

Federal Reserve Bank of Minneapolis
Research Department Staff Report 206

January 1996

Econometric Evaluation of Asset Pricing Models

Wayne E. Ferson*

University of Washington

Ravi Jagannathan*

Federal Reserve Bank of Minneapolis,
University of Minnesota, and Hong Kong
University of Science and Technology

ABSTRACT

We provide a brief review of the techniques that are based on the Generalized Method of Moments (GMM) and used for evaluating capital asset pricing models. We first develop the CAPM and multi-beta models and discuss the classical two-stage regression method originally used to evaluate them. We then describe the pricing kernel representation of a generic asset pricing model; this representation facilitates use of the GMM in a natural way for evaluating the conditional and unconditional versions of most asset pricing models. We also discuss diagnostic methods that provide additional insights.

*Ferson acknowledges financial support from the Pigott-PACCAR Professorship at the University of Washington. Jagannathan acknowledges financial support from the National Science Foundation, grant SBR-9409824. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

STAFF REPORT 206 – Errata

Replace equation (2.18) on page 12 with:

$$W = \sum_{l,k=1,\dots,K_2} (x'x)^{-1} x' (\gamma_{2k} \gamma_{2l} \{ \Omega_{ll}^{-1} \Pi_{kl} \Omega_{ll}^{-1} \}) x (x'x)^{-1}$$

where, Π_{kl} is a matrix whose i,j 'th element is the limiting value of $Cov(\sqrt{T} f_k' \varepsilon_{ik}, \sqrt{T} f_l' \varepsilon_{jl})$ as $T \rightarrow \infty$.

Replace

$$E[(b_k - \beta_k)(R_l - \mu_l)|f]$$

in the middle of page 41 by

$$E[(b_k - \beta_k)(R_l - \mu_l)|f_k]$$

Replace the material in Page 42 with the following:

Since $\sqrt{T}(b_{ik} - \beta_{ik})$ converges in distribution to the random variable $\Omega_{kk}^{-1} \sqrt{T} f_k' \varepsilon_{ik}$, we have,

$$Var(h\gamma_2) = \sum_{l,k=1,\dots,K_2} \gamma_{2k} \gamma_{2l} \{ \Omega_{ll}^{-1} \Pi_{kl} \Omega_{ll}^{-1} \}$$

and

$$W = (x'x)^{-1} x' Var(h\gamma_2) x (x'x)^{-1}$$

$$= \sum_{l,k=1,\dots,K_2} (x'x)^{-1} x' (\gamma_{2k} \gamma_{2l} \{ \Omega_{ll}^{-1} \Pi_{kl} \Omega_{ll}^{-1} \}) x (x'x)^{-1}.$$

1. Introduction

A major part of the research effort in finance is directed toward understanding why we observe a variety of financial assets with different expected rates of return. For example, the U.S. stock market as a whole earned an average annual return of 11.94% during the period from January of 1926 to the end of 1991. U.S. Treasury bills, in contrast, earned only 3.64%. The inflation rate during the same period was 3.11% (see Ibbotson Associates 1992).

To appreciate the magnitude of these differences, note that in 1926 a nice dinner for two in New York would have cost about \$10. If the same \$10 had been invested in Treasury bills, by the end of 1991 it would have grown to \$110, still enough for a nice dinner for two. Yet ten dollars invested in stocks would have grown to \$6,756. The point is that the average return differentials among financial assets are both substantial and economically important.

A variety of asset pricing models have been proposed to explain this phenomenon. Asset pricing models describe how the price of a claim to a future payoff is determined in securities markets. Alternatively, we may view asset pricing models as describing the expected rates of return on financial assets, such as stocks, bonds, futures, options, and other securities. Differences among the various asset pricing models arise from differences in their assumptions that restrict investors' preferences, endowments, production, and information sets; the stochastic process governing the arrival of news in the financial markets; and the type of frictions allowed in the markets for real and financial assets.

While there are differences among asset pricing models, there are also important commonalities. All asset pricing models are based on one or more of three central concepts. The first is the law of one price, according to which the prices of any two claims which promise the same future payoff must be the same. The law of one price arises as an implication of the second concept, the no-arbitrage principle. The no-arbitrage principle states that market forces tend to align the prices of

financial assets to eliminate arbitrage opportunities. Arbitrage opportunities arise when assets can be combined, by buying and selling, to form portfolios that have zero net cost, no chance of producing a loss, and a positive probability of gain. Arbitrage opportunities tend to be eliminated by trading in financial markets, because prices adjust as investors attempt to exploit them. For example, if there is an arbitrage opportunity because the price of security A is too low, then traders' efforts to purchase security A will tend to drive up its price. The law of one price follows from the no-arbitrage principle, when it is possible to buy or sell two claims to the same future payoff. If the two claims do not have the same price, and if transaction costs are smaller than the difference between their prices, then an arbitrage opportunity is created. The arbitrage pricing model (APT, Ross 1976) is the most well-known asset pricing model based on arbitrage principles.

The third central concept behind asset pricing models is financial market equilibrium. Investors' desired holdings of financial assets are derived from an optimization problem. A necessary condition for financial market equilibrium in a market with no frictions is that the first-order conditions of the investors' optimization problem be satisfied. This requires that investors be indifferent at the margin to small changes in their asset holdings. Equilibrium asset pricing models follow from the first-order conditions for the investors' portfolio choice problem and from a market-clearing condition. The market-clearing condition states that the aggregate of investors' desired asset holdings must equal the aggregate "market portfolio" of securities in supply.

The earliest of the equilibrium asset pricing models is the Sharpe-Lintner-Black-Mossin Capital Asset Pricing Model (CAPM), developed in the early 1960s. The CAPM states that expected asset returns are given by a linear function of the assets' *betas*, which are their regression coefficients against the market portfolio. Merton (1973) extended the CAPM, which is a single-period model, to an economic environment where investors make consumption, savings, and investment decisions repetitively over time. Econometrically, Merton's model generalizes the CAPM from a model with

a single beta to one with multiple betas. A *multiple-beta model* states that assets' expected returns are linear functions of a number of betas. The APT model of Ross (1976) is another example of a multiple-beta asset pricing model, although in the APT the expected returns are only approximately a linear function of the relevant betas.

In this paper we emphasize (but not exclusively) the econometric evaluation of asset pricing models using the Generalized Method of Moments (GMM, Hansen 1982). We focus on the GMM because, in our opinion, it is the most important innovation in empirical methods in finance within the past fifteen years. The approach is simple, flexible, valid under general statistical assumptions, and often powerful in financial applications. One reason why the GMM is "general" is that many empirical methods used in finance and other areas can be viewed as special cases of GMM.

The rest of this paper is organized as follows. In Section 2 we develop the CAPM and multiple-beta models and discuss the classical two-stage regression procedure that was originally used to evaluate these models. This material provides an introduction to the various statistical issues involved in the empirical study of the models; it also motivates the need for multivariate estimation methods. In Section 3 we describe an alternative representation of the asset pricing models which facilitates the use of the GMM. We show that most asset pricing models can be represented in this *stochastic discount factor* form. In Section 4 we describe the GMM procedure and illustrate how to use it to estimate and test conditional and unconditional versions of asset pricing models. In Section 5 we discuss model diagnostics that provide additional insight into the causes for statistical rejections and that help assess specification errors in the models. In order to avoid a proliferation of symbols, we sometimes use the same symbols to mean different things in different subsections. The definitions should be clear from the context. We conclude with a summary in Section 6.

2. Cross-Sectional Regression Methods for Testing Beta Pricing Models

In this section we first derive the CAPM and generalize its empirical specification to include multiple-beta models. We then describe the intuitively appealing cross-sectional regression method that was first employed by Black, Jensen, and Scholes (1972, abbreviated here as BJS) and discuss its shortcomings.

2.1 The Capital Asset Pricing Model

The CAPM was the first equilibrium asset pricing model, and it remains one of the foundations of financial economics. The model was developed by Sharpe (1964), Lintner (1965), Mossin (1966), and Black (1972). There are a huge number of theoretical papers which refine the necessary assumptions and provide derivations of the CAPM. Here we provide a brief review of the theory.

Let R_{it} denote one plus the return on asset i during period t , $i = 1, 2, \dots, N$. Let R_{mt} denote the corresponding gross return for the market portfolio of all assets in the economy. The return on the market portfolio envisioned by the theory is not observable. In view of this, empirical studies of the CAPM commonly assume that the market return is an exact linear function of the return on an observable portfolio of common stocks.¹ Then, according to the CAPM,

$$(2.1) \quad E(R_{it}) = \delta_0 + \delta_1 \beta_i$$

where

$$\beta_i = \text{Cov}(R_{it}, R_{mt}) / \text{Var}(R_{mt}).$$

According to the CAPM, the market portfolio with return R_{mt} is on the *minimum-variance frontier* of returns. A return is said to be on the minimum-variance frontier if there is no other portfolio with the

¹ When this assumption fails, it introduces market proxy error. This source of error is studied by Roll (1977), Stambaugh (1982), Kandel (1984), Kandel and Stambaugh (1987), Shanken (1987), Hansen and Jagannathan (1994), and Jagannathan and Wang (1995), among others. We will ignore proxy error in our discussion.

same expected return but lower variance. If investors are risk averse, the CAPM implies that R_{mt} is on the positively sloped portion of the minimum-variance frontier, which implies that the coefficient $\delta_1 > 0$. In equation (2.1), $\delta_0 = E(R_{0t})$, where the return R_{0t} is referred to as a *zero-beta asset* to R_{mt} because of the condition $\text{Cov}(R_{0t}, R_{mt}) = 0$.

To derive the CAPM, assume that investors choose asset holdings at each date t so as to maximize the following one-period objective function:

$$(2.2) \quad V[E(R_{pt}|I), \text{Var}(R_{pt}|I)]$$

where R_{pt} denotes the date- t return on the optimally chosen portfolio and $E(\cdot|I)$ and $\text{var}(\cdot|I)$ denote the expectation and variance of return, conditional on the information set I of the investor as of time $t - 1$. We assume that the function $V[\cdot, \cdot]$ is increasing and concave in its first argument, decreasing in its second argument, and time-invariant. For the moment we assume that the information set I includes only the unconditional moments of asset returns, and we drop the symbol I to simplify the notation. The first-order conditions for the optimization problem given above can be manipulated to show that the following must hold:

$$(2.3) \quad E(R_{it}) = E(R_{0t}) + \beta_{ip}E(R_{pt} - R_{0t})$$

for every asset $i = 1, 2, \dots, N$, where R_{pt} is the return on the optimally chosen portfolio, R_{0t} is the return on the asset that has zero covariance with R_{pt} , and $\beta_{ip} = \text{Cov}(R_{it}, R_{pt}) / \text{Var}(R_{pt})$.

To get from the first-order condition for an investor's optimization problem, as stated in equation (2.3), to the CAPM, it is useful to understand some of the properties of the minimum-variance frontier, that is, the set of portfolio returns with the minimum variance, given their expected returns. It can be readily verified that the optimally chosen portfolio of the investor is on the minimum-variance frontier.

One property of the minimum-variance frontier is that it is closed to portfolio formation. That is, portfolios of frontier portfolios are also on the frontier. Suppose that all investors have the same beliefs. Then every investor's optimally chosen portfolio will be on the same frontier, and hence the market portfolio of all assets in the economy—which is a portfolio of every investor's optimally chosen portfolio—will also be on the frontier. It can be shown (Roll 1977) that equation (2.3) will hold if R_{pt} is replaced by the return of any portfolio on the frontier and R_{0t} is replaced by its corresponding zero-beta return. Hence we can replace an investor's optimal portfolio in equation (2.3) with the return on the market portfolio to get the CAPM, as given by equation (2.1).

2.2 Testable Implications of the CAPM

Given an interesting collection of assets, and if their expected returns and market-portfolio betas β_i are known, a natural way to examine the CAPM would be to estimate the empirical relation between the expected returns and the betas and see if that relation is linear. However, neither betas nor expected returns are observed by the econometrician. Both must be estimated. The finance literature first attacked this problem by using a two-step, time-series, cross-sectional approach.

Consider the following sample analogue of the population relation given in (2.1):

$$(2.4) \quad R_i = \delta_0 + \delta_1 b_i + e_i, \quad i = 1, \dots, N$$

which is a cross-sectional regression of R_i on b_i , with regression coefficients equal to δ_0 and δ_1 . In equation (2.4), R_i denotes the sample average return of the asset, i , and b_i is the (OLS) slope coefficient estimate from a regression of the return, R_{it} , over time on the market index return, R_{mt} ; b_i is a constant. Let $u_i = R_i - E(R_{it})$ and $v_i = \beta_i - b_i$. Substituting these relations for $E(R_{it})$ and β_i in (2.1) leads to (2.4) and specifies the composite error as $e_i = u_i + \delta_1 v_i$. This gives rise to a classic errors-in-variables problem, as the regressor b_i in the cross-sectional regression model (2.4) is measured with error. Using finite time-series samples for the estimate of b_i , the regression (2.4) will deliver

inconsistent estimates of δ_0 and δ_1 , even with an infinite cross-sectional sample. However, the cross-sectional regression will provide consistent estimates of the coefficients as the time-series sample size T (which is used in the first step to estimate the beta coefficient β_i) becomes very large. This is because the first-step estimate of β_i is consistent, so as T becomes large, the errors-in-variables problem of the second-stage regression vanishes.

The measurement error in beta may be large for individual securities, but it is smaller for portfolios. In view of this fact, early research focused on creating portfolios of securities in such a way that the betas of the portfolios could be estimated precisely. Hence one solution to the errors-in-variables problem is to work with portfolios instead of individual securities. This creates another problem. Arbitrarily chosen portfolios tend to exhibit little dispersion in their betas. If all the portfolios available to the econometrician have the same betas, then equation (2.1) has no empirical content as a cross-sectional relation. Black, Jensen, and Scholes (BJS) came up with an innovative solution to overcome this difficulty. At every point in time for which a cross-sectional regression is run, they estimate betas on individual securities based on past history, sort the securities based on the estimated values of beta, and assign individual securities to beta groups. This results in portfolios with a substantial dispersion in their betas. Similar portfolio formation techniques have become standard practice in the empirical finance literature.

Suppose that we can create portfolios in such a way that we can view the errors-in-variables problem as being of second-order importance. We still have to determine how to assess whether there is empirical support for the CAPM. A standard approach in the literature is to consider specific alternative hypotheses about the variables which determine expected asset returns. According to the CAPM, the expected return for any asset is a linear function of its beta only. Therefore, one natural test would be to examine if any other cross-sectional variable has the ability to explain the deviations from equation (2.1). This is the strategy that Fama and MacBeth (1973) followed by incorporating

the square of beta and measures of nonmarket (or residual time-series) variance as additional variables in the cross-sectional regressions. More recent empirical studies have used the relative size of firms, measured by the market value of their equity, the ratio of book-to-market-equity, and related variables.² For example, the following model may be specified:

$$(2.5) \quad E(R_{it}) = \delta_0 + \delta_1 \beta_i + \delta_{\text{size}} LME_i$$

where LME_i is the natural logarithm of the total market value of the equity capital of firm i . In what follows we will first show that these ideas extend easily to the general multiple-beta model. We will then develop a sampling theory for the cross-sectional regression estimators.

2.3 Multiple-Beta Pricing Models and Cross-Sectional Regression Methods

According to the CAPM, the expected return on an asset is a linear function of its market beta. A multiple-beta model asserts that the expected return is a linear function of several betas, i.e.,

$$(2.6) \quad E(R_{it}) = \delta_0 + \sum_{k=1, \dots, K} \delta_k \beta_{ik}$$

where β_{ik} , $k = 1, \dots, K$, are the multiple regression coefficients of the return of asset i on K economy-wide pervasive risk factors, f_k , $k = 1, \dots, K$. The coefficient δ_0 is the expected return on an asset that has $\beta_{0k} = 0$, for $k = 1, \dots, K$; i.e., it is the expected return on a zero- (multiple-) beta asset. The coefficient δ_k , corresponding to the k th factor, has the following interpretation: it is the expected return differential, or premium, for a portfolio that has $\beta_{ik} = 1$ and $\beta_{ij} = 0$ for all $j \neq k$, measured in excess of the zero-beta asset's expected return. In other words, it is the expected return premium per unit of beta risk for the risk factor, k . Ross (1976) showed that an approximate version of (2.6) will hold in an arbitrage-free economy. Connor (1984) provided sufficient conditions for (2.6) to hold exactly in an economy with an infinite number of assets, in general equilibrium. This version of the

² Fama and French (1992) is a prominent recent example of this approach. Berk (1994) provides a justification for using relative market value and book-to-price ratios as measures of expected returns.

multiple-beta model, the exact APT, has received wide attention in the finance literature. When the factors, f_k , are observed by the econometrician, the cross-sectional regression method can be used to empirically evaluate the multiple-beta model.³ For example, the alternative hypothesis that the size of the firm is related to expected returns, given the factor betas, may be examined by using cross-sectional regressions of returns on the K factor betas and the LME_i , similar to equation (2.5), and by examining whether the coefficient δ_{size} is different from zero.

2.4 Sampling Distributions for Coefficient Estimators: The Two-Stage, Cross-Sectional Regression Method

In this section we follow Shanken (1992) and Jagannathan and Wang (1993, 1995) in deriving the asymptotic distribution of the coefficients that are estimated using the cross-sectional regression method. For the purposes of developing the sampling theory, we will work with the following generalization of equation (2.6):

$$(2.7) \quad E(R_{it})^{K_1} = \sum_{k=0}^{K_2} \gamma_{1k} A_{ik}^{K_2} + \sum_{k=1}^{K_2} \gamma_{2k} \beta_{ik}$$

where $\{A_{ik}\}$ are observable characteristics of firm i , which are assumed to be measured without error (the first “characteristic,” when $k = 0$, is the constant 1.0). One of the attributes may be the size variable LME_i . The β_i are regression betas on a set of K_2 economic risk factors, which may include the market index return. Equation (2.7) can be written more compactly using matrix notation as

$$(2.8) \quad \mu = X\gamma$$

where $R_i = [R_{1i}, \dots, R_{N_i}]$, $\mu = E(R_i)$, $X = [A:\beta]$, and the definition of the matrices A and β and the vector γ follow from (2.7).

³ See Chen (1983), Connor and Korajczyk (1986), Lehmann and Modest (1987), and McElroy and Burmeister (1988) for discussions on estimating and testing the model when the factor realizations are not observable, under some additional auxiliary assumptions.

The cross-sectional method proceeds in two stages. First, β is estimated by time-series regressions of R_{it} on the risk factors and a constant. The estimates are denoted by b . Let $x = [A:b]$, and let R denote the time-series average of the return vector R_t . Let g denote the estimator of the coefficient vector obtained from the following cross-sectional regression:

$$(2.9) \quad g = (x'x)^{-1}x'R$$

where we assume that x is of rank $1 + K_1 + K_2$. If b and R converge respectively to β and $E(R_t)$ in probability, then g will converge in probability to γ .

Black, Jensen, and Scholes (BJS) suggest estimating the sampling errors associated with the estimator, g , as follows. Regress R_t on x at each date t to obtain g_t , where

$$(2.10) \quad g_t = (x'x)^{-1}x'R_t$$

The BJS estimate of the covariance matrix of $T^{1/2}(g - \gamma)$ is given by

$$(2.11) \quad v = T^{-1} \sum_t (g_t - g)(g_t - g)'$$

which uses the fact that g is the sample mean of the g_t s. Substituting the expression for g_t given in (2.10) into the expression for v given in (2.11) gives

$$(2.12) \quad v = (x'x)^{-1}x'[T^{-1}\sum_t(R_t - R)(R_t - R)']x(x'x)^{-1}.$$

To analyze the BJS covariance matrix estimator, we write the average return vector, R , as:

$$(2.13) \quad R = x\gamma + (R - \mu) - (x - X)\gamma.$$

Substitute this expression for R into the expression for g in (2.9) to obtain

$$(2.14) \quad g - \gamma = (x'x)^{-1}x'[(R - \mu) - (b - \beta)\gamma_2].$$

Assume that b is a consistent estimate of β and that $T^{1/2}(R - \mu) \rightarrow_d u$ and $T^{1/2}(b - \beta) \rightarrow_d h$, where u and h are random variables with well-defined distributions and \rightarrow_d indicates convergence in distribution. We then have

$$(2.15) \quad T^{1/2}(g - \gamma) \rightarrow_d (x'x)^{-1}x'u - (x'x)^{-1}x'h\gamma_2.$$

In (2.15) the first term on the right side is that component of the sampling error that arises from replacing μ by the sample average R . The second term is the component of the sampling error that arises due to replacing β by its estimate b .

The usual consistent estimate of the asymptotic variance of u is given by

$$(2.16) \quad T^{-1} \sum_t (R_t - R)(R_t - R)'$$

Therefore, a consistent estimate of variance of the first term in (2.15) is given by

$$(x'x)^{-1}x'[T^{-1} \sum_t (R_t - R)(R_t - R)']x(x'x)^{-1}$$

which is the same as the expression for the BJS estimate for the covariance matrix of the estimated coefficients v , given in (2.12). Hence if we ignore the sampling error that arises from using estimated betas, then the BJS covariance estimator provides a consistent estimate of the variance of the estimator g . However, if the sampling error associated with the betas is not small, then the BJS covariance estimator will have a bias. While it is not possible to determine the magnitude of the bias in general, Shanken (1992) provides a method to assess the bias under additional assumptions.⁴

Consider the following univariate time-series regression for the return of asset i on a constant and the k th economic factor:

⁴ Shanken (1992) uses betas computed from multiple regressions. The derivation which follows uses betas computed from univariate regressions, for simplicity of exposition. The two sets of betas are related by an invertible linear transformation. Alternatively, the factors may be orthogonalized without loss of generality.

$$(2.17) \quad R_{it} = \alpha_{ik} + \beta_{ik} f_{kt} + \varepsilon_{ikt}$$

We make the following additional assumptions about the error terms in (2.17): (1) the errors are mean zero, conditional on the time series of the economic factors f_k ; (2) the conditional covariance of ε_{ikt} and ε_{jlt} , given the factors, is a fixed constant σ_{ijkl} . We denote the matrix of the $\{\sigma_{ijkl}\}_{ij}$ by Σ_{kl} . Finally, we assume that (3) the sample covariance matrix of the factors exists and converges in probability to a constant positive definite matrix Ω , with the typical element Ω_{kl} .

Theorem 2.1. (Shanken, 1992/Jagannathan-Wang, 1995)

$T^{1/2}(g - \gamma)$ converges in distribution to a normally distributed random variable with zero mean and covariance matrix $V + W$, where V is the probability limit of the matrix v given in (2.12) and

$$(2.18) \quad W = \sum_{l,k=1, \dots, K} \gamma_{2k} \gamma_{2l}' (\Omega_{kk}^{-1} \Omega_{kl} \Omega_{ll}^{-1}) (x'x)^{-1} x' \Sigma_{kl} x (x'x)^{-1}.$$

Proof. See the appendix.

Theorem 2.1 shows that in order to obtain a consistent estimate of the covariance matrix of the BJS two-step estimator g , we first estimate v (a consistent estimate of V) by using the BJS method. We then estimate W by its sample analogue.

Although the cross-sectional regression method is intuitively very appealing, the above discussion shows that in order to assess the sampling errors associated with the parameter estimators, we need to make rather strong assumptions. In addition, the econometrician must take a stand on a particular alternative hypothesis against which to reject the model. The general approach developed in Section 4 below has, among its advantages, weaker statistical assumptions and the ability to handle both unspecified as well as specific alternative hypotheses.

3. Asset Pricing Models and Stochastic Discount Factors

Virtually all financial asset pricing models imply that any gross asset return $R_{i,t+1}$, multiplied by some market-wide random variable m_{t+1} , has a constant conditional expectation:

$$(3.1) \quad E_t\{m_{t+1}R_{i,t+1}\} = 1, \text{ all } i.$$

The notation $E_t\{\cdot\}$ will be used to denote the conditional expectation, given a market-wide information set. Sometimes it will be convenient to refer to expectations conditional on a subset Z_t of the market information, which are denoted as $E(\cdot|Z_t)$. For example, Z_t can represent a vector of instrumental variables for the public information set, which are unavailable to the econometrician. When Z_t is the null information set, the unconditional expectation is denoted as $E(\cdot)$. If we take the expected values of equation (3.1), it follows that versions of the same equation must hold for the expectations $E(\cdot|Z_t)$ and $E(\cdot)$.

The random variable m_{t+1} has various names in the literature. It is known as a stochastic discount factor, an equivalent martingale measure, a Radon-Nicodým derivative, or an intertemporal marginal rate of substitution. We will refer to an m_{t+1} which satisfies (3.1) as a valid *stochastic discount factor*. The motivation for use of this term arises from the following observation. Write equation (3.1) as

$$P_{it} = E_t\{m_{t+1}X_{i,t+1}\}$$

where $X_{i,t+1}$ is the payoff of asset i at time $t + 1$ (the market value plus any cash payments) and $R_{i,t+1} = X_{i,t+1} / P_{it}$. Equation (3.1) says that if we multiply a future payoff $X_{i,t+1}$ by the stochastic discount factor m_{t+1} and take the expected value, we obtain the present value of the future payoff.

The existence of an m_{t+1} that satisfies (3.1) says that all assets with the same payoffs have the same price (i.e., *the law of one price*). With the restriction that m_{t+1} is a strictly positive random variable, equation (3.1) becomes equivalent to a *no-arbitrage* condition. The condition is that all

portfolios of assets with payoffs that can never be negative, but are positive with positive probability, must have positive prices.

The no-arbitrage condition does not uniquely identify m_{t+1} unless markets are complete, which means that there are as many linearly independent payoffs available in the securities markets as there are states of nature at date $t + 1$. To obtain additional insights about the stochastic discount factor and the no-arbitrage condition, assume for the moment that the markets are complete. Given complete markets, positive state prices are required to rule out arbitrage opportunities.⁵ Let q_{ts} denote the time- t price of a security that pays one unit at date $t + 1$ if and only if the state of nature at $t + 1$ is s . Then the time- t price of a security that promises to pay $\{X_{i,s,t+1}\}$ units at date $t + 1$, as a function of the state of nature s , is given by

$$\sum_s q_{ts} X_{i,s,t+1} = \sum_s \pi_{ts} (q_{ts} / \pi_{ts}) X_{i,s,t+1}$$

where π_{ts} is the probability, as assessed at time t , that state s occurs at time $t + 1$. Comparing this expression with equation (3.1) shows that $m_{s,t+1} = q_{ts} / \pi_{ts}$ is the value of the stochastic discount factor in state s , under the assumption that the markets are complete. Since the probabilities are positive, the condition that the random variable defined by $\{m_{s,t+1}\}$ is strictly positive is equivalent to the condition that all state prices are positive.

Equation (3.1) is convenient for developing econometric tests of asset pricing models. Let R_{t+1} denote the vector of gross returns on the N assets on which the econometrician has observations. Then (3.1) can be written as

$$(3.2) \quad E\{R_{t+1} m_{t+1} - \underline{1}\} = \underline{0}$$

⁵ See Debreu (1959) and Arrow (1970) for models of complete markets. See Beja (1971), Rubinstein (1976), Ross (1977), Harrison and Kreps (1979), and Hansen and Richard (1987) for further theoretical discussions.

where $\underline{1}$ denotes the N vector of ones and $\underline{0}$ denotes the N vector of zeros. The set of N equations given in (3.2) will form the basis for tests using the generalized method of moments. It is the specific form of m_{t+1} implied by a model that gives the equation empirical content.

3.1 Stochastic Discount Factor Representations of the CAPM and Multiple-Beta Asset Pricing Models

Consider the CAPM, as given by equation (2.1)

$$E(R_{it+1}) = \delta_0 + \delta_1 \beta_i$$

where

$$\beta_i = \text{Cov}(R_{it+1}, R_{mt+1}) / \text{Var}(R_{mt+1}).$$

The CAPM can also be expressed in the form of equation (3.1), with a particular specification of the stochastic discount factor. To see this, expand the expected product in (3.1) into the product of the expectations plus the covariance, and then rearrange to obtain

$$(3.3) \quad E(R_{it+1}) = 1 / E(m_{t+1}) + \text{Cov}(R_{it+1}; -m_{t+1} / E(m_{t+1})).$$

Equating terms in equations (2.1) and (3.3) shows that the CAPM of equation (2.1) implies a version of equation (3.1), where

$$E(R_{it+1} m_{t+1}) = 1$$

where

$$m_{t+1} = c_0 - c_1 R_{mt+1}$$

$$(3.4) \quad c_0 = [1 + E(R_{mt+1}) \delta_1 / \text{Var}(R_{mt+1})] / \delta_0$$

and

$$c_1 = \delta_1 / [\delta_0 \text{Var}(R_{mt+1})].$$

Equation (3.4) was originally derived by Dybvig and Ingersoll (1982).

Now consider the following multiple-beta model, which was given in equation (2.6):

$$E(R_{it+1}) = \delta_0 + \sum_{k=1, \dots, K} \delta_k \beta_{ik}$$

It can be readily verified by substitution that this model implies the following stochastic discount factor representation.

$$E(R_{it+1} m_{it+1}) = 1$$

where

$$m_{it+1} = c_0 + c_1 f_{1t+1} + \dots + c_K f_{Kt+1}$$

with

$$(3.5) \quad c_0 = [1 + \sum_k \{\delta_k E(f_k) / \text{Var}(f_k)\}] / \delta_0$$

and

$$c_j = -\{\delta_j / \delta_0 \text{Var}(f_j)\}, j = 1, \dots, K.$$

The preceding results apply to the CAPM and multiple-beta models, interpreted as statements about the unconditional expected returns of the assets. These models are also interpreted as statements about conditional expected returns in some tests, where the expectations are conditioned on predetermined, publicly available information. All of the analysis of this section can be interpreted as applying to conditional expectations, with the appropriate changes in notation. In this case, the parameters c_0 , c_1 , δ_0 , δ_1 , etc. will be functions of the time- t information set.

3.2 Other Examples of Stochastic Discount Factors

In equilibrium asset pricing models, equation (3.1) arises as a first-order condition for a consumer-investor's optimization problem. The agent maximizes a lifetime utility function of consumption (including possibly a bequest to heirs). Denote this function by $V(\cdot)$. If the allocation of resources to consumption and to investment assets is optimal, it is not possible to obtain higher utility by changing the allocation. Suppose that an investor considers reducing consumption at time t to purchase more of (any) asset. The utility cost at time t of the forgone consumption is the marginal

utility of consumption expenditures C_t , denoted by $(\partial V / \partial C_t) > 0$, multiplied by the price $P_{i,t}$ of the asset, measured in the same units as the consumption expenditures. The expected utility gain of selling the share and consuming the proceeds at time $t + 1$ is

$$E_t\{(P_{i,t+1} + D_{i,t+1})(\partial V / \partial C_{t+1})\}$$

where $D_{i,t+1}$ is the cash flow or dividend paid at time t . If the allocation maximizes expected utility, the following must hold:

$$P_{i,t}E_t\{(\partial V / \partial C_t)\} = E_t\{(P_{i,t+1} + D_{i,t+1})(\partial V / \partial C_{t+1})\}.$$

This intertemporal Euler equation is equivalent to equation (3.1), with

$$(3.6) \quad m_{t+1} = (\partial V / \partial C_{t+1}) / E_t\{(\partial V / \partial C_t)\}.$$

The m_{t+1} in equation (3.6) is the *intertemporal marginal rate of substitution* (IMRS) of the representative consumer. The rest of this section shows how many models in the asset pricing literature are special cases of (3.1), where m_{t+1} is defined by equation (3.6).⁶

If a representative consumer's lifetime utility function $V(\cdot)$ is time-separable, the marginal utility of consumption at time t , $(\partial V / \partial C_t)$, depends only on variables dated at time t . Lucas (1978) and Breeden (1979) derived consumption-based asset pricing models of the following type, assuming that the preferences are time-separable and additive:

$$V = \sum_t \beta^t u(C_t)$$

⁶ Asset pricing models typically focus on the relation of security returns to aggregate quantities. It is therefore necessary to aggregate the Euler equations of individuals to obtain equilibrium expressions in terms of aggregate quantities. Theoretical conditions which justify the use of aggregate quantities are discussed by Gorman (1953), Wilson (1968), Rubinstein (1974), Constantinides (1982), Lewbel (1989), Luttmer (1993), and Constantinides and Duffie (1994).

where β is a time discount parameter and $u(\cdot)$ is increasing and concave in current consumption C_t .

A convenient specification for $u(\cdot)$ is

$$(3.7) \quad u(C) = C^{1-\alpha} - 1 / 1 - \alpha.$$

In equation (3.7), $\alpha > 0$ is the concavity parameter of the period utility function. This function displays constant relative risk aversion equal to α .⁷ Based on these assumptions and using aggregate consumption data, a number of empirical studies test the consumption-based asset pricing model.⁸

Dunn and Singleton (1986) and Eichenbaum, Hansen, and Singleton (1988), among others, model consumption expenditures that may be durable in nature. Durability introduces *nonseparability* over time, since the flow of consumption services depends on the consumer's previous expenditures, and the utility is defined over the services. Current expenditures increase the consumer's future utility, if the expenditures are durable. The consumer optimizes over the expenditures C_t ; this accounts for the fact that durability implies that the marginal utility, $(\partial V / \partial C_t)$, depends on variables dated other than date t .

Another form of time-nonseparability arises if the utility function exhibits *habit persistence*. Habit persistence means that consumption at two points in time are complements. For example, the utility of current consumption is evaluated relative to what was consumed in the past. Such models are derived by Ryder and Heal (1973), Becker and Murphy (1988), Sundaresan (1989), Constantinides (1990), Detemple and Zapatero (1991), and Novales (1992), among others.

⁷ *Relative risk aversion* in consumption is defined as $-Cu''(C) / u'(C)$. Absolute risk aversion is $-u''(C) / u'(C)$, where a prime (') denotes a derivative. Ferson (1983) studies a consumption-based asset pricing model with constant absolute risk aversion.

⁸ Substituting (3.7) into (3.6) shows that $m_{t+1} = \beta(C_{t+1} / C_t)^{-\alpha}$. Empirical studies of this model include Hansen and Singleton (1982, 1983), Ferson (1983), Brown and Gibbons (1985), Jagannathan (1985), Ferson and Merrick (1987), and Wheatley (1988).

Ferson and Constantinides (1991) model both the durability of consumption expenditures and habit persistence in consumption services. They show that the two combine as opposing effects. In an example where the effect is truncated at a single lag, the derived utility of expenditures is

$$(3.8) \quad V = (1 - \alpha)^{-1} \sum_t \beta^t (C_t + bC_{t-1})^{1-\alpha}.$$

The marginal utility at time t is

$$(3.9) \quad (\partial V / \partial C_t) = \beta^t (C_t + bC_{t-1})^{-\alpha} + \beta^{t+1} b E_t \{ (C_{t+1} + bC_t)^{-\alpha} \}.$$

The coefficient b is positive and measures the rate of depreciation if the good is durable and there is no habit persistence. If habit persistence is present and the good is nondurable, this implies that the lagged expenditures enter with a negative effect ($b < 0$).

Heaton (1995) and Ferson and Harvey (1992) consider a form of time-nonseparability which emphasizes seasonality. The utility function is

$$(1 - \alpha)^{-1} \sum_t \beta^t (C_t + bC_{t-4})^{1-\alpha}$$

where the consumption expenditure decisions are assumed to be quarterly. The subsistence level (in the case of habit persistence) or the flow of services (in the case of durability) is assumed to depend only on the consumption expenditure in the same quarter of the previous year.

Abel (1990) studies a form of habit persistence in which the consumer evaluates current consumption relative to the aggregate consumption in the previous period, consumption that he or she takes as exogenous. The utility function is like equation (3.8), except that the “habit stock,” bC_{t-1} , refers to the aggregate consumption. The idea is that people care about “keeping up with the Joneses.” Campbell and Cochrane (1995) also develop a model in which the habit stock is taken as exogenous by the consumer. This approach results in a simpler and more tractable model, since the consumer’s

optimization does not have to take account of the effects of current decisions on the future habit stock.

Epstein and Zin (1989, 1991) consider a class of recursive preferences which can be written as $V_t = F(C_t, CEQ_t(V_{t+1}))$. $CEQ_t(\cdot)$ is a time- t “certainty equivalent” for the future lifetime utility V_{t+1} . The function $F(\cdot, CEQ_t(\cdot))$ generalizes the usual expected utility function of lifetime consumption and may be non-time-separable.

Epstein and Zin (1989) study a special case of the recursive preference model in which the preferences are

$$(3.10) \quad V_t = [(1-\beta)C_t^p + \beta E_t(V_{t+1}^{1-\alpha})^{p/(1-\alpha)}]^{1/p}.$$

They show that when $p \neq 0$ and $1 - \alpha \neq 0$, the IMRS for a representative agent becomes

$$(3.11) \quad [\beta(C_{t+1} / C_t)^{p-1}]^{(1-\alpha)/p} \{R_{m,t+1}\}^{((1-\alpha-p)/p)}$$

where $R_{m,t+1}$ is the gross market portfolio return. The coefficient of relative risk aversion for timeless consumption gambles is α , and the elasticity of substitution for deterministic consumption is $(1 - p)^{-1}$. If $\alpha = 1 - p$, the model reduces to the time-separable, power utility model. If $\alpha = 1$, the log utility model of Rubinstein (1976) is obtained.

In summary, many asset pricing models are special cases of the equation (3.1). Each model specifies that a particular function of the data and the model parameters is a valid stochastic discount factor. We now turn to the issue of estimating the models stated in this form.

4. The Generalized Method of Moments

In this section we provide an overview of the Generalized Method of Moments and a brief review of the associated asymptotic test statistics. We then show how the GMM is used to estimate and test various specifications of asset pricing models.

4.1 An Overview of the Generalized Method of Moments in Asset Pricing Models

Let x_{t+1} be a vector of observable variables. Given a model which specifies $m_{t+1} = m(\theta, x_{t+1})$, estimation of the parameters θ and tests of the model can then proceed under weak assumptions, using the GMM as developed by Hansen (1982) and illustrated by Hansen and Singleton (1982) and Brown and Gibbons (1985). Define the following model error term:

$$(4.1) \quad u_{i,t+1} = m(\theta, x_{t+1})R_{i,t+1} - 1.$$

The equation (3.1) implies that $E_t\{u_{i,t+1}\} = 0$ for all i . Given a sample of N assets and T time periods, combine the error terms from (4.1) into a $T \times N$ matrix u , with typical row u_{t+1}' . By the law of iterated expectations, the model implies that $E(u_{i,t+1} | Z_t) = 0$ for all i and t (for any Z_t in the information set at time t), and therefore $E(u_{t+1}Z_t) = 0$ for all t . The condition $E(u_{t+1}Z_t) = 0$ says that u_{t+1} is orthogonal to Z_t , and is therefore called an *orthogonality condition*. These orthogonality conditions are the basis of tests of asset pricing models using the GMM.

A few points deserve emphasis. First, GMM estimates and tests of asset pricing models are motivated by the implication that $E(u_{i,t+1} | Z_t) = 0$, for any Z_t in the information set at time t . However, the weaker condition $E(u_{t+1}Z_t) = 0$, for a given set of instruments Z_t , is actually used in the estimation. Therefore, GMM tests of asset pricing models have not exploited all of the predictions of the theories. We believe that further refinements to exploit the implications of the theories more fully will be useful.

Empirical work on asset pricing models relies on *rational expectations*, interpreted as the assumption that the expectation terms in the model are mathematical conditional expectations. For example, the rational expectations assumption is used when the expected value in equation (3.1) is treated as a mathematical conditional expectation to obtain expressions for $E(\cdot | Z)$ and $E(\cdot)$. Rational

expectations implies that the difference between observed realizations and the expectations in the model should be unrelated to the information that the expectations in the model are conditioned on.

Equation (3.1) says that the conditional expectation of the product of m_{t+1} and $R_{i,t+1}$ is the constant 1.0. Therefore, the error term $1 - m_{t+1}R_{i,t+1}$ in equation (4.1) should not be predictably different from zero when we use any information available at time t . If there is variation over time in a return $\tilde{R}_{i,t+1}$ that is predictable using instruments Z_t , the model implies that the predictability is removed when $R_{i,t+1}$ is multiplied by a valid stochastic discount factor, m_{t+1} . This is the sense in which conditional asset pricing models are asked to “explain” predictable variation in asset returns. This idea generalizes the “random walk” model of stock values, which implies that stock returns should be completely unpredictable. That model is a special case which can be motivated by risk neutrality. Under risk neutrality the IMRS is a constant. In this case, equation (3.1) implies that the return $R_{i,t+1}$ should not differ predictably from a constant.

GMM estimation proceeds by defining an $N \times L$ matrix of sample mean orthogonality conditions, $G = (u'Z / T)$, and letting $g = \text{vec}(G)$, where Z is a $T \times L$ matrix of observed instruments with typical row Z_t' , a subset of the available information at time t .⁹ The $\text{vec}(\cdot)$ operator means to partition G into row vectors, each of length L : $(\underline{h}_1, \underline{h}_2, \dots, \underline{h}_N)$. Then one stacks the h s into a vector, g , with length equal to the number of orthogonality conditions, NL . Hansen’s (1982) GMM estimates of θ are obtained by searching for parameter values that make g close to zero by minimizing a quadratic form $g'Wg$, where W is an $NL \times NL$ weighting matrix.

Somewhat more generally, let $u_{t+1}(\theta)$ denote the random N vector $R_{t+1}m(\theta, x_{t+1}) - 1$, and define $g_T(\theta) = T^{-1} \sum_t (u_t(\theta) \otimes Z_{t-1})$. Let θ_T denote the parameter values that minimize the quadratic form

⁹ This section assumes that the same instruments are used for each of the asset equations. In general, each asset equation could use a different set of instruments, which complicates the notation.

$g_T' A_T g_T$, where A_T is any positive definite $NL \times NL$ matrix that may depend on the sample, and let J_T denote the minimized value of the quadratic form $g_T' A_T g_T$. Jagannathan and Wang (1995) show that J_T will have a weighted Chi-Square distribution which can be used for testing the hypothesis that (3.1) holds.

Theorem 4.1 (Jagannathan-Wang, 1993). Suppose that the matrix A_T converges in probability to a constant positive definite matrix A . Assume also that $\sqrt{T}g_T(\theta_0) \rightarrow_d N(0, S)$, where $N(.,.)$ denotes the multivariate normal distribution, θ_0 are the true parameter values, and S is a positive definite matrix. Let

$$D = E[\partial g_T / \partial \theta] | \theta = \theta_0$$

and let

$$Q = (S^{1/2})(A^{1/2})[I - (A^{1/2})'D(D'AD)^{-1}D'(A^{1/2})](A^{1/2})(S^{1/2})$$

where $A^{1/2}$ and $S^{1/2}$ are the upper triangular matrices from the Cholesky decompositions of A and S . Then the matrix Q has $NL\text{-dim}(\theta)$ nonzero, positive eigenvalues. Denote these eigenvalues by λ_i , $i = 1, 2, \dots, NL\text{-dim}(\theta)$. Then J_T converges to

$$\lambda_1 \chi_{1} + \dots + \lambda_{NL\text{-dim}(\theta)} \chi_{NL\text{-dim}(\theta)}$$

where χ_i , $i = 1, 2, \dots, NL\text{-dim}(\theta)$ are $NL\text{-dim}(\theta)$ independent random variables, each with a Chi-Square distribution with one degree of freedom.

Proof. See Jagannathan and Wang (1993).

Notice that when the matrix A is $W \equiv S^{-1}$, the matrix Q is idempotent, of rank $NL\text{-dim}(\theta)$. Hence the nonzero eigenvalues of Q are unity. In this case, the asymptotic distribution reduces to a simple Chi-Square distribution with $NL\text{-dim}(\theta)$ degrees of freedom. This is the special case

considered by Hansen (1982), who originally derived the asymptotic distribution of the J_T -statistic. The J_T -statistic and its extension, as provided in theorem 4.1, provide a goodness-of-fit test for models estimated by the GMM.

Hansen (1982) shows that the estimators of θ that minimize $g'Wg$ are consistent and asymptotically normal, for any fixed W . If the weighting matrix W is chosen to be the inverse of a consistent estimate of the covariance matrix of the orthogonality conditions S , the estimators are asymptotically efficient in the class of estimators that minimize $g'Wg$ for fixed W s. The asymptotic variance matrix of this optimal GMM estimator of the parameter vector is given as

$$(4.2) \quad \text{Cov}(\theta) = [E(\partial g / \partial \theta)'WE(\partial g / \partial \theta)]^{-1}$$

where $\partial g / \partial \theta$ is an $NL \times \dim(\theta)$ matrix of derivatives. A consistent estimator for the asymptotic covariance of the sample mean of the orthogonality conditions is used in practice. That is, we replace W in (4.2) with $\hat{\text{Cov}}(g)^{-1}$ and replace $E(\partial g / \partial \theta)$ with its sample analogue. An example of a consistent estimator for the optimal weighting matrix is given by Hansen (1982) as

$$(4.3) \quad \hat{\text{Cov}}(g) = [(1/T) \sum_t \sum_j (u_{t+1} u_{t+1-j})' \otimes (Z_t Z_{t-j}')]]$$

where \otimes denotes the Kronecker product. A special case that often proves useful arises when the orthogonality conditions are not serially correlated. In that special case, the optimal weighting matrix is the inverse of the matrix $\hat{\text{Cov}}(g)$, where

$$(4.4) \quad \hat{\text{Cov}}(g) = [(1/T) \sum_t (u_{t+1} u_{t+1}') \otimes (Z_t Z_t')].$$

The GMM weighting matrices originally proposed by Hansen (1982) have some drawbacks. The estimators are not guaranteed to be positive definite, and they may have poor finite sample properties in some applications. A number of studies have explored alternative estimators for the GMM weighting matrix. Prominent examples include Newey and West (1987a), who suggest weighting the

autocovariance terms in (4.3) with Bartlett weights to achieve a positive semi-definite matrix. Additional refinements to improve the finite sample properties are proposed by Andrews (1991), Andrews and Monahan (1992), and Ferson and Foerster (1994).

4.2 Testing Hypotheses with the GMM

As we noted above, the J_T -statistic provides a goodness-of-fit test for a model that is estimated by the GMM, when the model is overidentified. Hansen's J_T -statistic is the most commonly used test in the finance literature that has used the GMM. Other standard statistical tests based on the GMM are also used in the finance literature for testing asset pricing models. One is a generalization of the Wald test, and a second is analogous to a likelihood ratio test statistic. Additional test statistics based on the GMM are reviewed by Newey (1985) and Newey and West (1987b).

For the Wald test, consider the hypothesis to be tested as expressed in the M -vector valued function $H(\theta) = 0$, where $M \leq \dim(\theta)$. The GMM estimates of θ are asymptotically normal, with mean θ and variance matrix $\hat{\text{Cov}}(\theta)$. Given standard regularity conditions, it follows that the estimates of \hat{H} are asymptotically normal, with mean zero and variance matrix $\hat{H}_\theta \hat{\text{Cov}}(\theta) \hat{H}_\theta'$, where subscripts denote partial derivatives, and that the quadratic form

$$T\hat{H}'[\hat{H}_\theta \hat{\text{Cov}}(\theta) \hat{H}_\theta']^{-1}\hat{H}$$

is asymptotically Chi-Square, providing a standard Wald test.

A likelihood ratio type test is described by Newey and West (1987b), Eichenbaum, Hansen, and Singleton (1988, appendix C), and Gallant (1987). Newey and West (1987b) call this the D test. Assume that the null hypothesis implies that the orthogonality conditions $E(g^*) = 0$ hold, while, under the alternative, a subset $E(g) = 0$ hold. For example, $g^* = (g, h)$. When we estimate the model under the null hypothesis, the quadratic form $g^{*'}W^*g^*$ is minimized. Let W_{11}^* be the inverse of the upper left block of W^{*-1} ; that is, let it be the estimate of $\text{Cov}(g)$ under the null. When we hold this matrix fixed

the model can be estimated under the alternative by minimizing $g'W_{11}^*g$. The difference of the two quadratic forms

$$T[g^{*\prime}Wg^* - g'W_{11}^*g]$$

is asymptotically Chi-Square, with degrees of freedom equal to M if the null hypothesis is true. Newey and West (1987b) describe additional variations on these tests.

4.3 Illustrations: Using the GMM to Test the Conditional CAPM

The CAPM imposes nonlinear overidentifying restrictions on the first and second moments of asset returns. These restrictions can form a basis for econometric tests. To see these restrictions more clearly, notice that if an econometrician knows or can estimate $\text{Cov}(R_{it}, R_{mt})$, $E(R_{mt})$, $\text{Var}(R_{mt})$, and $E(R_{0t})$, it is possible to compute $E(R_{it})$ from the CAPM, using equation (2.1). Given a direct sample estimate of $E(R_{it})$, the expected return is overidentified. It is possible to use the overidentification to construct a test of the CAPM by asking if the expected return on the asset is different from the expected return assigned by the model. In this section we illustrate such tests by using both the traditional, return-beta formulation and the stochastic discount factor representation of the CAPM. These examples extend easily to the multiple-beta models.

4.3.1 Static or Unconditional CAPMs

If we make the assumption that all the expectation terms in the CAPM refer to the unconditional expectations, we have an *unconditional* version of the CAPM. It is straightforward to estimate and then test an unconditional version of the CAPM, using equation (3.1) and the stochastic discount factor representation given in equation (3.4). The stochastic discount factor is

$$m_{t+1} = c_0 + c_1 R_{m,t+1}$$

where c_0 and c_1 are fixed parameters. Using only the unconditional expectations, the model implies that

$$E\{(c_0 + c_1 R_{m,t+1})R_{t+1} - 1\} = 0$$

where R_{t+1} is the vector of gross asset returns. The vector of sample orthogonality conditions is

$$g_T = g_T(c_0, c_1) = (1/T) \sum_t \{(c_0 + c_1 R_{m,t+1})R_{t+1} - 1\}.$$

With assets $N > 2$, the number of orthogonality conditions is N and the number of parameters is 2, so the J_T -statistic has $N - 2$ degrees of freedom. Tests of the unconditional CAPM using the stochastic discount factor representation are conducted by Jagannathan and Wang (1995) and Carhart et al. (1995), who reject the model after using monthly data for the postwar United States.

Tests of the unconditional CAPM may also be conducted using the linear, return-beta formulation of equation (2.1) and the GMM. Let $r_t = R_t - R_{0t}\mathbf{1}$ be the vector of excess returns, where R_{0t} is the gross return on some reference asset and $\mathbf{1}$ is an N vector of ones; also let $u_t = r_t - \beta r_{mt}$, where β is the N vector of the betas of the excess returns, relative to the market, and $r_{mt} = R_{mt} - R_{0t}$ is the excess return on the market portfolio. The model implies that

$$E(u_t) = E(u_t r_{mt}) = 0.$$

Let the instruments be $Z_t = (1, r_{mt})'$. The sample orthogonality condition is then

$$g_T(\beta) = T^{-1} \sum_t (r_t - \beta r_{mt}) \otimes Z_{t-1}.$$

The number of orthogonality conditions is $2N$ and the number of parameters is N , so the model is overidentified and may be tested using the J_T -statistic.

An alternative approach to testing the model using the return-beta formulation is to estimate the model under the hypothesis that expected returns depart from the predictions of the CAPM by vector of parameters α , which are called *Jensen's alphas*. Redefining $u_t = r_t - \alpha - \beta r_{mt}$, the model

has $2N$ parameters and $2N$ orthogonality conditions, so it is exactly identified. It is easy to show that the GMM estimators of α and β are the same as the OLS estimators, and equation (4.4) delivers White's (1980) heteroskedasticity-consistent standard errors. The CAPM may be tested using a Wald test or the D -statistic, as described above.

Tests of the unconditional CAPM using the linear return-beta formulation are conducted with the GMM by MacKinlay and Richardson (1991), who reject the model for monthly U.S. data.

4.3.2 Conditional CAPMs

Empirical studies that rejected the unconditional CAPM, as well as mounting evidence of predictable variation in the distribution of security rates of return, led to empirical work on conditional versions of the CAPM starting in the early 1980s. In a *conditional asset pricing model* it is assumed that the expectation terms in the model are conditional expectations, given a public information set that is represented by a vector of predetermined instrumental variables Z_t . The multiple-beta models of Merton (1973) and Cox, Ingersoll, and Ross (1985) are intended to accommodate conditional expectations. Merton (1973, 1980) and Cox-Ingersoll-Ross also showed how a conditional version of the CAPM may be derived as a special case of their intertemporal models. Hansen and Richard (1987) describe theoretical relations between conditional and unconditional versions of mean-variance efficiency.

The earliest empirical formulations of conditional asset pricing models were the *latent variable models*, developed by Hansen and Hodrick (1983) and Gibbons and Ferson (1985) and later refined by Campbell (1987) and Ferson, Foerster, and Keim (1993). These models allow time-varying expected returns, but maintain the assumption that the conditional betas are fixed parameters. Consider the linear, return-beta representation of the CAPM under these assumptions, writing $E(r_t|Z_{t-1}) =$

$\beta E(r_{m_t}|Z_{t-1})$. The returns are measured in excess of a risk-free asset. Let r_{1t} be some reference asset with nonzero β_1 , so that

$$E(r_{1t}|Z_{t-1}) = \beta_1 E(r_{m_t}|Z_{t-1}).$$

Solving this expression for $E(r_{m_t}|Z_{t-1})$ and substituting, we have

$$E(r_t|Z_{t-1}) = CE(r_{1t}|Z_{t-1})$$

where $C = (\beta / \beta_1)$ and $/$ denotes element-by-element division. With this substitution, the expected market risk premium is the latent variable in the model, and C is the N vector of the model parameters. When we form the error term $u_t = r_t - Cr_{1t}$, the model implies $E(u_t|Z_{t-1}) = 0$ and we can estimate and test the model by using the GMM. Gibbons and Ferson (1985) argued that the latent variable model is attractive in view of the difficulties in measuring the true market portfolio, but Wheatley (1989) emphasized that it remains necessary to assume that ratios of the betas, measured with respect to the unobserved market portfolio, are constant parameters.

Campbell (1987) and Ferson and Foerster (1995) show that a single-beta latent variable model is rejected by the data. This finding rejects the hypothesis that there is a (conditional) minimum variance portfolio such that the ratios of conditional betas on this portfolio are fixed parameters. Therefore, the empirical evidence suggests that conditional asset pricing models should be consistent with either (1) a time-varying beta or (2) more than one beta for each asset.¹⁰

Conditional, multiple-beta models with constant betas are examined empirically by Ferson and Harvey (1991), Evans (1994), and Ferson and Korajczyk (1995). They reject such models after the usual statistical tests but find that they still capture a large fraction of the predictability of stock and

¹⁰ A model with more than one fixed beta, and with time-varying risk premiums, is generally consistent with a single, time-varying beta for each asset. For example, assume that there are two factors with constant betas and time-varying risk premiums, where a time-varying combination of the two factors is a minimum-variance portfolio.

bond returns over time. By allowing for time-varying betas, these studies find that the time-variation in betas contributes a relatively small amount to the time-variation in expected asset returns. Intuition for this finding can be obtained by considering the following approximation. Suppose that time-variation in expected excess returns is $E(r|Z) = \lambda\beta$, where λ is a vector of time-varying expected risk premiums for the factors and β is a matrix of time-varying betas. Using a Taylor series, we can approximate

$$\text{Var}[E(r|Z)] \approx E(\beta)' \text{Var}[\lambda] E(\beta) + E(\lambda)' \text{Var}[\beta] E(\lambda).$$

The first term in the decomposition reflects the contribution of the time-varying risk premiums; the second reflects the contribution of time-varying betas. Since the average beta $E(\beta)$ is on the order of 1.0 in monthly data, while the average risk premium $E(\lambda)$ is typically less than 0.01, the first term dominates the second term. This means that time-variation in conditional betas is less important than time-variation in expected risk premiums, from the perspective of modeling predictable variation in expected asset returns.

While from the perspective of modeling predictable time-variation in asset returns, time-variation in conditional betas is not as important as time-variation in expected risk premiums, this does not imply that beta variation is empirically unimportant. From the perspective of modeling the cross-sectional variation in expected asset returns, beta variation over time may be very important. To see this, consider the unconditional expected excess return vector, obtained from the model as

$$E\{E(r|Z)\} = E\{\lambda\beta\} = E(\lambda)E(\beta) + \text{Cov}(\lambda,\beta).$$

Viewed as a cross-sectional relation, the term $\text{Cov}(\lambda,\beta)$ may differ significantly in a cross section of assets. Therefore, the implications of a conditional version of the CAPM for the cross section of unconditional expected returns may depend importantly on common time-variation in betas and

expected market risk premiums. The empirical tests of Jagannathan and Wang (1995) suggest that this is the case.

Harvey (1989) replaced the constant beta assumption with the assumption that the ratio of the expected market premium to the conditional market variance is a fixed parameter, as in

$$E(r_{mt}|Z_{t-1}) / \text{Var}(r_{mt}|Z_{t-1}) = \gamma.$$

The conditional expected returns may then be written according to the conditional CAPM as

$$E(r_t|Z_{t-1}) = \gamma \text{Cov}(r_p, r_{mt}|Z_{t-1}).$$

Harvey's version of the conditional CAPM is motivated by Merton's (1980) model, in which the ratio γ , called the *market price of risk*, is equal to the relative risk aversion of a representative investor in equilibrium. Harvey also assumes that the conditional expected risk premium on the market (and the conditional market variance, given fixed γ) is a linear function of the instruments, as in

$$E(r_{mt}|Z_{t-1}) = \delta_m' Z_{t-1}$$

where δ_m is a coefficient vector. Define the error terms $v_t = r_{mt} - \delta_m' Z_{t-1}$ and $w_t = r_t(1 - v_t\gamma)$. The model implies that the stacked error term $u_t = (v_t, w_t)$ satisfies $E(u_t|Z_{t-1}) = 0$, so it is straightforward to estimate and then test the model using the GMM. Harvey (1989) rejects this version of the conditional CAPM for monthly data in the U.S. In Harvey (1991) the same formulation is rejected when applied using a world market portfolio and monthly data on the stock markets of 21 developed countries.

The conditional CAPM may be tested using the stochastic discount factor representation given by equation (3.4): $m_{t+1} = c_{0t} - c_{1t} R_{mt+1}$. In this case the coefficients c_{0t} and c_{1t} are measurable functions of the information set Z_t . To implement the model empirically it is necessary to specify functional forms for the c_{0t} and c_{1t} . From the expression (3.4) it can be seen that these coefficients are nonlinear functions of the conditional expected market return and its conditional variance. As yet there is no

theoretical guidance for specifying the functional forms. Cochrane (1992) suggests approximating the coefficients using linear functions, and this approach is followed by Carhart et al. (1995), who reject the conditional CAPM for monthly U.S. data.

Jagannathan and Wang (1993) show that the conditional CAPM implies an *unconditional* two-factor model. They show that

$$m_{t+1} = a_0 + a_1 E(r_{m,t+1} | I_t) + R_{m,t+1}$$

(where I_t denotes the information set of investors and a_0 and a_1 are fixed parameters) is a valid stochastic discount factor in the sense that $E(R_{i,t+1} m_{t+1}) = 1$ for this choice of m_{t+1} . Using a set of observable instruments Z_t , and assuming that $E(r_{m,t+1} | Z_t)$ is a linear function of Z_t , they find that their version of the model explains the cross section of unconditional expected returns better than does an unconditional version of the CAPM. Bansal and Viswanathan (1993) develop conditional versions of the CAPM and multiple-factor models in which the stochastic discount factor m_{t+1} is a nonlinear function of the market or factor returns. Using nonparametric methods, they find evidence to support the nonlinear versions of the models. Bansal, Hsieh, and Viswanathan (1993) compare the performance of nonlinear models with linear models, using data on international stocks, bonds, and currency returns, and they find that the nonlinear models perform better. Additional empirical tests of the conditional CAPM and multiple-beta models, using stochastic discount factor representations, are beginning to appear in the literature. We expect that future studies will further refine the relations among the various empirical specifications.

5. Model Diagnostics

We have discussed several examples of stochastic discount factors corresponding to particular theoretical asset pricing models, and we have shown how to test whether these models assigned the right expected returns to financial assets. The stochastic discount factors corresponding to these

models are particular parametric functions of the data observed by the econometrician. While empirical studies based on these parametric approaches have led to interesting insights, the parametric approach makes strong assumptions about the economic environment. In this section, we discuss some alternative econometric approaches to the problem of asset pricing models.

5.1 Moment Inequality Restrictions

Hansen and Jagannathan (1991) derive restrictions from asset pricing models while assuming as little structure as possible. In particular, they assume that the financial markets obey the law of one price and that there are no arbitrage opportunities. These assumptions are sufficient to imply that there exists a stochastic discount factor m_{t+1} (which is almost surely positive, if there is no arbitrage) such that equation (3.1) is satisfied.

Note that if the stochastic discount factor is a degenerate random variable (i.e., a constant), then equation (3.1) implies that all assets must earn the same expected return. If assets earn different expected returns, then the stochastic discount factor cannot be a constant. In other words, cross-sectional differences in expected asset returns carry implications for the variance of any valid stochastic discount factor, which satisfies equation (3.1). Hansen and Jagannathan make use of this observation to derive lower bounds on the volatility of stochastic discount factors. Shiller (1979, 1981), Singleton (1980), and Leroy and Porter (1981) derive related volatility bounds in specific models, and their empirical work suggests that the stochastic discount factors implied by these simple models are not volatile enough to explain expected returns across assets. Hansen and Jagannathan (1991) show how to use volatility bounds as a general diagnostic device.

In what follows we derive the Hansen and Jagannathan (1991) bounds and discuss their empirical application. To simplify the exposition, we focus on an unconditional version of the bounds, using only the unconditional expectations. We posit a hypothetical, unconditional risk-free asset with

return $R_f = E(m_{t+1})^{-1}$. We take the value of R_f , or equivalently $E(m_{t+1})$, as a parameter to be varied as we trace out the bounds.

The law of one price guarantees the existence of some stochastic discount factor which satisfies equation (3.1). Consider the following projection of any such m_{t+1} on the vector of gross asset returns, R_{t+1} :

$$(5.1) \quad m_{t+1} = R_{t+1}'\beta + \varepsilon_{t+1}$$

where

$$E(\varepsilon_{t+1}R_{t+1}) = 0$$

and where β is the projection coefficient vector. Multiply both sides of equation (5.1) by R_{t+1} and take the expected value of both sides of the equation, using $E[R_{t+1}\varepsilon_{t+1}] = 0$, to arrive at an expression which may be solved for β . Substituting this expression back into (5.1) gives the “fitted values” of the projection as

$$(5.2) \quad m_{t+1}^* = R_{t+1}'\beta = R_{t+1}'E(R_{t+1}R_{t+1}')^{-1}\underline{1}.$$

By inspection, the m_{t+1}^* given by equation (5.2) is a valid stochastic discount factor, in the sense that equation (3.1) is satisfied when m_{t+1}^* is used in place of m_{t+1} . We have therefore constructed a stochastic discount factor m_{t+1}^* that is also a payoff on an investment position in the N given assets, where the vector $E(R_{t+1}R_{t+1}')^{-1}\underline{1}$ provides the weights. This payoff is the unique linear least squares approximation of every admissible stochastic discount factor in the space of available asset payoffs.

Substituting m_{t+1}^* for $R_{t+1}'\beta$ in equation (5.1) shows that we may write any stochastic discount factor, m_{t+1} , as

$$m_{t+1} = m_{t+1}^* + \varepsilon_{t+1}$$

where $E(\epsilon_{t+1} m_{t+1}^*) = 0$. It follows that $\text{Var}(m_{t+1}) \geq \text{Var}(m_{t+1}^*)$. This expression is the basis of the Hansen-Jagannathan bound on the variance of m_{t+1} . Since m_{t+1}^* depends only on the second moment matrix of the N returns, the lower bound depends only on the assets available to the econometrician and not on the particular asset pricing model that is being studied. To obtain an explicit expression for the variance bound in terms of the underlying asset-return moments, substitute from the previous expressions to obtain

$$\begin{aligned}
 \text{Var}(m_{t+1}) &\geq \text{Var}(m_{t+1}^*) \\
 (5.3) \quad &= \beta' \text{Var}(R_{t+1}) \beta \\
 &= [\text{Cov}(m, R') \text{Var}(R)^{-1}] \text{Var}(R) [\text{Var}(R)^{-1} \text{Cov}(m, R')] \text{Cov}(m, R') \\
 &= [1 - E(m)E(R')] \text{Var}(R)^{-1} [1 - E(m)E(R)]
 \end{aligned}$$

where the time subscripts are suppressed to conserve notation and the last line follows from $E(mR) = 1 = E(m)E(R) + \text{Cov}(m, R)$. As we vary the hypothetical values of $E(m) = R_f^{-1}$, the equation (5.3) traces out a parabola in $E(m)$, $\sigma(m)$ space, where $\sigma(m)$ is the standard deviation of m_{t+1} . If we place $\sigma(m)$ on the y axis and $E(m)$ on the x axis, the Hansen-Jagannathan bounds resemble a cup, and the implication is that any valid stochastic discount factor m_{t+1} must have a mean and standard deviation that places it within the cup.

The lower bound on the volatility of a stochastic discount factor, as given by equation (5.3), is closely related to the standard mean-variance analysis that has long been used in the financial economics literature. To see this, recall that if $r = R - R_f$ is the vector of excess returns, then (3.1) implies that

$$0 = E(mR) = E(m)E(r) + \rho \sigma(m)\sigma(r).$$

Since $-1 \leq \rho \leq 1$, we have that

$$\sigma(m) / E(m) \geq E(r) / \sigma(r)$$

for all i . The right side of this expression is the *Sharpe ratio* for asset i . The Sharpe ratio is defined as the expected excess return on an asset, divided by the standard deviation of the excess return (see Sharpe 1994 for a recent discussion of this ratio). Consider plotting every portfolio that can be formed from the N assets, in the Standard Deviation (x axis) – Mean (y axis) plane. The set of such portfolios with the smallest possible standard deviation for a given mean return is the minimum-variance boundary. Consider the tangent to the minimum-variance boundary from the point $1/E(m)$ on the y axis. The tangent point is a portfolio of the asset returns, and the slope of this tangent line is the maximum Sharpe ratio that can be attained with a given set of N assets and a given risk-free rate, $R_f = 1/E(m)$. The slope of this line is also equal to R_f multiplied by the Hansen-Jagannathan lower bound on $\sigma(m)$ for a given $E(m) = R_f^{-1}$. That is, we have that

$$\sigma(m) \geq E(m) |\text{Max}\{E(r_i) / \sigma(r_i)\}|$$

for the given R_f .

The preceding analysis is based on equation (3.1), which is equivalent to the law of one price. If there are no arbitrage opportunities, it implies that m_{t+1} is a strictly positive random variable. Hansen and Jagannathan (1991) show how to obtain tighter bounds on the standard deviation of m_{t+1} by making use of the restriction that there are no arbitrage opportunities. They also show how to incorporate conditioning variables into the analysis. Snow (1991) extends the Hansen-Jagannathan analysis to include higher moments of the asset returns. His extension is based on the Holder inequality, which implies that for given values of δ and p such that

$$1/\delta + 1/p = 1$$

then

$$E(mR) \leq E(m^\delta)^{1/\delta} E(R^p)^{1/p}.$$

Cochrane and Hansen (1992) refine the Hansen-Jagannathan bounds to consider information about the correlation between a given stochastic discount factor and the vector of asset returns. This provides a tighter set of restrictions than the original bounds, which only make use of the fact that the correlation must be between -1 and $+1$.

5.2 Statistical Inference for Moment Inequality Restrictions

Cochrane and Hansen (1992), Burnside (1994), and Cecchetti, Lam, and Mark (1994) show how to take sampling errors into account when examining whether a particular candidate stochastic discount factor satisfies the Hansen-Jagannathan bound. In what follows we will outline a computation which allows for sampling errors, following the discussion in Cochrane and Hansen (1992).

Assume that the econometrician has a time series of T observations on a candidate for the stochastic discount factor, denoted by y_t , and the N asset returns R_t . We also assume that the risk-free asset is not one of the N assets. Hence $v = E(m) = 1/R_f$ is an unknown parameter to be estimated. Consider a linear regression of m_{t+1} onto the unit vector and the vector of asset returns as $m_{t+1} = \alpha + R_t'\beta + u_{t+1}$. We use the regression function in the following system of population moment conditions:

$$(5.4) \quad \begin{aligned} E(\alpha + R_t'\beta) &= v \\ E(R_t\alpha + R_tR_t'\beta) &= \mathbf{1}_N \\ E(y_t) &= v \\ E[(\alpha + R_t'\beta)^2] - E[y_t^2] &\leq 0. \end{aligned}$$

The first equation says that the expected value of $m_t \equiv \alpha + R_t'\beta = v$. The second equation says that the regression function for m_t is a valid stochastic discount factor. The third equation says that v is the expected value of the particular candidate discount factor that we wish to test. The fourth equation

states that the Hansen-Jagannathan bound is satisfied by the particular candidate stochastic discount factor.

We can estimate the $N + 2$ parameters ν and α , and the N vector β , using the $N + 3$ equations in (5.4), by treating the last inequality as an equality and using the GMM. Treating the last equation as an equality corresponds to the null hypothesis that the mean and variance of y_t place it on the Hansen-Jagannathan boundary. Under the null hypothesis that the last equation of (5.4) holds as an equality, the minimized value of the GMM criterion function J_T , multiplied by T , has a Chi-Square distribution with one degree of freedom. Cochrane and Hansen (1992) suggest testing the inequality relation using the one-sided test.

5.3 Specification Error Bounds

The methods we have examined so far are developed, for the most part, under the null hypothesis that the asset pricing model under consideration by the econometrician assigns the right prices (or expected returns) to all assets. An alternative is to assume that the model is wrong and examine how wrong the model is. In this section we will follow Hansen and Jagannathan (1994) and discuss one possible way to examine what is missing in a model and assign a scalar measure of the model's misspecification.¹¹

Let y_t denote the candidate stochastic discount factor corresponding to a given asset pricing model, and let m^* denote the unique stochastic discount factor that we constructed earlier, as a combination of asset payoffs. We assume that $E[y_t R_t]$ does not equal $\underline{1}_N$, the N vector of ones; i.e.,

¹¹ GMM-based model specification tests are examined in a general setting by Newey (1985). Other related work includes Boudoukh, Richardson, and Smith (1993), who compute approximate bounds on the probabilities of the test statistics in the presence of inequality restrictions. Chen and Knez (1992) develop nonparametric measures of market integration by using related methods. Hansen, Heaton, and Luttmer (1995) show how to compute specification error and volatility bounds when there are market frictions such as short-sale constraints and proportional transaction costs.

the model does not correctly price all of the gross returns. We can project y_t on the N asset returns to get $y_t = R_t'\alpha + u_t$, and project m_t^* on the vector of asset returns to get $m_t^* = R_t'\beta + \varepsilon_t$. Since the candidate y_t does not correctly price all of the assets, then α and β will not be the same. Define $p_t = (\beta - \alpha)'R_t$ as the *modifying payoff* to the candidate stochastic discount factor y_t . Clearly, $(y_t + p_t)$ is a valid stochastic discount factor, satisfying equation (3.1). Hansen and Jagannathan (1994) derive specification tests based on the size of the modifying payoff, which measures how far the model's candidate for a stochastic discount factor y_t is from a valid stochastic discount factor. Hansen and Jagannathan (1994) show that a natural measure of this distance is $\delta = E(p_t^2)$, which provides an economic interpretation for the model's misspecification. Payoffs that are orthogonal to p_t are correctly priced by the candidate y_t , and $E(p_t^2)$ is the maximum amount of mispricing by using y_t for any payoff normalized to have a unit second moment. The modifying payoff p_t is also the minimal modification that is sufficient to make y_t a valid stochastic discount factor.

Hansen and Jagannathan (1994) consider an estimator of the distance measure δ given as the solution to the following maximization problem:

$$(5.5) \quad \delta_T = \text{Max}_{\alpha} T^{-1} \sum_t [y_t^2 - (y_t + \alpha'R_t)^2 + 2\alpha'1_N]^{1/2} .$$

If α_T is the solution to (5.5), then the estimate of the modifying payoff is $\alpha_T'R_t$. It can be readily verified that the first-order condition to (5.5) implies that $\alpha_T'R_t$ satisfies the sample counterpart to the asset pricing equation (3.1).

To obtain an estimate of the sampling error associated with the estimated value δ_T , consider

$$u_t = y_t^2 - (y_t + \alpha_T'R_t)^2 + 2\alpha_T'1_N.$$

The sample mean of u_t is δ_T^2 . We can obtain a consistent estimator of the variance of δ_T^2 by the frequency zero spectral density estimators described in Newey and West (1987a) or Andrews (1991) and applied to the time series $\{u_t - \delta_T^2\}_{t=1\dots T}$. Let s_T denote the estimated standard deviation of δ_T^2

obtained this way. Then, under standard assumptions, we have that $T^{1/2}(\delta_T - \delta) / s_T$ converges to a normal (0,1) random variable. Hence, using the delta method, we obtain

$$(5.6) \quad T^{1/2}\delta_T / 2s_T(\delta_T - \delta) - \text{prob} \rightarrow N(0,1).$$

6. Conclusions

In this article we have reviewed econometric tests of a wide range of asset pricing models, where the models are based on the law of one price, the no-arbitrage principle, and models of market equilibrium with investor optimization. Our review included the earliest of the equilibrium asset pricing models, the CAPM, and also considered dynamic multiple-beta and arbitrage pricing models. We provided some results for the asymptotic distribution of traditional two-pass estimators for asset pricing models stated in the linear, return-beta formulation. We emphasized the econometric evaluation of asset pricing models by using Hansen's (1982) Generalized Method of Moments. Our examples illustrate the simplicity and flexibility of the GMM approach. We showed that most asset pricing models can be represented in a *stochastic discount factor* form, which makes the application of the GMM straightforward. Finally, we discussed model diagnostics that provide additional insight into the causes of the statistical rejections in GMM tests and which help assess the specification errors in these models.

Appendix

Proof of Theorem 2.1 The proof comes from Jagannathan and Wang (1995). We first introduce some additional notation. Let I_N be the N -dimensional identity matrix and 1_T be a T -dimensional vector of ones. It follows from equation (2.17) that

$$R - \mu = T^{-1}(I_N \otimes 1_T')\varepsilon_k, \quad k = 1, \dots, K_2$$

where

$$\varepsilon_k = (\varepsilon_{1k1}, \dots, \varepsilon_{1kT}, \dots, \varepsilon_{Nk1}, \dots, \varepsilon_{NkT})'.$$

By the definition of b_k , we have that

$$b_k - \beta_k = [I_N \otimes ((f_k' f_k)^{-1} f_k')] \varepsilon_k$$

where f_k is the vector-demeaned factor realizations, conformable to the vector ε_k . In view of the assumption that the conditional covariance of ε_{ik} and ε_{jl} , given the time series of the factors (denoted by f), is a fixed constant σ_{ijkl} , we have that

$$\begin{aligned} & E[(b_k - \beta_k)(R_1 - \mu_1) | f] \\ &= T^{-1} [I_N \otimes ((f_k' f_k)^{-1} f_k')] E[\varepsilon_k \varepsilon_1' | f] (I_N \otimes 1_T) \\ &= T^{-1} [I_N \otimes ((f_k' f_k)^{-1} f_k')] \Sigma_{kl} (I_N \otimes 1_T) \\ &= T^{-1} \Sigma_{kl} \otimes [(f_k' f_k)^{-1} f_k' 1_T] = 0 \end{aligned}$$

where we denote the matrix of the $\{\sigma_{ijkl}\}_{ij}$ by Σ_{kl} . The last line follows from the fact that $f_k' 1_T = 0$. Hence we have shown that $(b_k - \beta_k)$ is uncorrelated with $(R - \mu)$. Therefore, the terms u and $h\gamma_2$ should be uncorrelated, and the asymptotic variance of $T^{1/2}(g - \gamma)$ in equation (2.15) is given by

$$(x'x)^{-1} x' [\text{Var}(u) + \text{Var}(h\gamma_2)] x(x'x)^{-1}.$$

We assume that the sample covariance matrix of the factors exists and converges in probability to a constant positive definite matrix Ω , with typical element Ω_{kl} . We therefore have

$$\begin{aligned}
& TE[(b_k - \beta_k)(b_l - \beta_l) | f] \\
&= T^{-1}[I_N \otimes ((f_k' f_k)^{-1} f_k')] (\Sigma_{kl} \otimes T_T) [I_N \otimes ((f_k' f_k)^{-1} f_k')] \\
&= \Sigma_{kl} \otimes [(T^{-1} f_k' f_k)^{-1} (T^{-1} f_k' f_k)^{-1} (T^{-1} f_l' f_l)^{-1}] \\
&\rightarrow_d (\Omega_{kk}^{-1} \Omega_{kl} \Omega_{ll}^{-1}) \Sigma_{kl} \text{ (as } T \rightarrow \infty \text{)}.
\end{aligned}$$

Thus

$$\text{Var}(h\gamma_2) = \sum_{l, k=1, \dots, K_2} \gamma_{2k} \gamma_{2l} (\Omega_{kk}^{-1} \Omega_{kl} \Omega_{ll}^{-1}) \Sigma_{kl}$$

and

$$\begin{aligned}
W &= (x'x)^{-1} x' \text{Var}(h\gamma_2) x (x'x)^{-1} \\
&= \sum_{l, k=1, \dots, K_2} \gamma_{2k} \gamma_{2l} (\Omega_{kk}^{-1} \Omega_{kl} \Omega_{ll}^{-1}) (x'x)^{-1} x' \Sigma_{kl} x (x'x)^{-1}
\end{aligned}$$

Q.E.D.

References

- Abel, A. (1990). Asset prices under habit formation and catching up with the Jones, *American Economic Review Papers and Proceedings* 80, 38–42.
- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Andrews, D. W. K., and J. C. Monahan. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.
- Arrow, K. J. (1970). *Essays in the Theory of Risk-Bearing*. Amsterdam: North-Holland.
- Bansal, R., and S. Viswanathan. (1993). No arbitrage and arbitrage pricing: A new approach. *Journal of Finance* 57, 1231–1262.
- Bansal, R., D. A. Hsieh, and S. Viswanathan. (1993). A new approach to international arbitrage pricing. *Journal of Finance* 57, 1719–1747.
- Becker, G. S., and K. M. Murphy. (1988). A theory of rational addiction. *Journal of Political Economy* 96, 675–700.
- Beja, A. (1971). The structure of the cost of capital under uncertainty. *Review of Economic Studies* 4, 359–369.
- Berk, J. B. (1994). A critique of size-related anomalies. *Review of Financial Studies* 8, 275–286.
- Black, F. (1972). Capital market equilibrium with restricted borrowing. *Journal of Business* 45, 444–455.
- Black, F., M. C. Jensen, and M. Scholes. (1972). The Capital Asset Pricing Model: Some Empirical Tests. In *Studies in the Theory of Capital Markets*, M. C. Jensen, ed. New York: Praeger, 79–121.

- Boudoukh, J., M. Richardson, and T. Smith. (1993). Is the ex ante risk premium always positive? A new approach to testing conditional asset pricing models. *Journal of Financial Economics* 34, 387–408.
- Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics* 7, 265–296.
- Brown, D. P., and M. R. Gibbons. (1985). A simple econometric approach for utility-based asset pricing models. *Journal of Finance* 40, 359–381.
- Burnside, C. (1994). Hansen-Jagannathan bounds as classical tests of asset-pricing models. *Journal of Business and Economic Statistics* 12, 57–79.
- Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics* 18, 373–399.
- Campbell, J. Y., and J. Cochrane. (1995). By force of habit. Manuscript. Harvard Institute of Economic Research, Harvard University.
- Carhart, M., K. Welch, R. Stevens, and R. Krail. (1995). Testing the conditional CAPM. Working paper. University of Chicago.
- Cecchetti, S. G., P. Lam, and N. C. Mark. (1994). Testing volatility restrictions on intertemporal marginal rates of substitution implied by Euler equations and asset returns. *Journal of Finance* 49, 123– .
- Chen, N. (1983). Some empirical tests of the theory of arbitrage pricing. *Journal of Finance* 38, 1393–1414.
- Chen, Z., and P. Knez. (1992). A measurement framework of arbitrage and market integration. Working paper. University of Wisconsin.
- Cochrane, John H. (1992). A cross-sectional test of a production based asset pricing model. Working paper. University of Chicago.

- Cochrane, J., and L. Hansen. (1992). Asset pricing lessons for macroeconomics. In *The Macroeconomics Annual*, O. Blanchard and S. Fisher, eds. Cambridge, Mass.: MIT Press.
- Connor, G. (1984). A unified beta pricing theory. *Journal of Economic Theory* 34, 13–31.
- Connor, G., and R. A. Korajczyk. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics* 15, 373–394.
- Constantinides, G. (1982). Intertemporal asset pricing with heterogeneous consumers and without demand aggregation. *Journal of Business* 55, 253–267.
- Constantinides, G. M. (1990). Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy* 98, 519–543.
- Constantinides, G. M., and D. Duffie. (1994). Asset pricing with heterogeneous consumers. Working paper. University of Chicago and Stanford University.
- Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross. (1985). A theory of the term structure of interest rates. *Econometrica* 53, 385–407.
- Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. New York: Wiley.
- Detemple, J. B., and F. Zapatero. (1991). Asset prices in an exchange economy with habit formation. *Econometrica* 59, 1633–1657.
- Dunn, K. B., and K. J. Singleton. (1986). Modeling the term structure of interest rates under non-separable utility and durability of goods. *Journal of Financial Economics* 17, 27–55.
- Dybvig, P. H., and J. E. Ingersoll, Jr. (1982). Mean-variance theory in complete markets. *Journal of Business* 55, 233–251.
- Eichenbaum, M. S., L. P. Hansen, and K. J. Singleton. (1988). A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *Quarterly Journal of Economics* 103, 51–78.

- Epstein, L. G., and S. E. Zin. (1989). Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica* 57, 937–969.
- Epstein, L. G., and S. E. Zin. (1991). Substitution, risk aversion and the temporal behavior of consumption and asset returns. *Journal of Political Economy* 99, 263–286.
- Evans, M. D. D. (1994). Expected returns, time-varying risk, and risk premia. *Journal of Finance* 49, 655–679.
- Fama, E. F., and K. R. 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.
- Ferson, W. E. (1983). Expectations of real interest rates and aggregate consumption: Empirical tests. *Journal of Financial and Quantitative Analysis* 18, 477–497.
- Ferson, W. E., and G. M. Constantinides. (1991). Habit persistence and durability in aggregate consumption: Empirical tests. *Journal of Financial Economics* 29, 199–240.
- Ferson, W. E., and S. R. Foerster. (1994). Finite sample properties of the generalized methods of moments tests of conditional asset pricing models. *Journal of Financial Economics* 36, 29–55.
- Ferson, W., and S. Foerster. (1995). Further results on the small-sample properties of the generalized method of moments: Tests of latent variable models. In *Research in Finance*, vol. 13. Greenwich, Conn.: JAI Press, pp. 91–114.
- Ferson, W. E., S. R. Foerster, and D. B. Keim. (1993). General tests of latent variable models and mean-variance spanning. *Journal of Finance* 48, 131–156.
- Ferson, W. E., and C. R. Harvey. (1991). The variation of economic risk premiums. *Journal of Political Economy* 99, 385–415.

- Ferson, W. E., and C. R. Harvey. (1992). Seasonality and consumption-based asset pricing. *Journal of Finance* 47, 511–552.
- Ferson, W. E., and R. A. Korajczyk. (1995). Do arbitrage pricing models explain the predictability of stock returns? *Journal of Business* 68, 309–349.
- Ferson, W. E. and J. J. Merrick, Jr. (1987). Non-stationarity and stage-of-the-business-cycle effects in consumption-based asset pricing relations. *Journal of Financial Economics* 18, 127–146.
- Gallant, R. (1987). *Nonlinear statistical models*. New York: Wiley.
- Gibbons, M. R., and W. Ferson. (1985). Testing asset pricing models with changing expectations and an unobservable market portfolio. *Journal of Financial Economics* 14, 217–236.
- Gorman, W. M. (1953). Community preference fields. *Econometrica* 21, 63–80.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.
- Hansen, L. P., J. Heaton, and E. G. J. Luttmer. (1995). Econometric evaluation of asset pricing models. *Review of Financial Studies* 8, 237–274.
- Hansen, L. P., and R. Hodrick. (1983). Risk averse speculation in the forward foreign exchange market: An econometric analysis of linear models. In *Exchange Rates and International Macroeconomics*, J. A. Frenkel, ed. Chicago: University of Chicago Press.
- Hansen, L. P., and R. Jagannathan. (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy* 99, 225–262.
- Hansen, L. P., and R. Jagannathan. (1994). Assessing specification errors in stochastic discount factor models. NBER Technical Working Paper No. 153.
- 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.

- Hansen, L. P., and K. J. Singleton. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50, 1269–1286.
- Hansen, L. P., and K. J. Singleton. (1983). Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy* 91, 249–265.
- Harrison, M., and D. Kreps. (1979). Martingales and arbitrage in multi-period securities markets. *Journal of Economic Theory* 20, 381–408.
- Harvey, C. R. (1989). Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics* 24, 289–318.
- Harvey, C. R. (1991). The world price of covariance risk. *Journal of Finance* 46, 111–158.
- Heaton, J. (1995). An empirical investigation of asset pricing with temporally dependent preference specifications. *Econometrica* 63, 681– .
- Ibbotson Associates. (1992). Stocks, bonds, bills, and inflation. *1992 Yearbook*. Chicago: Ibbotson Associates.
- Jagannathan, R. (1985). An investigation of commodity futures prices using the consumption-based intertemporal capital asset pricing model. *Journal of Finance* 40, 175–191.
- Jagannathan, R., and Z. Wang. (1993). The CAPM is alive and well. Federal Reserve Bank of Minneapolis Research Department Staff Report 165.
- Jagannathan, R., and Z. Wang. (1995). The conditional-CAPM and the cross-section of expected returns. *Journal of Finance* (forthcoming).
- Kandel, S. (1984). On the exclusion of assets from tests of the mean-variance efficiency of the market portfolio. *Journal of Finance* 39, 63–75.
- Kandel, S., and R. F. Stambaugh. (1987). On correlations and inferences about mean-variance efficiency. *Journal of Financial Economics* 18, 61–90.

- Lehmann, B. N., and D. M. Modest. (1987). Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons. *Journal of Finance* 42, 233–265.
- Leroy, S., and M. Porter. (1981). The present value relation: Tests based on implied variance bounds. *Econometrica* 49, 555–574.
- Lewbel, A. (1989). Exact aggregation and a representative consumer. *Quarterly Journal of Economics* 104, 621–633.
- Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and Capital budgets. *Review of Economics and Statistics* 47, 13–37.
- Lucas, R. E., Jr. (1978). Asset prices in an exchange economy. *Econometrica* 46, 1429–1445.
- Luttmer, E. (1993). Asset pricing in economies with frictions. Working paper, Northwestern University.
- McElroy, M. B., and E. Burmeister. (1988). Arbitrage pricing theory as a restricted nonlinear multivariate regression model. *Journal of Business and Economic Statistics* 6, 29–42.
- MacKinlay, A. C., and M. P. Richardson. (1991). Using generalized method of moments to test mean-variance efficiency. *Journal of Finance* 46, 511–527,
- Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica* 41, 867–887.
- Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8, 323–361.
- Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica* 34, 768–783.
- Newey, W., (1985). Generalized method of moments specification testing. *Journal of Econometrics* 29, 199–256.
- Newey, W. K. and K. D. West. (1987a). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

- Newey, W. K. and K. D. West. (1987b). Hypothesis testing with efficient method of moments estimation. *International Economic Review* 28, 777–787.
- Novalés, A. (1992). Equilibrium interest-rate determination under adjustment costs. *Journal of Economic Dynamics and Control* 16, 1–25.
- Roll, R. (1977). A critique of the asset pricing theory's tests: Part 1—on past and potential testability of the theory. *Journal of Financial Economics* 4, 129–176.
- Ross, S. A. (1976). The arbitrage pricing theory of capital asset pricing. *Journal of Economic Theory* 13, 341–360.
- Ross, S., (1977). Risk, return and arbitrage. In *Risk and Return in Finance*, I. Friend and J. L. Bicksler, eds. *Risk and Return in Finance*. Cambridge, Mass.: Ballinger.
- Rubinstein, M. (1974). An aggregation theorem for securities markets. *Journal of Financial Economics* 1, 225–244.
- Rubinstein, M. (1976). The valuation of uncertain income streams and the pricing of options. *Bell Journal of Economics and Management Science* 7, 407–425.
- Ryder, and Heal (1973). Optimal growth with intertemporally dependent preferences. *Review of Economic Studies* 40, 1–31.
- Shanken, J. (1987). Multivariate proxies and asset pricing relations: Living with the Roll critique. *Journal of Financial Economics* 18, 91–110.
- Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial Studies* 5, 1–33.
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19, 425–442.
- Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management* 21, 49–59.
- Shiller, R. J. (1979). The volatility of long-term interest rates and expectations models of the term structure. *Journal of Political Economy* 87, 1190–1219.

- Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 71, 421–436.
- Singleton, K. J. (1980). Expectations models of the term structure and implied variance bounds. *Journal of Political Economy* 88, 1159–1176.
- Snow, K. N. (1991). Diagnosing asset pricing models using the distribution of asset returns. *Journal of Finance* 46, 955–983.
- Stambaugh, R. F. (1982). On the exclusion of assets from tests of the two-parameter model: a sensitivity analysis. *Journal of Financial Economics* 10, 237–268.
- Sundaresan, S. M. (1989). Intertemporally dependent preferences and the volatility of consumption and wealth. *Review of Financial Studies* 2, 73–89.
- Wheatley, S. (1988). Some tests of international equity integration. *Journal of Financial Economics* 21, 177–212.
- Wheatley, S. M. (1989). A critique of latent variable tests of asset pricing models. *Journal of Financial Economics* 23, 325–338.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.
- Wilson, R. (1968). The theory of syndicates. *Econometrica* 36, 119–132.