Assessment of Asset-Pricing Models Using Cross-Sectional Regressions

Georgios Skoulakis*
Finance Department
Northwestern University

May 14, 2005

ABSTRACT

This paper provides a comprehensive treatment of the econometric evaluation of linear beta asset-pricing models based on the cross-sectional regression (CSR) method. The asymptotic distribution of the risk premia estimator is obtained, along with the optimal choice of the CSR weighting matrix. Building on a theoretical mispricing measure defined in the spirit of Hansen and Jagannathan (HJ) (1997), we develop a specification test that generalizes the CSR test of Shanken (1985) allowing for arbitrary choice of the weighting matrices in the CSR and the quadratic form. When the weighting matrix in the quadratic form is the inverse of the second moment matrix of the returns, the distance equals the suitably defined maximum pricing error per unit norm of portfolio return. The test based on the corresponding statistic has the theoretical appeal of the HJ distance test and robust small-sample statistical properties. Finally, comparison of the CSR method with the generalized method of moments (GMM) approach shows that when the two methods are applied optimally they yield asymptotically equivalent estimators of the parameters of interest.

*Department of Finance, Kellogg School of Management, Northwestern University. I am grateful to Torben Andersen, Daniel Cohen, Bob Korajczyk, Ronnie Sadka, Costis Skiadas, Arne Staal, and, especially Ravi Jagannathan for discussions, comments and suggestions. Any errors are my responsibility. Comments and suggestions are welcome at g-skoulakis@northwestern.edu.

1 Introduction

One of the main themes of the asset-pricing literature is the wide cross-sectional variation of historical average returns on financial assets. For example, during the 1926-1999 period, large capitalization stocks earned an annualized average return of 13.0 percent whereas long-term bonds earned only 5.6 percent. At the same time, small capitalization stocks earned 18.9 percent which is substantially higher than the return on large stocks. These differences are statistically and economically significant as documented, for instance, by Jagannathan and McGrattan (1995).[1] Furthermore, such significant differences in average returns are observed among other classes of stocks. The natural question that arises is: What drives these pronounced differences? Many authors have addressed this very question. A variety of asset-pricing models have been advanced that seek to enhance our understanding as to why different assets generate different rates of return. Linear factor pricing models form a prominent class of such models. The theoretical proposition delivered by these models is, in a nutshell, that a small number of pervasive factors are adequate to represent economy-wide aggregate risk, and that the expected return on an asset is a linear function of its factor betas.[2]

The main difference among the various models lies in the way they determine the important factors. There are, broadly speaking, three main approaches to the issue of factor selection. Some models specify factors based on equilibrium arguments. The most important factor of this type is the return on the market portfolio, which is based on the Capital Asset Pricing Model (CAPM) first derived by Sharpe (1964). Other models specify factors based on economic intuition. Examples of such factors are term premium, default premium, the growth rate of industrial production, and inflation as suggested by Chen, Roll and Ross (1986), and the size and book-to-market factors as proposed by Fama and French (1992). Another important approach to factor specification is based on empirical analysis. Systematic econometric procedures using principal components and factor analysis are established by Connor and Korajczyk (1988) and Lehmann and Modest (1988).[3] The aim of this paper is to provide a comprehensive procedure for econometric evaluation of linear factor models based on the cross-sectional regression (CSR) method.

Specification tests of linear beta-pricing models are conducted through a number of different procedures. Typically, researchers examine whether factor betas have statistically significant non-zero risk premiums or whether firm characteristics not related to factor covariance risk have non-zero rewards. A successful approach to this exercise is to use the CSR method, which is both intuitive and

[1] The documentation of the "small-firm" effect goes back to Banz (1981).
[2] See, for instance, the classic works of Ross (1976), Chamberlain and Rothschild (1983), and Connor (1984).
[3] In a more recent work, Jones (2001) shows how to extract factors from heteroscedastic asset returns.

easy to implement. These two features constitute its main advantages and explain its widespread popularity. In this paper, we also demonstrate that the CSR methodology is competitive with respect to approaches based on the generalized method of moments (GMM) in terms of efficiency and can be used to develop specification tests for the asset-pricing model under consideration. The CSR method was first employed for testing the validity of the CAPM in the early 1970s. Leading examples of this line of research are the seminal papers by Fama and MacBeth (1973) and Black, Jensen and Scholes (1972). According to the CSR method, estimation and testing of linear factor pricing models is conducted in two passes. In the first pass, factor betas are estimated using time-series regression. In the second pass, the factor risk premiums are computed using cross-sectional regression of stock returns on the factor betas estimated from the first pass. The use of portfolios instead of individual stocks substantially increases the precision with which factor betas are estimated in the first pass, as pointed out by Black, Jensen and Scholes (1972).

Fama and MacBeth (1973) suggest obtaining an estimate of the parameters for each period using CSR, and treating the obtained estimates as being independent samples of the population factor risk premiums and characteristics. This procedure, however, ignores the resulting error-in-variables problem that results from the fact that in the second pass we use beta estimates instead of the true betas. Shanken (1992) provides the first comprehensive analysis of the statistical properties of the CSR two-pass estimator, under the assumption that the time-series regression innovations exhibit conditional homoscedasticity given the factors. He shows how to take into account the sampling errors in the betas estimated in the first pass. Jagannathan and Wang (1998) extend the analysis of Shanken (1992) allowing for conditional heteroscedasticity and incorporating security-specific characteristics as a tool for detecting of model misspecification.

There are at least two different ways to represent linear factor pricing models including the classical beta representation and the stochastic discount factor (SDF) representation. The model parameters can be estimated and specification tests can be performed by using cross-sectional regressions and the beta representation, or by applying GMM to either representation. The GMM approach has the advantage that it allows for estimation of model parameters in a single pass, thereby avoiding the error-in-variables problem. In this paper, we show how to take into account the error-in-variables problem and reproduce the results of Jagannathan and Wang (1998) in a more convenient representation that allows us to obtain the optimal CSR weighting matrix and develop a new CSR specification test. This test relates to the Hansen-Jagannathan (1997) (HJ) distance and generalizes the CSR test of Shanken (1985). The advantage of our test over Shanken's test lies in the fact that it does not use a model-specific weighting matrix, such as the inverse of the covariance matrix of the time-series regression residuals, and thus allows a fair comparison across different models. This point has been forcefully emphasized by Hansen and Jagannathan (1997) and Jagannathan and Wang (1996). Our CSR test relates to two GMM-based tests. The first is the test that builds on the HJ distance and uses the inverse of the second moment matrix of returns as the weighting matrix and the second is the standard GMM test of overidentifying restrictions of Hansen (1982) that uses the efficiency-optimal weighting matrix. By means of a simulation, we compare all the aforementioned tests and examine their finite sample performance.

A popular device used to gauge the explanatory power of an asset-pricing model in explaining the cross-sectional dispersion in expected returns has been the cross-sectional R-square. Although R-square is an intuitive measure of fit widely used in the linear regression paradigm, its usefulness in the context of pricing model evaluation is questionable. Indeed, in a recent paper, Lewellen and Nagel (2004) argue that, despite its widespread popularity, the cross-sectional R-square does not seem to be a very meaningful metric. They point out that R-square is not very informative when comparing multifactor models since almost any multifactor model will produce a high R-square. In addition, they observe that, typically, researchers do not report confidence intervals and thus the interpretation of R-square is not clear. The simulation in Ahn and Gadarowski (1999a) also shows that the standard deviation of R-square can be rather high, and that R-square is sensitive to the signal-to-noise ratio.[4] Sharing the view of Lewellen and Nagel (2004), we suggest that asset-pricing models should be evaluated by means of econometric tools with solid foundation. The cross-sectional test we develop offers such an alternative.

In two recent contributions to the literature, Kan and Zhou (2004) and Chen and Kan (2004) provide a finite sample analysis of the HJ distance test and the CSR method, respectively, under the assumption of i.i.d. multivariate normality.[5] In contrast, our approach does not make distributional assumptions and is based on asymptotic arguments. In particular, we allow for serial correlation and conditional heteroscedasticity of unknown form. While the multivariate normality assumption does not seem plausible for asset return data and can be rejected by standard tests as shown in Tu and Zhou (2004), we view this approach as valuable. The normality assumption can be thought of as an approximation of the true data generating process. If the approximation is close enough, then finite sample inference will be valid. However, when the data distribution deviates significantly from normality, our approach will be more accurate as long as a long time series of data is available. Therefore, we consider the two approaches complementary and useful in the pricing model evaluation.

[4]Figure 1 in Ahn and Gadarowski (1999a) shows that R-square monotonically increases with the signal-to-noise ratio.

[5]Earlier work on asset-pricing tests based on such assumptions includes Shanken (1985), MacKinlay (1987), and Gibbons, Ross, and Shanken (1989).

While our methodology is primarily concerned with inference and hypothesis testing for unconditional asset-pricing models, our results could be useful for evaluating conditional models as well. As explained in Cochrane (2001a), one can derive a number of unconditional moment restrictions as implications of conditional pricing equations. This can be achieved by simply conditioning down or using conditioning information and managed portfolios. An alternative popular approach is to model the stochastic discount factor with coefficients expressed as affine functions of a small number of instruments. Examples of such practice include the papers by Cochrane (1996) and Lettau and Ludvigson (2001), among others. Following a different approach, Jagannathan and Wang (1996) impose a number of restrictions on their conditional pricing structure to obtain an unconditional beta factor model. Regardless of how one obtains unconditional from conditional restrictions, the techniques advanced in this paper are readily applicable to testing the unconditional implications of a conditional asset-pricing model.

Practical implementation of the methods developed in this paper is straightforward. Consistent estimates of the asymptotic covariance matrix of the risk premium estimates can be easily computed, and thus asymptotically valid t-statistics and p-values for the specification test can be calculated. The computation of asymptotically valid standard errors involves using sample analogues of the relevant population quantities and heteroscedasticity and autocorrelation consistent (HAC) estimates of certain covariance matrices. Methods for obtaining HAC covariance estimates are advanced in Newey and West (1987), Andrews (1991), and den Haan and Levin (1997).[6]

The rest of the paper is organized as follows. In Section 2, we discuss a number of alternative representations of linear factor asset-pricing models and their interrelations. Section 3 briefly reviews related GMM tests based on the stochastic discount factor representation of the beta-pricing model, focusing on the HJ test and the standard Hansen GMM test. In Section 4, we develop the econometric framework for the CSR method and derive a number of results. We provide the asymptotic theory for the risk premium estimators based on the CSR method, show that Shanken's CSR test obtains as a special case of our test under suitable assumptions, and obtain the optimal CSR weighting matrix. Section 5 compares the CSR method with its GMM counterparts and shows that the two approaches, when applied optimally, are equivalent in terms of asymptotic efficiency. In section 6, we define a measure of model misspecification and use it to construct a CSR specification test. A simulation study, presented in Section 7, is used to examine the small sample behavior of the CSR-based tests developed in this paper and provide a comparison with related GMM-based tests. In the last section, we provide some concluding remarks. All the proofs not contained in the main text are collected in the Appendix.

[6] See also Newey and West (1994), Andrews and Monahan (1992), and den Haan and Levin (2000).

2 Model Representations and Pricing Errors

Asset-pricing models prescribe functional relations between expected asset returns and measures of risk. Under the assumption that there exists a riskless asset with an observable rate of return, researchers use returns in excess of the riskless rate; otherwise, simple or gross returns are employed. In addition, risk can be represented in a number of different ways. Traditional models, such as the CAPM, measure risk by the factor betas. Equivalently, factor pricing models can be stated in terms of covariances of the basis assets with the factors. Examples of such models are developed in Campbell (1993, 1996). In addition, a factor model can be described by its stochastic discount factor (SDF) representation involving either gross or excess asset returns. The term *stochastic discount factor* was introduced by Hansen and Richard (1987), while the derivation of the SDF representation of the CAPM can be traced back to Dybvig and Ingersoll (1982).[7] In summary, there are a number of ways to represent an asset-pricing factor model depending on whether we use simple, gross or excess returns and on which measure of risk is employed. In the following subsections, we describe the model formulations that are relevant for our purposes and are used in the subsequent analysis.

Before doing so, however, we need to make some assumptions and fix some notation that is used throughout the paper. We assume that the time series involved in the following analysis are stationary, possess finite first and second moments, and satisfy the standard law of large numbers. Some notation that is heavily used in the sequel is introduced next: 1_L (0_L) denotes the $L \times 1$ vector of ones (zeros), I_L denotes the $L \times L$ identity matrix and $0_{L\times J}$ denotes the $L \times J$ matrix of zeros. Furthermore, for any two L-dimensional vector time series a_T and b_T that satisfy $a_T - b_T \xrightarrow{p} 0_L$, we denote $a_T = b_T + o_p(1)$. Note that in this case, if a_T possesses a limiting distribution, then b_T possesses the same limiting distribution and vice versa. Finally, for any stationary time series $\{x_t\}$ and $\{y_t\}$, we denote $\mu_x = E[x_t]$, $\Sigma_{xy} = Cov[x_t, y_t] = E[(x_t - \mu_x)(y_t - \mu_y)']$, and $\Sigma_x \equiv \Sigma_{xx} = Var[x_t]$.

2.1 Gross returns

We consider a linear factor pricing model with observable factors f_t, such as returns to traded portfolios or macroeconomic variables. Examples of traded factors include market return proxies and spread portfolios[8], while examples of macroeconomic variables used as factors include consumption-

[7] Ingersoll (1987) obtains the SDF representation for a number of asset-pricing models and Cochrane (2001a) presents a comprehensive textbook account of the various asset-pricing model formulations and their interrelations.

[8] The most widely used spread portfolios in empirical research are the SMB and HML portfolios of Fama and French (1992).

growth proxies and term-structure variables.[9] Let K be the number of variables contained in f_t, and suppose there are N basis assets with gross returns at time t denoted by R_t. The associated asset-pricing relation, including a zero-beta rate λ_0, is

$$\mu_R = \lambda_0 1_N + B\lambda_1 = X\lambda, \tag{1}$$

where

$$X = \begin{bmatrix} 1_N & B \end{bmatrix}, \tag{2}$$

$$\lambda = \begin{bmatrix} \lambda_0 & \lambda_1' \end{bmatrix}', \tag{3}$$

λ_1 is the $K \times 1$ vector of risk premiums and the beta matrix B is given by

$$B = \Sigma_{Rf}\Sigma_f^{-1}. \tag{4}$$

Associated with the above beta representation is the standard time-series regression of the returns on the factors. Defining the disturbances $\varepsilon_t = R_t - \mu_R - B(f_t - \mu_f)$, we obtain the regression

$$R_t = \mu_R + B(f_t - \mu_f) + \varepsilon_t. \tag{5}$$

By definition, it then follows that the disturbances ε_t satisfy

$$E[\varepsilon_t] = 0_N \text{ and } E\left[\varepsilon_t f_t'\right] = 0_{N\times K}. \tag{6}$$

Next, we introduce the SDF representation of linear factor asset-pricing models. The majority of known asset-pricing models admit such a representation. It is well-known that beta-pricing models admit an SDF representation, with the SDF being an affine function of the factors. Specifically, the beta-pricing equation (1) is equivalent to the statement that there exists an SDF of the form

$$M_t \equiv M_t(\delta) = \delta_0 + \delta_1' f_t = \delta' Y_t,$$

where

$$\delta = \begin{bmatrix} \delta_0 & \delta_1' \end{bmatrix}', \tag{7}$$

$$Y_t = \begin{bmatrix} 1 & f_t' \end{bmatrix}', \tag{8}$$

for some appropriate scalar δ_0 and K-dimensional vector δ_1. This means that gross returns R_t satisfy the asset-pricing equation

$$E[M_t(\delta)R_t] = 1_N. \tag{9}$$

[9] Other nontraded factors include the labor income growth used by Jagannathan and Wang (1996) and the proprietary business income growth of Heaton and Lucas (2000).

There is a simple correspondence between the parameters λ and δ. Indeed, it is a matter of standard algebraic manipulation[10] to show that the parameter δ can be expressed in terms of the parameter λ as follows

$$\begin{bmatrix} \delta_0 \\ \delta_1 \end{bmatrix} = \frac{1}{\lambda_0} \begin{bmatrix} 1+\lambda_1' \Sigma_f^{-1} \mu_f \\ -\Sigma_f^{-1} \lambda_1 \end{bmatrix}, \tag{10}$$

while the inverse relation is given by

$$\begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} = \frac{1}{\delta_0 + \delta_1' \mu_f} \begin{bmatrix} 1 \\ -\Sigma_f \delta_1 \end{bmatrix}. \tag{11}$$

The subsequent econometric analysis focuses on the moment conditions implied by the pricing equations (1) and (9).

In what follows, we describe two measures of model fit based on the beta and the SDF representations. We first consider the familiar HJ distance that uses the SDF representation. Define the vector of pricing errors associated with the pricing equation (9) by

$$e_t^{\mathrm{S}}(\delta) = M_t(\delta) R_t - 1_N \tag{12}$$

with the superscript S standing for SDF. If the asset-pricing model is correctly specified, then the expected pricing errors should equal zero, namely $E\left[e_t^{\mathrm{S}}(\delta)\right] = 0_N$. These moment conditions form the null hypothesis of correct specification in the context of GMM estimation. Hansen and Jagannathan (1997) propose the following related measure of model fit, the so-called HJ distance

$$\Delta^{\mathrm{S}}(\delta) = \sqrt{E\left[e_t^{\mathrm{S}}(\delta)\right]' G^{-1} E\left[e_t^{\mathrm{S}}(\delta)\right]},$$

where G is the second moment matrix of the returns, that is, $G = E\left[R_t R_t'\right]$. Hansen and Jagannathan (1997) demonstrate that the HJ distance possesses a number of economically appealing properties. Specifically, they show that the HJ distance equals the maximum pricing error per unit norm of portfolio payoff, as well as the least-squares distance between an SDF implied by a model and the space of all valid SDFs.

Next, we turn our attention to the beta representation. Let

$$\theta = \begin{bmatrix} vec(B)' & \lambda' \end{bmatrix}' \tag{13}$$

denote the parameter used by the beta representation, where *vec* is the column stacking operator. Then, analogously to (12), we can define the pricing errors corresponding to the beta-pricing equation (1) by

$$e_t^{\mathrm{B}}(\theta) = R_t - X\lambda, \tag{14}$$

[10] For a detailed derivation, see Section 6.3 in Cochrane (2001a).

where the superscript B indicates the association with the beta representation. If the asset-pricing model accurately describes expected returns, then the expected pricing errors should equal zero, namely $E\left[e_t^{\mathrm{B}}(\theta)\right] = 0_N$.

One of the goals in this paper is to develop an econometric specification test based on the beta representation that utilizes the CSR method. Before doing so, we need to take a stand on a theoretical measure of model (mis)specification. We propose a measure that relates to two well-known procedures for testing the validity of linear asset-pricing models: the cross-sectional regression test of Shanken (1985), and the HJ distance test (see Hansen and Jagannathan (1997) and Jagannathan and Wang (1996)). The intuition behind the misspecification measure is simple. Under the null hypothesis that the model is correctly specified, the pricing equation $\mu_R = X\lambda$ holds. Therefore, we can measure the degree of misspecification by looking at how large the difference $\mu_R - X\lambda$ is. Since the difference is an N-dimensional vector, a natural way to measure its size is to use a norm. To this end, let G be a positive definite and symmetric $N \times N$ matrix and define the distance $\Delta_G^{\mathrm{B}}(\theta)$ as follows

$$\Delta_G^{\mathrm{B}}(\theta) = \sqrt{\left(\mu_R - X\lambda\right)' G^{-1} \left(\mu_R - X\lambda\right)}. \tag{15}$$

Shanken (1985) developed the CSR test based on the sample analogue of the square of distance $\Delta_G^{\mathrm{B}}(\theta)$, specifying the weighting matrix G to be the covariance matrix Σ_ε of the residuals in the time-series regression of the portfolio returns on the factors. This choice seems to be a natural one in the context of the conditionally homoscedastic and normally distributed residuals in Shanken (1985). It is worth emphasizing that the misspecification measure given by (15) is well-defined and admits the same interpretation for either gross, simple, or excess returns.

It is has been argued by Hansen and Jagannathan (1997) that employing a weighting matrix that is not model-specific provides a robust way for equally treating all models under consideration. In their context of stochastic discount factors, they further argue that the appropriate choice is the second-moment matrix of gross returns, which corresponds to specifying $G = E[R_t R_t']$. We argue here that the choice $G = E[R_t R_t']$ in definition (15) admits an economically appealing interpretation, as in the case of the HJ distance. The following proposition shows that the square of $\Delta_G^{\mathrm{B}}(\theta)$, when G is the second-moment matrix of the risk-adjusted returns, is equal to the maximum pricing error per unit norm of payoff on any portfolio. We follow Hansen and Jagannathan (1997) in using the L_2 norm $||p|| = \sqrt{E[p^2]}$ as the norm of the payoff p. It is clear that, when trying to assess pricing errors, one has to normalize by the size of the payoff. Moreover, the choice of the L_2 norm seems to be the most analytically convenient.

Proposition 1 *Let $G = E[RR']$, $w = (w_1, \ldots, w_N)'$ denote a vector of investment in the N assets and $p = w'R$ be the corresponding payoff. Then*

$$\Delta_G^B(\theta) = \max \left\{ \frac{|w'(\mu_R - X\lambda)|}{||p||} : p = w'R, w \neq 0_N \right\}.$$

Although the two choices for the weighting matrix G discussed above, namely Σ_ε and $E[R_t R_t']$, can be justified on theoretical grounds, it is their performance in practice that is important. In addition, it has been argued in the literature that using $G = \Sigma_\varepsilon$, which is model-specific, might give an advantage to models that produce very volatile residuals, leading to potentially erroneous inferences. Therefore, we proceed without taking a stand on the choice of G, and argue that econometrically, it makes sense to use different weighting matrices and examine their robustness and performance in finite samples.

2.2 Excess returns

When there exists an observable riskless asset, we can use excess returns in tests of asset-pricing models. Then the beta representation of a factor pricing model with factors f_t is

$$\mu_R = B\lambda, \tag{16}$$

where the beta matrix is given by $B = \Sigma_{Rf}\Sigma_f^{-1}$. Slightly abusing the notation, we denote excess returns by R_t. Whether we refer to gross or excess returns when we write R_t should be clear from the context. As in the case of gross returns, there exists an equivalent SDF-based representation of the model. More specifically, there exists an SDF, which is affine in the factors, given by

$$M_t(\delta) = 1 + \delta' f_t, \tag{17}$$

such that the beta-pricing equation is equivalent to

$$E\left[M_t(\delta)R_t\right] = 0_N. \tag{18}$$

Note that a vector of zeros replaces a vector of ones in the right-hand side of the above equation as we move from gross to excess returns. Furthermore, it is clear that any multiple of $M_t(\delta)$ qualifies for an SDF. For identification purposes, we maintain the assumption that the constant term in the affine specification is equal to 1. Again, as in the case of gross returns, there exists a simple relation that links the beta parameter λ with the SDF parameter δ. Specifically, one can easily show that

$$\lambda = -\frac{\Sigma_f \delta}{1 + \delta' \mu_f} \tag{19}$$

under the assumption that $E[M_t(\delta)] = 1 + \delta' \mu_f \neq 0$, and, inversely, that

$$\delta = -\left(\Sigma_f + \lambda \mu_f'\right)^{-1} \lambda \tag{20}$$

under the assumption that the matrix $\Sigma_f + \lambda \mu_f'$ is invertible. Note that this is indeed the case when the factors are excess returns to traded portfolios. Then $\lambda = \mu_f$, and, therefore, $\Sigma_f + \lambda \mu_f'$ is positive definite since it is the sum of a positive definite and a positive semidefinite matrix.

An alternative representation of linear factor models can be cast in terms of covariance risk as we illustrate next. Defining

$$b = \Sigma_f^{-1} \lambda \tag{21}$$

we can rewrite the pricing equation (16) as

$$\mu_R = \Sigma_{Rf} b. \tag{22}$$

It then becomes clear that b measures covariance risk, whereas λ measures beta risk. We refer to parameter b as the covariance risk premia and to pricing equation (22) as the covariance risk (CR) representation of the linear factor model.[11]

3 GMM Estimation and Testing

3.1 Gross Returns

The econometric analysis based on the SDF representation of the asset-pricing model, given by (9), uses the GMM procedure as follows. Recall that the vector of pricing errors associated with the SDF $M_t(\delta)$ is given by $e_t^{\mathrm{S}}(\delta) = M_t(\delta) R_t - 1_N$. If the asset-pricing model is correctly specified, then the expected pricing errors should equal zero, namely $E[e_t^{\mathrm{S}}(\delta)] = 0_N$. These moment conditions form the null hypothesis of correct specification in the GMM estimation. Hansen and Jagannathan (1997) proposed the following related measure of model fit, the HJ distance

$$\Delta^{\mathrm{S}}(\delta) = \sqrt{E[e_t^{\mathrm{S}}(\delta)]' \, G^{-1} E[e_t^{\mathrm{S}}(\delta)]},$$

where G is the second moment matrix of the returns, that is, $G = E[R_t R_t']$.

Empirical evaluation of a model based on its SDF representation uses the sample analogue of the HJ distance, defined by

$$\Delta_T^{\mathrm{S}}(\delta_T) = \sqrt{\bar{e}_T^{\mathrm{S}}(\delta_T)' G_T^{-1} \bar{e}_T^{\mathrm{S}}(\delta_T)},$$

[11] Examples of models cast in the CR representation include Campbell (1993, 1996).

where δ_T is the estimator of δ based on the minimization of the HJ distance, G_T^{-1} is a consistent estimator of G^{-1}, and

$$\bar{e}_T^{\mathrm{S}}(\delta) = \frac{1}{T} \sum_{t=1}^{T} e_t^{\mathrm{S}}(\delta) = D_T\delta - 1_N.$$

Jagannathan and Wang (1996) develop the econometric framework, obtain the asymptotic distribution of $T\,[\Delta_T^{\mathrm{S}}(\delta_T)]^2$, and show how to calculate p-values for the corresponding hypothesis test. We briefly describe their method. First, we define the following statistics

$$\begin{aligned} D_T &= \frac{1}{T} \sum_{t=1}^{T} R_t Y_t', \\ G_T &= \frac{1}{T} \sum_{t=1}^{T} R_t R_t' \end{aligned}$$

so that $\bar{e}_T^{\mathrm{S}}(\delta) = D_T\delta - 1_N$. The sample analogue of the HJ distance is minimized by

$$\delta_T^{\mathrm{HJ}} = \left(D_T' G_T^{-1} D_T\right)^{-1} D_T' G_T^{-1} 1_N,$$

which is the GMM estimator based on the moment conditions $E\left[e_t^{\mathrm{S}}(\delta)\right] = 0_N$ and the weighting matrix G_T^{-1}.

Since the HJ distance test is not using the optimal GMM weighting matrix, it does not follow a chi-square distribution asymptotically. Instead, it follows a mixture of independent chi-square random variables with one degree of freedom, as shown by Jagannathan and Wang (1996). To state the precise result we need some additional notation. Under general conditions, we have that $\sqrt{T}\bar{e}_T^{\mathrm{S}}(\delta) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_t^{\mathrm{S}}(\delta)$ converges to a multivariate normal distribution. Let S be the asymptotic covariance matrix of $\sqrt{T}\bar{e}_T^{\mathrm{S}}(\delta)$, so that $\sqrt{T}\bar{e}_T^{\mathrm{S}}(\delta) \xrightarrow{d} N(0_N, S)$, and S_T be a consistent estimator of S. If the data are serially uncorrelated, we can use $S_T = \frac{1}{T} \sum_{t=1}^{T} e_t^{\mathrm{S}}(\delta) e_t^{\mathrm{S}}(\delta)'$ where δ_T is a consistent estimator of δ. If, however, the data are serially correlated then a heteroscedasticity and autocorrelation robust estimator of S is called for. Such estimators are provided by Andrews (1991) and Newey and West (1994) among other papers. We further assume that $D_T \xrightarrow{p} D = E[R_t Y_t']$ and $G_T \xrightarrow{p} G = E[R_t R_t']$ as $T \to \infty$. Let $S^{1/2}$ denote the upper triangular matrix in the Cholesky decomposition of S such that $S = (S^{1/2})' S^{1/2}$. Then the asymptotic distribution of the HJ distance is given by

$$T\,[\Delta_T^{\mathrm{S}}(\delta_T^{\mathrm{HJ}})]^2 \xrightarrow{d} Z'AZ,$$

where Z follows a $N(0_N, I_N)$ distribution and

$$A = S^{1/2}(G^{-1} - G^{-1}D(D'G^{-1}D)^{-1}D'G^{-1})(S^{1/2})',$$

as shown by Jagannathan and Wang (1996) (see their equations (C17) and (C18)). It turns out that the matrix A has exactly $N-(1+K)$ positive eigenvalues, which we denote by $\xi_1,\ldots,\xi_{N-(1+K)}$. Then, an alternative representation of the asymptotic distribution of the HJ distance is given by $T\,[\Delta_T^{\mathrm{S}}(\delta_T^{\mathrm{HJ}})]^2 \xrightarrow{d} \sum_{j=1}^{N-(1+K)} \xi_j v_j$, where $v_j, j=1,\ldots,N-(1+K)$ are i.i.d. $\chi^2(1)$ random variables. Jagannathan and Wang (1996) also show how to use simulation to obtain asymptotically valid p-values for the specification test based on the HJ distance. In this paper, we refer to this test as the GMM-HJ test.

In addition to the HJ distance test, we briefly review Hansen's standard GMM test of overidentifying restrictions. The only difference is that instead of choosing G to be the second moment matrix of gross returns, we need to use the optimal GMM weighting matrix when minimizing the quadratic from. Hansen (1982) shows that the optimal matrix is S^{-1}. In practice, we can consistently estimate S by S_T as discussed above. Note that this requires a consistent estimate δ_T of δ. One natural choice, in the present context, is to use δ_T^{HJ}. Using the efficiency-optimal weighting matrix S_T^{-1}, the optimal GMM estimator is then given by

$$\delta_T^{\mathrm{OP}} = \left(D_T' S_T^{-1} D_T\right)^{-1} D_T' S_T^{-1} 1_N$$

where the superscript OP stands for optimal. The GMM test uses the so-called J-statistic

$$J_T = T \bar{e}_T^{\mathrm{S}}(\delta_T^{\mathrm{OP}})' S_T^{-1} \bar{e}_T^{\mathrm{S}}(\delta_T^{\mathrm{OP}}),$$

which has an asymptotic chi-square distribution with $N-(1+K)$ degrees of freedom under the null hypothesis of correct model specification. In this paper, we refer to this test as the GMM-H test.

Evidence on the small sample behavior of the GMM-HJ and GMM-H tests is provided by Ahn and Gadarowski (2004). In Section 7, we perform a simulation experiment to examine the small-sample properties of the GMM tests described in this section and the CSR-based tests in Section 6.

3.2 Excess Returns

Recall that the SDF representation of the model is $E\left[R_t M_t(\delta)\right] = 0_N$ where $M_t(\delta) = 1 + \delta' f_t$. The SDF-GMM method uses the preceding moment condition to obtain the following estimate of δ:

$$\delta_T^{\mathrm{G}} = -\left(D_T' G_T^{-1} D_T\right)^{-1} D_T' G_T^{-1} \bar{R}_T,$$

where

$$D_T = \frac{1}{T} \sum\nolimits_{t=1}^{T} R_t f_t'$$

and the superscript G indicates that the estimator is based on the GMM approach. The matrix G_T^{-1} is a consistent estimate of the weighting matrix G^{-1}, which is to be chosen appropriately. By the moment condition above, we have $\mu_R = -D\delta$ where $D = E[R_t f_t']$. By the law of large numbers, $D_T \xrightarrow{p} D$, so that δ_T^{G} is a consistent estimator of δ:

$$\delta_T^{\mathrm{G}} = -\left(D_T' G_T^{-1} D_T\right)^{-1} D_T' G_T^{-1} \bar{R}_T \xrightarrow{p} -\left(D' G^{-1} D\right)^{-1} D' G^{-1} \mu_R = \delta.$$

The asymptotic distribution of δ_T^{G} can be described as follows. Note that

$$\begin{aligned} \sqrt{T}(\delta_T^{\mathrm{G}} - \delta) &= -\left(D_T' G_T^{-1} D_T\right)^{-1} D_T' G_T^{-1} \sqrt{T}(\bar{R}_T + D_T \delta) \\ &= -\left(D_T' G_T^{-1} D_T\right)^{-1} D_T' G_T^{-1} \frac{1}{\sqrt{T}} \sum\nolimits_{t=1}^{T} n_t, \end{aligned}$$

where

$$n_t = R_t(1 + \delta' f_t). \tag{23}$$

Under standard regularity conditions, it can be verified that

$$\frac{1}{\sqrt{T}} \sum\nolimits_{t=1}^{T} n_t \xrightarrow{d} N(0_N, \Omega_n),$$

with $\Omega_n = \Sigma_{n,0} + \sum\nolimits_{j=1}^{\infty} (\Sigma_{n,j} + \Sigma_{n,j}')$ where $\Sigma_{n,j} = E[n_t n_{t+j}]$ for $j = 0, \ldots, \infty$. The efficiency-optimal choice of G^{-1} is equal to Ω_n^{-1} according to Hansen (1982). Under this choice, it follows that

$$\sqrt{T}\,(\delta_T^{\mathrm{G}} - \delta) = -\left(D' \Omega_n^{-1} D\right)^{-1} D' \Omega_n^{-1} \frac{1}{\sqrt{T}} \sum\nolimits_{t=1}^{T} n_t + o_p(1), \tag{24}$$

and thus the asymptotic distribution of δ_T^{G} is described by

$$\sqrt{T}\,(\delta_T^{\mathrm{G}} - \delta) \xrightarrow{d} N(0_K, \left(D' \Omega_n^{-1} D\right)^{-1}).$$

4 Cross-Sectional Regression Method

4.1 Description of the CSR Method

We now turn our attention to the CSR method. The method consists of two steps. The beta matrix is estimated in the first pass using a standard time-series regression of the returns on the factors.

Equivalently, it can be seen that the estimation of the beta matrix is based on the sample analogue principle. Recall that $B = \Sigma_{Rf}\Sigma_f^{-1}$. Proceeding in the standard fashion and using all the data available, we obtain the following estimate of the beta matrix

$$B_T = \Sigma_{Rf,T}\Sigma_{f,T}^{-1}, \tag{25}$$

where

$$\Sigma_{Rf,T} = \frac{1}{T}\sum_{t=1}^{T}\left(R_t - \bar{R}_T\right)\left(f_t - \bar{f}_T\right)', \tag{26}$$

$$\Sigma_{f,T} = \frac{1}{T}\sum_{t=1}^{T}\left(f_t - \bar{f}_T\right)\left(f_t - \bar{f}_T\right)', \tag{27}$$

and

$$\bar{R}_T = \tfrac{1}{T}\sum_{t=1}^{T} R_t, \tag{28}$$

$$\bar{f}_T = \tfrac{1}{T}\sum_{t=1}^{T} f_t. \tag{29}$$

The CSR method proceeds in the second pass by estimating the vector of risk premiums λ at each time $t = 1, \ldots, T$ as follows

$$\hat{\lambda}_t = \left(X_T' Q_T^{-1} X_T\right)^{-1} X_T' Q_T^{-1} R_t, \tag{30}$$

where

$$X_T = \left[\begin{array}{cc} 1_N & B_T \end{array}\right], \tag{31}$$

and Q_T^{-1} is a consistent estimator of a CSR matrix Q^{-1}. It is assumed that Q and Q_T for large T are symmetric and positive definite. Further, we assume that $N > 1 + K$, as is typically the case in practice. Then the CSR estimate of λ is the time-series average

$$\lambda_T = \frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_t = H_T\bar{R}_T, \tag{32}$$

where

$$H_T = \left(X_T' Q_T^{-1} X_T\right)^{-1} X_T' Q_T^{-1}. \tag{33}$$

Let us define the probability limit of H_T as

$$H = \left(X' Q^{-1} X\right)^{-1} X' Q^{-1}. \tag{34}$$

It is easily shown that λ_T is a consistent estimator of λ. Since $B_T \xrightarrow{p} B$ we have $X_T \xrightarrow{p} X$ as $T \to \infty$, and thus $H_T\bar{R}_T \xrightarrow{p} H\mu_R$ by Slutsky's theorem. Finally, we have $\lambda_T \xrightarrow{p} HX\lambda = \lambda$ as $T \to \infty$ according to the asset-pricing relation (1), which states that $\mu_R = X\lambda$.

In a more general context, incorporating firm characteristics, Jagannathan and Wang (1998) show that λ_T is a consistent estimator of λ and asymptotically follows a multivariate normal distribution. Without the assumption of conditional homoscedasticity of the disturbance terms given the factors, they provide the expression for the asymptotically valid covariance matrix of λ_T. Next, we reproduce their result under a different representation and use it in the sequel to develop a specification test for the asset-pricing model under examination. One of the advantages of our representation is that it allows us to obtain the optimal CSR weighting matrix in terms of asymptotic efficiency. Risk premia estimators that use this CSR matrix have minimum asymptotic variance among all CSR risk premia estimators.

4.2 Asymptotic Distribution of the CSR Risk-Premia Estimator

In this subsection we study the asymptotic behavior of the CSR estimator λ_T defined by (32). Consider the following quantities

$$m_{1,t} = R_t - \mu_R, \; m_{2,t} = \left[\left(f_t - \mu_f\right)' \Sigma_f^{-1} \lambda_1 \right] \varepsilon_t, \; m_t = \left[\begin{array}{cc} (m_{1,t})' & (m_{2,t})' \end{array} \right]', \tag{35}$$

where ε_t is the disturbance term in the time-series regression (5). Recall that ε_t and f_t are uncorrelated by the definition of ε_t and, therefore, it follows that $E[m_t] = 0_{2N}$ (see equation (6)). In order to proceed, we make the following assumption on the asymptotic behavior of the vector time series m_t.

Assumption 2 *For all $\lambda_1 \in \mathbb{R}^K$, the vector time series m_t defined by (35) satisfies the central limit theorem, that is*

$$\frac{1}{\sqrt{T}} \sum\nolimits_{t=1}^{T} m_t \xrightarrow{d} N(0_{2N}, \Sigma_m),$$

with

$$\Sigma_m = \left[\begin{array}{cc} \Psi_m & \Gamma_m \\ \Gamma_m' & \Pi_m \end{array} \right], \tag{36}$$

where Ψ_m, Γ_m and Π_m are $N \times N$ matrices, $\Sigma_m = \sum_{j=-\infty}^{\infty} \Sigma_{m,j}$, and $\Sigma_{m,j} = E\left[m_t m_{t+j}'\right]$, for all j.

Assumption 2 is rather mild and, in fact, can be obtained under standard stationarity, mixing and moment conditions. Related results can be found in, among others, Hall and Heyde (1980), Davidson (1994), and White (2001).[12]

[12]An assumption of this nature is made in Jagannathan and Wang (1998).

Now, we are in a position to state the main result of this section describing the asymptotic distribution of the CSR risk premia estimator. Note that the result is obtained without assumptions such as normality or homoscedasticity of the disturbances given the factors.

Theorem 3 *Assume that the matrix $X = [\ 1_N \quad B\ ]$ is of full rank equal to $1+K$. Then, under Assumption 2, as $T \to \infty$,*

$$\sqrt{T}\,(\lambda_T - \lambda) = \frac{1}{\sqrt{T}} \sum\nolimits_{t=1}^{T} H\,(m_{1,t} - m_{2,t}) + o_p(1),$$

and therefore $\sqrt{T}\,(\lambda_T - \lambda)$ converges to a multivariate normal distribution with zero mean and covariance matrix given by

$$\Sigma_\lambda = H\Omega_m H', \tag{37}$$

where

$$\Omega_m = \Psi_m + \Pi_m - (\Gamma_m + \Gamma_m'), \tag{38}$$

H is defined by (34), and Ψ_m, Π_m and Γ_m are defined by (36).

We devote the next two subsections to take a closer look at the special case of conditionally homoscedastic time-series disturbances ε_t, and then address the issue of asymptotic efficiency of the CSR risk-premia estimator.

4.3 Conditional Homoscedasticity

In this subsection, we examine the asymptotic behavior of the CSR estimator when the residuals ε_t in the time-series regression (5) are conditionally homoscedastic given the factors f_t. Shanken (1992), using this assumption, was the first to obtain the asymptotic distribution of the CSR estimator. The formal statement of the assumption is as follows.

Assumption 4 *Let $\mathcal{F}$ be the information set generated by the entire factor sequence $\{f_t : t = 1, 2, \ldots\}$. Given the information set $\mathcal{F}$, the time-series regression residuals ε_t have zero conditional mean, i.e., $E[\varepsilon_t|\mathcal{F}] = 0_N$. Furthermore, given $\mathcal{F}$, the residuals ε_t have constant conditional covariance equal to Σ_ε and are conditionally serially uncorrelated, i.e. $E[\varepsilon_t\varepsilon_t'|\mathcal{F}] = \Sigma_\varepsilon$ and $E[\varepsilon_t\varepsilon_{t+j}'|\mathcal{F}] = 0_{N\times N}$ for all nonzero integers j.*

When returns and factors are jointly independently and identically distributed and follow a normal distribution, then the foregoing assumption is satisfied. However, as acknowledged in Shanken

(1992), the assumption of conditional homoscedasticity is rather restrictive. The following proposition captures the asymptotic distribution of the CSR estimator under Assumption 4. Assume that $\bar{f}_T = \frac{1}{T}\sum_{t=1}^{T} f_t$ satisfies the central limit theorem and denote by $\Sigma_{\bar{f}}$ its asymptotic covariance matrix, that is

$$\Sigma_{\bar{f}} = \sum_{j=-\infty}^{\infty} \Sigma_{f,j},\ \Sigma_{f,j} = E[\left(f_t - \mu_f\right)\left(f_{t+j} - \mu_f\right)'] \text{ for all } j,$$

and let $\Sigma_{\bar{f}}^{*}$ be the so-called bordered version of $\Sigma_{\bar{f}}$, defined as follows

$$\Sigma_{\bar{f}}^{*} = \begin{bmatrix} 0 & 0_K' \\ 0_K & \Sigma_{\bar{f}} \end{bmatrix}. \tag{39}$$

Proposition 5 *Assume that the matrix $X = [\ 1_N \quad B\ ]$ is of full rank equal to $1+K$. Under Assumption 4, we have*

$$\Psi_m = B\Sigma_{\bar{f}}B' + \Sigma_\varepsilon,\ \Pi_m = (\lambda_1'\Sigma_f^{-1}\lambda_1)\Sigma_\varepsilon,\ \Gamma_m = 0_{N\times N},$$

where Ψ_m, Π_m and Γ_m are defined by (36). Therefore, the asymptotic distribution of the CSR estimator λ_T is given by

$$\sqrt{T}\,(\lambda_T - \lambda) \xrightarrow{d} N(0_{1+K}, \Sigma_\lambda^{CH}),$$

where

$$\Sigma_\lambda^{CH} = \Sigma_{\bar{f}}^{*} + (1 + \lambda_1'\Sigma_f^{-1}\lambda_1)H\Sigma_\varepsilon H',$$

with $\Sigma_{\bar{f}}^{*}$ and H being defined by (39) and (34), respectively.

Note that a version of the preceding proposition is obtained in Shanken (1992) when $Q^{-1} = \Sigma_\varepsilon^{-1}$ under the slightly stronger assumption that the residuals ε_t are i.i.d. given $\mathcal{F}$. (see Theorem 3.1 in his paper).

4.4 Efficiency-Optimal CSR weighting matrix

According to Theorem 3 and definition (34), the asymptotic covariance matrix of the CSR risk premia estimator is

$$\Sigma_\lambda = H\Omega_m H' = \left(X'Q^{-1}X\right)^{-1} X'Q^{-1}\Omega_m Q^{-1}X\left(X'Q^{-1}X\right)^{-1},$$

where we make use of the fact that the CSR weighting matrix Q^{-1} is symmetric by assumption. Note that the structure of the matrix Σ_λ is identical to that of the asymptotic covariance matrix of a generic GMM estimator, where Q^{-1} and Ω_m correspond to the weighting GMM matrix and the asymptotic covariance matrix of the time series defining the moment condition, respectively. As shown by Hansen (1982), the optimal weighting GMM with respect to asymptotic efficiency equals the inverse of the moment condition covariance matrix. The exact matrix-algebraic result is stated as fact (F2) in the Appendix. In our context, this means that the efficiency-optimal CSR weighting matrix is given by

$$Q_{\mathrm{OP}}^{-1} = \Omega_m^{-1},$$

under which choice the asymptotic covariance matrix of the CSR estimator becomes

$$\Sigma_\lambda^{\mathrm{OP}} = \left(X'\Omega_m^{-1}X\right)^{-1}.$$

When researchers apply the CSR method in practice, they typically use the identity matrix or the inverse of an estimator of the covariance matrix of the time-series disturbances Σ_ε. The preceding result implies that using these matrices could lead to some efficiency loss. The claim might appear surprising, or even erroneous, in the light of the fact that Σ_ε^{-1} has been shown to be asymptotically efficient when factors and disturbances are i.i.d. and jointly normally distributed by Shanken (1992) (see Theorem 4 in his paper). When Assumption 4 holds, it follows from the proof of Proposition (5) that

$$\Omega_m = (1 + \lambda_1'\Sigma_f^{-1}\lambda_1)\Sigma_\varepsilon + B\Sigma_{\bar{f}}B' \tag{40}$$

which is clearly different than Σ_ε. It turns out that, under Assumption 4, the two approaches are essentially equivalent. The following proposition clarifies the issue. To be concrete, let us denote by λ_T^{OP} and λ_T^{SH} the CSR estimators that use Ω_m^{-1} and Σ_ε^{-1} as CSR matrices, respectively. More specifically,

$$\lambda_T^{\mathrm{OP}} = (X_T'\Omega_{m,T}^{-1}X_T)^{-1}X_T'\Omega_{m,T}^{-1}\bar{R}_T, \tag{41}$$

$$\lambda_T^{\mathrm{SH}} = (X_T'\Sigma_{\varepsilon,T}^{-1}X_T)^{-1}X_T'\Sigma_{\varepsilon,T}^{-1}\bar{R}_T \tag{42}$$

where $\Omega_{m,T}$ and $\Sigma_{\varepsilon,T}$ are consistent estimators of Ω_m and Σ_ε respectively. In the light of expression (40), a consistent estimator of Ω_m is given by

$$\Omega_{m,T} = (1 + \lambda_{1,T}'\Sigma_{f,T}^{-1}\lambda_{1,T})\Sigma_{\varepsilon,T} + B_T\Sigma_{\bar{f},T}B_T' \tag{43}$$

where $\lambda_{1,T}$, $\Sigma_{f,T}$, B_T and $\Sigma_{\bar{f},T}$ are consistent estimators of λ_1, Σ_f, B and $\Sigma_{\bar{f}}$ respectively and $\Sigma_{\varepsilon,T}$ is the same consistent estimator used in (42). We now state the result that shows the equivalence of the two methods under Assumption 4.[13] However, when the assumption of conditional homoscedasticity is violated, use of Ω_m would produce more precise estimates of the risk premia.

Proposition 6 *Assume that the matrix $X = [\ 1_N \quad B\ ]$ is of full rank equal to $1 + K$ and that Assumptions 2 and 4 are in effect. Then, the cross-sectional regression risk premium estimators λ_T^{OP} and λ_T^{SH} are identical when the estimator $\Omega_{m,T}$ given by (43) is used. In general, the two estimators λ_T^{OP} and λ_T^{SH} possess equal asymptotic covariance matrices as long as they are based on consistent estimators $\Omega_{m,T}$ and $\Sigma_{\varepsilon,T}$.*

Regardless of whether Assumption 4 holds, one can obtain consistent estimators of the optimal CSR matrix Ω_m using that fact that $T^{-1/2}\sum_{t=1}^{T}(m_{1,t} - m_{2,t}) \xrightarrow{d} N(0_N, \Omega_m)$ as it follows from definition (38), Assumption 2, and definition (35). Note that $m_{1,t} - m_{2,t}$ depends on the unknown parameter λ. The rest of the input needed for the computation of a consistent estimator of Ω_m can be obtained by replacing μ_R, μ_f and Σ_f^{-1} by their sample analogues and using the estimated residuals from the time-series regression. Similar to the GMM setting, one has to follow a two-step procedure to deal with the fact that λ is unknown. In the first step, an initial CSR estimate of λ is obtained using a CSR matrix such as the identity matrix or an estimate of Σ_ε^{-1}. Then, in the second step, the estimate of λ is used to construct consistent estimators of the time series $m_{1,t} - m_{2,t}$, from which one can obtain consistent estimators of Ω_m. The final estimator of λ is computed using the estimator of Ω_m^{-1} as the CSR matrix.

5 Comparison of the CSR and GMM Methods

The purpose of this section is to compare the CSR method, which is based on the beta representation, with the GMM approach based on the SDF representation or the CR representation. Comparison of different approaches to empirical testing of asset-pricing models has been a subject of interest in recent years. Kan and Zhou (1999) initiated a debate on the usefulness of the SDF methodology by comparing the GMM method applied to the SDF representation to a maximum likelihood (ML) estimation based on the beta representation. They concluded that the SDF methodology is rather unreliable based on their finding that the standard error of the GMM estimated risk premium can be more than 40 times greater than that of the traditional ML methodology. This appeared to be a surprising result given the fact that the SDF and beta representations are mathematically equivalent. It was soon pointed out by Cochrane (2001b) and Jagannathan and Wang (2002) that the results of Kan and Zhou (1999) stem from their assumption that the factor

[13] Ahn and Gadarowski (1999b) state this result but do not provide a proof. See Section 3.2 in their paper.

mean and variance are assumed to be known or predetermined by the econometrician. Making such an assumption gives an advantage to the beta method, which actually uses this information, but not to the SDF method. Cochrane (2001b) and Jagannathan and Wang (2002) demonstrate that, when the comparison is cast in the right framework, the two approaches are equivalent in terms of asymptotic efficiency. Our contribution to the existing literature on this topic is to extend the set of methods under examination by including the CSR method. In our analysis, we focus on excess returns over an observable risk-free rate in order to reduce notational complexity. In contrast to the aforementioned papers on the subject, we allow for more than one factor and do not make strong distributional assumptions such as i.i.d. joint normality of the asset returns and factors. Instead, we rely on mild assumptions that guarantee the validity of a central limit theorem for appropriate time series. Our main finding is that, when the CSR method is used optimally to estimate the SDF parameter or the CR risk premia, it is as asymptotically efficient as the GMM method which uses Hansen's optimal weighting matrix.

5.1 Using the SDF Representation

We next show how to use the CSR method to obtain an estimator of the SDF parameter δ. Recall equation (20) which states that $\delta = -(\Sigma_f + \lambda\mu_f')^{-1}\lambda$. This relation suggests that, we can use the CSR estimate λ_T of λ to obtain an estimator of δ, provided that estimators of μ_f and Σ_f are available. The natural choice is to use the sample analogues $\bar{f}_T$ and $\Sigma_{f,T}$ given by expressions (29) and (27) respectively. Then the CSR estimate of δ is naturally defined by

$$\delta_T^{\mathrm{C}} = -(\Sigma_{f,T} + \lambda_T \bar{f}_T')^{-1}\lambda_T,$$

with the superscript C indicating that δ_T^{C} is a CSR estimator. Since we work with excess returns the CSR estimator is given by

$$\lambda_T = (B_T' Q_T^{-1} B_T)^{-1} B_T' Q_T^{-1} \bar{R}_T$$

Since λ_T, $\bar{f}_T$ and $\Sigma_{f,T}$ are consistent estimators of λ, μ_f and Σ_f, respectively, it follows by Slutsky's theorem that δ_T^{C} is a consistent estimator of δ: $\delta_T^{\mathrm{C}} \xrightarrow{p} \delta$ as $T \to \infty$.

Given the asymptotic distributions of λ_T, $\bar{f}_T$ and $\Sigma_{f,T}$, one can obtain the asymptotic distribution of δ_T^{C} using the delta method. We follow an essentially equivalent but more transparent approach that allows us to easily obtain the optimal CSR matrix Q^{-1}. It is worthwhile emphasizing that the optimal CSR matrix used in the estimation of the beta parameter λ does not lead to efficient estimation of the SDF parameter δ. Furthermore, our approach enables us not only to show that δ_T^{G} and δ_T^{C} have the same asymptotic distribution, but also, as shown by the following theorem, establish a more direct link between the two estimators.

Theorem 7 *Let δ_T^{G} be the optimal GMM estimator of the SDF parameter δ with weighting matrix equal to Ω_n^{-1}, and δ_T^{C} be the optimal CSR estimator of δ with CSR matrix equal to Ω_n^{-1}. Then, as $T \to \infty$,*

$$\sqrt{T}(\delta_T^G - \delta) = \sqrt{T}(\delta_T^C - \delta) + o_p(1).$$

We have demonstrated that the two estimators δ_T^{G} and δ_T^{C} possess equivalent asymptotic distributions, and thus share the same asymptotic variance. This result shows that the CSR method, when applied optimally, can estimate the SDF parameter with the same asymptotic precision as the GMM method.

5.2 Using the CR Representation

We devote this subsection to the estimation of the CR risk premia b. Recall that the factor model is described by the CR pricing equation $\mu_R = \Sigma_{Rf} b$ where $b = \Sigma_f^{-1}\lambda$. We describe two ways to estimate the vector of covariance risk premia b. The first is based on the CSR method while the second is an application of GMM. We show that when the methods are applied optimally they yield estimators of the parameter b that possess equal asymptotic covariance matrices. Therefore, use of the CSR method to estimate b does not entail any loss in asymptotic efficiency.

5.2.1 Cross-Sectional Regression

The variant of the CSR method we use to estimate b proceeds in two steps. In the first step, the covariance matrix Σ_{Rf} is estimated by its sample analogue

$$\Sigma_{Rf,T} = \frac{1}{T}\sum\nolimits_{t=1}^{T} R_t(f_t - \bar{f}_T)'.$$

In the second step, we run the cross-sectional regression

$$\bar{R}_T = \Sigma_{Rf,T} b + u$$

to obtain the CSR estimate

$$b_T^{\mathrm{C}} = (\Sigma_{Rf,T}' Q_T^{-1} \Sigma_{Rf,T})^{-1} \Sigma_{Rf,T}' Q_T^{-1} \bar{R}_T,$$

where Q_T^{-1} is a consistent estimate of a weighting matrix Q^{-1}. The superscript C indicates that b_T^{C} is a CSR estimator. We require Σ_{Rf} and $\Sigma_{Rf,T}$, for large enough T, to have full rank equal to K, so that the estimator b_T^{C} and its probability limit are well defined. Further, we assume that both Q^{-1} and Q_T^{-1}, for large enough T, are symmetric and positive definite. The optimal choice of the matrix Q^{-1} will be determined on grounds of asymptotic efficiency.

Note that, alternatively, one can utilize the identity $b = \Sigma_f^{-1}\lambda$ to provide an estimator of b. In a natural fashion, we can estimate Σ_f by its sample analogue $\Sigma_{f,T}$ and λ by the CSR estimator λ_T and define $\tilde{b}_T = \Sigma_{f,T}^{-1}\lambda_T$ as an estimator of b. However, since $B_T = \Sigma_{Rf,T}\Sigma_{f,T}^{-1}$ it immediately follows that $\tilde{b}_T = b_T^{\mathrm{C}}$. Therefore, there is no need to consider $\tilde{b}_T$ separately.

Routine application of Slutsky's theorem, in the light of the pricing equation $\mu_R = \Sigma_{Rf}b$, yields that $b_T^{\mathrm{C}} \xrightarrow{p} b$ as $T \to \infty$; that is, the CSR estimate b_T^{C} is a consistent estimate of b. The next step is to describe the limiting distribution of b_T^{C}. To state and prove the result, which is captured in the forthcoming theorem, we need to make the following assumption.

Assumption 8 *The time series*

$$\phi_t = R_t(1 - (f_t - \mu_f)'b) + \mu_R b'(f_t - \mu_f)$$

satisfies the central limit theorem, so that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_t \xrightarrow{d} N(0_N, \Omega_\phi),$$

where $\Omega_\phi = \sum_{j=-\infty}^{\infty}\Sigma_{\phi,j}$ *and* $\Sigma_{\phi,j} = E[\phi_t\phi_{t+j}']$.

Theorem 9 *The asymptotic distribution of the CSR estimate of the covariance risk premia is given by*

$$\sqrt{T}(b_T^C - b) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}P\phi_t + o_p(1),$$

where

$$P = (\Sigma_{Rf}'Q^{-1}\Sigma_{Rf})^{-1}\Sigma_{Rf}'Q^{-1},$$

and so

$$\sqrt{T}(b_T^C - b) \xrightarrow{d} N(0_K, P\Omega_\phi P').$$

Since the asymptotic variance of b_T^{C} is equal to

$$P\Omega_\phi P' = (\Sigma_{Rf}'Q^{-1}\Sigma_{Rf})^{-1}\Sigma_{Rf}'Q^{-1}\Omega_\phi Q^{-1}\Sigma_{Rf}(\Sigma_{Rf}'Q^{-1}\Sigma_{Rf})^{-1},$$

it follows that the optimal weighting matrix is $Q_{\mathrm{OP}}^{-1} = \Omega_\phi^{-1}$. The argument leading to this conclusion is identical to the one used to determine the optimal weighting matrix in a GMM framework (see fact (F2) in the Appendix). The asymptotic variance of b_T^{C} under this choice is then equal to $\Omega_b^{\mathrm{C}} = (\Sigma_{Rf}'\Omega_h^{-1}\Sigma_{Rf})^{-1}$.

5.2.2 GMM Approach

The GMM approach to estimating b is based on the asset-pricing equation $\mu_R = \Sigma_{Rf} b$, which is equivalent to

$$E[R_t - R_t(f_t - \mu_f)'b] = 0_N.$$

Since the factor mean μ_f is typically unknown, we also need the moment condition

$$E[f_t - \mu_f] = 0_K.$$

Estimation and inference then proceeds by applying the GMM on the vector moment condition

$$E[g_t(\theta)] = 0_{N+K}, \tag{44}$$

where $\theta = [\ b' \quad \mu_f'\ ]'$, and

$$g_t(\theta) = \begin{bmatrix} R_t - R_t(f_t - \mu_f)'b \\ f_t - \mu_f \end{bmatrix} \equiv \begin{bmatrix} h_t(\theta) \\ f_t - \mu_f \end{bmatrix}.$$

Let us denote the GMM estimator by $\theta_T^{\mathrm{G}} = (b_T^{\mathrm{G}\prime}, \mu_T^{\mathrm{G}\prime})'$. To proceed with the GMM estimation, we make the standard assumption that the time series $g_t(\theta)$ satisfies the central limit theorem, so that

$$\frac{1}{\sqrt{T}} \sum\nolimits_{t=1}^{T} g_t(\theta) \xrightarrow{d} N(0_N, \Omega_g)$$

for some positive definite $(N+K) \times (N+K)$ matrix Ω_g. We decompose the matrix Ω_g in a natural fashion as follows

$$\Omega_g = \begin{bmatrix} \Omega_h & \Omega_{hf} \\ \Omega_{fh} & \Omega_f \end{bmatrix}. \tag{45}$$

Then, under the optimal choice for the GMM weighting matrix, namely Ω_g^{-1}, it follows that

$$\sqrt{T}(\theta_T^{\mathrm{G}} - \theta) \xrightarrow{d} N(0_N, (D_g'\Omega_g^{-1} D_g)^{-1}),$$

where the $(N+K) \times 2K$ matrix D_g is given by

$$D_g = E\left[\frac{\partial g_t(\theta)}{\partial \theta}\right] = \begin{bmatrix} -\Sigma_{Rf} & \mu_R b' \\ 0_{K\times K} & -I_K \end{bmatrix}. \tag{46}$$

This implies that

$$\sqrt{T}(b_T^{\mathrm{G}} - b) \xrightarrow{d} N(0_N, \Omega_b^{\mathrm{G}}),$$

where Ω_b^{G} is the upper-left $K \times K$ submatrix of $\Omega_\theta^{\mathrm{G}} = (D_g'\Omega_g^{-1} D_g)^{-1}$. In the next theorem, we show that the GMM estimator b_T^{G} possesses the same asymptotic covariance matrix as its CSR counterpart b_T^{C}.

Theorem 10 *Let b_T^G be the estimator of b obtained by using GMM to estimate the parameter $\theta = [\ b' \quad \mu_f'\ ]'$ based on the moment condition (44) and the optimal weighting matrix Ω_g^{-1}, and b_T^C be the optimal CSR estimator with CSR matrix equal to Ω_ϕ^{-1}. Then, b_T^G and b_T^C possess the same asymptotic covariance matrix:*

$$avar[b_T^G] \equiv \Omega_b^G = \Omega_b^C \equiv avar\,[b_T^C]\,.$$

The message of Theorems 7 and 10 is perfectly in line with the results advanced in Cochrane (2001b) and Jagannathan and Wang (2002). That is, regardless of which pricing model representation is employed by the researcher, the approaches are equivalent in terms of efficiency as long as each method is applied optimally and no method is given an advantage a priori. It should be emphasized, however, that our results are more powerful since our analysis is rather general. Specifically, our theorems are valid for any number of pricing factors and, more importantly, do not rely on strong assumptions on the dynamics and distributions of returns and factors such as the i.i.d. normality assumption made in Jagannathan and Wang (2002).

6 A CSR Specification Test

At this point, we are in a position to develop an econometric (mis)specification test based on the distance $\Delta_G^{\mathrm{B}}(\theta)$ defined in (15). Since $\mu_R = X\lambda$ we have $\bar{R}_T - X_T\lambda_T \xrightarrow{p} 0_N$ where the risk premium estimator λ_T is given by (32). Hence, a sensible way to assess the specification would be to examine the distance between $\bar{R}_T$ and $X_T\lambda_T$. To do so, we consider the sample analogue of the distance $\Delta_G^{\mathrm{B}}(\theta)$ and seek to develop the related sampling theory. In other words, our objective is to obtain the asymptotic distribution of the quadratic form

$$T\Delta_T^2 = T\left(\bar{R}_T - X_T\lambda_T\right)' G_T^{-1}\left(\bar{R}_T - X_T\lambda_T\right), \tag{47}$$

where G_T is a consistent estimator of the weighting matrix G. We will return to the choice of G in the sequel. The normalization T reflects the fact that $\sqrt{T}(\bar{R}_T - X_T\lambda_T)$ will turn out to have a multivariate normal limiting distribution.

6.1 Derivation of the Test

Similar to the asymptotic result for the HJ distance test, derived in Jagannathan and Wang (1996), it is expected that the asymptotic distribution of $T\Delta_T^2$ is a mixture of $\chi^2(1)$ distributed random variables. The proof proceeds as follows. First, observe that

$$\begin{aligned}\sqrt{T}\left(\bar{R}_T - X_T\lambda_T\right) &= \sqrt{T}\left(\bar{R}_T - \mu_R\right) - \sqrt{T}(X_T\lambda_T - X\lambda)\\ &= \sqrt{T}\left(\bar{R}_T - \mu_R\right) - \sqrt{T}(X_T - X)\lambda - X_T\sqrt{T}(\lambda_T - \lambda).\end{aligned}$$

Furthermore, we have

$$(X_T - X)\lambda = [\begin{array}{cc} 0_N & B_T - B \end{array}]\lambda = (B_T - B)\,\lambda_1,$$

since $\lambda = [\begin{array}{cc} \lambda_0 & \lambda_1' \end{array}]'$. Hence, combining terms we obtain

$$\sqrt{T}\left(\bar{R}_T - X_T\lambda_T\right) = \sqrt{T}\left(\bar{R}_T - \mu_R\right) - \sqrt{T}\,(B_T - B)\,\lambda_1 - X\sqrt{T}(\lambda_T - \lambda) + o_p(1)$$

upon using $X_T \xrightarrow{p} X$, $\sqrt{T}(\lambda_T - \lambda) \xrightarrow{d} N(0_{1+K}, \Sigma_\lambda)$ (see Theorem 3), and the fact (F1) stated in the Appendix. Combining Lemma 17, Theorem 3, and definition (35) we obtain

$$\begin{aligned}\sqrt{T}\left(\bar{R}_T - X_T\lambda_T\right) &= \frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T}\left[(m_{1,t} - m_{2,t}) - XH\,(m_{1,t} - m_{2,t})\right] + o_p(1)\\ &= M\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T}(m_{1,t} - m_{2,t}) + o_p(1),\end{aligned}$$

where

$$M = I_N - XH = I_N - X(X'Q^{-1}X)^{-1}X'Q^{-1}. \tag{48}$$

It is easily verified that the matrix M is idempotent and therefore $rank(M) = tr(M)$. Further, M satisfies the property $MX = 0_{N\times(1+K)}$. Using the properties of the trace operator, we obtain

$$\begin{aligned}tr(M) &= tr(I_N - X(X'Q^{-1}X)^{-1}X'Q^{-1}) = N - tr(X(X'Q^{-1}X)^{-1}X'Q^{-1})\\ &= N - tr(X'Q^{-1}X(X'Q^{-1}X)^{-1}) = N - tr(I_{1+K}) = N - (1+K).\end{aligned}$$

Assumption 2 and definition (38) imply that

$$\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T}(m_{1,t} - m_{2,t}) \xrightarrow{d} N(0_N, \Omega_m). \tag{49}$$

It follows that, the limiting distribution of $\sqrt{T}\left(\bar{R}_T - X_T\lambda_T\right)$ is multivariate normal with mean 0_N and covariance matrix

$$\Xi = M\Omega_m M'. \tag{50}$$

It is clear that Ξ is a nonnegative definite matrix and so its eigenvalues are all nonnegative. However, it is not of full rank, since $rank(M) = N - (1+K) < N$. This means that $\sqrt{T}\left(\bar{R}_T - X_T\lambda_T\right)$ has a limiting normal distribution that is degenerate.

Let $\Omega_m^{1/2}$ be a matrix such that $\Omega_m = (\Omega_m^{1/2})'\Omega_m^{1/2}$. For instance, $\Omega_m^{1/2}$ could be the upper triangular matrix in the Cholesky decomposition of Ω_m. Furthermore, define $\Xi^{1/2} = \Omega_m^{1/2} M'$, so that $\Xi = (\Xi^{1/2})'\Xi^{1/2}$. Then, it is easy to see that the limiting distribution of $\sqrt{T}\left(\bar{R}_T - X_T\lambda_T\right)$ is the same as the distribution of $(\Xi^{1/2})'Z$, where Z follows a multivariate normal distribution with mean 0_N and covariance matrix I_N. Therefore, we are now able to state the theorem that characterizes the limiting distribution of the mispricing measure Δ_T.

Theorem 11 *Assume that the matrix $X = [\ 1_N \quad B\ ]$ is of full rank equal to $1+K$ and that Assumption 2 is in effect. Let $M = I_N - X(X'Q^{-1}X)^{-1}X'Q^{-1}$ and $\Xi^{1/2} = \Omega_m^{1/2}M'$ where $\Omega_m^{1/2}$ satisfies $\Omega_m = (\Omega_m^{1/2})'\Omega_m^{1/2}$, and Ω_m is defined by (38). Then, the asymptotic distribution of the mispricing measure Δ_T, defined in (47), is given by*

$$T\Delta_T^2 \xrightarrow{d} Z'\left[\Xi^{1/2}G^{-1}(\Xi^{1/2})'\right]Z, \tag{51}$$

where Z is an N-dimensional standard normal variable.

From the preceding theorem, it follows that, in general, the asymptotic distribution of the test statistic $T\Delta_T^2$ is not a standard tabulated distribution. Therefore, as is the case with the HJ distance test, a simulation procedure is required to obtain asymptotically valid p-values. This is the subject of the following discussion.

Let $G^{1/2}$ be a matrix so that $G = (G^{1/2})'G^{1/2}$ (for example, $G^{1/2}$ could be the upper triangular matrix in the Cholesky decomposition of G). Then, using properties 0.4.6 (b) and (d) in Horn and Johnson (1990), we have

$$\begin{aligned} rank(\Xi^{1/2}G^{-1}(\Xi^{1/2})') &= rank(\Xi^{1/2}G^{-1/2}(G^{-1/2})'(\Xi^{1/2})') = rank(\Xi^{1/2}G^{-1/2}) \\ &= rank(\Xi^{1/2}) = rank((\Xi^{1/2})'\Xi^{1/2}) = rank(\Xi), \end{aligned}$$

since $G^{1/2}$ is invertible. Since G is taken to be symmetric, it follows that $\Xi^{1/2}G^{-1}(\Xi^{1/2})'$ is symmetric as well, and thus diagonizable. Since $rank(\Xi^{1/2}G^{-1}(\Xi^{1/2})') = N-(1+K)$, the matrix $\Xi^{1/2}G^{-1}(\Xi^{1/2})'$ has $N-(1+K)$ positive eigenvalues, say $\rho_j, j = 1,\ldots,N-(1+K)$.[14] Then, the limiting distribution v in (51) can be represented as a mixture of $N-(1+K)$ independent $\chi^2(1)$ random variables v_j with weights ρ_j as follows: $v = \sum_{j=1}^{N-(1+K)} \rho_j v_j$. This representation is useful in comparing our testing procedure to the HJ distance approach as illustrated in Jagannathan and Wang (1996). However, for all practical purposes one does not need to compute the eigenvalues above in order to obtain p-values associated with the limiting distribution in (51). One can obtain

[14] See Theorem 3.11 in Schott (1997).

asymptotically valid p-values using the following procedure. First, we need to provide a consistent estimator, say P_T, of the matrix $\Xi^{1/2}G^{-1}(\Xi^{1/2})'$. It is typically the case that either G is known, or its choice allows one to obtain a consistent estimator. Thus, obtaining P_T is easy given that consistent estimates Ω_m and M are readily available. Let $\{Z_t : t = 1, \ldots T^*\}$ be independent simulated values from $N(0_N, I_N)$, where T^* is a large enough integer, and further let $z_t = Z_t' P_T Z_t$ for $t = 1, \ldots T^*$. It follows by the Monte Carlo principle that the distribution function of v, $F_v(a) = P[v \leq a]$, can be approximated by $\frac{1}{T^*}\sum_{t=1}^{T^*} 1_{[z_t \leq a]}$, with the approximation becoming better as T and T^* become larger.

6.2 Shanken's CSR Test

Cross-sectional regression tests of zero-beta factor models were first developed by Shanken (1985). His analysis assumes that the time-series regression innovation terms are normally distributed and serially independent as well as independent of the factors. In what follows, we show that Shanken's CSR test remains asymptotically valid under more relaxed assumptions on the dynamics of the innovation terms. The test developed in the previous section becomes Shanken's CSR test when we choose both the CSR matrix and the weighting matrix in the quadratic form to be equal to the inverse of the innovation covariance matrix, namely when $Q = \Sigma_\varepsilon$ and $G = \Sigma_\varepsilon$.

The formal set of assumptions used in the development of Shanken's CSR test is contained in the statement of Assumption 4. The same assumptions were used by Shanken (1992) in deriving the asymptotic properties of his CSR estimator. In this subsection, we examine how the test based on the distance Δ_T simplifies under Assumption 4. It follows from Proposition 5 that

$$\Omega_m = \Pi_m + \Psi_m = B\Sigma_{\bar{f}}B' + (1 + \lambda_1'\Sigma_f^{-1}\lambda_1)\Sigma_\varepsilon,$$

and so

$$\Xi = M\left[B\Sigma_{\bar{f}}B' + (1 + \lambda_1'\Sigma_f^{-1}\lambda_1)\Sigma_\varepsilon\right]M',$$

where M and Ξ are given by (48) and (50) respectively. Note that since $MX = 0_{N\times(1+K)}$ and $X = \left[\begin{array}{cc} 1_N & B \end{array}\right]$, we have $MB = 0_{N\times K}$, which, in turn, implies

$$\Xi = M\Omega_m M' = (1 + \lambda_1'\Sigma_f^{-1}\lambda_1)M\Sigma_\varepsilon M'.$$

Write $\Xi = (\Xi^{1/2})'\Xi^{1/2}$ where $\Xi^{1/2} = (1 + \lambda_1'\Sigma_f^{-1}\lambda_1)^{1/2}\Sigma_\varepsilon^{1/2}M'$ and $\Sigma_\varepsilon^{1/2}$ satisfies $\Sigma_\varepsilon = (\Sigma_\varepsilon^{1/2})'\Sigma_\varepsilon^{1/2}$.

Then

$$\begin{aligned}
\Xi^{1/2}G^{-1}(\Xi^{1/2})' &= (1+\lambda_1'\Sigma_f^{-1}\lambda_1)\Sigma_\varepsilon^{1/2}M'G^{-1}M(\Sigma_\varepsilon^{1/2})' \\
&= (1+\lambda_1'\Sigma_f^{-1}\lambda_1)\Sigma_\varepsilon^{1/2}M'G^{-1/2}(G^{-1/2})'M(\Sigma_\varepsilon^{1/2})' \\
&= (1+\lambda_1'\Sigma_f^{-1}\lambda_1)FF',
\end{aligned}$$

where

$$\begin{aligned}
F &= \Sigma_\varepsilon^{1/2}M'G^{-1/2} = \Sigma_\varepsilon^{1/2}\left(I_N - X(X'Q^{-1}X)^{-1}X'Q^{-1}\right)'G^{-1/2} \\
&= \Sigma_\varepsilon^{1/2}G^{-1/2} - \Sigma_\varepsilon^{1/2}Q^{-1}X(X'Q^{-1}X)^{-1}X'G^{-1/2}.
\end{aligned}$$

Hence, according to Theorem 11, as $T \to \infty$,

$$\frac{T\Delta_T^2}{1+\lambda_1'\Sigma_f^{-1}\lambda_1} \xrightarrow{d} Z'(FF')Z,$$

where $Z \sim N(0_N, I_{N\times N})$.

Next, we consider the test advanced by Shanken (1985). Using our context's notation, this amounts to specifying both the CSR matrix and the quadratic form weighting matrix to be equal to the covariance matrix of the time-series regression disturbances, namely $Q = G = \Sigma_\varepsilon$. Then

$$F = I_N - (\Sigma_\varepsilon^{-1/2})'X(X'\Sigma_\varepsilon^{-1}X)^{-1}X'\Sigma_\varepsilon^{-1/2},$$

and so F is clearly symmetric and easily seen to be idempotent

$$FF = \Sigma_\varepsilon^{1/2}M'\Sigma_\varepsilon^{-1/2}\Sigma_\varepsilon^{1/2}M'\Sigma_\varepsilon^{-1/2} = \Sigma_\varepsilon^{1/2}(MM)'\Sigma_\varepsilon^{-1/2} = F,$$

since $MM = M$. By the properties of the rank operator, it follows that $rank(F) = rank(M) = N-(1+K)$. Therefore, F has $N-(1+K)$ nonzero eigenvalues all equal to 1 and so $Z'(FF')Z$ is distributed as a χ^2 random variable with $N-(1+K)$ degrees of freedom. Collecting terms and using Slutsky's theorem, we obtain that, as $T \to \infty$,

$$\frac{T\Delta_T^2}{1+\lambda_{1,T}'\Sigma_{f,T}^{-1}\lambda_{1,T}} \xrightarrow{d} \chi^2_{N-(1+K)}.$$

This result describes the asymptotic behavior of the CSR test statistic as developed by Shanken (1985).

6.3 Optimal CSR Matrix Chi-Square Test

When we use the efficiency-optimal CSR matrix Q_{OP}^{-1} both in the cross-sectional regression and the quadratic form, i.e. $G^{-1} = Q^{-1} = Q_{\text{OP}}^{-1} \equiv \Omega_m^{-1}$, the test derived in Theorem 11 reduces to a chi-square test. This fact is demonstrated next. Note first that

$$\begin{aligned} M'Q_{\text{OP}}^{-1}M &= \left(I_N - X(X'Q_{\text{OP}}^{-1}X)^{-1}X'Q_{\text{OP}}^{-1}\right)' Q_{\text{OP}}^{-1} \left(I_N - X(X'Q_{\text{OP}}^{-1}X)^{-1}X'Q_{\text{OP}}^{-1}\right) \\ &= Q_{\text{OP}}^{-1} - Q_{\text{OP}}^{-1}X(X'Q_{\text{OP}}^{-1}X)^{-1}X'Q_{\text{OP}}^{-1}, \end{aligned}$$

and therefore

$$\begin{aligned} \Xi^{1/2}G^{-1}(\Xi^{1/2})' &= \Omega_m^{1/2}M'Q_{\text{OP}}^{-1}M(\Omega_m^{1/2})' \\ &= \Omega_m^{1/2}\left(\Omega_m^{-1} - \Omega_m^{-1}X(X'\Omega_m^{-1}X)^{-1}X'\Omega_m^{-1}\right)(\Omega_m^{1/2})' \\ &= I_N - (\Omega_m^{-1/2})'X(X'\Omega_m^{-1}X)^{-1}X'\Omega_m^{-1/2}. \end{aligned}$$

Routine algebra shows that the matrix $I_N - (\Omega_m^{-1/2})'X(X'\Omega_m^{-1}X)^{-1}X'\Omega_m^{-1/2}$ is idempotent with rank equal to $N-(1+K)$. This implies that if Z is a $N(0_N, I_N)$ variable then $Z'\left(\Xi^{1/2}G^{-1}(\Xi^{1/2})'\right)Z$ follows a chi-square distribution with $N-(1+K)$ degrees of freedom. We have demonstrated the following corollary to Theorem 11.[15]

Corollary 12 *When the quadratic form weighting matrix G^{-1} and the CSR matrix Q^{-1} both equal the efficiency-optimal CSR matrix $Q_{OP}^{-1} \equiv \Omega_m^{-1}$, we have*

$$T\Delta_T^2 \xrightarrow{d} \chi^2_{N-(1+K)} \text{ as } T \to \infty.$$

Note the analogy to the J-test of Hansen (1982) which uses the optimal GMM weighting matrix. While in the CSR method we have to choose two matrices Q^{-1} and G^{-1}, it turns out that, when they are both chosen equal to the optimal CSR matrix, the resulting test has an asymptotic chi-square distribution instead of a mixture of chi-square distributions.

7 Simulation study

The cross-sectional test studied in this paper as well as the related GMM tests are justified asymptotically. Nevertheless, researchers are faced with data sets of finite, and occasionally rather small, sample sizes. It is therefore imperative to obtain a sense of the small sample performance of

[15] This result has also been obtained by Ahn and Gadarowski (1999b) using the minimum distance approach. See their Theorem 3.

the various tests. Since finite-sample analytical results can be obtained only under certain distributional assumptions, it is customary to resort to Monte Carlo simulation. This allows us to alter the simulation input and develop an understanding of how sensitive the results are with respect to the various features of the data generating process.

Our setup is closely related to the simulation design of Ahn and Gadarowski (2004). Their paper is the only available source of evidence on the finite sample performance of the Hansen-Jagannathan distance test and Hansen's standard GMM test, and therefore serves as a natural benchmark for our results. In order to proceed with our Monte Carlo simulation experiment, we have to make a number of choices regarding the exact form of the linear asset-pricing model under consideration. Among the quantities we have to specify are the size of the cross section N, the length of the time series T, the number of factors K, the values of the zero-beta rate λ_0, the risk premia λ_1 and the distributions of the factors and the disturbances. Since the majority of empirical studies use monthly data, we will select the related parameters accordingly.

Two of the most influential empirical studies of the cross section of stock returns, Fama and French (1992) and Jagannathan and Wang (1996), both employ three pervasive factors. Therefore, following Ahn and Gadarowski (2004), we will use three factors ($K = 3$) in our simulations. When examining the cross section of expected returns, it is rather difficult to work with individual securities for several reasons, including data availability and high dimensionality. This leads most researchers to form portfolios, typically a small number, and use them in the empirical study of the model under examination. The number of portfolios used ranges from 25 (Fama and French (1992), sorted according to size and book-to-market ratio) to as large as 100 (Jagannathan and Wang (1996), sorted according to size and beta). To ensure coverage of the most relevant cases, we will consider three choices for the number of portfolios: $N = 25$, $N = 48$, and $N = 100$. The rationale behind our choice will be clear in the sequel. The time-series length will assume four values: $T = 240$, $T = 360$, $T = 480$, and $T = 600$. These numbers correspond to 20, 30, 40, and 50 years of monthly data respectively. We also consider the following three values for the level of significance (or size): 0.01, 0.05, and 0.10; these values are the most commonly used in practice. The results we report were obtained using 10,000 simulation repetitions.

Next, we describe the data generating process used in the simulation. The three factors are assumed to be normally distributed and mutually independent, as well as independent across time. We also assume that the factors are returns to traded portfolios so that the following pricing relation holds: $\mu_f = \lambda_0 1_K + \lambda_1$. Incorporating this pricing relation and the beta asset-pricing equation (1) into the time-series regression (5) we obtain

$$R_t = \lambda_0 1_N + B\left(f_t - \lambda_0 1_K\right) + \varepsilon_t$$

which, in turn, can be written as

$$R_{it} = \lambda_0 + \beta_{1i}\tilde{f}_{1t} + \beta_{2i}\tilde{f}_{2t} + \beta_{3i}\tilde{f}_{kt} + \varepsilon_{it},$$

where $\tilde{f}_{kt} = f_{kt} - \lambda_0$, $k = 1, 2, 3$, and $B = \left[\begin{array}{ccc} \beta_1 & \beta_2 & \beta_3 \end{array}\right]$. Note that $E[\tilde{f}_{kt}] = \lambda_k$ and let $\sigma_k^2 = Var[\tilde{f}_{kt}]$. For simplicity, we further assume that $Var[\varepsilon_t] = \sigma_\varepsilon^2 I_N$ for a scalar σ_ε^2. Following Ahn and Gadarowski (2004), we assume that factor variances are equal across factors and define the ratio $\sigma_k^2/\sigma_\varepsilon^2$ as the signal-to-noise (S/N) ratio. The reported results are based on a S/N ratio value of 1.

The parameter values used in our simulation experiment are determined as follows. The zero-beta rate λ_0 is set equal to 1.0033, the risk premia λ_k, $k = 1, 2, 3$ are all set equal to 0.0022, and the variances σ_k^2, $k = 1, 2, 3$ are all set equal to $7 \cdot 10^{-5}$. To obtain empirically relevant betas, we use the betas from 25 and 100 sorted portfolios based on size and book-to-market and 48 industry portfolios with respect to the three Fama-French factors over the period July 1963 - December 2003.[16]

We examine the behavior of two CSR tests and two GMM tests. The tests differ in the choice of weighting matrices used in their implementation. For both CSR and GMM, we use two choices for weighting matrix in the quadratic form: the second moment matrix (SM) and the asymptotically optimal (OP) for the method under consideration. We denote the two CSR tests by CSR-SM and CSR-OP. The GMM test that uses the second moment matrix, suggested by Hansen and Jagannathan (1997), is denoted by GMM-HJ while the optimal GMM test, due to Hansen (1982), is denoted by GMM-H.

The evidence in Ahn and Gadarowski (2004) suggests that the empirical p-value of the GMM-HJ test rejects the null hypothesis too often as compared to its asymptotic distribution when the number of test assets is large. We find similar evidence for the CSR-OP and CSR-SM tests. Ahn and Gadarowski (1999b) conjecture that using a small-sample correction on their minimum-distance test, which is equivalent to our CSR-OP test, by multiplying the test statistics by the factor $1 - N/T$ would lead to better small-sample properties. Note that this small-sample adjustment is asymptotically equivalent to the original test statistic. Their conjecture is based on theoretical arguments and simulation evidence by Amsler and Schmidt (1985). Indeed, we find that this adjustment dramatically improves the performance of the CSR-based tests and the GMM-HJ test. On the

[16] All data used in the beta computation are obtained from the website of Professor Kenneth French.

contrary, when this type of correction is applied to the GMM-H, it leads to gross underrejection in finite samples with respect to the asymptotic chi-square distribution. Therefore, this correction is not suitable for the GMM-H test. This is supported by the simulation evidence in Ahn and Gadarowski (2004). The simulation results we report use the small-sample correction only for the CSR-OP, CSR-SM and GMM-HJ tests.

We summarize the simulation by reporting the empirical p-values of the four tests for each combination of the time-series length T, the size of the cross section N, and the statistical level of significance. The first set of results, assuming a multivariate normal data generating process, is presented in Table 1. In general, as one would have expected, the performance of all tests improves as T increases, while it gets worse as N increases. Overall, the CSR-SM and GMM-HJ tests have essentially the same performance and they are the most accurate and robust. They produce reasonable empirical p-values across almost all combinations considered. For instance, for target size of 5 percent and $T = 360$, the CSR-SM rejects 6.0 percent, 6.4 percent and 8.6 percent of the time when $N = 25$, $N = 48$, and $N = 100$, respectively, while the corresponding rejection rates for the GMM-HJ test are 5.9 percent, 6.1 percent, and 8.2 percent. Both tests based on optimal weighting matrices, namely CSR-OP and GMM-H, systematically overreject the null hypothesis. However, the GMM-H test seems to produce better results across different combinations of N and T. In general, it is also clear that one should be cautious when the number of assets is large compared to the time-series length.

Next, we deviate from the normality assumption and employ a t distribution in order to evaluate the sensitivity of the finite sample performance of the tests in the presence of fat tails. In a recent paper, Tu and Zhou (2004) find that, for the set of portfolios they examine, the multivariate normality assumption is overwhelmingly rejected, while the assumption of a multivariate t distribution cannot be rejected by the data. Therefore, motivated by this finding, we use a multivariate t distribution with six degrees of freedom to model the joint distribution of the factors and the regression innovation terms. Table 2 reports the empirical p-values for the four tests considered. There are a few observations to be made. First, by comparing with Table 1, we observe that the tests reject the null hypothesis more frequently than they do under the normality assumption, with the CSR-SM and GMM-HJ tests still remaining the more accurate and robust among the tests considered. Specifically, for a target size equal to 5 percent, the CSR-SM test produces empirical p-values that range between 5.9 and 11.7 percent for $T = 240$, between 5.3 and 8.9 percent for $T = 360$, between 5.5 and 7.6 percent for $T = 480$ and between 5.5 and 6.3 percent for $T = 600$. The corresponding values for the GMM-HJ test are between 6.4 and 12.2 percent for $T = 240$, between 5.9 and 10.6 percent for $T = 360$, between 5.9 and 9.3 percent for $T = 480$ and between

5.9 and 7.7 percent for $T = 600$.

Overall, our simulation evidence suggests that the CSR-SM test exhibits the best performance among the tests under examination. Under normality the GMM-HJ performs equally well, but when we generate the data using a t distribution, the CSR-SM test does better. The CSR-OP test leads to severe size distortions, has consistently the poorest performance and, therefore, it should be avoided. The classical GMM-H test can be trusted only when the cross section of returns is small. Therefore, we suggest the use of the CSR-SM test as a reliable tool for econometric evaluation of linear factor models.

8 Conclusion

This paper offers a comprehensive treatment of the cross-sectional regression (CSR) method as a reliable tool for inference in asset pricing. The CSR method provides a simple and intuitive way for estimating and evaluating linear factor models. The method, first introduced by Black, Jensen and Scholes (1972), and Fama and MacBeth (1973), was developed into a rigorous econometric framework by Shanken (1985) and Shanken (1992). More recently, Jagannathan and Wang (1998) generalized the corresponding asymptotic theory to allow for conditional heteroscedasticity of the time-series regression innovations, incorporating at the same time the use of firm-specific characteristics. In a different line of research, Hansen and Jagannathan (1997) address the issue of model misspecification using the stochastic discount factor representation of asset-pricing models. The main tool in their analysis is a mispricing measure termed the HJ distance. Jagannathan and Wang (1996) develop an econometric test based on the HJ methodology, and use it to test a version of the conditional CAPM.

We build on the work of the aforementioned authors to obtain the sampling distribution of the risk premia CSR estimator under general conditions. To put our work in perspective, we review related estimators and tests based on the generalized method of moments (GMM). We show that CSR-based estimators are equivalent to GMM-based estimators in terms of asymptotic efficiency when both methods are applied optimally. This holds true regardless of which model representation we employ to obtain the moment conditions for the implementation of GMM. Using the asymptotic distribution of the risk premia CSR estimator, we develop a CSR specification test. The test statistic, expressed as a quadratic form, is based on a mispricing measure closely related to the HJ distance. Furthermore, the test is a generalization of the CSR test of Shanken (1985) that does not require the weighting matrix in the quadratic form to be model-specific and thus can be used for fair comparisons across models. When the weighting matrix equals the second moment matrix of the returns, the distance is shown to be equal to the maximum pricing error per unit norm of portfolio return. The specification test based on the corresponding statistic, which we call the CSR-SM test, possesses robust small-sample statistical properties. In a simulation study, that compares the proposed CSR-SM test, the CSR-OP test, the HJ distance test, and the standard GMM test of Hansen (1982), we demonstrate the practical advantage of the CSR-SM test.

9 Appendix

A number of well-known facts are used repeatedly in the main text and/or in the subsequent proofs. We state them here explicitly for the convenience of the reader.

(F1) Let x_T and y_T be two vector random sequences of dimension $L \times 1$. If $x_T \xrightarrow{p} 0_L$ (i.e., $x_T = o_p(1)$) and $y_T \xrightarrow{d} y$ then $x_T y_T \xrightarrow{p} 0_L$ (i.e., $x_T y_T = o_p(1)$) and $x_T + y_T \xrightarrow{d} y$.[17]

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

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

is a positive semidefinite matrix.[18]

(F3) The identity

$$(A + CBD)^{-1} = A^{-1} - A^{-1}C\left(B^{-1} + DA^{-1}C\right)^{-1}DA^{-1}$$

holds for conformable matrices A, B, C and D and whenever the inverses involved are well-defined.[19]

9.1 Pricing restrictions on traded factors

In this subsection, we deal with the case in which some of the factors are returns to traded portfolios. For the sake of brevity, we refer to such factors as traded factors. Application of the asset-pricing equation on the traded factors implies certain pricing restrictions that relate the risk premiums on the factors to their expected values. To facilitate the presentation, we need to introduce the

[17] A proof of fact (F1) can be found in standard econometrics textbooks.

[18] Fact (F2) is used in the seminal paper by Hansen (1982) to obtain the efficiency-optimal weighting GMM matrix used in the quadratic form. See also subsection 8.3.3 in Wooldridge (2002).

[19] Fact (F3) is stated in Theorem 1.7 in Schott (1997)).

following notation. Let r_t denote the K_r-dimensional vector of traded factors, s_t denote the K_s-dimensional vector of the remaining factors such that $f_t = [\ r_t' \quad s_t'\ ]'$ and $K = K_r + K_s$. We refer to s_t as the nontraded factors. The risk premium vectors on the traded and nontraded factors are denoted by λ_r and λ_s. We further let

$$\tilde{\lambda}_1 = [\ \lambda_r' \quad \lambda_s{}'\ ]', \ \tilde{\lambda} = [\ \lambda_0 \quad \lambda_s'\ ]'$$

and decompose the beta matrix in the natural fashion as $B = [\ B_r \quad B_s\ ]$. The asset-pricing equation (1) applied to the traded factors implies that

$$\mu_r = \lambda_0 1_{K_r} + \lambda_r. \tag{52}$$

A researcher can make use of the pricing restriction (52) in two ways. One approach is to impose (52) and then proceed to estimate the risk premiums λ_s on the nontraded factors. This approach implicitly makes the assumption that the traded factors are correctly priced by the asset-pricing model under examination. The other approach is to estimate the risk premiums $\lambda = [\ \lambda_0 \quad \lambda_r' \quad \lambda_s'\ ]'$ on all factors and then construct a specification test based on the pricing restriction (52). We describe both approaches in the next two subsections.

9.1.1 Imposing pricing restrictions on traded factors

Incorporating the pricing restriction described by (52) into (1) we obtain

$$\mu_R - B_r\mu_r = \lambda_0(1_N - B_r 1_{K_r}) + B_s\lambda_s = \tilde{X}\tilde{\lambda} \tag{53}$$

where

$$\tilde{X} = [\ 1_N - B_r 1_{K_r} \quad B_s\ ]. \tag{54}$$

Equation (53) suggests the following two-step CSR procedure to estimate $\tilde{\lambda}$. First, estimate the beta matrix B by B_T given by (25) so that $B_T = [\ B_{r,T} \quad B_{s,T}\ ]$ and then estimate $\tilde{X}$ by

$$\tilde{X}_T = [\ 1_N - B_{r,T} 1_{K_r} \quad B_{s,T}\ ]. \tag{55}$$

In the second step, $\tilde{\lambda}$ is estimated by

$$\tilde{\lambda}_T = \tilde{H}_T(\bar{R}_T - B_{r,T}\bar{r}_T) \tag{56}$$

where

$$\tilde{H}_T = (\tilde{X}_T'\tilde{Q}_T^{-1}\tilde{X}_T)^{-1}\tilde{X}_T'\tilde{Q}_T^{-1} \tag{57}$$

and $\tilde{Q}_T^{-1}$ is a consistent estimator of a CSR matrix $\tilde{Q}^{-1}$. The probability limit of $\tilde{H}_T$ is

$$\tilde{H} = (\tilde{X}'\tilde{Q}^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{Q}^{-1}. \tag{58}$$

Next, we define the returns adjusted for traded factor risk by

$$R_t^* = R_t - B_r r_t \tag{59}$$

and note that $\mu_{R^*} = \mu_R - B_r\mu_r = \tilde{X}\tilde{\lambda}$. To state the result describing the asymptotic behavior of the CSR estimator $\tilde{\lambda}_T$ defined by (56) we need to define the auxiliary time series

$$\tilde{m}_{1,t} = R_t^* - \mu_{R^*},\ \tilde{m}_{2,t} = [(f_t - \mu_f)'\Sigma_f^{-1}\tilde{\lambda}_1]\varepsilon_t, \tilde{m}_t = [\ (\tilde{m}_{1,t})' \quad (\tilde{m}_{2,t})'\ ]', \tag{60}$$

where ε_t is the disturbance term in the time-series regression (5). Recall that ε_t and f_t are uncorrelated by the definition of ε_t and therefore it follows that $E[\tilde{m}_t] = 0_{2N}$ (see equation (6)). The following assumption can be verified under standard regularity conditions.

Assumption 13 *For all $\tilde{\lambda}_1 \in \mathbb{R}^K$, the vector time series $\tilde{m}_t$ defined by (60) satisfies the central limit theorem, that is*

$$\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T}\tilde{m}_t \xrightarrow{d} N(0_{2N}, \Sigma_{\tilde{m}}),$$

with

$$\Sigma_{\tilde{m}} = \begin{bmatrix} \Psi_{\tilde{m}} & \Gamma_{\tilde{m}} \\ \Gamma'_{\tilde{m}} & \Pi_{\tilde{m}} \end{bmatrix}, \tag{61}$$

where $\Psi_{\tilde{m}}, \Gamma_{\tilde{m}}$ and $\Pi_{\tilde{m}}$ are $N \times N$ matrices, $\Sigma_{\tilde{m}} = \sum_{k=-\infty}^{\infty}\Sigma_{\tilde{m},k}$, and $\Sigma_{\tilde{m},k} = E[\tilde{m}_t\tilde{m}'_{t+k}]$, for all k.

Theorem 14 *Assume that the matrix $\tilde{X} = [\ 1_N - B_r 1_{K_r} \quad B_s\ ]$ is of full rank equal to $1 + K_s$. Then, under Assumption 13, as $T \to \infty$,*

$$\sqrt{T}(\tilde{\lambda}_T - \tilde{\lambda}) = \frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T}\tilde{H}\,(\tilde{m}_{1,t} - \tilde{m}_{2,t}) + o_p(1),$$

and therefore $\sqrt{T}(\tilde{\lambda}_T - \tilde{\lambda}) \xrightarrow{d} N(0_{1+K_s}, \Sigma_{\tilde{\lambda}})$ where

$$\Sigma_{\tilde{\lambda}} = \tilde{H}\Omega_{\tilde{m}}\tilde{H}', \tag{62}$$

and

$$\Omega_{\tilde{m}} = \Psi_{\tilde{m}} + \Pi_{\tilde{m}} - (\Gamma_{\tilde{m}} + \Gamma'_{\tilde{m}}), \tag{63}$$

$\tilde{H}$ is defined by (58), and $\Psi_{\tilde{m}}, \Pi_{\tilde{m}}$ and $\Gamma_{\tilde{m}}$ are defined by (61).

9.1.2 Testing pricing restrictions on traded factors

Instead of imposing the pricing restriction (52), one can estimate the risk premiums on all factors, $\lambda = [\ \lambda_0 \quad \lambda_r' \quad \lambda_s'\ ]'$, as described in subsection 4.2 and then test whether (52) holds. To develop the associated test we need the following $K_r \times (1+K)$ auxiliary matrix

$$\Lambda = [\ 1_{K_r} \quad I_{K_r} \quad 0_{K_r \times K_s}\ ].$$

Then (52) is written as

$$\mu_r = \Lambda\lambda$$

and thus a test can be based on the quadratic form

$$\Theta_T = T(\bar{r}_T - \Lambda\lambda_T)' W_T^{-1}(\bar{r}_T - \Lambda\lambda_T)$$

where W_T^{-1} is a consistent estimator of a weighting matrix W^{-1}. Define the vector time series

$$y_t = (r_t - \mu_r) - \Lambda H(m_{1,t} - m_{2,t}) \tag{64}$$

and make the following

Assumption 15 *The vector time series y_t satisfies the central limit theorem, that is*

$$\frac{1}{\sqrt{T}} \sum\nolimits_{t=1}^{T} y_t \xrightarrow{d} N(0_{K_r}, \Sigma_y)$$

where $\Sigma_y = \sum_{j=-\infty}^{\infty} \Sigma_{y,j}$, and $\Sigma_{y,j} = E\left[y_t y_{t+j}'\right]$, for all j.

Theorem 16 *The asymptotic distribution of the test statistic Θ_T is given*

$$\Theta_T = T(\bar{r}_T - \Lambda\lambda_T)' W_T^{-1}(\bar{r}_T - \Lambda\lambda_T) \xrightarrow{d} Z'[\Sigma_y^{1/2} W^{-1} (\Sigma_y^{1/2})']Z$$

where $\Sigma_y^{1/2}$ is determined by $\Sigma_y = (\Sigma_y^{1/2})'\Sigma_y^{1/2}$ and Z is a K_r-dimensional standard normal variable. If $W = \Sigma_y$ the test reduces to a chi-square test, namely

$$\Theta_T \xrightarrow{d} \chi^2_{K_r}.$$

9.2 Proofs

Proof of Proposition 1: Let $\xi = \mu_R - X\lambda$ and $\phi = G^{-1}\xi$. If $\xi = 0_N$ the result holds trivially. So, we only consider the case $\xi \neq 0_N$. Using the inner product $(x, y) = x'Gy$ on $\mathbb{R}^N$ we project w on ϕ to obtain $w = \gamma\phi + \omega$ with $\gamma = (w, \phi)/(\phi, \phi)$ and $(\omega, \phi) = 0$. Then note that $(\phi, \phi) = \xi' G^{-1}\xi$ $= [\Delta_G^{\mathrm{B}}(\theta)]^2$ and further

$$\begin{aligned}\frac{|w'(\mu_R - X\lambda)|}{||p||} &= \frac{|w'\xi|}{||w'R||} = \frac{|(\gamma\phi+\omega)'G\phi|}{\sqrt{w'E[RR']w'}} = \frac{|\gamma(\phi,\phi)+(\omega,\phi)|}{\sqrt{(w,w)}} \\ &= \frac{|\gamma||(\phi,\phi)|}{\sqrt{\gamma^2(\phi,\phi)+(\omega,\omega)}} \leq \frac{|\gamma||(\phi,\phi)|}{\sqrt{\gamma^2(\phi,\phi)}} = \sqrt{(\phi,\phi)} = \Delta_G^{\mathrm{B}}(\theta)\end{aligned}$$

with the equality achieved when $\omega = 0_N \Leftrightarrow w = \gamma G^{-1}\xi$ for some γ. Q.E.D.

The following lemma will be used in the subsequent proofs.

Lemma 17 *Let c be a vector of dimension $K \times 1$. Then*

$$\sqrt{T}\,(B_T - B)\,c = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\left(f_t - \mu_f\right)'\Sigma_f^{-1}c\right]\varepsilon_t + o_p(1).$$

Proof of Lemma 17: From the time-series regression (5) we obtain

$$R_t - \bar{R}_T = B(f_t - \bar{f}_T) + (\varepsilon_t - \bar{\varepsilon}_T)$$

Using the last expression and equation (27), we observe that (26) delivers

$$\Sigma_{Rf,T} = B\Sigma_{f,T} + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\left(f_t - \bar{f}_T\right)'.$$

Assumption 2 implies that the columns of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t\left(f_t - \mu_f\right)'$ are asymptotically jointly normally distributed. Thus, it follows from fact (F1) that

$$\begin{aligned}\sqrt{T}\,(B_T - B)\,c &= \sqrt{T}\left(\Sigma_{Rf,T}\Sigma_{f,T}^{-1} - B\right)c \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t\left(f_t - \bar{f}_T\right)'\Sigma_{f,T}^{-1}c \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\left(f_t - \mu_f\right)'\Sigma_f^{-1}c\right]\varepsilon_t + o_p(1)\end{aligned}$$

since $\bar{f}_T \xrightarrow{p} \mu_f$ and $\Sigma_{f,T} \xrightarrow{p} \Sigma_f$ as $T \to \infty$. Q.E.D.

Proof of Theorem 14: From equations (56) and (59) we have,

$$\begin{aligned}\tilde{\lambda}_T - \tilde{\lambda} &= \tilde{H}_T(\bar{R}_T - B_{r,T}\bar{r}_T) - \tilde{\lambda} \\ &= H_T[\bar{R}_T^* - (B_{r,T} - B_r)\bar{r}_T - \tilde{X}_T\tilde{\lambda}] \\ &= H_T[(\bar{R}_T^* - \tilde{X}\tilde{\lambda}) - (B_{r,T} - B_r)\bar{r}_T - (\tilde{X}_T - \tilde{X})\tilde{\lambda}].\end{aligned}$$

Since $\tilde{X}_T - \tilde{X} = \begin{bmatrix} -(B_{r,T} - B_r)1_{K_r} & B_{s,T} - B_s \end{bmatrix}$ it follows that

$$(\tilde{X}_T - \tilde{X})\tilde{\lambda} = -\lambda_0(B_{r,T} - B_r)1_{K_r} + (B_{s,T} - B_s)\lambda_s.$$

Using Lemma 17 we obtain

$$\sqrt{T}(\tilde{\lambda}_T - \tilde{\lambda}) = H[\sqrt{T}(\bar{R}_T^* - \mu_{R^*}) - \sqrt{T}(B_{r,T} - B_r)\mu_r - \sqrt{T}(\tilde{X}_T - \tilde{X})\tilde{\lambda}] + o_p(1)$$

Furthermore,

$$(B_{r,T} - B_r)\mu_r = (B_{r,T} - B_r)(\lambda_0 1_{K_r} + \lambda_r) = \lambda_0(B_{r,T} - B_r)1_{K_r} + (B_{r,T} - B_r)\lambda_r.$$

Collecting terms we obtain

$$\begin{aligned}\sqrt{T}(\tilde{\lambda}_T - \tilde{\lambda}) &= H[\sqrt{T}(\bar{R}_T^* - \mu_{R^*}) - \sqrt{T}(B_{r,T} - B_r)\lambda_r - \sqrt{T}(B_{s,T} - B_s)\lambda_s] + o_p(1) \\ &= H[\sqrt{T}(\bar{R}_T^* - \mu_{R^*}) - \sqrt{T}(B_T - B)\tilde{\lambda}_1] + o_p(1).\end{aligned}$$

upon using the definition $\tilde{\lambda}_1 = \begin{bmatrix} \lambda_r' & \lambda_s' \end{bmatrix}'$. Lemma 17 implies that

$$\sqrt{T}\,(B_T - B)\,\tilde{\lambda}_1 = \frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T} \tilde{m}_{2,t} + o_p(1)$$

where $\tilde{m}_{2,t}$ is defined by (60). Thus, collecting terms together and using definition (60) yields

$$\begin{aligned}\sqrt{T}(\tilde{\lambda}_T - \tilde{\lambda}) &= \frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T} \tilde{H}\,(\tilde{m}_{1,t} - \tilde{m}_{2,t}) + o_p(1) \\ &= \tilde{H}\begin{bmatrix} I_N & -I_N \end{bmatrix}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T} \tilde{m}_t + o_p(1).\end{aligned}$$

It now follows that $\sqrt{T}(\tilde{\lambda}_T - \tilde{\lambda})$ has a limiting normal distribution with mean zero and covariance matrix given by

$$\Sigma_{\tilde{m}} = \tilde{H}\begin{bmatrix} I_N & -I_N \end{bmatrix}\begin{bmatrix} \Psi_{\tilde{m}} & \Gamma_{\tilde{m}} \\ \Gamma_{\tilde{m}}' & \Pi_{\tilde{m}} \end{bmatrix}\begin{bmatrix} I_N & -I_N \end{bmatrix}'\tilde{H}' = \tilde{H}\left[\Psi_{\tilde{m}} + \Pi_{\tilde{m}} - (\Gamma_{\tilde{m}} + \Gamma_{\tilde{m}}')\right]\tilde{H}'$$

as it follows from Assumption 13. This completes the proof. Q.E.D.

Proof of Theorem 16: From (32) it follows that

$$\begin{aligned}
\bar{r}_T - \Lambda\lambda_T &= \bar{r}_T - \Lambda H_T \bar{R}_T \\
&= (\bar{r}_T - \mu_r) - \Lambda H(\bar{R}_T - \mu_R) - \Lambda(H_T - H)\bar{R}_T \\
&= \frac{1}{T}\sum\nolimits_{t=1}^{T}[(r_t - \mu_r) - \Lambda H m_{1,t}] - \Lambda(H_T - H)\bar{R}_T
\end{aligned}$$

since $\Lambda H \mu_R = \Lambda H X \lambda = \Lambda\lambda = \mu_r$. Next, observe that

$$\begin{aligned}
\sqrt{T}(H_T - H)\bar{R}_T &= \sqrt{T}(H_T - H)X\lambda + (H_T - H)[\sqrt{T}(\bar{R}_T - \mu_R)] \\
&= \sqrt{T}(H_T - H)X\lambda + o_p(1).
\end{aligned}$$

Moreover, we have $(H_T - H)X = H_T(X - X_T) + H_T X_T - HX = -H_T[\begin{array}{cc} 0_N & B_T - B \end{array}]$ and thus, using Lemma 17 we obtain

$$\sqrt{T}(H_T - H)\bar{R}_T = -H_T[\sqrt{T}(B_T - B)\lambda_1] + o_p(1) = -H\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T} m_{2,t} + o_p(1)$$

where $m_{2,t}$ is given by (35). Collecting terms yields that, as $T \to \infty$,

$$\sqrt{T}(\bar{r}_T - \Lambda\lambda_T) = \frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T} y_t + o_p(1) \xrightarrow{d} (\Sigma_y^{1/2})' Z$$

where y_t is defined by (64), $\Sigma_y^{1/2}$ satisfies $\Sigma_y = (\Sigma_y^{1/2})'\Sigma_y^{1/2}$ and Z has a $N(0_{K_r}, I_{K_r})$ distribution. The desired conclusion follows upon using Slutsky's theorem. Q.E.D.

Proof of Proposition 5: Definition (35) and the time-series regression (5) imply

$$\begin{aligned}
E[m_{1,t}(m_{2,t+j})'] &= E[(R_t - \mu_R)[\left(f_{t+j} - \mu_f\right)' \Sigma_f^{-1}\lambda_1]\varepsilon'_{t+j}] \\
&= E[(B(f_t - \mu_f) + \varepsilon_t)[\left(f_{t+j} - \mu_f\right)' \Sigma_f^{-1}\lambda_1]\varepsilon'_{t+j}].
\end{aligned}$$

Thus, Assumption 4 and the law of iterated expectations yield $E[m_{1,t}(m_{2,t+j})'] = 0_{N\times N}$ for all j, and therefore $\Gamma_m = 0_{N\times N}$. Furthermore, since $E[\varepsilon_t\varepsilon'_{t+j}|\mathcal{F}] = 0_{N\times N}$ for $j \neq 0$ we have

$$\begin{aligned}
E[m_{2,t}(m_{2,t+j})'] &= I_{[j=0]}E\left[\left[\left(f_t - \mu_f\right)' \Sigma_f^{-1}\lambda_1\right]^2 \varepsilon_t\varepsilon'_t\right] \\
&= I_{[j=0]}E\left[\lambda'_1\Sigma_f^{-1}\left(f_t - \mu_f\right)\left(f_t - \mu_f\right)' \Sigma_f^{-1}\lambda_1\right]\Sigma_\varepsilon \\
&= I_{[j=0]}(\lambda'_1\Sigma_f^{-1}\lambda_1)\Sigma_\varepsilon
\end{aligned}$$

and

$$E[m_{1,t}(m_{1,t+j})'] = BE\left[(f_t - \mu_f)(f_{t+j} - \mu_f)'\right]B' + I_{[j=0]}E\left[\varepsilon_t\varepsilon'_t\right]$$

using the law of iterated expectations once again. Hence, we obtain

$$\Psi_m = B\Sigma_{\bar{f}}B' + \Sigma_\varepsilon \text{ and } \Pi_m = (\lambda_1'\Sigma_f^{-1}\lambda_1)\Sigma_\varepsilon$$

and so $\Omega_m = \Psi_m + \Pi_m = B\Sigma_{\bar{f}}B' + (1 + \lambda_1'\Sigma_f^{-1}\lambda_1)\Sigma_\varepsilon$. According the Theorem 3 the asymptotic variance of λ_T equals

$$\Sigma_\lambda^{\text{CH}} = H\Omega_m H' = HB\Sigma_{\bar{f}}B'H' + (1 + \lambda_1'\Sigma_f^{-1}\lambda_1)H\Sigma_\varepsilon H'.$$

Moreover, $HX = (X'Q^{-1}X)^{-1}X'Q^{-1}X = I_{1+K}$ and thus $HB = [\ 0_K \quad I_K\ ]'$ which, in the light of definition (39), implies $HB\Sigma_{\bar{f}}B'H' = \Sigma_{\bar{f}}^*$ completing the proof. Q.E.D.

Proof of Proposition 6: From the proof of Proposition 5 it follows that $\Omega_m = \tau\Sigma_\varepsilon + B\Sigma_{\bar{f}}B'$ where $\tau = 1 + \lambda_1'\Sigma_f^{-1}\lambda_1$. Thus, using fact (F3), we obtain

$$\Omega_m^{-1} = \tau^{-1}\left[\Sigma_\varepsilon^{-1} - \Sigma_\varepsilon^{-1}B(\tau\Sigma_{\bar{f}}^{-1} + B'\Sigma_\varepsilon^{-1}B)^{-1}B'\Sigma_\varepsilon^{-1}\right],$$

and so

$$\begin{aligned} & (X'\Omega_m^{-1}X)\left(X'\Sigma_\varepsilon^{-1}X\right)^{-1} \\ = \quad & \tau^{-1}\left[I_{K+1} - X'\Sigma_\varepsilon^{-1}B(\tau\Sigma_{\bar{f}}^{-1} + B'\Sigma_\varepsilon^{-1}B)^{-1}B'\Sigma_\varepsilon^{-1}X\left(X'\Sigma_\varepsilon^{-1}X\right)^{-1}\right]. \end{aligned}$$

Since $X = [\ 1_N \quad B\ ]$, it follows that

$$\left(B'\Sigma_\varepsilon^{-1}X\right)\left(X'\Sigma_\varepsilon^{-1}X\right)^{-1}X' = [\ 0_K \quad I_K\ ]X' = B'$$

which, in turn, implies

$$\begin{aligned} & (X'\Omega_m^{-1}X)\left(X'\Sigma_\varepsilon^{-1}X\right)^{-1}X'\Sigma_\varepsilon^{-1}\Omega_m \\ = \quad & \tau^{-1}\left[X'\Sigma_\varepsilon^{-1}\Omega_m - X'\Sigma_\varepsilon^{-1}B(\tau\Sigma_{\bar{f}}^{-1} + B'\Sigma_\varepsilon^{-1}B)^{-1}B'\Sigma_\varepsilon^{-1}\Omega_m\right] \\ = \quad & X'\tau^{-1}\left[\Sigma_\varepsilon^{-1} - \Sigma_\varepsilon^{-1}B(\tau\Sigma_{\bar{f}}^{-1} + B'\Sigma_\varepsilon^{-1}B)^{-1}B'\Sigma_\varepsilon^{-1}\right]\Omega_m \\ = \quad & X' \end{aligned}$$

using the expression for Ω_m^{-1} above. Hence, pre- and post-multiplying by appropriate inverses we obtain

$$(X'\Omega_m^{-1}X)^{-1}X'\Omega_m^{-1} = (X'\Sigma_\varepsilon^{-1}X)^{-1}X'\Sigma_\varepsilon^{-1}.$$

If the estimator $\Omega_{m,T}$ given by (43) is used in (41), we can repeat the above arguments to obtain

$$(X_T'\Omega_{m,T}^{-1}X_T)^{-1}X_T'\Omega_{m,T}^{-1} = (X_T'\Sigma_{\varepsilon,T}^{-1}X_T)^{-1}X_T'\Sigma_{\varepsilon,T}^{-1}$$

and thus $\lambda_T^{\text{OP}} = \lambda_T^{\text{SH}}$ according to definitions (41) and (42). The proof of the proposition is complete since the asymptotic covariance matrices of λ_T^{OP} and λ_T^{SH} are the same regardless of which consistent estimators $\Omega_{m,T}$ and $\Sigma_{\varepsilon,T}$ we use. Q.E.D.

Proof of Theorem 7: First note that

$$\begin{aligned}
\delta_T^{\text{C}} - \delta &= -\left[(\Sigma_{f,T} + \lambda_T \bar{f}_T')^{-1}\lambda_T - (\Sigma_f + \lambda\mu_f')^{-1}\lambda\right] \\
&= -(\Sigma_{f,T} + \lambda_T \bar{f}_T')^{-1}\left[\lambda_T - \lambda + \left[(\Sigma_f + \lambda\mu_f') - (\Sigma_{f,T} + \lambda_T\bar{f}_T')\right](\Sigma_f + \lambda\mu_f')^{-1}\lambda\right] \\
&= -(\Sigma_{f,T} + \lambda_T \bar{f}_T')^{-1}(\lambda_T - \lambda) \\
&\quad -(\Sigma_{f,T} + \lambda_T \bar{f}_T')^{-1}\left[-(\lambda_T - \lambda)\bar{f}_T' - \lambda(\bar{f}_T - \mu_f)' - (\Sigma_{f,T} - \Sigma_f)\right](\Sigma_f + \lambda\mu_f')^{-1}\lambda \\
&= -[1 - \bar{f}_T'(\Sigma_f + \lambda\mu_f')^{-1}\lambda](\Sigma_{f,T} + \lambda_T \bar{f}_T')^{-1}(\lambda_T - \lambda) \\
&\quad +(\Sigma_{f,T} + \lambda_T \bar{f}_T')^{-1}(\Sigma_{f,T} - \Sigma_f)(\Sigma_f + \lambda\mu_f')^{-1}\lambda \\
&\quad +(\Sigma_{f,T} + \lambda_T \bar{f}_T')^{-1}\lambda\lambda'(\Sigma_f + \mu_f\lambda')^{-1}(\bar{f}_T - \mu_f).
\end{aligned}$$

Recall that

$$\sqrt{T}\,(\lambda_T - \lambda) = \frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T} H k_t$$

where

$$H = (B'Q^{-1}B)^{-1}B'Q^{-1}$$

and

$$k_t = R_t - \mu_R - [(f_t - \mu_f)'\Sigma_f^{-1}\lambda]\varepsilon_t.$$

Thus

$$\begin{aligned}
\sqrt{T}\,(\delta_T^{\text{C}} - \delta) &= -[1 - \mu_f'(\Sigma_f + \lambda\mu_f')^{-1}\lambda](\Sigma_f + \lambda\mu_f')^{-1}\sqrt{T}\,(\lambda_T - \lambda) \\
&\quad +(\Sigma_f + \lambda\mu_f')^{-1}\sqrt{T}\,(\Sigma_{f,T} - \Sigma_f)(\Sigma_f + \lambda\mu_f')^{-1}\lambda \\
&\quad +(\Sigma_f + \lambda\mu_f')^{-1}\lambda\lambda'(\Sigma_f + \mu_f\lambda')^{-1}\sqrt{T}(\bar{f}_T - \mu_f) + o_p(1).
\end{aligned}$$

Now since $HB = I_K$, $\delta = -(\Sigma_f + \lambda\mu_f')^{-1}\lambda$, $1 - \mu_f'(\Sigma_f + \lambda\mu_f')^{-1}\lambda = 1 + \delta'\mu_f$ (see equation (20)) and

$$\begin{aligned}
\sqrt{T}(\bar{f}_T - \mu_f) &= \frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T}(f_t - \mu_f), \\
\sqrt{T}\,(\Sigma_{f,T} - \Sigma_f)\,a &= \frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T}\left[(f_t - \mu_f)(f_t - \mu_f)' - \Sigma_f\right]a + o_p(1),
\end{aligned}$$

we obtain

$$\sqrt{T}\,(\delta_T^{\mathrm{C}}-\delta)=(\Sigma_f+\lambda\mu_f')^{-1}H\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T}\xi_t+o_p(1)$$

where

$$\xi_t=-(1+\delta'\mu_f)k_t-B\left[(f_t-\mu_f)(f_t-\mu_f)'-\Sigma_f\right]\delta-B\lambda\delta'(f_t-\mu_f).$$

Next we establish the link between the variables ξ_t and n_t. Note that the time-series regression (5) implies

$$k_t=B(f_t-\mu_f)+[1-(f_t-\mu_f)'\Sigma_f^{-1}\lambda]\varepsilon_t=B(f_t-\mu_f)+\frac{1+\delta' f_t}{1+\delta'\mu_f}\varepsilon_t$$

since $1-(f_t-\mu_f)'\Sigma_f^{-1}\lambda=\frac{1+\delta' f_t}{1+\delta'\mu_f}$ as it follows from (19). Thus

$$\begin{aligned}\xi_t &= -(1+\delta'\mu_f)B(f_t-\mu_f)-(1+\delta' f_t)\varepsilon_t\\ &\quad -B(f_t-\mu_f)(f_t-\mu_f)'\delta+B\Sigma_f\delta-B\lambda\delta'(f_t-\mu_f).\end{aligned}$$

By definition $n_t=R_t(1+\delta' f_t)$ and thus the time-series regression (5) and the beta-pricing equation (16) imply

$$\begin{aligned}n_t &= B(\lambda+f_t-\mu_f)(1+\delta' f_t)+\varepsilon_t(1+\delta' f_t)\\ &= B(\lambda+f_t-\mu_f)\left[1+\delta'\mu_f+\delta'(f_t-\mu_f)\right]+\varepsilon_t(1+\delta' f_t)\\ &= (1+\delta'\mu_f)B(f_t-\mu_f)+B(f_t-\mu_f)(f_t-\mu_f)'\delta\\ &\quad +B\lambda\delta'(f_t-\mu_f)+B\lambda(1+\delta'\mu_f)+\varepsilon_t(1+\delta' f_t)\\ &= -\xi_t\end{aligned}$$

since $\lambda(1+\delta'\mu_f)=-\Sigma_f\delta$ (see equation (19)). Therefore,

$$\sqrt{T}\,(\delta_T^{\mathrm{C}}-\delta)=-(\Sigma_f+\lambda\mu_f')^{-1}(B'Q^{-1}B)^{-1}B'Q^{-1}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T}n_t+o_p(1).$$

Applying the argument that yields the asymptotically efficient choice of the GMM matrix, we obtain that the optimal choice of the CSR matrix is $Q^{-1}=\Omega_n^{-1}$. Under this choice we immediately have

$$\sqrt{T}(\delta_T^{\mathrm{C}}-\delta)=-(\Sigma_f+\lambda\mu_f')^{-1}(B'\Omega_n^{-1}B)^{-1}B'\Omega_n^{-1}\frac{1}{\sqrt{T}}\sum\nolimits_{t=1}^{T}n_t+o_p(1)\tag{65}$$

and thus the asymptotic distribution of δ_T^{C} is given by

$$\sqrt{T}(\delta_T^{\mathrm{C}}-\delta)\xrightarrow{d}N(0_K,(\Sigma_f+\lambda\mu_f')^{-1}(B'\Omega_n^{-1}B)^{-1}(\Sigma_f+\mu_f\lambda')^{-1}).$$

By the definition of B and the asset-pricing equation (16), we have

$$B\Sigma_f = E[R_t\left(f_t - \mu_f\right)'] = E[R_t f_t'] - B\lambda\mu_f'$$

and thus

$$B(\Sigma_f + \lambda\mu_f') = D \Leftrightarrow B = D(\Sigma_f + \lambda\mu_f')^{-1} \tag{66}$$

under the assumption that $\Sigma_f + \lambda\mu_f'$ is invertible. Thus, in the light of equation (66), combining the expressions (24) and (65) completes the proof. Q.E.D.

<u>Proof of Theorem 9</u>: Defining $P_T = (\Sigma_{Rf,T}' Q_T^{-1}\Sigma_{Rf,T})^{-1}\Sigma_{Rf,T}' Q_T^{-1}$ we have $P_T\Sigma_{Rf,T} = I_K$ so that

$$\sqrt{T}(b_T^{\mathrm{C}} - b) = \sqrt{T}(P_T\overline{R_T} - b) = P_T\sqrt{T}(\overline{R_T} - \Sigma_{Rf,T}b).$$

Furthermore,

$$\begin{aligned}
\sqrt{T}(\overline{R_T} - \Sigma_{Rf,T}b) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[R_t - R_t(f_t - \bar{f}_T)'b\right] \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[R_t(1 - (f_t - \mu_f)'b) + R_t b'(\bar{f}_T - \mu_f)\right] \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[R_t(1 - (f_t - \mu_f)'b)\right] + \bar{R}_T b'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(f_t - \mu_f) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_t + o_p(1).
\end{aligned}$$

The result now follows by Slutsky's theorem since $P_T \xrightarrow{p} P$ as $T \to \infty$. Q.E.D.

<u>Proof of Theorem 10</u>: First, recall that the asymptotic variance of b_T^{C} is given by

$$\Omega_b^{\mathrm{C}} = (\Sigma_{Rf}'\Omega_\phi^{-1}\Sigma_{Rf})^{-1}.$$

Moreover, it follows from (45) and the formula for the inverse of a partitioned matrix that

$$\Omega_g^{-1} = \begin{bmatrix} \Omega^h & \Omega^{hf} \\ \Omega^{fh} & \Omega^f \end{bmatrix}$$

where

$$\Omega^h = (\Omega_h - \Omega_{hf}\Omega_f^{-1}\Omega_{fh})^{-1} = \Omega_h^{-1} + \Omega_h^{-1}\Omega_{hf}\Omega^f\Omega_{fh}\Omega_h^{-1},$$

$$\Omega^f = (\Omega_f - \Omega_{fh}\Omega_h^{-1}\Omega_{hf})^{-1},$$

$$\Omega^{hf} = -\Omega_h^{-1}\Omega_{hf}(\Omega_f - \Omega_{fh}\Omega_h^{-1}\Omega_{hf})^{-1} = -\Omega^h\Omega_{hf}\Omega_f^{-1},$$
$$\Omega^{fh} = -\Omega_f^{-1}\Omega_{fh}(\Omega_h - \Omega_{hf}\Omega_f^{-1}\Omega_{fh})^{-1} = -\Omega^f\Omega_{fh}\Omega_h^{-1}.$$

This implies

$$D_g'\Omega_g^{-1}D_g = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}$$

where

$$\begin{aligned} U_{11} &= \Sigma_{Rf}'\Omega^h\Sigma_{Rf}, \\ U_{12} &= -\Sigma_{Rf}'(\Omega^h\Sigma_{Rf}bb' - \Omega^{hf}), \\ U_{21} &= U_{12}' = -(bb'\Sigma_{Rf}'\Omega^h - \Omega^{fh})\Sigma_{Rf}, \\ U_{22} &= bb'(\Sigma_{Rf}'\Omega^h\Sigma_{Rf})bb' - bb'\Sigma_{Rf}'\Omega^{hf} - \Omega^{fh}\Sigma_{Rf}bb' + \Omega^f. \end{aligned}$$

Using the formula for the inverse of a partitioned matrix and the formula for the inverse of a sum of matrices, we obtain that

$$\Omega_b^{\mathrm{G}} = (U_{11} - U_{12}U_{22}^{-1}U_{21})^{-1}.$$

Since

$$U_{11} - U_{12}U_{22}^{-1}U_{21} = \Sigma_{Rf}'[\Omega^h - (\Omega^h\Sigma_{Rf}bb' - \Omega^{hf})U_{22}^{-1}(bb'\Sigma_{Rf}'\Omega^h - \Omega^{fh})]\Sigma_{Rf}$$

to show that $\Omega_b^{\mathrm{C}} = \Omega_b^{\mathrm{G}}$ it suffices to show

$$\begin{aligned} \Omega_\phi^{-1} &= \Omega^h - (\Omega^h\Sigma_{Rf}bb' - \Omega^{hf})U_{22}^{-1}(bb'\Sigma_{Rf}'\Omega^h - \Omega^{fh}) \Leftrightarrow \\ \Omega_\phi^{-1} &= \Omega^h - \Omega^h(\Sigma_{Rf}bb' - (\Omega^h)^{-1}\Omega^{hf})U_{22}^{-1}(bb'\Sigma_{Rf}' - \Omega^{fh}(\Omega^h)^{-1})\Omega^h \Leftrightarrow \\ \Omega_\phi^{-1} &= \Omega^h - \Omega^h(\Sigma_{Rf}bb' + \Omega_{hf}\Omega_f^{-1})U_{22}^{-1}(bb'\Sigma_{Rf}' + \Omega_f^{-1}\Omega_{fh})\Omega^h \end{aligned}$$

which is equivalent to

$$\Omega_\phi^{-1} = \Omega^h - \Omega^h G_1'\Omega_f^{-1}U_{22}^{-1}\Omega_f^{-1}G_1\Omega^h \tag{68}$$

where

$$G_1 = \Omega_f bb'\Sigma_{Rf}' + \Omega_{fh}.$$

Note that use was made of the identity $(\Omega^h)^{-1}\Omega^{hf} = -\Omega_{hf}\Omega_f^{-1}$. Next, we proceed to show that the equality stated in (68) is indeed valid. First, note that

$$\begin{aligned}
\Omega_f^{-1} &= \Omega^f - \Omega^{fh}(\Omega^h)^{-1}\Omega^{hf} \\
&= \Omega^f - \Omega^{fh}(\Omega^h)^{-1}\Omega^h(\Omega^h)^{-1}\Omega^{hf} \\
&= \Omega^f - \Omega^{fh}(\Omega^h)^{-1}\Omega^h(\Omega^h)^{-1}\Omega^{hf} \\
&= \Omega^f - \Omega_f^{-1}\Omega_{fh}\Omega^h\Omega_{hf}\Omega_f^{-1}
\end{aligned}$$

implying

$$\Omega^f - \Omega_f^{-1} = \Omega_f^{-1}\Omega_{fh}\Omega^h\Omega_{hf}\Omega_f^{-1}.$$

Therefore,

$$U_{22} = bb'(\Sigma_{Rf}'\Omega^h\Sigma_{Rf})bb' - bb'\Sigma_{Rf}'\Omega^{hf} - \Omega^{fh}\Sigma_{Rf}bb' + \Omega^f \Rightarrow$$

$$U_{22} - \Omega_f^{-1} = bb'(\Sigma_{Rf}'\Omega^h\Sigma_{Rf})bb' - bb'\Sigma_{Rf}'\Omega^{hf} - \Omega^{fh}\Sigma_{Rf}bb' + \Omega_f^{-1}\Omega_{fh}\Omega^h\Omega_{hf}\Omega_f^{-1} \Rightarrow$$

$$U_{22} - \Omega_f^{-1} = (bb'\Sigma_{Rf}' + \Omega_f^{-1}\Omega_{fh})\Omega^h(\Sigma_{Rf}bb' + \Omega_{hf}\Omega_f^{-1}) \Rightarrow$$

$$\Omega_f U_{22}\Omega_f - \Omega_f = G_1\Omega^h G_1' \Rightarrow$$

$$\Omega_f - U_{22}^{-1} = G_1\Omega^h G_1'\Omega_f^{-1}U_{22}^{-1} \Rightarrow$$

$$\Omega_f^{-1} - \Omega_f^{-1}U_{22}^{-1}\Omega_f^{-1} = \Omega_f^{-1}G_1\Omega^h G_1'\Omega_f^{-1}U_{22}^{-1}\Omega_f^{-1}$$

from which, after pre-multiplication by G_1' and post-multiplication by G_1 we obtain

$$\begin{aligned}
G_1'\Omega_f^{-1}G_1 - G_1'\Omega_f^{-1}U_{22}^{-1}\Omega_f^{-1}G_1 &= G_1'\Omega_f^{-1}G_1\Omega^h G_1'\Omega_f^{-1}U_{22}^{-1}\Omega_f^{-1}G_1 \Rightarrow \\
G_2 - G_3 &= G_2\Omega^h G_3
\end{aligned}$$

where

$$G_2 = G_1'\Omega_f^{-1}G_1,\ G_3 = G_1'\Omega_f^{-1}U_{22}^{-1}\Omega_f^{-1}G_1.$$

Hence,

$$\begin{aligned}
\Omega^h G_2 - \Omega^h G_3 - \Omega^h G_2\Omega^h G_3 &= 0_{N\times N} \Rightarrow \\
(\Omega^h + \Omega^h G_2\Omega^h)((\Omega^h)^{-1} - G_3) &= I_N \Rightarrow \\
((\Omega^h)^{-1} + G_2)\Omega^h((\Omega^h)^{-1} - G_3)\Omega^h &= I_N.
\end{aligned}$$

Since $\phi_t = \Phi g_t$ where $\Phi = [\ I_N \quad b\mu_R'\ ]$ we have

$$\begin{aligned}
\Omega_\phi &= \Phi\Omega_g\Phi' = \Omega_h + \mu_R b'\Omega_{fh} + \Omega_{hf}b\mu_R' + \mu_R b'\Omega_f b\mu_R \Rightarrow \\
\Omega_\phi &= \Omega_h - \Omega_{hf}\Omega_f^{-1}\Omega_{fh} + \left(\Sigma_{Rf}bb'\Omega_f + \Omega_{hf}\right)\Omega_f^{-1}\left(\Omega_f bb'\Sigma_{Rf}' + \Omega_{fh}\right) \Rightarrow \\
\Omega_\phi &= (\Omega^h)^{-1} + G_2
\end{aligned}$$

Therefore, it follows that

$$\Omega_\phi^{-1} = \Omega^h((\Omega^h)^{-1} - G_3)\Omega^h = \Omega^h - \Omega^h G_3 \Omega^h$$

and so the equality in (68) is established. This completes the proof. Q.E.D.

REFERENCES

Ahn, S.C., and C. Gadarowski, 1999a, Small Sample Properties of the Model Specification Test Based on the Hansen-Jagannathan Distance, Working Paper, Arizone State University.

Ahn, S.C., and C. Gadarowski, 1999b, Two-Pass Cross-Sectional Regression of Factor Pricing Models: Minimum Distance Approach, Working Paper, Arizona State University.

Ahn, S.C., and C. Gadarowski, 2004, Small Sample Properties of the Model Specification Test Based on the Hansen-Jagannathan Distance, *Journal of Empirical Finance* 11, 109-132.

Amsler, C.E., and P. Schimdt, 1985, A Monte Carlo investigation of the accuracy of multivariate CAPM tests, *Journal of Financial Economics* 14, 359-375.

Andrews, D. W. K., 1991, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* 59, 817-858.

Andrews, D. W. K., and J.C. Monahan, 1992, An Improved Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica* 60, 953-966.

Banz, R.W., 1981, The Relationship between Return and Market Value of Common Stocks, *Journal of Financial Economics* 9, 3-18.

Black, F., M. Jensen, and M. S. Scholes, 1972, The Capital Asset Pricing Model: Some Empirical Tests, in Jensen M. (ed.), *Studies in the Theory of Capital Markets*, Praeger, New York.

Campbell, J., 1993, Intertemporal Asset Pricing without Consumption Data, *American Economic Review* 83, 487-512.

Campbell, J., 1996, Understanding Risk and Return, *Journal of Political Economy* 104, 298-344.

Chamberlain, G., and M. Rothschild, 1983, Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets, *Econometrica* 51, 1281-1304.

Chen, R., and R. Kan, 1986, Finite Sample Analysis of Two-Pass Cross-Sectional Regressions, Working Paper, University of Toronto.

Chen, N., R. Roll, and S. 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.

Cochrane, J.H., 2001a, *Asset Pricing*, Princeton University Press, Princeton.

Cochrane, J.H., 2001b, A Rehabilitation of Stochastic Discount Factor Methodology, Manuscript, University of Chicago.

Connor, G., 1984, A Unified Beta Pricing Theory, *Journal of Economic Theory* 34, 13-31.

Connor, G., and R. Korajczyk, 1988, Risk and Return in an Equilibrium APT: Application of a New Test Methodology, *Journal of Financial Economics* 15, 373-394.

Davidson, J., 1994, *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford University Press.

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

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

Dybvig, P. H., and J. E. Ingersoll, 1982, Mean-variance Theory in Complete Markets, *Journal of Business* 55, 233-252.

Fama, E. F., and K. French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance* 47, 427-465.

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

Hall, P., and C.C. Heyde, 1980, *Martingale Limit Theory and its Application*, Academic Press.

Hansen, L.P., 1982, Large Sample Properties of the Generalized Method of Moments Estimators, *Econometrica* 50, 1029-1054.

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.

Heaton, J., and D. Lucas, 2000, Portfolio Choice and Asset Prices: The Importance of Enterpreneurial Risk, *Journal of Finance* 55, 1163-1198.

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

Ingersoll, J., 1987, *Theory of Financial Decision Making*, Rowman & Littlefield.

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., 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, Asymptotic Theory for Estimating Beta Pricing Models Using Cross-Sectional Regressions, *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, 2337-2367.

Jones, C.S., Extracting Factors from Heteroscedastic Asset Returns, *Journal of Financial Economics* 62, 293-325.

Kan, R., and G. Zhou, 1999, A Critique of the Stochastic Discount Factor Methodology, *Journal of Finance* 54, 1221-1248.

Kan, R., and G. Zhou, 2003, Hansen-Jagannathan Distance: Geometry and Exact Distribution, Working Paper, Washington University.

Lehmann, B. and D. Modest, 1988, The Empirical Foundations of the Arbitrage Pricing Theory, *Journal of Financial Economics* 21, 213-254.

Lettau, M., and S. Ludvigson, 2001, Ressurecting the (C)CAPM: A Cross-Sectional Test when Risk Premia are Time-Varying, *Journal of Political Economy* 109, 1238-1287.

Lewellen, J., and S. Nagel, 2004, The Conditional CAPM Does Not Explain Asset-Pricing Anomalies, Working Paper, MIT.

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

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

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

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.

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

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

Sharpe, W.F., 1964, Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance* 19, 425-442.

Tu, J., and G. Zhou, 2004, Data-generating Process Uncertainty - What Difference Does it Make in Portfolio Decisions?, *Journal of Financial Economics* 72, 385-421.

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

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

Table 1: Empirical p-values under normal distribution.

This table reports the empirical p-values (in percentages) for T = 240, 360, 480 and 600, N = 25, 48 and 100 and level of significance (size) equal to 1 percent, 5 percent and 10 percent. The factors and the time-series regression disturbances are generated from a normal distribution.

	Test	N=25			N=48			N=100		
Size		1	5	10	1	5	10	1	5	10
T=240	CSR-SM	1.2	5.8	10.3	1.6	6.6	11.5	3.9	10.5	16.3
	CSR-OP	2.6	8.6	14.1	7.6	17.7	26.8	35.5	51.5	60.3
	GMM-HJ	1.2	5.6	10.3	1.5	6.4	11.1	3.3	9.4	15.4
	GMM-H	1.5	7.3	13.1	2.7	11.9	21.1	5.8	24.6	41.4
T=360	CSR-SM	1.4	6.0	11.0	1.6	6.4	11.6	2.7	8.6	14.2
	CSR-OP	2.3	7.9	14.2	4.8	13.6	21.4	18.4	34.3	45.1
	GMM-HJ	1.3	5.9	10.8	1.6	6.1	11.4	2.4	8.2	13.9
	GMM-H	1.6	6.8	13.3	2.4	10.2	18.2	5.8	20.1	33.2
T=480	CSR-SM	1.1	5.1	10.3	1.4	6.0	11.1	1.9	7.5	13.6
	CSR-OP	1.6	6.9	12.7	3.3	10.9	18.2	12.3	26.3	37.1
	GMM-HJ	1.1	5.1	10.3	1.3	5.8	10.9	1.7	7.4	13.1
	GMM-H	1.2	6.2	11.9	2.0	8.7	15.9	4.6	17.3	28.9
T=600	CSR-SM	1.3	5.8	11.1	1.2	5.6	11.0	1.7	6.9	13.2
	CSR-OP	1.8	6.9	12.7	2.7	10.0	16.8	9.0	22.3	32.3
	GMM-HJ	1.3	5.7	11.1	1.2	5.4	11.0	1.6	6.7	12.7
	GMM-H	1.4	6.4	12.2	1.7	8.1	15.1	4.0	15.6	26.4

Table 2: Empirical p-values under t distribution.

This table reports the empirical p-values (in percentages) for T = 240, 360, 480 and 600, N = 25, 48 and 100 and level of significance (size) equal to 1 percent, 5 percent and 10 percent. The factors and the time-series regression disturbances are generated jointly from a t distribution with 6 degrees of freedom.

	Test	N=25			N=48			N=100		
Size		1	5	10	1	5	10	1	5	10
T=240	CSR-SM	1.2	5.9	11.3	1.3	7.0	13.1	3.9	11.7	18.8
	CSR-OP	3.9	11.8	19.3	13.1	27.3	37.5	44.9	62.9	72.1
	GMM-HJ	1.5	6.4	12.1	1.5	7.8	14.3	3.9	12.2	20.2
	GMM-H	2.7	10.1	18.5	5.6	19.5	31.7	8.6	32.4	51.2
T=360	CSR-SM	1.1	5.3	11.0	1.1	5.5	11.7	2.2	8.9	16.5
	CSR-OP	3.0	10.1	17.3	7.8	20.2	29.6	30.3	50.2	61.1
	GMM-HJ	1.3	5.9	11.8	1.3	6.4	12.9	2.7	10.6	18.6
	GMM-H	2.2	9.4	16.9	4.4	16.1	26.8	10.9	33.1	49.1
T=480	CSR-SM	0.9	5.5	10.6	0.7	4.7	10.6	1.5	7.6	14.6
	CSR-OP	2.5	8.8	15.4	5.6	16.4	25.6	22.0	40.9	52.4
	GMM-HJ	1.1	5.9	11.3	0.9	5.5	11.8	2.0	9.3	16.8
	GMM-H	2.1	8.5	15.5	3.6	14.0	23.9	10.1	30.3	45.0
T=600	CSR-SM	1.0	5.5	10.8	1.1	4.7	10.2	1.4	6.3	13.1
	CSR-OP	2.2	8.5	15.2	4.3	13.8	22.9	17.1	35.0	47.1
	GMM-HJ	1.1	5.9	11.5	1.1	5.3	11.5	1.8	7.7	15.8
	GMM-H	1.9	8.3	15.5	3.2	12.6	21.7	8.8	27.6	41.4