathbf{B}]$ and $\widetilde{\boldsymbol{\Psi}}, \widetilde{\boldsymbol{\Gamma}}$ and $\widetilde{\boldsymbol{\Pi}}$ are defined by (59).

3.6.2 Traded factors with imposed pricing restrictions

In this subsection, we study the behavior of the cross-sectional regression two-pass estimator when all factors are returns on traded portfolios and the relevant pricing restrictions are imposed on the factors. In particular, we examine the approach that Brennan, Chordia and Subrahmanyam (1988) proposed to handle the errors-in-variables problem without the requirement of grouping securities into portfolios. As in subsection 3.5, we assume that there are N risky securities, K economy-wide factors, M security-specific characteristics with the addition of a riskless asset. The expected excess return on the jth asset can be written as

$$E[R_j - R_f] = a_0 + + \sum_{m=1}^{M} E[Z_m^j] a_m + \sum_{k=1}^{K} \beta_{jk}\lambda_k,\ \ j = 1, \ldots, N \tag{60}$$

where λ_k is the risk premium of factor k, β_{jk} is the factor loading of factor k for the jth security, Z_m^j is the value of the mth characteristic specific to the jth security, and a_m is the reward or premium per unit of the mth characteristic. Examples of factors that are returns on portfolios of traded securities include the first five principal components of Connor-Korajckyk (1988) and the three factors of Fama-French (1993). Examples of security characteristics used include relative firm size, relative book-to-market ratio, dividend yield, relative strength and turnover. To test the validity of the factor pricing model one needs to construct a test of the null hypothesis $H_0 : a_0 = a_1 = \cdots = a_M = 0$.

The notation of the previous subsections is employed except that we now denote $\mathbf{c} = [a_0 \ \ \mathbf{a}']'$ where $\mathbf{a} = (a_1, \ldots, a_M)'$ is the vector of characteristics rewards. We further use superscript e to denote excess returns on the assets and the factors: $\mathbf{R}_t^e = \mathbf{R}_t - R_{ft}\mathbf{1}_N$ and $\mathbf{f}_t^e = \mathbf{f}_t - R_{ft}\mathbf{1}_K$. Then we can define the factor loading matrix as

$$\mathbf{B}^e = E\left[\left(\mathbf{R}_t^e - E[\mathbf{R}_t^e]\right)\left(\mathbf{f}_t^e - E[\mathbf{f}_t^e]\right)'\right] \boldsymbol{\Sigma}_{\mathrm{F}}^{e\,-1} \tag{61}$$

where the covariance matrix of the factor excess returns is given

$$\boldsymbol{\Sigma}_{\mathrm{F}}^e = E[(\mathbf{f}_t^e - E[\mathbf{f}_t^e])(\mathbf{f}_t^e - E[\mathbf{f}_t^e])']. \tag{62}$$

Similarly to (13), but now using excess returns, we let $\mathbf{u}_t^e = \mathbf{R}_t^e - E[\mathbf{R}_t^e] - \mathbf{B}^e(\mathbf{f}_t^e - E[\mathbf{f}_t^e])$ to obtain the time-series regression

$$\mathbf{R}_t^e = E[\mathbf{R}_t^e] + \mathbf{B}(\mathbf{f}_t^e - E[\mathbf{f}_t^e]) + \mathbf{u}_t^e \ \ \text{with} \ \ E[\mathbf{u}_t^e] = \mathbf{0}_N \ \ \text{and} \ \ E[\mathbf{u}_t^e \mathbf{f}_t^{e\prime}] = \mathbf{0}_{N\times K}. \tag{63}$$

Therefore, the model can be written in vector-matrix notation form as $E[\mathbf{R}_t^e] = a_0\mathbf{1}_N + E[\mathbf{Z}_t]\mathbf{a} + \mathbf{B}^e\boldsymbol{\lambda}$ which allows to rewrite equation (63) as

$$\mathbf{R}_t^e = a_0\mathbf{1}_N + E[\mathbf{Z}_t]\mathbf{a} + \mathbf{B}^e(\mathbf{f}_t^e - E[\mathbf{f}_t^e] + \boldsymbol{\lambda}) + \mathbf{u}_t^e.$$

Since the factors are assumed to be returns on traded assets, it follows that the factor risk premia equal expected excess returns, that is $\boldsymbol{\lambda} = E[\mathbf{f}_t^e]$, and so the last equation becomes

$$\mathbf{R}_t^e = a_0 \mathbf{1}_N + \mathbf{B}^e \mathbf{f}_t^e + \mathbf{Z}_t \mathbf{a} + \mathbf{e}_t \quad \text{where} \quad \mathbf{e}_t = -(\mathbf{Z}_t - E[\mathbf{Z}_t])\mathbf{a} + \mathbf{u}_t^e. \tag{64}$$

Following Brennan et al (1998) we define the risk-adjusted returns as

$$\mathbf{R}_t^* = \mathbf{R}_t^e - \mathbf{B}^e \mathbf{f}_t^e \tag{65}$$

and write the previous equation as

$$\mathbf{R}_t^* = \mathbf{X}_t \mathbf{c} + \mathbf{e}_t \quad \text{where} \quad \mathbf{X}_t = [\mathbf{1}_N \quad \mathbf{Z}_t] \quad \text{and} \quad \mathbf{c} = [a_0 \quad \mathbf{a}']'. \tag{66}$$

The estimation and testing procedure proposed in Brennan et al (1998) consists of the following steps. First, a set of factors is selected. Two sets of factors are used, the principal component factors of Connor-Korajczyk (1988) and the Fama-French (1993) factors. Second, the factor betas are estimated using standard time-series regressions as follows

$$\widehat{\mathbf{B}}_T^e = \left[\frac{1}{T} \sum_{t=1}^{T} (\mathbf{R}_t^e - \overline{\mathbf{R}_T^e})(\mathbf{f}_t^e - \overline{\mathbf{f}_T^e})' \right] \widehat{\boldsymbol{\Sigma}_{\mathrm{F}T}^{e}}^{-1} \quad \text{where} \quad \widehat{\boldsymbol{\Sigma}_{\mathrm{F}T}^{e}} = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{f}_t^e - \overline{\mathbf{f}_T^e})(\mathbf{f}_t^e - \overline{\mathbf{f}_T^e})'. \tag{67}$$

Third, using the estimates of the factor betas from the second step the estimated risk-adjusted returns are formed:

$$\widehat{\mathbf{R}}_t^* = \mathbf{R}_t^e - \widehat{\mathbf{B}}_T^e \mathbf{f}_t^e. \tag{68}$$

Then the risk-adjusted returns are used in the following cross-sectional regressions

$$\widehat{R}_{jt}^* = c_0 + \sum_{m=1}^{M} c_m Z_{m,t}^j + \tilde{e}_{jt}, \ j = 1, \ldots, N \tag{69}$$

or in vector-matrix notation $\widehat{\mathbf{R}}_t^* = \mathbf{X}_t \mathbf{c} + \tilde{\mathbf{e}}_t$ for all $t = 1, \ldots, T$. Thus, in the fourth step, for each t, cross-sectional simple regression is used to obtain the following estimates $\widehat{\mathbf{c}}_t$ of the vector $\mathbf{c}$ of characteristic rewards

$$\widehat{\mathbf{c}}_t = (\mathbf{X}_t' \mathbf{X}_t)^{-1} \mathbf{X}_t' \widehat{\mathbf{R}}_t^*, \ t = 1, \ldots, T. \tag{70}$$

Finally, in the spirit of Fama and MacBeth, the time-series average of the estimates is formed to obtain what Brennan et al (1998) term the raw estimate of $\mathbf{c}$ as follows

$$\overline{\widehat{\mathbf{c}}}_T = \frac{1}{T} \sum_{t=1}^{T} \widehat{\mathbf{c}}_t. \tag{71}$$

Alternatively, the use of GLS would allow us obtain more efficient estimates. Employing the previously used notation, we can write the GLS raw estimate of $\mathbf{c}$ as

$$\bar{\widehat{\mathbf{c}}}_T = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{X}_t'\widehat{\mathbf{Q}}_T\mathbf{X}_t)^{-1}\mathbf{X}_t'\widehat{\mathbf{Q}}_T\widehat{\mathbf{R}}_t^*. \tag{72}$$

Next, we describe the probability limit and the asymptotic distribution of the GLS version of the estimator proposed by Brennan et al (1998). We make the standard assumption that the vector process $\left[(\mathbf{R}_t^e)' \;\; (\mathbf{f}_t^e)' \;\; (\mathrm{vec}(\mathbf{Z}_t))'\right]'$ is stationary and ergodic so that the law of large numbers applies. As before, some additional assumptions will ensure the validity of the results. The first assumption, which we state next, is the suitable modification of Assumption C that was made in subsection 3.6.1.

Assumption E. Consider an arbitrarily small $\delta > 0$ and assume that, for all t, the smallest eigenvalue of the matrix $\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t$ is greater than δ, where $\mathbf{X}_t = [\mathbf{1}_N \;\; \mathbf{Z}_t]$. In addition, assume that all elements of the characteristics matrix $\mathbf{Z}_t$ are bounded uniformly in t.

It turns out that the cross-sectional regression estimator $\bar{\widehat{\mathbf{c}}}_T$, obtained by regressing the estimated risk-adjusted returns on a constant and the time-varying firm characteristics as described by equation(69), is not a consistent estimator of the parameter $\mathbf{c}$. This is a property we encountered in Proposition 3.10 under a different setting. The next proposition presents the asymptotic bias of $\bar{\widehat{\mathbf{c}}}_T$.

Proposition 3.12 *Let* $\mathbf{c} = [a_0 \;\; \mathbf{a}']'$. *Under Assumption E, the probability limit of the cross-sectional regression raw estimate* $\bar{\widehat{\mathbf{c}}}_T = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{X}_t'\widehat{\mathbf{Q}}_T\mathbf{X}_t)^{-1}\mathbf{X}_t'\widehat{\mathbf{Q}}_T\widehat{\mathbf{R}}_t^*$ *is given by*

$$\bar{\widehat{\mathbf{c}}}_T \xrightarrow{P} \mathbf{c} + \boldsymbol{\gamma}^*, \quad \text{as } T \to \infty$$

where the asymptotic bias is given by

$$\boldsymbol{\gamma}^* = E[(\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Q}\mathbf{e}_t] \tag{73}$$

assuming that the expectation that defines $\boldsymbol{\gamma}^*$ *exists and is finite. Under the null hypothesis* H_0 : $\mathbf{c} = \mathbf{0}$ *the bias is given by* $\boldsymbol{\gamma}^* = E[(\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Q}\mathbf{u}_t]$.

The description of the asymptotic distribution of the raw estimate now follows. This will require an additional assumption. To this end, we define

$$\mathbf{g}_t^1 = (\mathbf{X}_t'\mathbf{Q}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Q}\mathbf{e}_t - \boldsymbol{\gamma}^*,$$

$$\mathbf{g}_t^2 = \begin{bmatrix} \mathrm{vec}\Big((\mathbf{f}_t^e - \overline{\mathbf{f}_T^e})(\mathbf{f}_t^e - \overline{\mathbf{f}_T^e})' - E\left[(\mathbf{f}_t^e - \mathrm{E}[\mathbf{f}_t^e])(\mathbf{f}_t^e - E[\mathbf{f}_t^e])'\right]\Big) \\ \mathrm{vec}\Big((\mathbf{R}_t^e - \overline{\mathbf{R}_T^e})(\mathbf{f}_t^e - \overline{\mathbf{f}_T^e})' - E\left[(\mathbf{R}_t^e - \mathrm{E}[\mathbf{R}_t^e])(\mathbf{f}_t^e - \mathrm{E}[\mathbf{f}_t^e])'\right]\Big) \end{bmatrix}, \tag{74}$$

and $\mathbf{g}_t = [(\mathbf{g}_t^1)' \;\; (\mathbf{g}_t^2)']'$.

Note that clearly $E[\mathbf{g}_t^2] = \mathbf{0}_{K(N+K)}$ and that (73) implies $E[\mathbf{g}_t^1] = \mathbf{0}_{M+1}$. The following mild assumption is similar to Assumptions A and D.1 made earlier.

Assumption F. The central limit theorem applies to the random sequence $\mathbf{g}_t$, that is $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{g}_t$ converges in distribution to a multivariate normal with zero mean and covariance matrix given by

$$\begin{bmatrix} \mathbf{\Psi}^* & \mathbf{\Gamma}^* \\ \mathbf{\Gamma}^{*\prime} & \mathbf{\Pi}^* \end{bmatrix}$$

where

$$\mathbf{\Psi}^* = \sum_{k=-\infty}^{+\infty} E\left[\mathbf{g}_t^1(\mathbf{g}_{t+k}^1)'\right], \; \mathbf{\Gamma}^* = \sum_{k=-\infty}^{+\infty} E\left[\mathbf{g}_t^1(\mathbf{g}_{t+k}^2)'\right] \text{ and } \mathbf{\Pi}^* = \sum_{k=-\infty}^{+\infty} E\left[\mathbf{g}_t^2(\mathbf{g}_{t+k}^2)'\right]. \tag{75}$$

The following notation will also be used in the statement of the next theorem.

$$\begin{aligned} \mathbf{H}^* &= E\left[\left[(\mathbf{f}_t^e)'(\mathbf{\Sigma}_{\mathrm{F}}^e)^{-1}\right] \otimes \left[(\mathbf{Z}_t'\mathbf{Q}\mathbf{Z}_t)^{-1}\mathbf{Z}_t'\mathbf{Q}\mathbf{B}\right]\right], \\ \mathbf{K}^* &= \mathrm{E}\left[\left[(\mathbf{f}_t^e)'(\mathbf{\Sigma}_{\mathrm{F}}^e)^{-1}\right] \otimes \left[(\mathbf{Z}_t'\mathbf{Q}\mathbf{Z}_t)^{-1}\mathbf{Z}_t'\mathbf{Q}\right]\right] \\ \text{and} \;\; \mathbf{D}^* &= [-\mathbf{H}^* \;\; \mathbf{K}^*]. \end{aligned} \tag{76}$$

We are now able to state the result providing the asymptotic distribution.

Theorem 3.13 *Let* $\mathbf{c} = [a_0 \;\; \mathbf{a}']'$ *and* $\bar{\bar{\mathbf{c}}}_T = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{X}_t'\widehat{\mathbf{Q}}_T\mathbf{X}_t)^{-1}\mathbf{X}_t'\widehat{\mathbf{Q}}_T\widehat{\mathbf{R}}_t^*$ *where* $\mathbf{X}_t = [\mathbf{1}_N \;\; \mathbf{Z}_t]$. *Under Assumptions E and F, as* $T \to \infty$, $\sqrt{T}(\bar{\bar{\mathbf{c}}}_T - [\mathbf{c} + \boldsymbol{\gamma}^*])$ *converges in distribution to a multivariate normal with zero mean and covariance matrix*

$$\mathbf{\Sigma}^* = \mathbf{\Psi}^* + \mathbf{D}^*\mathbf{\Pi}^*\mathbf{D}^{*\prime} - (\mathbf{\Gamma}^*\mathbf{D}^{*\prime} + \mathbf{D}^*\mathbf{\Gamma}^{*\prime}) \tag{77}$$

where $\mathbf{\Psi}^*, \mathbf{\Gamma}^*$ *and* $\mathbf{\Pi}^*$ *are defined in (75) and* $\mathbf{D}^*$ *is defined in (76).*

Even if one ignores the asymptotic bias $\boldsymbol{\gamma}^*$ of the raw estimator $\bar{\bar{\mathbf{c}}}_T$, the method proposed by Brennan et al (1998) does not solve the errors-in-variables problem. That is, the Fama-MacBeth covariance estimator $\widehat{\mathbf{V}}_T = \frac{1}{T}\sum_{t=1}^{T}(\widehat{\mathbf{c}}_t - \bar{\bar{\mathbf{c}}}_T)(\widehat{\mathbf{c}}_t - \bar{\bar{\mathbf{c}}}_T)'$ is not generally a consistent estimator of the asymptotic covariance of the CSR raw estimate. The following proposition addresses this point.

Proposition 3.14 *The probability limit of the Fama-MacBeth covariance estimator* $\widehat{\mathbf{V}}_T$ *is the following covariance matrix*

$$\mathbf{V}^* = E\left[\left((\mathbf{Z}_t'\mathbf{Q}\mathbf{Z}_t)^{-1}\mathbf{Z}_t'\mathbf{Q}\mathbf{e}_t - \boldsymbol{\gamma}^*\right)\left((\mathbf{Z}_t'\mathbf{Q}\mathbf{Z}_t)^{-1}\mathbf{Z}_t'\mathbf{Q}\mathbf{e}_t - \boldsymbol{\gamma}^*\right)'\right] \tag{78}$$

assuming that the expectation that defines $\mathbf{V}^*$ *exists and is finite and where, as in (73),* $\boldsymbol{\gamma}^* = E\left[(\mathbf{Z}_t'\mathbf{Q}\mathbf{Z}_t)^{-1}\mathbf{Z}_t'\mathbf{Q}\mathbf{e}_t\right]$.

Note that, when the observations $\left[(\mathbf{R}_t^e)' \quad (\mathbf{f}_t^e)' \quad (\mathrm{vec}(\mathbf{Z}_t))'\right]'$ are serially independent we have $\mathbf{\Psi}^* = \mathrm{E}\left[\mathbf{g}_t^1(\mathbf{g}_t^1)'\right] = \mathbf{V}^*$, as it follows from (74) and (78). Thus, in the case of serially independent observations the Fama-MacBeth covariance converges in probability to $\mathbf{\Psi}^*$ which is the first term in the expression (77) of the asymptotic covariance of the CSR estimate. The second term in (78) is a positive semidefinite matrix but the last term in (78) is neither positive nor negative semidefinite with certainty. Thus, as already noted in the subsection 3.3, it is not obvious whether the Fama-MacBeth estimator underestimates or overestimates the asymptotic covariance of $\overline{\widehat{\mathbf{c}}}_T$ in the absence of additional assumptions even in the case of serially independent observations.

3.6.3 Using time-average characteristics to avoid the bias

In this section we reconsider the framework of subsection 3.6.1 and demonstrate how we can overcome the problem of the asymptotic bias by using time-average firm characteristics instead of the spot values of characteristics in the cross-sectional regression. We adopt the notation of subsection 3.6.1, namely, we denote $\mathbf{X}_t = [\mathbf{1}_N \quad \mathbf{Z}_t \quad \mathbf{B}]$ where $\mathbf{Z}_t$ is the $N \times M$ matrix of t-time characteristics and $\mathbf{c} = [a_0 \quad \mathbf{a}' \quad \boldsymbol{\lambda}']'$. Cross-sectional regression using the time-average $\overline{\mathbf{Z}}_T$ of the firm characteristics yields the following estimates of $\mathbf{c}$

$$\widehat{\mathbf{c}}_t = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\mathbf{R}_t, \quad t = 1, \ldots, T \tag{79}$$

where

$$\widehat{\mathbf{X}}_T = [\mathbf{1}_N \quad \overline{\mathbf{Z}}_T \quad \widehat{\mathbf{B}}_T] \tag{80}$$

which is assumed to be of full rank equal to $L = 1 + M + K$. The use of time-average $\overline{\mathbf{Z}}_T$ ensures the consistency of the CSR estimator as the next proposition illustrates.

Proposition 3.15 *The time-series average $\overline{\widehat{\mathbf{c}}}_T$ of the cross-sectional estimates*

$$\widehat{\mathbf{c}}_t = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\mathbf{R}_t, \ t = 1, \ldots, T$$

where $\widehat{\mathbf{X}}_T = [\mathbf{1}_N \quad \overline{\mathbf{Z}}_T \quad \widehat{\mathbf{B}}_T]$ is a consistent estimator of $\mathbf{c} = \left[a_0 \quad \mathbf{a}' \quad \boldsymbol{\lambda}'\right]'$, that is

$$\overline{\widehat{\mathbf{c}}}_T \xrightarrow{P} \mathbf{c} \text{ as } T \to \infty. \tag{81}$$

Next, we describe the asymptotic distribution of the CSR estimator using the time-average characteristics. For this we need some additional assumptions and notation. We assume that $E[\mathbf{X}_t] = [\mathbf{1}_N \quad E[\mathbf{Z}_t] \quad \mathbf{B}]$ is of full rank equal to L and define

$$\mathbf{D} = (E[\mathbf{X}_t]'\mathbf{Q}E[\mathbf{X}_t])^{-1}E[\mathbf{X}_t]'\mathbf{Q} \tag{82}$$

and

$$\mathbf{k}_t^1 = \mathbf{R}_t - E[\mathbf{R}_t] + (\mathbf{Z}_t - E[\mathbf{Z}_t])\mathbf{a},\ \mathbf{k}_t^2 = \left[(\mathbf{f}_t - E[\mathbf{f}_t])'\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}\boldsymbol{\lambda}\right]\mathbf{u}_t \text{ and } \mathbf{k}_t = [(\mathbf{k}_t^1)' \ \ (\mathbf{k}_t^2)']'. \quad (83)$$

By definition, we have $E[\mathbf{k}_t^1] = \mathbf{0}_N$. Moreover, from the decomposition given in (13) it follows that $E[\mathbf{k}_t^2] = \mathbf{0}_N$. In the present context, we consider the following assumption which is the counterpart of Assumptions A, D.1 and F made earlier.

Assumption G. The central limit theorem applies to the random sequence $\mathbf{k}_t$ defined in (83), that is $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{k}_t$ converges in distribution to a multivariate normal with zero mean and covariance matrix given by

$$\begin{bmatrix} \mathbf{K} & \boldsymbol{\Xi} \\ \boldsymbol{\Xi}' & \mathbf{N} \end{bmatrix}$$

where

$$\mathbf{K} = \sum_{k=-\infty}^{+\infty} E\left[\mathbf{k}_t^1 \ \ (\mathbf{k}_{t+k}^1)'\right],\ \boldsymbol{\Xi} = \sum_{k=-\infty}^{+\infty} E\left[\mathbf{k}_t^1(\mathbf{k}_{t+k}^2)'\right] \text{ and } \mathbf{N} = \sum_{k=-\infty}^{+\infty} E\left[\mathbf{k}_t^2(\mathbf{k}_{t+k}^2)'\right]. \quad (84)$$

We are now in a position to state the theorem that gives the asymptotic distribution of the cross-sectional regression estimator using time-average characteristics.

Theorem 3.16 *Let* $\mathbf{c} = [a_0 \ \ \mathbf{a}' \ \ \boldsymbol{\lambda}']'$ *and* $\overline{\widehat{\mathbf{c}}}_T = (\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\widehat{\mathbf{X}}_T)^{-1}\widehat{\mathbf{X}}_T'\widehat{\mathbf{Q}}_T\overline{\mathbf{R}}_T$ *where* $\widehat{\mathbf{X}}_T = [\mathbf{1}_N \ \ \overline{\mathbf{Z}}_T \ \ \widehat{\mathbf{B}}_T]$. *Under Assumption G, as* $T \to \infty$, $\sqrt{T}(\overline{\widehat{\mathbf{c}}}_T - \mathbf{c})$ *converges in distribution to a multivariate normal with zero mean and covariance*

$$\boldsymbol{\Sigma}_c = \mathbf{DKD}' + \mathbf{DND}' - \mathbf{D}\left(\boldsymbol{\Xi} + \boldsymbol{\Xi}'\right)\mathbf{D}' \quad (85)$$

where $\mathbf{D} = (E[\mathbf{X}_t]'\mathbf{Q}E[\mathbf{X}_t])^{-1}E[\mathbf{X}_t]'\mathbf{Q}$ *with* $\mathbf{X}_t = [\mathbf{1}_N \ \ \mathbf{Z}_t \ \ \mathbf{B}]$ *and* $\mathbf{K}, \boldsymbol{\Xi}$ *and* $\mathbf{N}$ *are defined in (84).*

3.7 N-consistency of the CSR estimator

The preceding subsections present an extensive discussion of the asymptotic properties, as $T \to \infty$, of the cross-sectional regression estimator in a number of different formulations. In this subsection, we examine the CSR estimator from perspective. We assume that the length of the available time series T is fixed and consider limiting behavior of the estimator as the number of individual assets N increases without bound. This perspective is of particular interest given the availability of rather large cross sections of return data.

Consider an economy with N traded assets and a linear pricing factor model with K factors $f^1, \ldots, f^K$. Consider the following pricing equation

$$E[\mathbf{R}_t] = \gamma_0\mathbf{1}_N + \mathbf{B}\boldsymbol{\gamma}_1 = \mathbf{X}_N\boldsymbol{\Gamma} \quad (86)$$

where the beta matrix $\mathbf{B}$ is given by $\mathbf{B} = E[(\mathbf{R}_t - E[\mathbf{R}_t])\,(\mathbf{f}_t - E[\mathbf{f}_t])']\boldsymbol{\Sigma}_{\mathrm{F}}^{-1}$ with $\boldsymbol{\Sigma}_{\mathrm{F}}$ denoting the factor covariance matrix $E[(\mathbf{f}_t - E[\mathbf{f}_t])\,(\mathbf{f}_t - E[\mathbf{f}_t])']$, $\mathbf{X}_N = [\mathbf{1}_N \quad \mathbf{B}]$ ($N \times (1+K)$ matrix) and $\boldsymbol{\Gamma} = (\gamma_0, \boldsymbol{\gamma}_1')'$ ($(1+K) \times 1$ vector). The factors could be either traded or non-traded. However, even if they are traded we do not impose any pricing restriction implied by the model.

Recall from (13) the following time series regression equation $\mathbf{R}_t = E[\mathbf{R}_t] + \mathbf{B}\,(\mathbf{f}_t - E[\mathbf{f}_t]) + \mathbf{u}_t$, $t = 1, \ldots, T$ with $E[\mathbf{u}_t] = \mathbf{0}_N$ and $E[\mathbf{u}_t\mathbf{f}_t'] = \mathbf{0}_{N\times K}$. Taking time average in the last regression and incorporating (86) we obtain

$$\overline{\mathbf{R}}_T = \gamma_0 \mathbf{1}_N + \mathbf{B}\left(\boldsymbol{\gamma}_1 + \overline{\mathbf{f}}_T - E[\mathbf{f}_t]\right) + \overline{\mathbf{u}}_T = \mathbf{X}_N \overline{\boldsymbol{\Gamma}}_T + \overline{\mathbf{u}}_T \tag{87}$$

where $\overline{\boldsymbol{\Gamma}}_T = (\gamma_0, \overline{\boldsymbol{\gamma}}_{1T}')'$ and $\overline{\boldsymbol{\gamma}}_{1T} = \boldsymbol{\gamma}_1 + \overline{\mathbf{f}}_T - E[\mathbf{f}_t]$. In Shanken (1992) the vector $\overline{\boldsymbol{\Gamma}}_T$ is referred to as the vector of "ex-post prices of risk".

Assume the time series length T is fixed and consider limiting behavior as the number of assets N goes to ∞. Since T is fixed it is clear that we cannot estimate the vector of "ex-ante prices of risk" $\boldsymbol{\Gamma}$. Instead we can estimate the vector of "ex-post prices of risk" $\overline{\boldsymbol{\Gamma}}_T$. Following Shanken (1992), we define an estimator to be N-consistent if it converges in probability to the ex-post parameter vector $\overline{\boldsymbol{\Gamma}}_T$ as $N \to \infty$.

To ease the exposition let us suppose for a moment that the beta matrix $\mathbf{B}$ is known. Then equation (87) suggests $(\mathbf{X}_N'\mathbf{X}_N)^{-1}\mathbf{X}_N'\overline{\mathbf{R}}_T$ as an estimator for $\overline{\boldsymbol{\Gamma}}_T$. Indeed from (87) we have

$$\left(\mathbf{X}_N'\mathbf{X}_N\right)^{-1}\mathbf{X}_N'\overline{\mathbf{R}}_T = \overline{\boldsymbol{\Gamma}}_T + \left(\mathbf{X}_N'\mathbf{X}_N\right)^{-1}\mathbf{X}_N'\overline{\mathbf{u}}_T. \tag{88}$$

Then, under appropriate assumptions which are stated below, we have that as $N \to \infty$

$$\frac{1}{N}\mathbf{X}_N'\mathbf{X}_N = \begin{bmatrix} 1 & \frac{1}{N}\sum_{i=1}^N \boldsymbol{\beta}_i' \\ \frac{1}{N}\sum_{i=1}^N \boldsymbol{\beta}_i & \frac{1}{N}\sum_{i=1}^N \boldsymbol{\beta}_i\boldsymbol{\beta}_i' \end{bmatrix} \xrightarrow{P} \begin{bmatrix} 1 & \boldsymbol{\mu}_\beta' \\ \boldsymbol{\mu}_\beta & \boldsymbol{\Sigma}_\beta + \boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta' \end{bmatrix} \tag{89}$$

and

$$\frac{1}{N}\mathbf{X}_N'\overline{\mathbf{u}}_T = \begin{bmatrix} 1 & \cdots & 1 \\ \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_N \end{bmatrix} \begin{bmatrix} \overline{u}_{1T} \\ \vdots \\ \overline{u}_{NT} \end{bmatrix} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^T \left(\frac{1}{N}\sum_{i=1}^N u_{it}\right) \\ \frac{1}{T}\sum_{t=1}^T \left(\frac{1}{N}\sum_{i=1}^N u_{it}\boldsymbol{\beta}_i\right) \end{bmatrix} \xrightarrow{P} \mathbf{0}_{1+K} \tag{90}$$

and so

$$\left(\mathbf{X}_N'\mathbf{X}_N\right)^{-1}\mathbf{X}_N'\overline{\mathbf{R}}_T \xrightarrow{P} \overline{\boldsymbol{\Gamma}}_T.$$

However, in practice the beta matrix $\mathbf{B}$ is unknown and has to be estimated using the available data. The two-pass procedure uses time series regressions to estimate $\mathbf{B}$. The estimate $\widehat{\mathbf{B}}_T$ is given in (17). Replacing $\mathbf{B}$ by $\widehat{\mathbf{B}}_T$ we obtain $(\widehat{\mathbf{X}}'_N\widehat{\mathbf{X}}_N)^{-1}\widehat{\mathbf{X}}'_N\overline{\mathbf{R}}_T$ as an estimator for $\overline{\boldsymbol{\Gamma}}_T$ where $\widehat{\mathbf{X}}'_N = [\mathbf{1}_N \ \ \widehat{\mathbf{B}}_T]$. However, this estimator is subject to the well-known errors-in-variables problem and needs to be modified in order to maintain the property of N-consistency. Before we proceed with the illustration of the appropriate modification, we present a set of conditions which will guarantee the validity of our claim. The first condition ensures that there is sufficiently weak cross-sectional dependence as required in the statement of Theorem 5 in Shanken (1992). The second condition requires that average betas and their squares converge to well-defined quantities as $N \to \infty$.

Assumption 3.1 *The time series residuals $\mathbf{u}_t$ satisfy the following cross-sectional properties:*

$$(a) \quad \frac{1}{N}\sum_{i=1}^{N} u_{it} \xrightarrow{P} 0 \ \textit{as } N \to \infty \textit{ for all } t = 1, 2, \ldots, \tag{91}$$

$$(b) \quad \frac{1}{N}\sum_{i=1}^{N} u_{it}\boldsymbol{\beta}_i \xrightarrow{P} \mathbf{0}_K \ \textit{as } N \to \infty \textit{ for all } t = 1, 2, \ldots, \tag{92}$$

$$(c) \quad \frac{1}{N}\sum_{i=1}^{N} u_{is}u_{it} \xrightarrow{P} 0 \ \textit{as } N \to \infty \textit{ for all } t, s = 1, 2, \ldots \textit{ with } s \neq t. \tag{93}$$

Denote by $\boldsymbol{\Sigma}_U$ the covariance matrix of $\mathbf{u}_t$ and by σ_i^2 the ith diagonal element of $\boldsymbol{\Sigma}_U$, $i = 1, \ldots, N$. Define $v_{it} = u_{it}^2 - \sigma_i^2$ which implies $E[v_{it}] = 0$ for all i and t. Then the following hold

$$(d) \quad \frac{1}{N}\sum_{i=1}^{N} v_{it} \xrightarrow{P} 0 \ \textit{as } N \to \infty \textit{ for all } t = 1, 2, \ldots, \tag{94}$$

$$(e) \quad \frac{1}{N}\sum_{i=1}^{N} \sigma_i^2 \to \bar{\sigma}^2 \ \textit{as } N \to \infty. \tag{95}$$

Assumption 3.2 *The sequence of betas $\{\boldsymbol{\beta}_i : i = 1, 2, \ldots\}$ satisfies the following two conditions:*

$$(a) \quad \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\beta}_i \to \boldsymbol{\mu}_\beta \ \textit{as } N \to \infty \tag{96}$$

$$(b) \quad \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\beta}_i\boldsymbol{\beta}_i' \to \boldsymbol{\Sigma}_\beta + \boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta' \ \textit{as } N \to \infty \tag{97}$$

where $\boldsymbol{\mu}_\beta$ is a $K \times 1$ vector and $\boldsymbol{\Sigma}_\beta$ is positive-definite symmetric $K \times K$ matrix.

Returning to our calculation we have

$$\frac{1}{N}\widehat{\mathbf{X}}_N'\overline{\mathbf{R}}_T = \begin{bmatrix} 1 & \cdots & 1 \\ \widehat{\boldsymbol{\beta}}_{1T} & \cdots & \widehat{\boldsymbol{\beta}}_{NT} \end{bmatrix} \begin{bmatrix} \overline{R}_{1T} \\ \vdots \\ \overline{R}_{NT} \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^N \overline{R}_{iT} \\ \\ \frac{1}{N}\sum_{i=1}^N \overline{R}_{iT}\widehat{\boldsymbol{\beta}}_{iT} \end{bmatrix}. \tag{98}$$

It follows from (87) that

$$\frac{1}{N}\sum_{i=1}^N \overline{R}_{iT} = \gamma_0 + \left(\frac{1}{N}\sum_{i=1}^N \boldsymbol{\beta}_i\right)'\left(\boldsymbol{\gamma}_1 + \overline{\mathbf{f}}_T - E[\mathbf{f}_t]\right) + \frac{1}{T}\sum_{t=1}^T\left(\frac{1}{N}\sum_{i=1}^N u_{it}\right)$$

and so

$$\frac{1}{N}\sum_{i=1}^N \overline{R}_{iT} \xrightarrow{P} \gamma_0 + \boldsymbol{\mu}_\beta'\left(\boldsymbol{\gamma}_1 + \overline{\mathbf{f}}_T - E[\mathbf{f}_t]\right) \text{ as } N \to \infty \tag{99}$$

as it follows from Assumptions 3.1 and 3.2. Recall that

$$\widehat{\boldsymbol{\beta}}_{iT} = \widehat{\boldsymbol{\Sigma}}_{\mathrm{FT}}^{-1}\left[\frac{1}{T}\sum_{t=1}^T\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)\left(R_{it} - \overline{R}_{iT}\right)\right]$$

and

$$R_{it} - \overline{R}_{iT} = \boldsymbol{\beta}_i'\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right) + \left(u_{it} - \overline{u}_{iT}\right) \tag{100}$$

which combined deliver

$$\widehat{\boldsymbol{\beta}}_{iT} = \boldsymbol{\beta}_i + \boldsymbol{\xi}_{iT} \tag{101}$$

where

$$\boldsymbol{\xi}_{iT} = \widehat{\boldsymbol{\Sigma}}_{\mathrm{FT}}^{-1}\left[\frac{1}{T}\sum_{t=1}^T u_{it}\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)\right]. \tag{102}$$

Thus

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N \overline{R}_{iT}\widehat{\boldsymbol{\beta}}_{iT} &= \frac{1}{N}\sum_{i=1}^N(\boldsymbol{\beta}_i + \boldsymbol{\xi}_{iT})\left[\gamma_0 + \boldsymbol{\beta}_i'\left(\boldsymbol{\gamma}_1 + \overline{\mathbf{f}}_T - E[\mathbf{f}_t]\right) + \overline{u}_{iT}\right] \\ &= \gamma_0\left(\tfrac{1}{N}\textstyle\sum_{i=1}^N \boldsymbol{\beta}_i + \tfrac{1}{N}\sum_{i=1}^N \boldsymbol{\xi}_{iT}\right) \\ &\quad + \left(\tfrac{1}{N}\textstyle\sum_{i=1}^N \boldsymbol{\beta}_i\boldsymbol{\beta}_i' + \tfrac{1}{N}\sum_{i=1}^N \boldsymbol{\xi}_{iT}\boldsymbol{\beta}_i'\right)\left(\boldsymbol{\gamma}_1 + \overline{\mathbf{f}}_T - E[\mathbf{f}_t]\right) \\ &\quad + \tfrac{1}{T}\textstyle\sum_{t=1}^T\left(\tfrac{1}{N}\sum_{i=1}^N u_{it}\boldsymbol{\beta}_i + \tfrac{1}{N}\sum_{i=1}^N u_{it}\boldsymbol{\xi}_{iT}\right).\end{aligned} \tag{103}$$

We consider the limiting behavior of each term in the last equation. It follows from (102) and assumption (91) that $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\xi}_{iT} \xrightarrow{P} \mathbf{0}_K$ as $N \to \infty$. Similarly, from (102) and assumption (92) we obtain $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\xi}_{iT}\boldsymbol{\beta}_i' \xrightarrow{P} \mathbf{0}_{K\times K}$ as $N \to \infty$. Further we have

$$\begin{aligned}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}u_{it}\boldsymbol{\xi}_{iT} &= \widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}u_{is}u_{it}\right)\left(\mathbf{f}_s-\overline{\mathbf{f}}_T\right)\right] \\ &\xrightarrow{P} \widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\bar{\sigma}^2\left(\mathbf{f}_t-\overline{\mathbf{f}}_T\right)\right] = \mathbf{0}_K \text{ as } N\to\infty\end{aligned} \tag{104}$$

which follows from assumptions (93), (94) and (95). Using assumptions (96), (97) and (92), we now obtain from equation (103) that

$$\frac{1}{N}\sum_{i=1}^{N}\overline{R}_{iT}\widehat{\boldsymbol{\beta}}_{iT} \xrightarrow{P} \gamma_0\boldsymbol{\mu}_\beta + (\boldsymbol{\Sigma}_\beta+\boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta')(\boldsymbol{\gamma}_1+\overline{\mathbf{f}}_T-E[\mathbf{f}_t]) \text{ as } N\to\infty. \tag{105}$$

Finally, the combination (98), (99) and (105) yields

$$\frac{1}{N}\widehat{\mathbf{X}}_N'\overline{\mathbf{R}}_T \xrightarrow{P} \begin{bmatrix} \gamma_0+\boldsymbol{\mu}_\beta'\left(\boldsymbol{\gamma}_1+\overline{\mathbf{f}}_T-E[\mathbf{f}_t]\right) \\ \gamma_0\boldsymbol{\mu}_\beta + (\boldsymbol{\Sigma}_\beta+\boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta')(\boldsymbol{\gamma}_1+\overline{\mathbf{f}}_T-E[\mathbf{f}_t])\end{bmatrix} \text{ as } N\to\infty. \tag{106}$$

Next we study the limiting behavior of $\frac{1}{N}\widehat{\mathbf{X}}_N'\widehat{\mathbf{X}}_N$. This will enable us to appropriately modify the proposed estimator to achieve N-consistency. We have

$$\frac{1}{N}\widehat{\mathbf{X}}_N'\widehat{\mathbf{X}}_N = \begin{bmatrix} 1 & \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_i' \\ \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_i & \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_i\widehat{\boldsymbol{\beta}}_i'\end{bmatrix}. \tag{107}$$

The calculations above imply that, as $N\to\infty$, $\frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_i \xrightarrow{P} \boldsymbol{\mu}_\beta$ and $\frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_i\widehat{\boldsymbol{\beta}}_i' \xrightarrow{P} \boldsymbol{\Sigma}_\beta + \boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta' + \text{p-}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}'$. Using (102) and assumptions (94) and (95) we obtain

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}' &= \widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1}\left[\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}u_{it}u_{is}\right)\left(\mathbf{f}_t-\overline{\mathbf{f}}_T\right)\left(\mathbf{f}_s-\overline{\mathbf{f}}_T\right)'\right]\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1} \\ &\xrightarrow{P} \widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1}\left[\frac{1}{T^2}\sum_{t=1}^{T}\bar{\sigma}^2\left(\mathbf{f}_t-\overline{\mathbf{f}}_T\right)\left(\mathbf{f}_t-\overline{\mathbf{f}}_T\right)'\right]\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1} = \bar{\sigma}^2\frac{1}{T}\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1} \text{ as } N\to\infty.\end{aligned} \tag{108}$$

Thus, from (107) it follows that

$$\frac{1}{N}\widehat{\mathbf{X}}_N'\widehat{\mathbf{X}}_N \xrightarrow{P} \begin{bmatrix} 1 & \boldsymbol{\mu}_\beta' \\ \boldsymbol{\mu}_\beta & \boldsymbol{\Sigma}_\beta+\boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta' + \bar{\sigma}^2\frac{1}{T}\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1}\end{bmatrix}. \tag{109}$$

Using the expression for the inverse of a partitioned matrix and a few steps of algebra, we now obtain

$$\begin{bmatrix} 1 & \boldsymbol{\mu}_\beta' \\ \boldsymbol{\mu}_\beta & \boldsymbol{\Sigma}_\beta+\boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta'\end{bmatrix}^{-1}\begin{bmatrix} \gamma_0+\boldsymbol{\mu}_\beta'\left(\boldsymbol{\gamma}_1+\overline{\mathbf{f}}_T-E[\mathbf{f}_t]\right) \\ \gamma_0\boldsymbol{\mu}_\beta + (\boldsymbol{\Sigma}_\beta+\boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta')(\boldsymbol{\gamma}_1+\overline{\mathbf{f}}_T-E[\mathbf{f}_t])\end{bmatrix}$$

$$= \begin{bmatrix} 1+\boldsymbol{\mu}_\beta'\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta & -\boldsymbol{\mu}_\beta'\boldsymbol{\Sigma}_\beta^{-1} \\ -\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta & \boldsymbol{\Sigma}_\beta^{-1} \end{bmatrix} \begin{bmatrix} \gamma_0 + \boldsymbol{\mu}_\beta'\left(\boldsymbol{\gamma}_1 + \overline{\mathbf{f}}_T - E[\mathbf{f}_t]\right) \\ \gamma_0\boldsymbol{\mu}_\beta + (\boldsymbol{\Sigma}_\beta + \boldsymbol{\mu}_\beta\boldsymbol{\mu}_\beta')(\boldsymbol{\gamma}_1 + \overline{\mathbf{f}}_T - E[\mathbf{f}_t]) \end{bmatrix} = \overline{\boldsymbol{\Gamma}}_T.$$

Thus, to obtain an N-consistent estimator we need to subtract a term from $\widehat{\mathbf{X}}_N'\widehat{\mathbf{X}}_N$ in order to eliminate the term $\bar{\sigma}^2\frac{1}{T}\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1}$ in the probability limit in (109). Assumption (95) states that

$$\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{\mathrm{U}})}{N} = \frac{\sum_{i=1}^N \sigma_i^2}{N} \to \bar{\sigma}^2 \text{ as } N \to \infty$$

which, in the light of (109), suggests that we should replace $\widehat{\mathbf{X}}_N'\widehat{\mathbf{X}}_N$ by

$$\widehat{\mathbf{X}}_N'\widehat{\mathbf{X}}_N - \begin{bmatrix} 0 & \mathbf{0}_K' \\ \mathbf{0}_K & \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_{\mathrm{U}T})\frac{1}{T}\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1} \end{bmatrix} = \widehat{\mathbf{X}}_N'\widehat{\mathbf{X}}_N - \frac{\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_{\mathrm{U}T})}{T}\mathbf{M}'\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1}\mathbf{M}$$

where $\mathbf{M} = [\mathbf{0}_K \;\; \mathbf{I}_K]$ ($K \times (1+K)$ matrix) and

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{U}T} = \frac{1}{T-K-1}\sum_{t=1}^T \mathbf{e}_t\mathbf{e}_t', \quad \mathbf{e}_t = \mathbf{R}_t - \widehat{\mathbf{a}} - \widehat{\mathbf{B}}_T\mathbf{f}_t = \left(\mathbf{R}_t - \overline{\mathbf{R}}_T\right) - \widehat{\mathbf{B}}_T\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right). \tag{110}$$

That is, $\widehat{\boldsymbol{\Sigma}}_{\mathrm{U}T}$ is the unbiased estimator of $\boldsymbol{\Sigma}_{\mathrm{U}}$ based on the sample of size T. It remains to show that

$$\frac{\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_{\mathrm{U}T})}{N} \to \bar{\sigma}^2 \text{ as } N \to \infty. \tag{111}$$

Using (110) and the decomposition in (100) we obtain

$$\frac{\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_{\mathrm{U}T})}{N} = \frac{1}{N}\mathrm{tr}\left(\frac{1}{T-K-1}\sum_{t=1}^T \mathbf{e}_t\mathbf{e}_t'\right)$$

$$= \frac{1}{N(T-K-1)}\sum_{t=1}^T\sum_{i=1}^N \left[\left(R_{it} - \overline{R}_{iT}\right) - \widehat{\boldsymbol{\beta}}_{iT}'\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)\right]^2$$

$$= \frac{1}{N(T-K-1)}\sum_{t=1}^T\sum_{i=1}^N \left[\left(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_{iT}\right)'\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right) + (u_{it} - \overline{u}_{iT})\right]^2$$

$$\begin{aligned} &= \tfrac{1}{T-K-1}\textstyle\sum_{t=1}^T \left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)'\left(\frac{1}{N}\sum_{i=1}^N \boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}'\right)\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right) \\ &\quad -\tfrac{2}{T-K-1}\textstyle\sum_{t=1}^T \left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)'\left(\frac{1}{N}\sum_{i=1}^N u_{it}\boldsymbol{\xi}_{iT}\right) \\ &\quad +\tfrac{T}{T-K-1}\textstyle\frac{1}{N}\sum_{i=1}^N \left(\frac{1}{T}\sum_{t=1}^T (u_{it} - \overline{u}_{iT})^2\right) \end{aligned} \tag{112}$$

Following the calculation in (104) we obtain that $\frac{1}{N}\sum_{i=1}^{N} u_{it}\boldsymbol{\xi}_{iT} \xrightarrow{P} \frac{\bar{\sigma}^2}{T}\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1}\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)$ as $N \to \infty$. Further, using assumptions (93), (94) and (95), we have

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}(u_{it} - \overline{u}_{iT})^2\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}u_{it}^2 - \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{it}u_{is}\right) \xrightarrow{P} \frac{T-1}{T}\bar{\sigma}^2$$

Combining the above probability limits and the result in (108) we obtain from (112) that

$$\frac{\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_{\mathrm{U}T})}{N} \xrightarrow{P} \frac{T-1}{T-K-1}\bar{\sigma}^2 - \frac{1}{T-K-1}\bar{\sigma}^2\left[\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)'\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1}\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)\right] = \bar{\sigma}^2 \tag{113}$$

where $\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right)'\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}^{-1}\left(\mathbf{f}_t - \overline{\mathbf{f}}_T\right) = K$ since $\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}T}$ is assumed to be positive definite. The last equality follows from the following matrix algebraic result which is based on the properties of the trace operator for matrices. Suppose $\mathbf{x}_t$, $t = 1, \ldots, T$ are K-dimensional vectors and let $\mathbf{X} = \sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'$. If $\mathbf{X}$ is nonsingular, then $\sum_{t=1}^{T}\mathbf{x}_t'\mathbf{X}^{-1}\mathbf{x}_t = K$. Thus we have completed the proof of the following theorem.

Theorem 3.17 *Let Assumptions 3.1 and 3.2 be in effect. Then*

$$\left(\widehat{\mathbf{X}}_N'\widehat{\mathbf{X}}_N - tr(\widehat{\boldsymbol{\Sigma}}_{UT})\mathbf{M}'\widehat{\boldsymbol{\Sigma}}_{FT}^{-1}\mathbf{M}/T\right)^{-1}\widehat{\mathbf{X}}_N'\overline{\mathbf{R}}_T \tag{114}$$

is an N-consistent estimator of the vector of the ex-post prices of risk $\overline{\boldsymbol{\Gamma}}_T$ where $\widehat{\mathbf{X}}_N = [\mathbf{1}_N \;\; \widehat{\mathbf{B}}_T]$, $\mathbf{M} = [\mathbf{1}_K \;\; \mathbf{I}_K]$, and $\widehat{\mathbf{B}}_T$, $\widehat{\boldsymbol{\Sigma}}_{FT}$, $\widehat{\boldsymbol{\Sigma}}_{UT}$ are defined by (17), (16), (110) respectively.

4 Maximum Likelihood Methods

When the researcher is willing to make distributional assumptions on the joint dynamics of the asset returns and the factors, the likelihood function is available and the maximum likelihood (ML) method provides a natural way to estimate and test the linear beta pricing model. The earliest application of the ML method was by Gibbons (1982). Given the computational difficulties involved at that time, Gibbons linearized the restriction imposed by the linear beta pricing model. He estimated the model and tested the hypothesis that the restrictions imposed by the pricing model holds using the likelihood ratio test statistic which has a central chi-square distribution with the number of degrees of freedom equal to the number of constraints. In what follows, we first describe the ML method for estimating and testing the general linear beta-pricing model under the assumption that none of the factors are traded and then relax the assumption to allow some factors to be traded. We then consider the more special case where there is a risk-free asset.

4.1 Nontraded factors

Consider the linear factor model given in equation (5) reproduced below for convenience

$$\mathbf{R}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \mathbf{u}_t \quad \text{with} \quad \boldsymbol{\alpha} = \boldsymbol{\mu}_{\mathrm{R}} - \mathbf{B}\boldsymbol{\mu}_{\mathrm{F}} \tag{115}$$

where the alternative notation $\boldsymbol{\mu}_{\mathrm{R}} = E[\mathbf{R}_t]$ and $\boldsymbol{\mu}_{\mathrm{F}} = E[\mathbf{f}_t]$ is used. We are first concerned with the case in which the factors are not returns on traded portfolios. Assume that the innovations $\mathbf{u}_t$ are i.i.d. multivariate normal conditional on contemporaneous and past realizations of the factors, and denote by $\boldsymbol{\Sigma}_{\mathrm{U}}$ the covariance matrix of the innovations and by $\mathcal{L}_{\mathrm{F}}$ the marginal likelihood of the factors.

The linear beta pricing model in equation (4) imposes the restriction that

$$\boldsymbol{\alpha} = a_0\mathbf{1} + \mathbf{B}(\boldsymbol{\lambda} - \boldsymbol{\mu}_{\mathrm{F}}) \tag{116}$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}_{\mathrm{F}}$ denote the vector of the risk premia and the vector of expected values of the K factors respectively. In the special case where there is only one factor and it is the return on the market portfolio, R_{mt}, this gives the well known intercept restriction,

$$\alpha_i = a_0(1 - \beta_{mi}), \ i = 1, \ldots, N.$$

The loglikelihood function of the unconstrained model is given by

$$\mathcal{L} = -\frac{NT}{2}\log(2\pi) - \frac{T}{2}\log(|\boldsymbol{\Sigma}_{\mathrm{U}}|) - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{R}_t - \boldsymbol{\alpha} - \mathbf{B}\mathbf{f}_t)'\boldsymbol{\Sigma}_{\mathrm{U}}^{-1}(\mathbf{R}_t - \boldsymbol{\alpha} - \mathbf{B}\mathbf{f}_t) + \log(\mathcal{L}_{\mathrm{F}}).$$

It can be verified that the maximum likelihood estimator of the parameter vector is given by

$$\begin{aligned}
&\widehat{\boldsymbol{\mu}}_{\mathrm{R}} = \tfrac{1}{T}\textstyle\sum_{t=1}^{T}\mathbf{R}_t; \quad \widehat{\boldsymbol{\mu}}_{\mathrm{F}} = \tfrac{1}{T}\textstyle\sum_{t=1}^{T}\mathbf{f}_t; \\
&\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}} = \tfrac{1}{T}\textstyle\sum_{t=1}^{T}(\mathbf{f}_t - \widehat{\boldsymbol{\mu}}_{\mathrm{F}})(\mathbf{f}_t - \widehat{\boldsymbol{\mu}}_{\mathrm{F}})'; \quad \widehat{\mathbf{B}} = \left[\tfrac{1}{T}\textstyle\sum_{1}^{T}(\mathbf{R}_t - \widehat{\boldsymbol{\mu}}_{\mathrm{R}})(\mathbf{f}_t - \widehat{\boldsymbol{\mu}}_{\mathrm{F}})'\right]\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}}^{-1}; \\
&\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\mu}}_{\mathrm{R}} - \widehat{\mathbf{B}}\widehat{\boldsymbol{\mu}}_{\mathrm{F}}; \quad \widehat{\boldsymbol{\Sigma}}_{\mathrm{U}} = \tfrac{1}{T}\textstyle\sum_{t=1}^{T}(\mathbf{R}_t - \widehat{\boldsymbol{\alpha}} - \widehat{\mathbf{B}}\mathbf{f}_t)(\mathbf{R}_t - \widehat{\boldsymbol{\alpha}} - \widehat{\mathbf{B}}\mathbf{f}_t)'.
\end{aligned} \tag{117}$$

When the beta pricing model holds $\boldsymbol{\alpha}$ is given by equation (116). Substituting the right side of equation (116) for $\boldsymbol{\alpha}$ into the linear factor model in equation (115) and rearranging the terms gives

$$\mathbf{R}_t - a_0\mathbf{1} = \mathbf{B}(\mathbf{a}_1 + \mathbf{f}_t) + \mathbf{u}_t \tag{118}$$

where $\mathbf{a}_1 = \boldsymbol{\lambda} - \boldsymbol{\mu}_{\text{F}}$. Let $\mathbf{a} = [a_0 \quad \mathbf{a}_1']'$. It can then be verified that the constrained maximum likelihood estimators are given by

$$\begin{aligned}
&\widehat{\boldsymbol{\alpha}}_{\text{c}} = \widehat{\boldsymbol{\mu}}_{\text{R}} - \widehat{\mathbf{B}}^{\text{c}}\widehat{\boldsymbol{\mu}}_{\text{F}}; \quad \widehat{\mathbf{a}}_{\text{c}} = (\widehat{\mathbf{B}}_{\text{c}}'\widehat{\boldsymbol{\Sigma}}_{\text{Uc}}^{-1}\widehat{\mathbf{B}}_{\text{c}})^{-1}(\widehat{\mathbf{B}}_{\text{c}}'\widehat{\boldsymbol{\Sigma}}_{\text{Uc}}^{-1}\widehat{\boldsymbol{\alpha}}_{\text{c}}) \\
&\widehat{\mathbf{B}}_{\text{c}} = \left[\tfrac{1}{T}\textstyle\sum_{t=1}^{T}(\mathbf{R}_t - \widehat{a}_{0\text{c}}\mathbf{1})(\widehat{\mathbf{a}}_{1\text{c}} + \mathbf{f}_t)'\right]\left[\tfrac{1}{T}\textstyle\sum_{t=1}^{T}(\widehat{\mathbf{a}}_{1\text{c}} + \mathbf{f}_t)(\widehat{\mathbf{a}}_{1\text{c}} + \mathbf{f}_t)'\right]^{-1} \\
&\widehat{\boldsymbol{\Sigma}}_{\text{Uc}} = \tfrac{1}{T}\textstyle\sum_{t=1}^{T}[\mathbf{R}_t - \widehat{a}_{0\text{c}}\mathbf{1} - \widehat{\mathbf{B}}_{\text{c}}(\widehat{\mathbf{a}}_{1\text{c}} + \mathbf{f}_t)][\mathbf{R}_t - \widehat{a}_{0\text{c}}\mathbf{1} - \widehat{\mathbf{B}}_{\text{c}}(\widehat{\mathbf{a}}_{1\text{c}} + \mathbf{f}_t)]'
\end{aligned} \tag{119}$$

where the subscript c indicates that these are the constrained ML estimates.

The above equations can be solved by successive iteration. The unconstrained estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}_{\text{U}}$ are used to obtain an initial estimate of $\mathbf{a}$. The asymptotic variance of the estimators can be obtained using standard maximum likelihood procedures. When the linear beta pricing model restriction holds, the likelihood ratio test statistic (minus twice the logarithm of the likelihood ratio) given by

$$J_{\text{LR}} = -T\left[\log(|\widehat{\boldsymbol{\Sigma}}_{\text{U}}|) - \log(|\widehat{\boldsymbol{\Sigma}}_{\text{Uc}}|)\right] \tag{120}$$

has an asymptotic chi-square distribution with $N-(K+1)$ degrees of freedom. Following Bartlett (1938), Jobson and Korkie (1982) suggest replacing T in equation (120) with $T - \frac{N}{2} - K - 1$ for faster convergence to the asymptotic chi-square distribution.

4.2 Some factors are traded

When all the factors are returns on some benchmark portfolios, $\boldsymbol{\lambda} - \boldsymbol{\mu}_{\text{F}}$ is the vector of expected return on the benchmark portfolios in excess of the zero-beta return. In this case, Shanken (1985) shows that the exact constrained maximum likelihood can be computed without iteration. It is an extension of the method worked out in Kandel (1984) for the standard CAPM. For the general case in which a subset of the factors are returns on portfolios of traded securities, Shanken (1992) shows that the ML and the two-pass GLS cross sectional regression estimators are asymptotically equivalent under the standard regularity conditions assumed in empirical studies as the number of observations T becomes large. Hence the two-pass cross sectional regression approach is asymptotically efficient as T becomes large.

Shanken (1985) establishes the connection between the likelihood ratio test for the restrictions imposed by the pricing model to the multivariate T^2 test that examines whether the vector of model expected return errors are zero after allowing for sampling errors. Zhou (1991) derives the exact finite sample distribution of the likelihood ratio test statistic and shows that the distribution depends on a nuisance parameter.

Suppose the mean-variance frontier of returns generated by a given set of N assets is spanned by the returns on a subset of K benchmark assets only. This is a stronger assumption than the assumption that the K factor linear beta pricing model holds. In this case, there is an additional restriction on the parameters in the linear factor model given in equation (115). The β_{ik}'s should sum to 1 for each asset i, that is $\sum_{k=1}^{K} \beta_{ik} = 1$ for all $i = 1, \ldots, N$ and the intercept term $\boldsymbol{\alpha}$ should equal the zero vector. Under the assumption that the returns are i.i.d. with a multivariate normal distribution, Huberman and Kandel (1987) and Kan and Zhou (2001) show that the statistic $\frac{T-N-K}{N}\left[\left(\frac{|\widehat{\Sigma}_{\mathrm{Uc}}|}{|\widehat{\Sigma}_{\mathrm{U}}|}\right)^{\frac{1}{2}} - 1\right]$ has a central F distribution with $(2N, 2(T-N-K))$ degrees of freedom.

When a factor is a return on a traded asset, the ML estimate of the factor risk premium is its average return in the sample minus the ML estimate of the zero-beta return. For a non traded factor, the ML estimate of the expected value of the factor is its sample average.

4.3 Single risk-free lending and borrowing rates with portfolio returns as factors

When there is a risk-free asset and borrowing and lending is available at the same rate, it is convenient to work with returns in excess of the risk-free return. Let us denote by $\mathbf{R}_t^e$ the vector of date t excess returns on the N assets and by $\mathbf{f}_t^e$ the vector of excess returns on the K factor portfolios. The linear factor model in this case assumes the form

$$\mathbf{R}_t^e = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t^e + \mathbf{u}_t. \tag{121}$$

As before, we assume that the innovations $\mathbf{u}_t$ are i.i.d. with a multivariate normal distribution conditionally on contemporaneous and past factor excess returns, and $\mathcal{L}_{\mathrm{F}}$ denotes the marginal likelihood of the factor excess returns. The beta pricing model implies that the vector of intercepts, $\boldsymbol{\alpha}$, equals the zero vector. Let us denote by $\widehat{\boldsymbol{\alpha}}, \widehat{\mathbf{B}}$ the OLS estimates of $\boldsymbol{\alpha}, \mathbf{B}$, by $\widehat{\boldsymbol{\mu}}_{\mathrm{F}}$ the sample mean of the vector of excess returns on the factor portfolios, and by $\widehat{\boldsymbol{\Sigma}}_{\mathrm{U}}, \widehat{\boldsymbol{\Sigma}}_{\mathrm{F}}$ the sample covariance matrices of the linear factor model residuals and the excess returns on the factor portfolios respectively. Under the null hypothesis the statistic

$$J_{\mathrm{LR}} = \frac{T-N-K}{N}\left(\frac{\widehat{\boldsymbol{\alpha}}'\widehat{\boldsymbol{\Sigma}}_{\mathrm{U}}^{-1}\widehat{\boldsymbol{\alpha}}}{1 + \widehat{\boldsymbol{\mu}}_{\mathrm{F}}'\widehat{\boldsymbol{\Sigma}}_{\mathrm{F}}\widehat{\boldsymbol{\mu}}_{\mathrm{F}}}\right) \tag{122}$$

has an exact F distribution with $(N, T-N-K)$ degrees of freedom. The reader is referred to Jobson and Korkie (1985), MacKinlay (1987) and Gibbons, Shanken and Ross (1989) for further details, analysis of the test statistic under the alternative that $\boldsymbol{\alpha} \neq 0$, a geometric interpretation of the test when $K = 1$ and a discussion of the power of the tests.

5 The Generalized Method of Moments

One major shortcoming of the ML method is that the econometrician has to make strong assumptions regarding the joint distribution of stock returns. The common practice is to assume that returns are drawn from an i.i.d. multivariate normal distribution. The generalized method of moments (GMM) has made econometric analysis of stock returns possible under more realistic assumptions regarding the nature of the stochastic process governing the temporal evolution of economic variables. In this section, we discuss how to use the GMM to analyze the cross section of stock returns. After giving an overview of the GMM, we discuss the estimation and testing of linear beta pricing models using their beta as well as the stochastic discount factor (SDF) representation. Finally, we discuss the various issues related to the GMM tests of conditional models with time-varying parameters.

The use of the GMM in finance started with Hansen and Hodrick (1980) and Hansen and Singleton (1982). Subsequent developments have made it a reliable and robust econometric methodology for studying the implications of not only linear beta pricing models but also dynamic asset-pricing models in general, allowing stock returns and other economic variables to be serially correlated, leptokurtic, and conditionally heteroscedastic. The works by Newey and West (1987), Andrews (1991), and Andrews and Monahan (1992) on estimating covariance matrices in the presence of autocorrelation and heteroscedasticity are the most significant among these developments. We refer the readers to Jagannathan, Skoulakis and Wang (2002) for a review of financial econometric applications of the GMM .

The major disadvantage of the GMM when compared to the maximum likelihood method is that the sampling theory has only asymptotic validity. We therefore require a long history of observations on returns relative to the number of assets. When the number of assets is large, it is difficult to estimate the covariance matrix of returns precisely. Two approaches have been suggested in the literature to address this issue. The first approach is to group the primitive assets into a small number of portfolios and then evaluate the pricing models using return data on the portfolios. However, even when the econometrician is working with only 10 or 20 portfolios, the length of the time series of observations available may not be sufficient for appealing to the law of large numbers. Therefore Monte Carlo simulation is often used to check if GMM estimators and test statistics that rely on asymptotic theory have any biases. Using Monte Carlo simulations, Ferson and Foerster (1994) found that the GMM tends to overreject models when the number of observations, T, corresponds to values typically used in empirical studies. The second approach is to make additional assumptions so that the covariance matrix of the residuals in the linear factor generating model for the vector of asset returns takes a diagonal or block diagonal form. Some even suggest assuming it to be a scaled identity matrix. These assumptions usually reduce the GMM estimator to an OLS estimator in linear regressions.

5.1 An overview of the GMM

Let $\mathbf{x}_t$ be a vector of m variables observed in the tth period. Let $\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta})$ be a vector of n functions, where $\boldsymbol{\theta}$ is a vector of k unknown parameters. Suppose when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ the following moment restriction holds

$$E[\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta}_0)] = \mathbf{0}_n \tag{123}$$

where $\mathbf{0}_n$ is the column vector of n zeros. For any $\boldsymbol{\theta}$, the sample analogue of $E[\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta})]$ is

$$\mathbf{g}_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta}) \ . \tag{124}$$

Suppose $\mathbf{x}_t$ satisfies the necessary regularity conditions so that the central limit theorem can be applied to $\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta}_0)$. Then

$$\sqrt{T}\mathbf{g}_T(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}_n, \mathbf{S}), \quad \text{as } T \to \infty, \tag{125}$$

where $\mathbf{S}$ is the spectral density matrix of $\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta}_0)$, i.e.,

$$\mathbf{S} = \sum_{j=-\infty}^{\infty} E[\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta}_0)\mathbf{g}(\mathbf{x}_{t+j}, \boldsymbol{\theta}_0)'] \ . \tag{126}$$

A natural estimation strategy for $\boldsymbol{\theta}_0$ would be to choose those values that make $\mathbf{g}_T(\boldsymbol{\theta})$ as close to the zero vector as possible. For that reason we choose $\boldsymbol{\theta}$ to solve

$$\min_{\boldsymbol{\theta}} \ \mathbf{g}_T(\boldsymbol{\theta})'\mathbf{S}_T^{-1}\mathbf{g}_T(\boldsymbol{\theta}) \tag{127}$$

where $\mathbf{S}_T$ is a consistent estimator of $\mathbf{S}$. The solution to the minimization problem, denoted by $\widehat{\boldsymbol{\theta}}_T$, is the GMM estimator of $\boldsymbol{\theta}$. We assume that $\mathbf{g}$ satisfies the regularity conditions laid out in Hansen (1982) so that $\boldsymbol{\theta}_0$ is identified. In that case the following probability limit

$$\mathbf{D} = \text{p-}\lim_{T\to\infty} \frac{\partial \mathbf{g}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} = E\left[\frac{\partial \mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right] \tag{128}$$

exists and has rank k. Hansen (1982) shows that the asymptotic distribution of the GMM estimator is given by

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N\left(\mathbf{0}_k, (\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}\right), \ \text{as } T \to \infty. \tag{129}$$

In general, $\mathbf{g}_T(\widehat{\boldsymbol{\theta}}_T)$ will be different from the zero vector due to sampling errors. A natural test for model misspecification would be to examine whether $\mathbf{g}_T(\widehat{\boldsymbol{\theta}}_T)$ is indeed different from the zero vector only because of sampling errors. For that, we need to know the asymptotic distribution of $\mathbf{g}_T(\widehat{\boldsymbol{\theta}}_T)$, which is provided by

$$\sqrt{T}\mathbf{g}_T(\hat{\theta}) \xrightarrow{\mathcal{D}} N\left(\mathbf{0}_n, \mathbf{S} - \mathbf{D}(\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}\mathbf{D}'\right). \tag{130}$$

The covariance matrix $\mathbf{S} - \mathbf{D}(\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}\mathbf{D}'$ is positive semidefinite and can be degenerate. To test the moment restriction, Hansen (1982) suggests using the following J-statistic

$$J_T = T\mathbf{g}_T(\widehat{\boldsymbol{\theta}}_T)'\mathbf{S}_T^{-1}\mathbf{g}_T(\widehat{\boldsymbol{\theta}}_T). \tag{131}$$

He shows that the asymptotic distribution of J_T is a central χ^2 distribution with $n-k$ degrees of freedom, that is

$$J_T \xrightarrow{\mathcal{D}} \chi^2(n-k), \text{ as } T \to \infty. \tag{132}$$

The key to using the GMM is the specification of the moment restrictions, involving the observable variables $\mathbf{x}_t$, the unknown parameters $\boldsymbol{\theta}$, and the function $\mathbf{g}(\cdot)$. Once a decision regarding which moment restrictions are to be used is made, estimating the model parameters and testing the model specifications are rather straightforward.

5.2 Evaluating beta pricing models using the beta representation

Let $\mathbf{R}_t$ be a vector of N stock returns during period t in excess of the risk-free rate. Let $\mathbf{f}_t$ be a vector of K economy-wide pervasive risk factors during period t. The mean and variance of the factors are denoted by $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$. The standard linear beta pricing model, also referred to as the beta representation, is given by:

$$E[\mathbf{R}_t] = \mathbf{B}\boldsymbol{\delta}\ , \tag{133}$$

where $\boldsymbol{\delta}$ is the vector of factor risk premia, and $\mathbf{B}$ is the matrix of factor loadings defined as

$$\mathbf{B} \equiv E[\mathbf{R}_t(\mathbf{f}_t - \boldsymbol{\mu})']\boldsymbol{\Omega}^{-1}. \tag{134}$$

The factor loadings matrix, $\mathbf{B}$, can be equivalently defined as a parameter in the time-series regression: $\mathbf{R}_t = \boldsymbol{\phi} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\varepsilon}_t$. The residual $\boldsymbol{\varepsilon}_t$ has zero mean and is uncorrelated with the factor $\mathbf{f}_t$. The beta pricing model (133) imposes the following restriction on the intercept: $\boldsymbol{\phi} = \mathbf{B}(\boldsymbol{\delta} - \boldsymbol{\mu})$.

By substituting this expression for ϕ in the regression equation, we obtain:

$$\mathbf{R}_t = \mathbf{B}(\boldsymbol{\delta} - \boldsymbol{\mu} + \mathbf{f}_t) + \boldsymbol{\varepsilon}_t \tag{135}$$

$$E[\boldsymbol{\varepsilon}_t] = \mathbf{0}_N \tag{136}$$

$$E[\boldsymbol{\varepsilon}_t \mathbf{f}_t'] = \mathbf{0}_{N\times K}. \tag{137}$$

As we have pointed out, the key step in using GMM is to specify the moment restrictions. Equations (135), (136) and (137) in addition to the definition of $\boldsymbol{\mu}$ yield the following equations

$$E[\mathbf{R}_t - \mathbf{B}(\boldsymbol{\delta} - \boldsymbol{\mu} + \mathbf{f}_t)] = \mathbf{0}_N \tag{138}$$

$$E\left[[\mathbf{R}_t - \mathbf{B}(\boldsymbol{\delta} - \boldsymbol{\mu} + \mathbf{f}_t)]\mathbf{f}_t'\right] = \mathbf{0}_{N\times K} \tag{139}$$

$$E[\mathbf{f}_t - \boldsymbol{\mu}] = \mathbf{0}_K. \tag{140}$$

Notice that we need equation (140) to identify the vector of risk premium $\boldsymbol{\delta}$. In this case, the vector of unknown parameters is $\boldsymbol{\theta} = [\boldsymbol{\delta}' \ \ \text{vec}(\mathbf{B})' \ \ \boldsymbol{\mu}']'$, the vector of observable variables is $\mathbf{x}_t = [\mathbf{R}_t' \ \ \mathbf{f}_t']'$, and the function $\mathbf{g}$ in the moment restriction is given by

$$\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta}) = \begin{pmatrix} \mathbf{R}_t - \mathbf{B}(\boldsymbol{\delta} - \boldsymbol{\mu} + \mathbf{f}_t) \\ \text{vec}\left([\mathbf{R}_t - \mathbf{B}(\boldsymbol{\delta} - \boldsymbol{\mu} + \mathbf{f}_t)]\mathbf{f}_t'\right) \\ \mathbf{f}_t - \boldsymbol{\mu} \end{pmatrix}. \tag{141}$$

Then, the precision of the GMM estimate can be assessed by using the sampling distribution (129) and the model can be tested by using the J-statistic with asymptotic χ^2 distribution given in (132). The degree of freedom is $N-K$ because there are $N+NK+K$ equations in the moment restriction and $K+NK+K$ unknown parameters.

Often we are interested in finding out the extent to which a beta pricing model assigns the correct expected return to a particular collection of assets. For that purpose it is helpful to examine the vector of pricing errors associated with a model, given by

$$\boldsymbol{\alpha} = E[\mathbf{R}_t] - \mathbf{B}\boldsymbol{\delta}. \tag{142}$$

The pricing error associated with the CAPM is usually referred to as Jensen's alpha. When the model holds, we should have $\boldsymbol{\alpha} = \mathbf{0}_N$. The sample analogue of the pricing error is

$$\boldsymbol{\alpha}_T = \frac{1}{T}\sum_{t=1}^{T} \mathbf{R}_t - \widehat{\mathbf{B}}\widehat{\boldsymbol{\delta}}, \tag{143}$$

which can be served as a consistent estimator of the pricing error. In order to obtain the sampling distribution of $\boldsymbol{\alpha}_T$, we express it in terms of $\mathbf{g}_T(\widehat{\boldsymbol{\theta}}_T)$ as follows:

$$\boldsymbol{\alpha}_T = \widehat{\mathbf{Q}}\mathbf{g}_T(\widehat{\boldsymbol{\theta}}_T) \quad \text{where} \quad \widehat{\mathbf{Q}} = \begin{bmatrix}\mathbf{I}_N & \mathbf{0}_{N\times NK} & \widehat{\mathbf{B}}\end{bmatrix}. \tag{144}$$

Obviously, we have

$$\text{p-}\lim_{T\to\infty} \widehat{\mathbf{Q}} = [\mathbf{I}_N \ \ \mathbf{0}_{N\times K} \ \ \mathbf{B}]\,. \tag{145}$$

Let us use $\mathbf{Q}$ to denote the matrix on the right-hand side of the above equation. It follows from (130) that, under the null hypothesis that the asset pricing model holds and so $\boldsymbol{\alpha} = \mathbf{0}_N$, the sampling distribution of $\boldsymbol{\alpha}_T$ should be

$$\sqrt{T}\boldsymbol{\alpha}_T \xrightarrow{\mathcal{D}} N(\mathbf{0}_N, \mathbf{Q}(\mathbf{S} - \mathbf{D}(\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}\mathbf{D}')\mathbf{Q}') \tag{146}$$

where $\mathbf{S}$ and $\mathbf{D}$, given by (126) and (128) respectively, should be estimated using the function $\mathbf{g}$ in (141).

When an economy-wide factor, say f_{jt}, is the return on a portfolio of traded assets, we call it a traded factor. An example of a traded factor would be the return on the value-weighted portfolio of stocks used in empirical studies of Sharpe's (1964) Capital Asset Pricing Model (CAPM). Examples of non-traded factors can be found in Chen, Roll, and Ross (1986), who use the growth rate of industrial production and the rate of inflation, and Breeden, Gibbons and Litzenberger (1989), who use the growth rate in per capita consumption as a factor. When a factor f_{jt}, is the excess return on a traded asset, equations (133) and (134) imply $\delta_j = \mu_j$, i.e., the risk premium is the mean of the factor. In that case it can be verified that the sample mean of the factor is the estimator of the risk premium. If the factor is not traded, this restriction does not hold, and we have to estimate the risk premium using stock returns.

If all the factors in the model are traded, equations (138) and (139) become

$$E[\mathbf{R}_t - \mathbf{B}\mathbf{f}_t] = \mathbf{0}_N \tag{147}$$

$$E[(\mathbf{R}_t - \mathbf{B}\mathbf{f}_t)\mathbf{f}_t'] = \mathbf{0}_{N\times K}. \tag{148}$$

In that case, we can estimate $\boldsymbol{\beta}$ and test the model restrictions (147)-(148) without the need to estimate the risk premium $\boldsymbol{\delta}$. If one needs the estimate of risk premium, it can be obtained from equation (140). When all the factors are traded, the restriction imposed on the risk premium gives extra degrees of freedom – the number of degrees of freedom in (147)-(148) is N while the number of degrees of freedom in (138)-(140) is $N - K$. With traded factors, we can evaluate pricing errors of the model in the same way as done with non-traded factors.

The traded factors, however, allow us to estimate the pricing errors more conveniently as unknown parameters. For this purpose, we use the following moment restriction

$$E[\mathbf{R}_t - \boldsymbol{\alpha} - \mathbf{B}\mathbf{f}_t] = \mathbf{0}_N \tag{149}$$

$$E[(\mathbf{R}_t - \boldsymbol{\alpha} - \mathbf{B}'\mathbf{f}_t)\mathbf{f}_t'] = \mathbf{0}_{N\times K}\ , \tag{150}$$

and investigate the hypothesis $\boldsymbol{\alpha} = \mathbf{0}_N$. The above system is exactly identified because there are $N + NK$ equations and $N + NK$ unknown parameters. With these moment restrictions, we have $\mathbf{x}_t = [\mathbf{R}_t' \;\; \mathbf{f}_t']'$, $\boldsymbol{\theta} = [\boldsymbol{\alpha}' \;\; \mathrm{vec}(\mathbf{B})']'$ and

$$\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta}) = \begin{pmatrix} \mathbf{R}_t - \boldsymbol{\alpha} - \mathbf{B}\mathbf{f}_t \\ \mathrm{vec}\left([\mathbf{R}_t - \boldsymbol{\alpha} - \mathbf{B}\mathbf{f}_t]\mathbf{f}_t'\right) \end{pmatrix}. \tag{151}$$

The above function $\mathbf{g}$ defines the matrix $\mathbf{S}$ and $\mathbf{D}$ through equations (126) and (128). Let $\mathbf{S}_T$ and $\mathbf{D}_T$ be their consistent estimators. Then, the GMM estimator $\widehat{\boldsymbol{\theta}}_T$ can be obtained and the GMM estimator of $\boldsymbol{\alpha}$ is $\widehat{\boldsymbol{\alpha}}_T = \mathbf{P}\widehat{\boldsymbol{\theta}}_T$, where $\mathbf{P} = [\mathbf{I}_N \;\; \mathbf{0}_{N\times NK}]$. Under the null hypothesis $\boldsymbol{\alpha} = \mathbf{0}_N$, it follows from (129) that

$$T\widehat{\boldsymbol{\alpha}}_T' \left(\mathbf{P}\left(\mathbf{D}_T{}'\mathbf{S}_T^{-1}\mathbf{D}_T\right)^{-1}\mathbf{P}'\right)^{-1} \widehat{\boldsymbol{\alpha}}_T \xrightarrow{\mathcal{D}} \chi^2(N), \text{ as } T \to \infty. \tag{152}$$

Some remarks are in order. As shown above, for the case of traded factors, we can use the J-statistic to test the moment restrictions given by (147)-(148). However, we cannot do so for the moment restrictions given by (149)-(150) because the exactly identified system implies $J_T \equiv 0$. In that case, we can only examine whether $\widehat{\boldsymbol{\alpha}}_T$ is statistically different from zero using the distribution in (152). Therefore, for the case of traded factors, we have two choices. The first is to test (147)-(148) using the J-statistics, and the second is to test $\boldsymbol{\alpha} = \mathbf{0}_N$ using (149)-(150). MacKinlay and Richardson (1991) refer to the first as restricted test and the second as unrestricted test. It is important to notice that we cannot add $\boldsymbol{\alpha}$ as unknown parameters to equations (138) and (139) when factors are non-traded because there will be more unknown parameters than equations. The pricing error has to be obtained from $\mathbf{g}_T(\widehat{\boldsymbol{\theta}}_T)$. In this case, we have to use the J-statistic to test the moment restriction (138)-(140). Therefore, we can only do the restricted test when factors are non-traded.

Note that the factor risk premium does not appear in equations (149) and (150). In order to estimate the factor risk premium it is necessary to add the additional moment restriction that the factor risk premium is the expected value of the factor, which is the excess return on some benchmark portfolio of traded assets. As mentioned earlier, it can be verified that the best estimate of the factor risk premium is the sample average of the factor realizations.

When returns and factors exhibit conditional homoscedasticity and independence over time, the GMM estimator is equivalent to the ML estimator suggested by Gibbons, Ross, and Shanken (1989). For details, we refer readers to MacKinlay and Richardson (1991). This implies that GMM estimator of the risk premia in the linear beta model is the most efficient unbiased estimator when stock returns and factors are homoscedastic and independent over time. More importantly, MacKinlay and Richardson (1991) demonstrate that the ML estimation and test are biased when stock returns are conditionally heteroscedastic. They argue that the advantage of GMM is its robustness to the presence of conditional heteroscedasticity and thus recommend estimating the parameters in beta models using GMM.

5.3 Evaluating beta pricing models using the stochastic discount factor representation

We can derive a different set of moment restrictions from the linear beta pricing model. Substituting equation (134) into equation (133) and appropriately rearranging terms, we obtain

$$E\left[\mathbf{R}_t\left(1+\boldsymbol{\delta}'\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}-\boldsymbol{\delta}'\boldsymbol{\Omega}^{-1}\mathbf{f}_t\right)\right]=\mathbf{0}_N. \tag{153}$$

If the factors are traded, we have $1+\boldsymbol{\delta}'\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}=1+\boldsymbol{\mu}'\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}\geq 1$. If the factors are not traded, we assume $1+\boldsymbol{\delta}'\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}\neq 0$. Then, we can use it to divide both sides of equation (153) and obtain

$$E\left[\mathbf{R}_t\left(1-\frac{\boldsymbol{\delta}'\boldsymbol{\Omega}^{-1}}{1+\boldsymbol{\delta}'\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}}\mathbf{f}_t\right)\right]=\mathbf{0}_N. \tag{154}$$

If we transform the vector of risk premia, $\boldsymbol{\delta}$, into a vector of new parameters $\boldsymbol{\lambda}$ as follows

$$\boldsymbol{\lambda}=\frac{\boldsymbol{\Omega}^{-1}\boldsymbol{\delta}}{1+\boldsymbol{\delta}'\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}}, \tag{155}$$

then equation (154) becomes

$$E\left[\mathbf{R}_t\left(1-\boldsymbol{\lambda}'\mathbf{f}_t\right)\right]=\mathbf{0}_N. \tag{156}$$

This moment restriction is often referred to as a stochastic discount factor (SDF) representation of the beta pricing model. The variable $m_t\equiv 1-\boldsymbol{\lambda}'\mathbf{f}_t$ is the stochastic discount factor because $E[\mathbf{R}_t m_t]=\mathbf{0}_N$. In general, a number of random variables m_t satisfying $E[\mathbf{R}_t m_t]=\mathbf{0}_N$ exist and thus, there are more than one stochastic discount factor. The linear factor pricing model (156) designates the random variable m_t to be a linear function of the factors $\mathbf{f}_t$. Dybvig and Ingersoll (1982) who derived the SDF representation of the CAPM and Ingersoll (1987) derives the SDF representation for a number of theoretical asset pricing models. Hansen and Richard (1987) coined the term "stochastic discount factor".

Most asset pricing models have a convenient SDF representation - not just the linear beta pricing model. The SDF representation typically follows from the Euler equations (first order condition) for portfolio choice problems faced by an investor. In general, the Euler equation can be written as, $E_{t-1}[\mathbf{R}_t m_t]=\mathbf{0}_N$ for *excess* returns, where E_{t-1} is the expectation conditional on the information at the end of period $t-1$. This is often referred to as conditional SDF representation. When the utility function in the economic model depends on a vector of parameters, denoted by $\boldsymbol{\theta}$, the stochastic discount factor m_t is a function of $\boldsymbol{\theta}$ and a vector of economic variables $\mathbf{f}_t$. Then, the stochastic discount factor has the form of $m(\boldsymbol{\theta}, \mathbf{f}_t)$. For example, in the representative-agent and endowment economy of Lucas (1978), $m(\boldsymbol{\theta}, f_t) = \rho c_t^a$, where $f_t = c_t$, the growth rate of aggregate consumption, and $\boldsymbol{\theta} = [\rho \quad a]'$. The linear beta model implies $m(\boldsymbol{\theta}, \mathbf{f}_t) = 1 - \boldsymbol{\lambda}'\mathbf{f}_t$, where $\boldsymbol{\theta} = \lambda$ and $\mathbf{f}_t$ is the vector of factors. Bansal, Hsieh and Viswanathan (1993) and Bansal and Viswanathan (1993) specify m as a polynomial function of the factors $\mathbf{f}_t$. We will write the general conditional SDF representation of the pricing model as follows

$$E_{t-1}[\mathbf{R}_t m(\boldsymbol{\theta}, \mathbf{f}_t)] = \mathbf{0}_N. \tag{157}$$

Taking the unconditional expectation, we have

$$E[\mathbf{R}_t m(\boldsymbol{\theta}, \mathbf{f}_t)] = \mathbf{0}_N \tag{158}$$

which includes equation (156) as a special case.

We can test the pricing model using its conditional SDF representation (157) by incorporating conditional information. Let $\mathbf{z}_{t-1}$ be a vector of economic variables observed by the end of period $t-1$. Consider a $M \times N$ matrix denoted $\mathbf{H}(\mathbf{z}_{t-1})$, the elements of which are functions of $\mathbf{z}_{t-1}$. Multiplying $\mathbf{H}(\mathbf{z}_{t-1})$ to both sides of (157) and taking the unconditional expectation, we obtain

$$E\left[\mathbf{H}(\mathbf{z}_{t-1})\mathbf{R}_t m(\boldsymbol{\theta}, \mathbf{f}_t)\right] = \mathbf{0}_N. \tag{159}$$

In testing (157) using $\mathbf{z}_{t-1}$, the common practice is to multiply the returns by the instrumental variables to get $\mathrm{vec}(\mathbf{z}_{t-1} \otimes \mathbf{R}_t)$, generally referred to as the vector of scaled returns. If we choose $\mathbf{H}(\mathbf{z}_{t-1}) = \mathbf{z}_{t-1} \otimes \mathbf{I}_N$, then $\mathbf{H}(\mathbf{z}_{t-1})\mathbf{R}_t = \mathrm{vec}(\mathbf{z}_{t-1} \otimes \mathbf{R}_t)$. If we are only interested in testing the unconditional SDF model (158), we choose the conditional portfolios to be the original assets by setting $\mathbf{H}(\mathbf{z}_{t-1}) = \mathbf{I}_N$. If $m(\boldsymbol{\theta}, \mathbf{f}_t)$ satisfies equation (159), we say $m(\boldsymbol{\theta}, \mathbf{f}_t)$ prices the portfolios correctly and call it a valid SDF. The idea of scaling stock returns by conditional variables was first proposed by Hansen and Singleton (1982).

If we normalize each row of $\mathbf{H}(\mathbf{z}_{t-1})$ to a vector of weights that sum to 1, then the vector $\widetilde{\mathbf{R}}_t = \mathbf{H}(\mathbf{z}_{t-1})\mathbf{R}_t$ is, in fact, the vector of returns on conditional portfolios of stocks. The weights in these portfolios are time-varying. We thus obtain an unconditional SDF representation of the beta pricing model, i.e., the SDF representation becomes $E[\widetilde{\mathbf{R}}_t m(\boldsymbol{\theta}, \mathbf{f}_t)] = \mathbf{0}_M$. Therefore, testing a conditional SDF representation using conditional variables is equivalent to testing an unconditional SDF representation using portfolios managed using conditioning information. For example, portfolios constructed using firm size or book-to-market ratio are portfolios whose weights are managed to vary over time in a particular way, and tests of the unconditional SDF representation using these portfolios can be viewed as tests of conditional SDF representation using firm size or book-to-market ratio as conditioning variables. Therefore, in principle we do not lose any information by testing unconditional moment restrictions because we can always augment the set of assets with portfolios that are managed in clever ways using conditioning information.

The application of GMM to the SDF representation is straightforward. Let $\mathbf{x}_t = \left[\mathbf{R}_t' \;\; \mathbf{f}_t' \;\; \mathbf{z}_{t-1}'\right]'$ and

$$\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta}) = \widetilde{\mathbf{R}}_t m(\boldsymbol{\theta}, \mathbf{f}_t) = \mathbf{H}(\mathbf{z}_{t-1})\mathbf{R}_t m(\boldsymbol{\theta}, \mathbf{f}_t) \; . \tag{160}$$

We can provide consistent estimates of matrices $\mathbf{S}$ and $\mathbf{D}$ using the definition of the function $\mathbf{g}$. We obtain the GMM estimator $\widehat{\boldsymbol{\theta}}_T$ by solving (127), and test the restrictions (159), using the framework in Section 5.1. The sample analogue of the vector of pricing errors, $\mathbf{g}_T(\widehat{\boldsymbol{\theta}})$, is defined by equation (124). This pricing error is analyzed extensively by Hansen and Jagannathan (1997). Jagannathan and Wang (2002) examine the relation between pricing error and Jensen's alpha. It is shown that the former is a linear transformation of the latter in linear beta pricing models. Hodrick and Zhang (2001) compare the pricing errors across a variety of model specifications. The statistic J_T is calculated as in (131). The approach outlined here applies to SDF representation of asset pricing models in general - not just the linear beta pricing model which is the focus of our study in this chapter. In addition, the entries of the matrix $\mathbf{H}(\mathbf{z}_{t-1})$ do not have to correspond to portfolio weights.

Since the SDF representation can be used to represent an arbitrary asset pricing model, not just the beta pricing model, it is of interest to examine the relative efficiency and power of applying the GMM to the SDF and beta representations of the pricing model. Jagannathan and Wang (2002) find that using the SDF representation provides as precise an estimate of the risk premium as that obtained by using the beta representation. Using Monte Carlo simulations, they demonstrate that the two methods provide equally precise estimates in finite samples as well. The sampling errors in the two methods are similar, even when returns have fatter tails relative to the normal distribution allowing for conditional heteroscedasticity. They also examine the specification tests associated with the two approaches and find that these tests have similar power.

Although the J-statistic is useful for testing for model misspecification, comparing the J-statistics across different model specifications may lead to wrong conclusions. One model may do better than another not because the vector of average pricing errors, $\mathbf{g}_T$, associated with it are smaller, but because the inverse of the optimal weighting matrix, $\mathbf{S}_T$, associated with it is smaller. To overcome this difficulty, Hansen and Jagannathan (1997) suggested examining the pricing error of the most mispriced portfolio among those whose second moments of returns are normalized to 1. This corresponds to using the inverse of the second moment matrix of returns,

$\mathbf{G}^{-1} = \left(E[\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t']\right)^{-1}$, as the weighting matrix under the assumption that $\mathbf{G}$ is positive definite. Hansen and Jagannathan (1997) demonstrate that

$$d(\boldsymbol{\theta}) = \left[E[\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta})]'\mathbf{G}^{-1}E[\mathbf{g}(\mathbf{x}_t, \boldsymbol{\theta})]\right]^{\frac{1}{2}} \tag{161}$$

equals the least-square distance between the candidate SDF and the set of all valid SDFs. Further, they show that $d(\boldsymbol{\theta})$ is the maximum pricing error on normalized portfolios of the N assets, generally referred to as the HJ-distance.

For given $\boldsymbol{\theta}$, the sample estimate of the HJ-distance is

$$\hat{d}(\boldsymbol{\theta}) = \left[\mathbf{g}_T(\boldsymbol{\theta})'\mathbf{G}_T^{-1}\mathbf{g}_T(\theta)\right]^{\frac{1}{2}} \tag{162}$$

where $\mathbf{G}_T$ is a consistent estimate of $\mathbf{G}$. An obvious choice for the consistent estimate is

$$\mathbf{G}_T = \frac{1}{T}\sum_{t=1}^{T}\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t' = \frac{1}{T}\sum_{t=1}^{T}\mathbf{H}(\mathbf{z}_{t-1})\mathbf{R}_t\mathbf{R}_t'\mathbf{H}(\mathbf{z}_{t-1})'. \tag{163}$$

To check the model's ability to price stock returns, it is natural to choose the parameter θ that minimizes the estimated HJ-distance. This leads us to choose

$$\widetilde{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta}} \mathbf{g}_T(\boldsymbol{\theta})'\mathbf{G}_T^{-1}\mathbf{g}_T(\boldsymbol{\theta}). \tag{164}$$

It is important to note that $\widetilde{\boldsymbol{\theta}}_T$ is not the same as the GMM estimator $\widehat{\boldsymbol{\theta}}$ because $\mathbf{G}$ is not equal to $\mathbf{S}$. Also, the T-scaled estimate of the square of HJ-distance, $T[\hat{d}(\widetilde{\boldsymbol{\theta}}_T)]^2$, is not equal to the J-statistic, and thus does not follow the distribution $\chi^2(M-K)$ asymptotically.

Following Jagannathan and Wang (1996), it can be shown that the limiting distribution of $T[\hat{d}(\widetilde{\boldsymbol{\theta}}_T)]^2$ is a linear combination of χ^2 distributions, each of which has one degree of freedom. More precisely,

$$T[\hat{d}(\widetilde{\boldsymbol{\theta}}_T)]^2 \xrightarrow{\mathcal{D}} \sum_{j=1}^{M-K} a_j\xi_j, \text{ as } T\to\infty, \tag{165}$$

where $\xi_1, \ldots, \xi_{M-K}$ are independent random variables following $\chi^2(1)$ distributions. The coefficients, $a_1, \ldots, a_{M-K}$, are the nonzero eigenvalues of the matrix

$$\mathbf{A} = \mathbf{S}^{1/2}\mathbf{G}^{-1/2}\left(\mathbf{I}_M - (\mathbf{G}^{-1/2})'\mathbf{D}(\mathbf{D}'\mathbf{G}^{-1}\mathbf{D})^{-1}\mathbf{D}'\mathbf{G}^{-1/2}\right)\mathbf{G}^{-1/2})'(\mathbf{S}^{1/2})', \tag{166}$$

where $\mathbf{S}^{1/2}$ and $\mathbf{G}^{1/2}$ are the upper triangular matrices in the Cholesky decompositions of $\mathbf{S}$ and $\mathbf{G}$. The distribution in (165) can be used to test the hypothesis $d(\boldsymbol{\theta}) = 0$. Applications of the sampling distribution in (165) can be found in Jagannathan and Wang (1996), Buraschi and Jackworth (2001), and Hodrick and Zhang (2001).

5.4 Models with time-varying betas and risk premia

In analysis of stock returns in the cross section, financial economists often consider conditional linear models with time-varying beta and factor risk premia. In that case, we need to consider conditional moments of stock returns and factors in generating the moment restrictions. We need the following additional notation to denote the conditional means and variances of the factors and the conditional betas of the assets.

$$\begin{aligned}\boldsymbol{\mu}_{t-1} &= E_{t-1}\left[\mathbf{f}_t\right] &(167)\\ \boldsymbol{\Omega}_{t-1} &= E_{t-1}\left[(\mathbf{f}_t-\boldsymbol{\mu}_{t-1})(\mathbf{f}_t-\boldsymbol{\mu}_{t-1})'\right] &(168)\\ \mathbf{B}_{t-1} &= E_{t-1}\left[\mathbf{R}_t(\mathbf{f}_t-\boldsymbol{\mu}_{t-1})'\right]\boldsymbol{\Omega}_{t-1}^{-1}\ . &(169)\end{aligned}$$

The conditional analogue of the linear beta pricing model with time-varying risk premia, $\boldsymbol{\delta}_{t-1}$, is then given by

$$E_{t-1}\left[\mathbf{R}_t\right] = \mathbf{B}_{t-1}\boldsymbol{\delta}_{t-1}\ . \tag{170}$$

Obviously, it is impossible to estimate and test the above time-varying model without imposing additional assumptions. Generally, we need to reduce the time-varying unknown parameters to a small number of constant parameters. In this subsection, we discuss five ways of introducing testable hypotheses for GMM in conditional linear beta models with time-varying parameters.

Suppose either the betas or the risk premia, but not both, are constants. To see the implication of this assumption, take the unconditional expectation of both sides of equation (170) to obtain

$$E\left[\mathbf{R}_t\right] = E\left[\mathbf{B}_{t-1}\right]E\left[\boldsymbol{\delta}_{t-1}\right] + Cov\left[\mathbf{B}_{t-1},\boldsymbol{\delta}_{t-1}\right] \tag{171}$$

where $E[\boldsymbol{\beta}_{t-1}]$ and $E[\boldsymbol{\delta}_{t-1}]$ can be interpreted as the average beta and the risk premium. The covariance between beta and the risk premium is $Cov(\mathbf{B}_{t-1},\boldsymbol{\delta}_{t-1}) = E[\mathbf{B}_{t-1}(\boldsymbol{\delta}_{t-1}-E[\boldsymbol{\delta}_{t-1}])]$. If either beta or premium is constant, the covariance term is zero. Then, the unconditional expected return is simply the product of the average beta and the average premium. More generally, if the conditional beta and premium are uncorrelated, the model is essentially equivalent to an unconditional or static model. However, it does not make good economic sense to assume that the conditional beta and the risk premium are uncorrelated because both risk and risk premia are affected by the same pervasive forces that affect the economy as a whole. As a matter of fact, Ang and Liu (2002) demonstrate that this correlation is rather high.

The unconditional correlation between conditional betas and conditional risk premia induces stock returns to be unconditionally correlated with conditional factor risk premia. Therefore, when a conditional linear beta pricing model holds, the unconditional expected return on an asset will be a linear function of the asset's unconditional factor betas and the asset's unconditional betas with respect to the conditional factor risk premia. For example, when the conditional version of the CAPM holds, Jagannathan and Wang (1996) show that the following unconditional two-beta model would obtain:

$$E[\mathbf{R}_t] = \mathbf{B}\mathbf{a} + \mathbf{B}_\delta \mathbf{b}, \tag{172}$$

where the two vector of betas are defined as

$$\mathbf{B} = E[\mathbf{R}_t(\mathbf{f}_t - \boldsymbol{\mu})]\,\boldsymbol{\Omega}^{-1} \tag{173}$$

$$\mathbf{B}_\delta = E[\mathbf{R}_t(\boldsymbol{\delta}_{t-1} - \boldsymbol{\mu}_\delta)]\,\boldsymbol{\Omega}_\delta^{-1}\ . \tag{174}$$

In above equations, $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_\delta$ are unconditional mean of the factors and risk premia, and $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_\delta$ are their unconditional variances. The vectors $\mathbf{a}$ and $\mathbf{b}$ are unknown coefficients.

Since the risk premia are not observable, we still cannot estimate and test the model without additional assumptions. A convenient assumption that is often made in empirical studies is that the risk premium is a function of a prespecified set of observable economic variables, i.e.,

$$\boldsymbol{\delta}_{t-1} = \mathbf{h}_\delta(\boldsymbol{\gamma}_\delta, \mathbf{z}_{t-1}) \tag{175}$$

where $\mathbf{z}_{t-1}$ is a vector of economic variables observed at the end of period $t-1$, and $\boldsymbol{\gamma}_\delta$ is a vector of unknown parameters. For example, when studying the conditional CAPM, Jagannathan and Wang (1996) choose $\mathbf{z}_{t-1}$ to be the default spread of corporate bonds and $\mathbf{h}_\delta$ a linear function. Then, equation (172)-(175) can be transformed into an unconditional SDF model:

$$\textbf{Hypothesis I}: \quad E\left[\mathbf{R}_t\left(1 + \boldsymbol{\vartheta}'\mathbf{f}_t + \boldsymbol{\vartheta}_\delta'\mathbf{h}_\delta(\boldsymbol{\gamma}_\delta, \mathbf{z}_{t-1})\right)\right] = \mathbf{0}_N, \tag{176}$$

where $\boldsymbol{\vartheta}$ and $\boldsymbol{\vartheta}_\delta$ are vectors of unknown parameters. This model can be estimated and tested using GMM as discussed in Section 5.3. When factors are traded, the time-varying risk premia are equal to the conditional expectation of factors, i.e., $\boldsymbol{\delta}_{t-1} = \boldsymbol{\mu}_{t-1}$. Then, we can replace equation (175) by a prediction model for factors such as

$$\boldsymbol{\mu}_{t-1} = \mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}), \tag{177}$$

where $\boldsymbol{\gamma}_\mu$ is a vector of unknown parameters. The prediction model can be estimated and tested together with equation (176). That is, we can use GMM to estimate and test the following moment restriction:

$$E\left[\begin{pmatrix} \mathbf{f}_t - \mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}) \\ \mathbf{R}_t\left(1 + \vartheta'\mathbf{f}_t + \vartheta_\delta'\mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1})\right) \end{pmatrix}\right] = \mathbf{0}_{K+N}. \tag{178}$$

The first part of the equation is the restriction implied by the prediction model $\boldsymbol{\mu}_{t-1} = \mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1})$.

Instead of testing unconditional SDF representation of the conditional linear beta pricing model using Hypothesis I, we can directly obtain a conditional SDF representation from equation (170) and derive an alternative test by making additional assumptions about the time-varying parameters. Equation (176) implies

$$E_{t-1}\left[\mathbf{R}_t\left(1 + \boldsymbol{\delta}'_{t-1}\boldsymbol{\Omega}_{t-1}^{-1}\boldsymbol{\mu}_{t-1} - \boldsymbol{\delta}'_{t-1}\boldsymbol{\Omega}_{t-1}^{-1}\mathbf{f}_t\right)\right] = \mathbf{0}_N. \tag{179}$$

To make this model testable, assume that (177) holds and,

$$\boldsymbol{\Omega}_{t-1}^{-1}\boldsymbol{\delta}_{t-1} = \mathbf{h}_\omega(\boldsymbol{\gamma}_\omega, \mathbf{z}_{t-1}), \tag{180}$$

where $\boldsymbol{\gamma}_\omega$ is a vector of unknown parameters. When there is only one factor, $\boldsymbol{\Omega}_{t-1}^{-1}\boldsymbol{\delta}_{t-1}$ is the reward-to-variability ratio. Then, equation (179) becomes

$$E_{t-1}\left[\mathbf{R}_t\left(1 + \mathbf{h}_\omega(\boldsymbol{\gamma}_\omega, \mathbf{z}_{t-1})'\mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}) + \mathbf{h}_\omega(\boldsymbol{\gamma}_\omega, \mathbf{z}_{t-1})'\mathbf{f}_t\right)\right] = \mathbf{0}_N \tag{181}$$

We can estimate and test the conditional moment restriction given below using the GMM as described in Section 5.3:

$$\textbf{Hypothesis II}: \quad E_{t-1}\left[\begin{pmatrix} \mathbf{f}_t - \mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}) \\ \mathbf{R}_t m_t \end{pmatrix}\right] = \mathbf{0}_{K+N}, \tag{182}$$

where $m_t = 1 + \mathbf{h}_\omega(\boldsymbol{\gamma}_\omega, \mathbf{z}_{t-1})'\mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}) + \mathbf{h}_\omega(\boldsymbol{\gamma}_\omega, \mathbf{z}_{t-1})'\mathbf{f}_t$. The advantage of this approach is that we make use of the information in the conditioning variables $\mathbf{z}_{t-1}$. The cost is that we have to make assumption (180), which is rather difficult to justify. In empirical studies using this approach it is commonly assumed that $\mathbf{h}_\mu$ is a linear function of $\mathbf{z}_{t-1}$ and that $\mathbf{h}_\omega$ is constant. When factors are traded, assumption (180) implies

$$\mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}) = \boldsymbol{\Omega}_{t-1}\mathbf{h}_\omega(\boldsymbol{\gamma}_\omega, \mathbf{z}_{t-1}). \tag{183}$$

Letting $\mathbf{u}_t = \mathbf{f}_t - \mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1})$, we can rewrite the above equation as

$$E_t\left[\mathbf{u}_t\mathbf{u}_t'\mathbf{h}_\omega(\boldsymbol{\gamma}_\omega, \mathbf{z}_{t-1}) - \mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1})\right] = \mathbf{0}_K, \tag{184}$$

which can be added to the moment restriction in Hypothesis II. Harvey (1989) rejects the hypothesis under the assumption that $\mathbf{h}_\mu$ is linear and $\mathbf{h}_\omega$ is constant.

The third approach uses the conditional SDF representation and simply assumes that all the time-varying parameters are functions of a few chosen observable conditioning variables observed by the econometrician. In that case the conditional beta model (170) implies the following conditional SDF representation

$$E_{t-1}\left[\mathbf{R}_t\left(1-\boldsymbol{\lambda}_{t-1}'\mathbf{f}_t\right)\right]=\mathbf{0}_N \quad \text{where} \quad \boldsymbol{\lambda}_{t-1}=\frac{\boldsymbol{\Omega}_{t-1}^{-1}\boldsymbol{\delta}_{t-1}}{1+\boldsymbol{\delta}_{t-1}'\boldsymbol{\Omega}_{t-1}^{-1}\boldsymbol{\mu}_{t-1}}. \tag{185}$$

In this approach, we assume that $\boldsymbol{\lambda}_{t-1}$ is a function of $\mathbf{z}_{t-1}$ up to some unknown parameters denoted by $\boldsymbol{\gamma}_\lambda$. That is, we assume

$$\frac{\boldsymbol{\delta}_{t-1}'\boldsymbol{\Omega}_{t-1}^{-1}}{1+\boldsymbol{\delta}_{t-1}'\boldsymbol{\Omega}_{t-1}^{-1}\boldsymbol{\mu}_{t-1}}=\mathbf{h}_\lambda(\boldsymbol{\gamma}_\lambda,\mathbf{z}_{t-1})\ . \tag{186}$$

Using GMM, we can test the conditional moment restriction:

$$\textbf{Hypothesis III}: \quad E_{t-1}\left[\mathbf{R}_t\left(1-\mathbf{h}_\lambda(\boldsymbol{\gamma}_\lambda,\mathbf{z}_{t-1})\mathbf{f}_t\right)\right]=\mathbf{0}_N. \tag{187}$$

Cochrane (1996) assumes a linear function $\boldsymbol{\lambda}'=\mathbf{a}'+\mathbf{z}_{t-1}'\mathbf{C}$, where $\mathbf{a}$ is a vector of unknown parameters and $\mathbf{C}$ is a matrix of unknown parameters. Then, the hypothesis becomes

$$E_{t-1}\left[\mathbf{R}_t\left(1-\mathbf{a}'\mathbf{f}_t-(\mathbf{f}_t'\otimes\mathbf{z}_{t-1})\mathrm{vec}(\mathbf{C})\right)\right]=\mathbf{0}_N. \tag{188}$$

In the empirical literature the convention is to refer to $\mathbf{f}_t'\otimes\mathbf{z}_{t-1}$ as the scaled factors. Equation (188) can be viewed as a linear conditional SDF representation with scaled factors included as a subset of the factors. Estimation and testing of this model can be carried out using GMM as described in Section 5.3. The advantage of this approach is that it is straightforward and simple. The disadvantage of this approach is that assumption (186) often lacks economic intuition and it is not clear how to test the assumption.

None of the approaches discussed so far involve making assumptions about the laws governing the temporal evolution of $\mathbf{B}_{t-1}$ and $E_{t-1}[\mathbf{R}_t]$. However, assumptions about the dynamics of the conditional factor risk premia impose restrictions on the joint dynamics of the conditional expected stock returns and the conditional factor betas. Here onwards, for convenience, we will assume that all the factors are traded. Then, we have $\boldsymbol{\delta}_{t-1}=\boldsymbol{\mu}_{t-1}$. Equation (177) and the conditional beta pricing model (170) implies

$$E_{t-1}\left[\mathbf{R}_{t-1}\right]=\ \mathbf{B}_{t-1}\mathbf{h}_\mu(\boldsymbol{\gamma}_\mu,\mathbf{z}_{t-1}) \tag{189}$$

$$\mathbf{B}_{t-1}=E_{t-1}\left[\mathbf{R}_t\mathbf{u}_{t-1}'\right]\left(E_{t-1}\left[\mathbf{u}_{t-1}\mathbf{u}_{t-1}'\right]\right)^{-1}, \tag{190}$$

where $\mathbf{u}_{t-1}=\mathbf{f}_t-\mathbf{h}_\mu(\boldsymbol{\gamma}_\mu,\mathbf{z}_{t-1})$. It would be rather difficult to verify whether assumptions about the dynamics of conditional betas conditional expected stock returns are consistent with the restrictions imposed by the above equations. Hence, it would be advisable to make assumptions regarding the dynamics of only two of the three groups of variables, conditional factor risk premia,conditional factor betas and conditional expected return on assets.

Some empirical studies assume that $E_{t-1}[\mathbf{R}_t]$ is a prespecified function of only a few conditioning variables, $\mathbf{z}_{t-1}$, i.e.,

$$E_{t-1}\left[\mathbf{R}_t\right] = \mathbf{h}_r(\boldsymbol{\gamma}_r, \mathbf{z}_{t-1}). \tag{191}$$

In empirical studies, $\mathbf{z}_{t-1}$ usually contains only a small number of macroeconomic variables that help predict future factor realizations. Note that equation (191) is a rather strong assumption, since it requires these variables to be sufficient for predicting future returns on every stock. Therefore, before testing a conditional beta pricing model we need to examine whether assumptions (191) and (177) are reasonable. For this purpose we may use the GMM to test the following moment restrictions:

$$E\left[\mathbf{H}(\mathbf{z}_{t-1})\begin{pmatrix} \mathbf{R}_t - \mathbf{h}_r(\boldsymbol{\gamma}_r, \mathbf{z}_{t-1}) \\ \mathbf{f}_t - \mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}) \end{pmatrix}\right] = \mathbf{0}_N, \tag{192}$$

where $\mathbf{H}(\mathbf{z}_{t-1})$ is an $M \times (N+K)$ matrix of functions of $\mathbf{z}_{t-1}$, as described in Section 5.3.

Equation (191) imposes the following restrictions on the dynamics of conditional factor betas through the beta pricing model (170)

$$\mathbf{h}_r(\boldsymbol{\gamma}_r, \mathbf{z}_{t-1}) = \mathbf{B}_{t-1}\mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}) \ . \tag{193}$$

Some empirical studies of conditional beta pricing models assume that conditional factor betas are specific functions of a few prespecified conditioning variables, i.e.,

$$\mathbf{B}_{t-1} = \mathbf{h}_\beta(\boldsymbol{\gamma}_\beta, \mathbf{z}_{t-1}) \ . \tag{194}$$

In that case it is necessary to ensure that this assumption does not conflict with equation (193). For example, Ferson and Korajczyk (1995) assume (177) and (191) and take $\mathbf{h}_\mu$ and $\mathbf{h}_r$ as linear functions. Then, the dynamics of conditional betas in (194) cannot be captured by a linear function of the conditioning variables $\mathbf{z}_{t-1}$ since $\mathbf{h}_r$ would then be a quadratic function. To understand the nature of the restrictions on the dynamics of conditional betas, let us consider the special case where $\mathbf{z}_{t-1}$ is a scalar and $\mathbf{h}_r(\boldsymbol{\gamma}_r, \mathbf{z}_{t-1}) = \boldsymbol{\gamma}_r\mathbf{z}_{t-1}$, and $\mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}) = \boldsymbol{\gamma}_\mu\mathbf{z}_{t-1}$, where $\boldsymbol{\gamma}_r$ and $\boldsymbol{\gamma}_\mu$ are $N \times 1$ and $K \times 1$ vectors respectively. In that case, $\mathbf{B}_{t-1}$ is a constant and does not vary with time since $\mathbf{z}_{t-1}$ can be cancelled out from both sides of equation (193) yielding $\boldsymbol{\gamma}_r = \mathbf{B}_{t-1}\boldsymbol{\gamma}_\mu$. We can use GMM to test the following moment restrictions implied by (177) and (170)

$$\textbf{Hypothesis IV}: \quad E_{t-1}\left[\begin{pmatrix} \mathbf{f}_t - \mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}) \\ \mathbf{R}_t - \mathbf{h}_r(\boldsymbol{\gamma}_r, \mathbf{z}_{t-1}) \\ \mathbf{h}_r(\boldsymbol{\gamma}_r, \mathbf{z}_{t-1}) - \mathbf{h}_\beta(\boldsymbol{\gamma}_\beta, \mathbf{z}_{t-1})\mathbf{h}_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}) \end{pmatrix}\right] = \mathbf{0}_{K+2N}. \tag{195}$$

The main problem with this approach is that the specification for $\mathbf{h}_\beta$ can be inconsistent with the specifications for $\mathbf{h}_\mu$ and $\mathbf{h}_r$. Using international stocks, Ferson and Harvey (1993) reject the above hypothesis under the assumption that all the three functions, $\mathbf{h}_r$, $\mathbf{h}_\delta$ and $\mathbf{h}_\beta$, are linear. For the one-factor CAPM, Ghysels (1998) compares the mean square prediction error of various model specifications. Not surprisingly, the assumption of linear dynamics for conditional beta gives the worst results. This suggests that assumption (191) may not be appropriate. We may therefore test equations (194) and (189) and omit the second group of equations in (195).

When there is only one factor in the model, the restriction on conditional beta can be tested without specifying beta dynamics - i.e., equations (193), (177) and (191) can be estimated and tested without using equation (194). Substituting (190) into equation (193) and using the fact that f_t is a scalar, we obtain

$$\mathbf{h}_r(\boldsymbol{\gamma}_r, \mathbf{z}_{t-1}) = E_{t-1}\left[\mathbf{R}_t u_t\right]\left(E_{t-1}\left[u_t^2\right]\right)^{-1} h_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}), \tag{196}$$

where $u_t = f_t - h_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1})$. Multiplying $E_{t-1}[u_t^2]$ to both sides and rearranging terms, we get

$$E_{t-1}\left[u_t^2\mathbf{h}_r(\boldsymbol{\gamma}_r, \mathbf{z}_{t-1}) - \mathbf{R}_t u_t h_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1})\right] = \mathbf{0}_N. \tag{197}$$

Therefore, we can use GMM to test the following conditional moment restrictions:

$$\textbf{Hypothesis V}: \quad E_{t-1}\left[\begin{pmatrix} f_t - h_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}) \\ \mathbf{R}_t - \mathbf{h}_r(\boldsymbol{\gamma}_r, \mathbf{z}_{t-1}) \\ u_t^2\mathbf{h}_r(\boldsymbol{\gamma}_r, \mathbf{z}_{t-1}) - \mathbf{R}_t u_t h_\mu(\boldsymbol{\gamma}_\mu, \mathbf{z}_{t-1}) \end{pmatrix}\right] = \mathbf{0}_{1+2N}. \tag{198}$$

Although this hypothesis has the advantage that a stand on beta dynamics does not have to be taken, it has two disadvantages. First, it assumes equation (191), which is a prediction model for stock returns. Second, the asset pricing model can only have one factor. Using both U.S. stock returns, Harvey (1989) rejects the hypothesis under the assumption that $\mathbf{h}_r$ and $\mathbf{h}_\mu$ are both linear functions.

Ghysels (1998) points out that we should be cautious about over fitting data when evaluating beta pricing models with time varying parameters. When over fitting happens, the GMM test would not able to reject the model. In that case, it is likely that the constant parameters specified in the model would exhibit structural breaks. Therefore Ghysels suggests using the supLM test developed by Andrews (1993) to check the stability of all constant coefficients when a model specification passes the GMM test, and as an illustration tests linear versions of Hypotheses IV and V. To ensure that over fitting does occur in the examples he considers, he uses only one asset in each test. He shows that most of the J-statistics are not significant but many supLM statistics are significant. These results suggest that it would be advisable to conduct GMM tests using a large number of assets to rule out over fitting, especially for models containing many factors. Since empirical studies in this area typically use returns on a large collection of assets, over fitting is unlikely to be an issue in practice. For example, Ferson and Harvey (1993) and Harvey (1989) reject the linear versions of Hypotheses IV and using GMM.

Table 1 compares features of the five types of hypothesis we discussed. As can be seen, Hypothesis I involves minimal assumptions, but the moment restriction does not make use of conditioning information. As we have argued earlier, this need not be a serious limitation when the collection of assets available to the econometrician include portfolios managed in clever ways using conditioning information. The most important distinction of Hypotheses I, II and III from Hypotheses IV and V is that they do not require additional assumptions regarding the dynamics of conditional expected stock returns $E_{t-1}[\mathbf{R}_t]$. Hypotheses I, II and III allow the conditional expected stock returns to be completely determined by the conditional beta pricing model relation. When Hypotheses IV and V are rejected by GMM, further investigation is necessary to identify whether the rejection is due to the prediction model or the asset pricing model being wrong. Another important difference is that Hypotheses I, II and III can be viewed as special cases of SDF representation of the conditional linear beta pricing model.

Table 1: Comparison of testable hypotheses for GMM.

Type of hypothesis	Additional assumption on	Moment restrictions	Factor type & dimension	Example of studies
I	$\boldsymbol{\delta}_{t-1}$	unconditional	nontraded/multi	Jagannathan & Wang (1996)
II	$\boldsymbol{\mu}_{t-1}$, $\boldsymbol{\Omega}_{t-1}^{-1}\boldsymbol{\delta}_{t-1}$	conditional	nontraded/multi	Harvey (1989)
III	$\frac{\boldsymbol{\delta}'_{t-1}\boldsymbol{\Omega}_{t-1}^{-1}}{1+\boldsymbol{\delta}'_{t-1}\boldsymbol{\Omega}_{t-1}^{-1}\boldsymbol{\mu}_{t-1}}$	conditional	nontraded/multi	Cochrane (1996)
IV	$\boldsymbol{\mu}_{t-1}$, $E_{t-1}[\mathbf{R}_t]$, $\mathbf{B}_{t-1}$	conditional	traded/multi	Ferson & Harvey (1993)
V	$\boldsymbol{\mu}_{t-1}$, $E_{t-1}[\mathbf{R}_t]$	conditional	traded/single	Harvey (1991)

6 Conclusions

Linear beta pricing models have received wide attention in the asset pricing literature. In this chapter we reviewed econometric methods that are available for empirical evaluation of linear beta pricing models using time series observations on returns and characteristics on a large collection of financial assets. The econometric methods can be grouped into three classes: the two stage cross sectional regression method; the maximum likelihood method; and the generalized method of moments. Shanken (1992) showed that the cross sectional method and the maximum likelihood method are asymptotically equivalent when returns are drawn from an i.i.d. joint normal distribution. Under that condition MacKinlay and Richardson (1991) showed that the maximum likelihood method and the generalized method of moments are also asymptotically equivalent. The generalized method of moments, however, has advantages when returns are not jointly normal and exhibit conditional heteroscedasticity.

In general the number of assets will be large relative to the length of the time series of return observations. The classical approach to reducing the dimensionality, without losing too much information, is to use the portfolio grouping procedure of Black, Jensen and Scholes (1972). Since portfolio betas are estimated more precisely than individual security betas, the portfolio grouping procedure attenuates the errors in variables problem faced by the econometrician when using the classical two stage cross sectional regression method. The portfolio formation method can highlight or mask characteristics in the data that have valuable information about the validity or otherwise of the asset pricing model being examined. Hence the econometrician has to exercise care to avoid the data snooping biases discussed in Lo and MacKinlay (1990).

As Brennan, Chordia and Subrahmanyam (1998) observe, it is not necessary to group securities into portfolios to minimize the errors in variables problem when all the factors are excess returns on traded assets by working with risk adjusted returns as dependent variables. However, the advantages to working directly with security returns instead of first grouping securities into portfolios have not been fully explored in the literature.

When the linear beta pricing model holds, expected returns on every asset is a linear function of factor betas. The common practice for examining model misspecification using the cross-sectional regression method is to include security characteristics like relative size and book to price ratios as additional explanatory variables. These characteristics should not have any explanatory power when the pricing model is correctly specified. An alternative approach would be to test for linearity using multivariate tests. The former tests have the advantage that the alternative hypotheses provide valuable information as to what may be missing in the model. The latter tests, on the other hand, will have more power in general to detect any type of model misspecification. Hence both types of tests may have to be used in conjunction, in order to understand the dimensions along which a given model performs well when confronted with data.

References

Amemiya, T., 1985, *Advanced Econometrics*, Harvard University Press, Cambridge, Mass.

Andersen, T.W., 1984, *An introduction to multivariate statistical analysis*, 2nd Edition, John Wiley, New York.

Andrews, D. W. K., 1991, Heteroscedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817-858.

Andrews, D. W. K., 1993, Tests for parameter instability and structural change with unknown change point, *Econometrica* 61, 821-856.

Andrews, D. W. K., and J. C. Monahan, 1992, An improved heteroscedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica* 60, 953-966.

Ang, A., and J. Liu, 2002, How to discount cash flows with time-varying expected returns, Working Paper, Graduate School of Business, Columbia University.

Barone-Adesi, P., and P. Talwar, 1983, Market models and heteroscedasticity of security returns, *Journal of Business and economic statistics* 4, 163-168.

Bansal, R., D. A. Hsieh, and S. Viswanathan, 1993, A new approach to international arbitrage pricing, *Journal of Finance* 48, 1719-1747.

Bansal, R., and S. Viswanathan, 1993, No arbitrage and arbitrage pricing: a new approach, *Journal of Finance* 48, 1231-1262.

Banz, R. W., 1981, The relationship between returns and market value of common stocks, *Journal of Financial Economics* 9, 3-18.

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

Black, F., M. C. Jensen, and M. Scholes, 1972, The capital asset pricing model: some empirical tests, *Studies in the Theory of Capital Markets*, M. C. Jensen, ed, New York: Praeger, 79-121.

Blattberg, R. C., and N. J. Gonedes, 1974, A comparison of the stable and student distributions as statistical models of stock prices, *Journal of Business* 47, 244-280.

Bollerslev, T., R. F. Engle, and Jeffrey M. Wooldridge, 1988, A capital asset pricing model with time varying covariances, *Journal of Political Economy* 96, 116-131.

Breeden, D., M. Gibbons, and R. Litzenberger, 1989, Empirical tests of the consumption-oriented CAPM, *Journal of Finance* 44, 231-262.

Brennan, M.J., T. Chordia, and A. Subrahmanyam, 1998, Alternative factor specifications, security characteristics, and the cross section of expected returns, *Journal of Financial Economics* 49, 345-373.

Buraschi A., and J. Jackworth, 2001, The price of a smile: hedging and spanning in option markets, *Review of Financial Studies* 14, 495-527.

Campbell, John Y., 1993,Intertemporal asset pricing without consumption data, *American Economic Review* 83, 487-512.

Campbell, John Y., 1996, Understanding risk and return, *Journal of Political Economy* 104, 298-345.

Campbell, S. L., and C. D. Meyer, Jr. 1979, *Generalized inverses of linear transformations*, Dover, New York.

Chan, L., Y. Hamao, and J. Lakonishok, 1991, Fundamentals and stock returns in Japan, *Journal of Finance* 46, 1739-1764.

Chen, N., R. Roll, and S. A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383-404.

Cochrane, J. H., 1996, A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy* 104, 572-621.

Connor, G., 1984, A unified beta pricing theory, *Journal of Economic Theory* 34, 13-39.

Connor, G., and R. A. Korajczyk, 1986, Performance measurement with the arbitrage pricing theory: A new framework for analysis, *Journal of Financial Economics* 15, 373-394.

Connor, G., and R. A. Korajczyk, 1993, A test for the number of factors in an approximate factor model, *Journal of Financial Economics* 48, 1263-1291.

Daniel, K., and S. Titman, 1997, Evidence on the characteristics of cross-sectional variation of stock returns, *Journal of Finance* 52, 1-33.

Davidson, J., 1994, *Stochastic limit theory: an introduction for econometricians*, Oxford University Press, New York.

Davis, J. L., 1994, The cross-section of realized stock returns: the pre-COMPUSTAT evidence, *Journal of Finance* 49, 1579-1593.

Dybvig, P. H., and J. E. Ingersoll, 1982, Mean-variance theory in complete markets, *Journal of Business* 55, 233-252.

Fama, E. F., 1965, The behavior of stock market prices, *Journal of Business* 38, 34-105.

Fama, E. F., and K. French, 1992, The cross-section of expected sock returns, *Journal of Finance* 47, 427-466.

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

Fama, E. F., and K. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55-84.

Fama, E. F., and J. D. MacBeth, 1973, Risk, return and equilibrium: empirical tests, *Journal of Political Economy* 81, 607-636.

Ferson, W. E., and S. R. Foster, 1995, Finite sample properties of the generalized method of moments tests of conditional asset pricing models, *Journal of Financial Economics* 36, 29-55.

Ferson, W. E., and C. R. Harvey, 1993, The risk and predictability of international equity returns, *Review of Financial Studies* 6, 527-566.

Ferson, W. E., and R. A. Korajczyk, 1995, Do arbitrage pricing models explain the predictability of stock returns?, *Journal of Business* 68, 309-349.

Ferson, W. E., and R. Jagannathan, 1996, Econometric Evaluation of Asset Pricing Models, Handbook of Statistics, Vol 14: Statistical Methods in Finance, Edited by G.S. Maddala, Elsevier.

Ghysels, E., 1998, On stable factor structures in the pricing of risk: do time varying betas help or hurt?, *Journal of Finance* 53, 549-574.

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

Gibbons, M, S. Ross, and J. Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121-1152.

Hall, P., and C.C. Heyde, 1980, *Martingale limit theory and its application*, Academic Press, New York.

Hamilton, J. D., 1994, *Time Series Analysis*, Princeton University Press.

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

Hansen, L. P., and R. J. Hodrick, 1980, Forward exchange rates as optimal predictors of future spot rates: an econometric analysis, *Journal of Political Economy* 88, 829-853.

Hansen, L. P., and R. Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225-262.

Hansen, L. P., and R. Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 62, 557-590.

Hansen, L. P., and S. F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587-613.

Harvey, C. R., 1989, Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics* 24, 289-318.

Harvey, C. R., 1991, The world price of covariance risk, *Journal of Finance* 46, 111-157.

Hodrick, R., and X. Zhang, 2001, Assessing specification errors in stochastic discount factor models, *Journal of Financial Economics* 62, 327-376.

Horn, R. A., and Johnson C.R., 1990, *Matrix analysis*, Cambridge University Press, New York.

Huberman, G., and S. Kandel, 1987, Mean-variance spanning, *Journal of Finance* 42, 873-888.

Ingersoll, J., 1987, *Theory of financial decision making*, Rowman & Littlefield, Totowa, New Jersey.

Jagadeesh, N., 1992, Does market risk really explain the size effects? *Journal of Financial and Quantitative Analysis* 27, 337-351.

Jagannathan, R., and E. McGrattan, 1995, The CAPM Debate, *Federal Reserve Bank of Minneapolis Quarterly Review*, Vol. 19, No 4, Fall 1995, 2-17.

Jagannathan, R., G. Skoulakis, and Z. Wang, 2002, Testing linear asset pricing factor models using cross-sectional regressions and security characteristics, Working Paper, Kellogg School of Management, Northwestern University.

Jagannathan, R., G. Skoulakis, and Z. Wang, 2002, Generalized Method of Moments: Applications in Finance, *Journal of Business and Economic Statistics* 20, 470-481.

Jagannathan, R., and Z. Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3-53.

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

Jagannathan, R., and Z. Wang, 2002, Empirical evaluation of asset pricing models: a comparison of the SDF and beta methods, *Journal of Finance* 57, forthcoming.

Jobson, D., and R. Korkie, 1982, Potential performance and tests of portfolio efficiency, *Journal of Financial Economics* 10, 433-436.

Jobson, D., and R. Korkie, 1985, Some tests of linear asset pricing with multivariate normality, *Canadian Journal of Administrative Sciences* 2, 114-138.

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

Kan, R., and G. Zhou, 2002, Tests of mean-variance spanning, Working paper, Olin School of Business, Washington University.

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

Kandel, S., and R. F. Stambaugh, 1995, Portfolio efficiency and the cross section of expected returns, *Journal of Finance* 50, 157-184.

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

Kothari, S. P., J. Shanken, and R. G. Sloan, 1995, Another look at the cross-section of expected stock returns, *Journal of Finance* 50, 185-224.

Lehmann, B. N., and D. M. Modest, 1985, The empirical foundations of the arbitrage pricing theory II: The optimal construction of basis portfolios, Working Paper No. 1726, NBER.

Lehmann, B.N., and D.M.Modest, 1988, The empirical foundations of the arbitrage pricing theory, *Journal of Financial Economics* 21, 213-254.

Lintner, J., 1965, The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13-37.

Litzenberger, R., and Ramaswamy K., 1979, Dividends, short-selling restrictions, tax induced investor clienteles and market equilibrium, *Journal of Financial Economics* 7, 163-196.

Lo, A. W., and A. C. MacKinley, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431-467.

MacBeth, J., 1975, Tests of the two parameter model of capital market equilibrium, Ph.D. Dissertation, University of Chicago.

MacKinlay, A. C., 1987, On multivariate tests of the CAPM, *Journal of Financial Economics* 18, 341-372.

MacKinlay, A. C., and M. P. Richardson, 1991, Using generalized method of moments to test mean-variance efficiency, *Journal of Finance* 46, 551-527.

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

Roll, R., 1977, A critique of the asset pricing theory's tests; part I: On past and potential testability of the theory, *Journal of Financial Economics* 4, 129-176.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market efficiency, *Journal of Portfolio Management* 11, 9-17.

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

Schott, J. R., 1997, *Matrix analysis for statistics*, John Wiley, New York.

Schwert, G. William, 1983, Size and stock returns and other empirical regularities, *Journal of Financial Economics* 12, 3-12.

Schwert, W. G., and P. Seguin, 1990, Heteroscedasticity in stock returns, *Journal of Finance* 45, 1129-1155.

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

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

Shanken, J., 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1-33.

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

Stattman, Dennis, 1980, Book values and stock returns, The Chicago MBA: A Journal of Selected Papers, 4, 25-45.

Zhou, G., 1991, Small sample tests of portfolio efficiency, *Journal of Financial Economics* 30, 165-191.