G. S. Maddala and C. R. Rao, eds., *Handbook of Statistics, Vol. 14*

2

Instrumental Variables Estimation of Conditional Beta Pricing Models

Campbell R. Harvey and Chris Kirby

A number of well-known asset pricing models imply that the expected return on an asset can be written as a linear function of one or more beta coefficients that measure the asset's sensitivity to sources of undiversifiable risk. This paper provides an overview of the econometric evaluation of such models using the method of instrumental variables. We present numerous examples that cover both single-beta and multi-beta models. These examples are designed to illustrate the various options available to researchers for estimating and testing beta pricing models. We also examine the implications of a variety of different assumptions concerning the time-series behavior of conditional betas, covariances, and reward-to-risk ratios. The techniques discussed in this paper have applications in other areas of asset pricing as well.

1. Introduction

Asset pricing models often imply that the expected return on an asset can be written as a linear combination of market-wide risk premia, where each risk premium is multiplied by a *beta* coefficient that measures the sensitivity of the return on the asset to a source of undiverifiable risk in the economy. Indeed, this type of tradeoff between risk and expected return is implied by some of the most famous models in financial economics. The Sharpe (1964) – Lintner (1965) capital asset pricing model (CAPM), the Black (1972) CAPM, the Merton (1973) intertemporal CAPM, the arbitrage pricing theory (APT) of Ross (1976), and the Breeden (1979) consumption CAPM can all be classified under the general heading of beta pricing models. Although these models differ in terms of underlying structural assumptions, each implies a pricing relation that is linear in one or more betas.

The fundamental difference between *conditional* and *unconditional* beta pricing models is the specification of the information environment that investors use to form expectations. Unconditional models imply that investors set prices based on an unconditional assessment of the joint probability distribution of future returns. Under such a scenario we can construct an estimate of an investor's expected return on an asset by taking an average of past returns. Conditional models, on the other hand, imply that investors have time-varying expectations concerning the joint probability distribution of future returns. In order to construct an estimate of an investor's conditional expected return on an asset we have to use the information available to the investor at time $t-1$ to forecast the return for time t.

Both conditional and unconditional models attempt to explain the cross-sectional variation in expected returns. Unconditional models imply that differences in average risk across assets determine differences in average returns. There are no time-series predictions other than expected returns are constant. Conditional models have similar cross-sectional implications: differences in conditional risk determine differences in conditional expected returns. But conditional models have implications concerning the time-series properties of expected returns as well. Conditional expected returns vary with changes in conditional risk and fluctuations in market-wide risk premiums. In theory, we can test a conditional beta pricing model using a single asset.

Empirical tests of beta pricing models can be interpreted within the familiar framework of mean-variance analysis. Unconditional tests seek to determine whether a certain portfolio is on the efficient portion of the unconditional mean-variance frontier. The unconditional frontier is determined by the unconditional means, variances and covariances of the asset returns. Conditional tests of beta pricing models are designed to answer a similar question: does a certain portfolio lie on the efficient portion of the mean-variance frontier at each point in time? In conditional tests, however, the mean-variance frontier is determined by the conditional means, conditional variances, and conditional covariances of asset returns.

As a general rule, the rejection of unconditional efficiency does not imply a rejection of conditional mean-variance efficiency. This is easily demonstrated using an example given by Dybvig and Ross (1985) and Hansen and Richard (1987). Suppose we are testing whether the 30-day Treasury bill is unconditionally efficient using monthly data. Unconditionally, the 30-day bill does not lie on the efficient frontier. It is a single risky asset (albeit low risk) whose return has non-zero variance. Thus it is surely dominated by an appropriately chosen portfolio. At the conditional level, however, the conclusion is much different. Conditionally, the 30-day bill is nominally risk free. At the end of each month we know precisely what the return will be over the next month. Because the conditional variance of the return on the T-bill is zero, it must be conditionally efficient.

A number of different methods have been proposed for testing beta pricing models. This paper focuses on one in particular: the method of instrumental variables. Instrumental variables are a set of data, specified by the econometrician, that proxy for the information that investors use to form expectations. The primary advantage of the instrumental variables approach is that it provides a highly tractable way of characterizing time-varying risk and expected returns. Our discussion of the instrumental variables methodology is organized along the following lines. Section 2 uses the conditional version of the Sharpe (1964) – Lintner (1965) CAPM to illustrate how the instrumental variables approach can be employed to estimate and test single beta models. Section 3 extends the analysis to multi-beta models. Section 4 introduces the technique of latent variables. Section 5 provides an overview of the estimation methodology. The final section offers some brief closing remarks.

2. Single beta models

A. The conditional CAPM

The conditional version of the Sharpe (1964) – Lintner (1965) CAPM is undoubtedly one of the most widely studied conditional beta pricing models. We can express the pricing relation associated with this model as:

$$\mathrm{E}[r_{jt}|\boldsymbol{\Omega}_{t-1}] = \frac{\mathrm{Cov}[r_{jt}, r_{mt}|\boldsymbol{\Omega}_{t-1}]}{\mathrm{Var}[r_{mt}|\boldsymbol{\Omega}_{t-1}]}\mathrm{E}[r_{mt}|\boldsymbol{\Omega}_{t-1}] \ , \tag{1}$$

where r_{jt} is the return on portfolio j from time $t-1$ to time t measured in excess of the risk free rate, r_{mt} is the excess return on the market portfolio, and $\boldsymbol{\Omega}_{t-1}$ represents the information set that investors use to form expectations. The ratio of the conditional covariance between the return on portfolio j and the return on the market, $\mathrm{Cov}[r_{jt}, r_{mt}|\boldsymbol{\Omega}_{t-1}]$, to the variance of the return on the market, $\mathrm{Var}[r_{mt}|\boldsymbol{\Omega}_{t-1}]$, is the conditional beta of portfolio j with respect to the market. Any cross-sectional variation in expected returns can be attributed solely to differences in conditional beta coefficients.

As it stands the pricing relation shown in (1) is untestable. To make it testable we have to impose additional structure on the model. In particular, we have to specify a model for conditional expectations. Thus any test of (1) will be a joint test of the conditional CAPM and the assumed specification for conditional expectations. In theory any functional form could be used. Let $f(\boldsymbol{Z}_{t-1})$ denote the statistical model that generates conditional expectations where $\boldsymbol{Z}$ is a set of instrumental variables. The function $f(\cdot)$ could be a linear regression model, a Fourier flexible form [Gallant (1982)], a nonparametric kernel estimator [Silverman (1986), Harvey (1991), and Beneish and Harvey (1995)], a seminonparametric density [Gallant and Tauchen (1989)], a neural net [Gallant and White (1990)], an entropy encoder [Glodjo and Harvey (1995)], or a polynomial series expansion [Harvey and Kirby (1995)].

Once we take a stand on the functional form of the conditional expectations operator it is straightforward to construct a test of the conditional CAPM. First we use $f(\cdot)$ to obtain fitted values for the conditional mean of r_{jt}. This nails down the left-hand side of the pricing relation in (1). Then we apply $f(\cdot)$ again to get fitted values for the three components on the right-hand side of (1). Combining the fitted values for the conditional mean of r_{mt}, those for the conditional covariance between r_{jt} and r_{mt}, and those for the conditional variance of r_{mt} yields fitted values for the right-hand side of (1). If the conditional CAPM is valid then the pricing errors – the difference between the fitted values for the left-hand and right-hand sides of (1) – should be small and unpredictable. This is the basic intuition behind all tests of conditional beta pricing models.

In the presentation that follows we focus on one particular specification for conditional expectations: the linear model. This model, though very simple, has distinct advantages over the many nonlinear alternatives. The linear model is exceedingly easy to implement, and Harvey (1991) shows that it performs well against nonlinear alternatives in out-of-sample forecasting of the market return. In addition, the linear specification is actually more general than it may seem. Recent work has shown that many nonlinear models can be consistently approximated via an expanding sequencing of finite-dimensional linear models. Harvey and Kirby (1995) exploit this fact to develop a simple procedure for constructing analytic tests of both single beta and multi-beta pricing models.

B. Linear conditional expectations

The easiest way to motivate the linear specification for conditional expectations is to assume that the joint distribution of the asset returns and instrumental variables is spherically invariant. This class of distributions is analyzed in Vershik (1964), who shows that it is sufficient for linear conditional expectations, and applied to tests of the conditional CAPM in Harvey (1991). Vershik (1964) provides the following characterization. Consider a set of random variables, $\{x_1, \ldots, x_n\}$, that have finite second moments. Let $\mathbf{H}$ denote a linear manifold spanned by this set. If all random variables in the linear manifold $\mathbf{H}$ that have the same variance have the same distribution then: (i) $\mathbf{H}$ is a *spherically invariant space*; (ii) $\{x_1, \ldots, x_n\}$ is *spherically invariant*; and (iii) every distribution function of any variable in $\mathbf{H}$ is a *spherically invariant distribution*. The above requirements are satisfied, for example, by both the multivariate normal and multivariate t distributions.

A potential disadvantage of Vershik's (1964) definition is that it does not encompass processes like Cauchy for which the variance is undefined. Blake and Thomas (1968) and Chu (1973) propose a definition for an elliptical class of distributions that addresses this shortcoming. A random vector $\boldsymbol{x}$ is said to have an elliptical distribution if and only if its probability density function $p(\boldsymbol{x})$ can be expressed as a function of a quadratic form, $p(\boldsymbol{x}) = f(\frac{1}{2}\boldsymbol{x}'\boldsymbol{C}^{-1}\boldsymbol{x})$, where $\boldsymbol{C}$ is positive definite. When the variance-covariance matrix of $\boldsymbol{x}$ exists it is proportional to $\boldsymbol{C}$ and the Vershik (1964), Blake and Thomas (1968) and Chu (1973) definitions are equivalent.[2] But the quadratic form of the density also covers processes like Cauchy that imply linear conditional expectations where the projection constants depend on the characteristic matrix.

[2] Implicit in Chu's (1973) definition is the existence of the density function. Kelker (1970) provides an alternative approach in terms of the characteristic function. See also Devlin, Gnanadesikan and Kettenring (1976).

C. A general framework for testing the CAPM

A linear specification for conditional expectations implies that the return on portfolio j can be written as:

$$r_{jt} = \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_j + u_{jt} \ , \tag{2}$$

where u_{jt} is the error in forecasting the return on portfolio j at time t, $\boldsymbol{Z}_{t-1}$ is a row vector of ℓ instrumental variables, and $\boldsymbol{\delta}_j$ is a $\ell \times 1$ set of time-invariant weights. Substituting the expression shown in (2) into equation (1) yields the restriction:

$$\boldsymbol{Z}_{t-1}\boldsymbol{\delta}_j = \frac{\boldsymbol{Z}_{t-1}\boldsymbol{\delta}_m}{\mathrm{E}[u_{mt}^2|\boldsymbol{Z}_{t-1}]}\mathrm{E}[u_{jt}u_{mt}|\boldsymbol{Z}_{t-1}] \ , \tag{3}$$

where u_{mt} is the error in forecasting the return on the market portfolio. Note that both the variance term, $\mathrm{E}[u_{mt}^2|\boldsymbol{Z}_{t-1}]$, and the covariance term, $\mathrm{E}[u_{jt}u_{mt}|\boldsymbol{Z}_{t-1}]$, are conditioned on $\boldsymbol{Z}_{t-1}$. Therefore, the pricing relation in (3) should be regarded as an approximation. This is the case because the expectation of the true conditional covariance is not the covariance conditioned on $\boldsymbol{Z}_{t-1}$. The two are connected via the relation: $\mathrm{E}[\mathrm{Cov}(r_{jt}, r_{mt}|\boldsymbol{\Omega}_{t-1})|\boldsymbol{Z}_{t-1}] = \mathrm{Cov}(r_{jt}, r_{mt}|\boldsymbol{Z}_{t-1}) - \mathrm{Cov}(\mathrm{E}[r_{jt}|\boldsymbol{\Omega}_{t-1}], \mathrm{E}[r_{mt} |\boldsymbol{\Omega}_{t-1}]|\boldsymbol{Z}_{t-1})$. An analogous relation holds for the true conditional variance of r_{mt} and the variance conditioned on $\boldsymbol{Z}_{t-1}$. There is no way to construct a test of the original version of pricing restriction given that the true information set $\boldsymbol{\Omega}$ is unobservable.

If we multiply both sides of (3) by the conditional variance of the return on the market portfolio we obtain the restriction:

$$\mathrm{E}[u_{mt}^2\boldsymbol{Z}_{t-1}\boldsymbol{\delta}_j|\boldsymbol{Z}_{t-1}] = \mathrm{E}[u_{jt}u_{mt}\boldsymbol{Z}_{t-1}\boldsymbol{\delta}_m|\boldsymbol{Z}_{t-1}] \ . \tag{4}$$

Notice that the conditional expected return on both the market portfolio and portfolio j have been moved inside the expectations operator. This can be done because both of these quantities are known conditional on $\boldsymbol{Z}_{t-1}$. As a result, we do not need to specify an explicit model for the conditional variance and covariance terms. We simply note that, under the null hypothesis, the disturbance:

$$e_{jt} \equiv u_{mt}^2\boldsymbol{Z}_{t-1}\boldsymbol{\delta}_j - u_{jt}u_{mt}\boldsymbol{Z}_{t-1}\boldsymbol{\delta}_m \ , \tag{5}$$

should have mean zero and be uncorrelated with the instrumental variables. If we divide e_{jt} by the conditional variance of the market return, then the resulting quantity can be interpreted as the deviation of the observed return from the return predicted by the model. Thus e_{jt} is essentially just a pricing error. A negative pricing error implies the model is overpricing while a positive pricing error indicates that the model is underpricing.

The generalized method of moments (GMM), which is discussed in detail in Section 5, provides a direct way to test the above restriction. Suppose we have a total of n assets. We can stack the disturbances in (2) and the pricing errors in (5) into the $(2n + 1) \times 1$ vector:

$$\boldsymbol{\varepsilon}_t \equiv (\boldsymbol{u}_t \quad u_{mt} \quad \boldsymbol{e}_t)' = \begin{pmatrix} [\boldsymbol{r}_t - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}]' \\ [r_{mt} - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_m]' \\ [u_{mt}^2 \boldsymbol{Z}_{t-1}\boldsymbol{\delta} - u_{mt}\boldsymbol{u}_t \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_m]' \end{pmatrix} , \tag{6}$$

where $\boldsymbol{u}$ is the innovation in the $1 \times n$ vector of conditional means and $\boldsymbol{e}$ is the $1 \times n$ vector of pricing errors. The conditional CAPM implies that $\boldsymbol{\varepsilon}_t$ should be uncorrelated with $\boldsymbol{Z}_{t-1}$. So if we form the Kronecker product of $\boldsymbol{\varepsilon}_t$ with the vector of instrumental variables:

$$\boldsymbol{\varepsilon}_t \otimes \boldsymbol{Z}'_{t-1} , \tag{7}$$

and take unconditional expectations, we obtain the vector of orthogonality conditions:

$$\mathrm{E}[\boldsymbol{\varepsilon}_t \otimes \boldsymbol{Z}'_{t-1}] = 0 . \tag{8}$$

With n assets there are $n+1$ columns of innovations for the conditional means and n columns of pricing errors. Thus, with ℓ instrumental variables we have $\ell(2n+1)$ orthogonality conditions. Note, however, that there are $\ell(n+1)$ parameters to estimate. This leaves $n\ell$ overidentifying restrictions.[3]

We can obtain consistent estimates of the $n\ell$ matrix of coefficients $\boldsymbol{\delta}$ and the $\ell \times 1$ vector of coefficients $\boldsymbol{\delta}_m$ by minimizing the quadratic objective function:

$$J_T \equiv g'_T \boldsymbol{S}_T^{-1} g_T , \tag{9}$$

where:

$$g_T \equiv \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t \otimes \boldsymbol{Z}'_{t-1} , \tag{10}$$

and $\boldsymbol{S}_T$ denotes a consistent estimate of:

$$\boldsymbol{S}_0 \equiv \sum_{j=-\infty}^{\infty} \mathrm{E}[(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{Z}'_{t-1})(\boldsymbol{\varepsilon}_{t-j} \otimes \boldsymbol{Z}'_{t-j-1})'] . \tag{11}$$

If the conditional CAPM is true then T times the minimized value of the objective function converges to a central chi-square random variable with $n\ell$ degrees of freedom. Thus we can use this criterion as a measure of the overall goodness-of-fit of the model.

[3] An econometric specification of this form is explored for New York Stock Exchange returns in Harvey (1989) and Huang (1989), for 17 international equity returns in Harvey (1991), for international bond returns in Harvey, Solnik and Zhou (1995), and for emerging equity market returns in Harvey (1995).

D. Constant conditional betas

The econometric specification shown in (6) assumes that all of the conditional moments – the means, variances and covariances – change through time. If some of these moments are constant then we can construct more powerful tests of the conditional CAPM by imposing this additional structure. Traditionally, tests of the CAPM have focused on whether expected returns are proportional to the expected return on a benchmark portfolio. We can construct the same type of test within our conditional pricing framework with a specification of the form:

$$\varepsilon_t = (\boldsymbol{r}_t - r_{mt}\boldsymbol{\beta})' \ , \tag{12}$$

where $\boldsymbol{\beta}$ is a row vector of n beta coefficients. The coefficient β_j represents the ratio of conditional covariance between the return on portfolio j and the return on the benchmark to the conditional variance of the benchmark return.

Typically, we think of r_{mt} as a proxy for the market portfolio. It is important to note, however, that the beta coefficients in (12) are left unrestricted. Thus (12) can also be interpreted as a test of a single factor latent variables model.[4] In the latent variables framework, β_j represents the ratio of conditional covariance between the return on portfolio j and an unobserved factor to the conditional covariance between the return on the benchmark portfolio and this factor. The testable implication is that $\mathrm{E}[\varepsilon_t|\boldsymbol{Z}_{t-1}] = 0$ where ε_t is the vector of pricing errors associated with the constant conditional beta model. There are $n\ell$ orthogonality conditions and n parameters to estimate so we have $\ell(n-1)$ overidentifying restrictions.

Of course we can easily incorporate the restrictions on the conditional beta coefficients by changing the specification to:

$$\varepsilon_t = (\,\boldsymbol{u}_t \quad u_{mt} \quad \boldsymbol{b}_t \quad \boldsymbol{e}_t\,)' = \begin{pmatrix} [\boldsymbol{r}_t - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}]' \\ [r_{mt} - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_m]' \\ [u_{mt}^2\boldsymbol{\beta} - u_{mt}\boldsymbol{u}_t]' \\ [\boldsymbol{r}_t - r_{mt}\boldsymbol{\beta}]' \end{pmatrix} \ , \tag{13}$$

where $\boldsymbol{b}$ is the disturbance vector associated with the constant conditional beta assumption. Tests based on this specification may shed additional light on the plausibility of the assumption of constant conditional betas. With n assets there are $n+1$ columns of innovations in the conditional means, n columns in $\boldsymbol{b}$ and n columns in $\boldsymbol{e}$. Thus there are $\ell(3n+1)$ orthogonality conditions, $\ell(n+1)+n$ parameters to estimate, and $n(2\ell-1)$ overidentifying restrictions.

E. Constant conditional reward-to-risk ratio

Another formulation of the conditional CAPM assumes that the conditional reward-to-risk ratio is constant. The conditional reward-to-risk ratio,

[4] See, for example, Hansen and Hodrick (1983), Gibbons and Ferson (1985) and Ferson (1990).

$\mathrm{E}[r_{mt}|\boldsymbol{\Omega}_{t-1}]/\mathrm{Var}[r_{mt}|\boldsymbol{\Omega}_{t-1}]$, is simply the price of covariance risk. This version of the conditional CAPM is examined in Campbell (1987) and Harvey (1989). The vector of pricing errors for the model becomes:

$$\boldsymbol{e}_t = \boldsymbol{r}_t - \lambda \boldsymbol{u}_t u_{mt} \ , \tag{14}$$

where λ is the conditional expected return on the market divided by its conditional variance. To complete the econometric specification we have to include models for the conditional means. The overall system is:

$$\boldsymbol{\varepsilon}_t = (\, \boldsymbol{u}_t \quad u_{mt} \quad \boldsymbol{e}_t\,)' = \begin{pmatrix} [\boldsymbol{r}_t - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}]' \\ [r_{mt} - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_m]' \\ [\boldsymbol{r}_t - \lambda(u_{mt}\boldsymbol{u}_t)]' \end{pmatrix} . \tag{15}$$

With n assets there are $n+1$ columns of innovations in the conditional means and n columns in $\boldsymbol{e}$. Thus with ℓ instrumental variables there are $\ell(2n+1)$ orthogonality conditions and $1+(\ell(n+1))$ parameters. This leaves $n\ell - 1$ overidentifying restrictions.

One way to simplify the estimation in (15) is to note that $\mathrm{E}[u_{mt}u_{jt}|\boldsymbol{Z}_{t-1}] = \mathrm{E}[u_{mt}r_{jt}|\boldsymbol{Z}_{t-1}]$. This follows from the fact that:

$$\begin{aligned} \mathrm{E}[u_{mt}u_{jt}|\boldsymbol{Z}_{t-1}] &= \mathrm{E}[u_{mt}(r_{jt} - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_j)|\boldsymbol{Z}_{t-1}] \\ &= \mathrm{E}[u_{mt}r_{jt}|\boldsymbol{Z}_{t-1}] - \mathrm{E}[u_{mt}\boldsymbol{Z}_{t-1}\boldsymbol{\delta}_j|\boldsymbol{Z}_{t-1}] \\ &= \mathrm{E}[u_{mt}r_{jt}|\boldsymbol{Z}_{t-1}] - \mathrm{E}[u_{mt}|\boldsymbol{Z}_{t-1}]\boldsymbol{Z}_{t-1}\boldsymbol{\delta}_j \\ &= \mathrm{E}[u_{mt}r_{jt}|\boldsymbol{Z}_{t-1}] \ . \end{aligned}$$

As a result, we can drop n of the conditional mean equations. The more parsimonious system is:

$$\boldsymbol{\varepsilon}_t = (\, u_{mt} \quad \boldsymbol{e}_t\,) = \begin{pmatrix} [r_{mt} - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_m]' \\ [\boldsymbol{r}_t - \lambda(u_{mt}\boldsymbol{r}_t)]' \end{pmatrix}' . \tag{16}$$

Now we have $n+1$ equations and $\ell(n+1)$ orthogonality conditions. With $\ell+1$ parameters there are $(n\ell)-1$ overidentifying restrictions. The specifications shown in (15) and (16) are asymptotically equivalent. But (16) is more computationally manageable.

The specifications in (15) and (16) do not restrict λ to be the conditional covariance to variance ratio. We can easily add this restriction:

$$\boldsymbol{\varepsilon}_t = (\, \boldsymbol{u}_t \quad u_{mt} \quad m_t \quad \boldsymbol{e}_t\,)' = \begin{pmatrix} [\boldsymbol{r}_t - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}]' \\ [r_{mt} - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_m]' \\ [u_{mt}^2\lambda - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_m]' \\ [\boldsymbol{r}_t - \lambda(u_{mt}\boldsymbol{r}_t)]' \end{pmatrix} , \tag{17}$$

where m is the disturbance associated with the constant reward-to-risk assumption. Tests of this specification should shed additional light on the plausibility of the assumption of a constant price of covariance risk. With n assets there are n columns in $\boldsymbol{u}$, one column in u_m, one column in m and n columns in $\boldsymbol{e}$. Thus there are $\ell(2n+2)$ orthogonality conditions, $\ell(n+1)+1$ parameters, and $n-1$ over-identifying restrictions.

F. Linear conditional betas

Ferson and Harvey (1994, 1995) explore specifications where the conditional betas are modelled as a linear functions of the instrumental variables. We could, for example, specify an econometric system of the form:

$$
\begin{aligned}
u_{1it} &= r_{it} - \boldsymbol{Z}^{i,w}_{t-1}\boldsymbol{\delta}_i \\
u_{2t} &= r_{mt} - \boldsymbol{Z}^{w}_{t-1}\boldsymbol{\delta}_m \\
u_{3it} &= [u^2_{2t}(\boldsymbol{Z}^{i,w}_{t-1}\boldsymbol{\kappa}_i)' - r_{mt}u_{1it}]' \\
u_{4it} &= \mu_i - \boldsymbol{Z}^{i,w}_{t-1}\boldsymbol{\delta}_i \\
u_{5it} &= (-\alpha_i + \mu_i) - \boldsymbol{Z}^{i,w}_{t-1}\boldsymbol{\kappa}_i(\boldsymbol{Z}^{w}_{t-1}\boldsymbol{\delta}_m)'
\end{aligned}
\tag{18}
$$

where the elements of $\boldsymbol{Z}^{i,w}\boldsymbol{\kappa}_i$ are the fitted conditional betas for portfolio i, μ_i is the mean return on portfolio i, and α_i is the difference between the unrestricted mean return and the mean return that incorporates the pricing restriction of the conditional CAPM. Note that (18) uses two sets of instruments. The set used to estimate the conditional mean return on portfolio i and the conditional beta for the portfolio, $\boldsymbol{Z}^{i,w}$, includes both asset specific (i) and market-wide (w) instruments. The conditional mean return on the market is estimated using only the market-wide instruments. This yields an exactly identified system of equations.[5]

The intuition behind the system shown in (18) is straightforward. The first two equations follow from our assumption of linear conditional expectations. They represent statistical models for expected returns. The third equation follows from the definition of the conditional beta:

$$
\beta_{it} = (\mathrm{E}[u^2_{2t}|\boldsymbol{Z}^{w}_{t-1}])^{-1}\mathrm{E}[r_{mt}u_{1it}|\boldsymbol{Z}^{i,w}_{t-1}] \ . \tag{19}
$$

In (18) the conditional beta is modelled as a linear function of both the asset-specific and market-wide information. The last two equations deliver the average pricing error for the conditional CAPM. Note that μ_i is the average fitted return from the statistical model. Thus α_i is the difference between the average fitted return from our statistical model and the fitted return implied by the pricing relation of conditional CAPM. It is analogous to the Jensen α. In the current analysis, however, both the betas and the risk premiums are changing through time.

Because of the complexity and size of the above system it is difficult to estimate from more one asset at a time. Thus, in general, not all the cross-sectional restrictions of conditional CAPM can be imposed, and it is not possible to report a multivariate test of whether the α_i are equal to zero. Note, however, that (18)

[5] For analysis of related systems see Ferson (1990), Shanken (1990), Ferson and Harvey (1991), Ferson and Harvey (1993), Ferson and Korajzcyk (1995), Ferson (1995), Harvey (1995) and Jagannathan and Wang (1996).

does impose one important cross-sectional restriction. Because the system is exactly identified, the market risk premium, $Z^{w}_{t-1}\delta_m$, will be identical for every asset examined. There are no overidentifying restrictions, so tests of the model are based on whether the coefficient α_i is significantly different from zero. Additional insights might be gained by analyzing the time-series properties of the disturbance:

$$u_{6it} = r_{it} - Z^{i,w}_{t-1}\kappa_i(Z^{w}_{t-1}\theta)' \ . \tag{20}$$

Under the null hypothesis, $\mathrm{E}[u_{6it}|Z^{i,w}_{t-1}]$ is equal to zero. Thus diagnostics can be conducted by regressing u_{6it} on various information variables. We could also construct tests for time-varying of betas based on the coefficient estimates associated with $Z^{i,w}\kappa_i$.

3. Models with multiple betas

A. The multi-beta conditional CAPM

The conditional CAPM can easily be generalized to a model that has multiple sources of risk. Consider, for example, a k-factor pricing relation of the form:

$$\mathrm{E}[r_t|Z_{t-1}] = \mathrm{E}[f_t|Z_{t-1}]\left(\mathrm{E}[u'_{ft}u_{ft}|Z_{t-1}]\right)^{-1}\mathrm{E}[u'_{ft}u_t|Z_{t-1}] \tag{21}$$

where r is a row vector of n asset returns, f is $1 \times K$ vector of factor realizations, u_f is a vector of innovations in the conditional means of the factors, and u is a vector of innovations in the conditional means of the returns. The first term on the right-hand side of (21) represents the conditional expectation of the factor realizations. It has dimension $1 \times k$. The second term is the inverse of the $k \times k$ conditional variance-covariance matrix of the factors. The final term measures the conditional covariance of the asset returns with the factors. Its dimension is $k \times n$.

The multi-beta pricing relation shown in (21) cannot be tested in the same manner as its single-beta counterpart. Recall that in our analysis of single-beta models it was possible to take the conditional variance of the market return to the left-hand side of the pricing relation. As a result, we could move the conditional means inside the expectations operator. This is not possible with a multi-beta specification. We can, however, get around this problem by focusing on specializations of the multi-beta model that parallel those discussed in the previous section. We begin by considering specifications that restrict the conditional betas to be linear functions of the instruments.

B. Linear conditional betas

The multi-beta analogue of the linear conditional beta specification shown in (18) takes the form:

$$\begin{aligned} u_{1it} &= r_{it} - \boldsymbol{Z}^{i,w}_{t-1}\boldsymbol{\delta}_i \\ \boldsymbol{u}_{2t} &= \boldsymbol{f}_t - \boldsymbol{Z}^{w}_{t-1}\boldsymbol{\delta}_f \\ \boldsymbol{u}_{3it} &= [\boldsymbol{u}'_{2t}\boldsymbol{u}_{2t}(\boldsymbol{Z}^{i,w}_{t-1}\boldsymbol{\kappa}_i)' - \boldsymbol{f}'_t u_{1it}]' \\ u_{4it} &= \mu_i - \boldsymbol{Z}^{i,w}_{t-1}\boldsymbol{\delta}_i \\ u_{5it} &= (-\alpha_i + \mu_i) - \boldsymbol{Z}^{i,w}_{t-1}\boldsymbol{\kappa}_i(\boldsymbol{Z}^{w}_{t-1}\boldsymbol{\delta}_f)' \end{aligned} \tag{22}$$

where the elements of $\boldsymbol{Z}^{i,w}\boldsymbol{\kappa}_i$ are the fitted conditional betas associated with the k sources of risk and $\boldsymbol{f}$ is a row vector of factor realizations. Note that as before the system is exactly identified, and the vector of conditional betas:

$$\boldsymbol{\beta}_{it} = (\mathrm{E}[\boldsymbol{u}'_{2t}\boldsymbol{u}_{2t}|\boldsymbol{Z}^{i,w}_{t-1}])^{-1}\mathrm{E}[\boldsymbol{f}'_t u_{1it}|\boldsymbol{Z}^{i,w}_{t-1}] \ . \tag{23}$$

is modelled as a linear function, $\boldsymbol{Z}^{i,w}\boldsymbol{\kappa}_i$, of the instruments. This specification can be tested by assessing the statistical significance of the pricing errors and checking to see whether the disturbance:

$$u_{6it} = r_{it} - \boldsymbol{Z}^{i,w}_{t-1}\boldsymbol{\kappa}_i(\boldsymbol{Z}^{w}_{t-1}\boldsymbol{\delta}_f)' \ , \tag{24}$$

is orthogonal to instruments. The primary advantage of the above formulation is that fitted values are obtained for the risk premiums, the expected returns, and the conditional betas. Thus it is simple to conduct diagnostics that focus on the performance of the model. Its main disadvantage is that it requires a heavy parameterization.

C. Constant conditional reward-to-risk ratios

Harvey (1989) suggests an alternative approach for testing multi-beta pricing relations. His strategy is to assume that the conditional reward-to-risk ratio is constant for each factor. This results is a multi-beta analogue of the specification shown in (15):

$$\boldsymbol{\varepsilon}_t = (\boldsymbol{u}_t \quad \boldsymbol{u}_{ft} \quad \boldsymbol{e}_t)' = \begin{pmatrix} [\boldsymbol{r}_t - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}]' \\ [\boldsymbol{f}_t - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_f]' \\ [\boldsymbol{r}_t - \boldsymbol{\lambda}(\boldsymbol{u}'_{ft}\boldsymbol{u}_t)]' \end{pmatrix} \ , \tag{25}$$

where $\boldsymbol{\lambda}$ is a row vector of k time-invariant reward-to-risk measures. The above system can be simplified to:

$$\boldsymbol{\varepsilon}_t = (\boldsymbol{u}_{ft} \quad \boldsymbol{e}_t)' = \begin{pmatrix} [\boldsymbol{f}_t - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_f]' \\ [\boldsymbol{r}_t - \boldsymbol{\lambda}(\boldsymbol{u}'_{ft}\boldsymbol{r}_t)]' \end{pmatrix} \ , \tag{26}$$

using the same approach that allowed us to simplify the single-beta specification discussed earlier.[6]

[6] Kan and Zhang (1995) generalize this formulation by modelling the conditional reward-to-risk ratios as linear functions of the instrumental variables. Their approach eliminates the need for asset-specific instruments and permits joint estimation of the pricing relation using multiple portfolios. But the type of diagnostics that fall out of the linear conditional beta model – fitted expected returns, betas, etc. – are no longer available.

4. Latent variables models

The latent variables technique introduced by Hansen and Hodrick (1983) and Gibbons and Ferson (1985) provides a rank restriction on the coefficients of the linear specifications that are assumed to describe expected returns. Suppose we assume that ratio formed by taking the conditional beta for one asset and dividing it by the corresponding conditional beta another asset is constant. Under these circumstances, the k-factor conditional beta pricing model implies that all of the variation in the expected returns is driven by changes in the k conditional risk premiums. We can still form our estimates of the conditional means by projecting returns on the ℓ-dimensional vector of instrumental variables. But if all the variation in expected returns is being driven changes in the k risk premiums then we should not need all $n\ell$ projection coefficients to characterize the time variation in the n returns. Thus the basic idea of the latent variables technique is to test restrictions on the rank of the projection coefficient matrix.

A. Constant conditional beta ratios

First we take the vector of excess returns on our set of portfolios and partition it as:

$$\boldsymbol{r}_t = (\boldsymbol{r}_{1t} \ \vdots \ \boldsymbol{r}_{2t}), \tag{34}$$

where $\boldsymbol{r}_{1t}$ is a $1 \times k$ vector of returns on the *reference assets* and $\boldsymbol{r}_{2t}$ is a $1 \times (n-k)$ vector of returns on the *test assets*. Then we partition the matrix of conditional beta coefficients associated with our multi-factor pricing model accordingly:

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1 \ \vdots \ \boldsymbol{\beta}_2), \tag{35}$$

where $\boldsymbol{\beta}_1$ is $k \times k$ and $\boldsymbol{\beta}_2$ is $k \times (n-k)$. The pricing relation for the multi-beta model tells us that:

$$\mathrm{E}[\boldsymbol{r}_{1t}|\boldsymbol{Z}_{t-1}] = \gamma_t \boldsymbol{\beta}_1 \tag{36}$$

and

$$\mathrm{E}[\boldsymbol{r}_{2t}|\boldsymbol{Z}_{t-1}] = \gamma_t \boldsymbol{\beta}_2 \ , \tag{37}$$

where γ_t is a $1 \times k$ vector of time-varying market-wide risk premiums. We can manipulate (36) to obtain the relation $\gamma_t = \mathrm{E}[\boldsymbol{r}_{1t}|\boldsymbol{Z}_{t-1}]\boldsymbol{\beta}_1^{-1}$. Substituting this expression for γ_t into (37) yields the pricing restriction:

$$\mathrm{E}[\boldsymbol{r}_{2t}|\boldsymbol{Z}_{t-1}] = \mathrm{E}[\boldsymbol{r}_{1t}|\boldsymbol{Z}_{t-1}]\boldsymbol{\beta}_1^{-1}\boldsymbol{\beta}_2 \ . \tag{38}$$

This says that the conditional expected returns on the test assets are proportional to the conditional expected returns on the reference assets. The constants of proportionality are determined by ratios of conditional betas.

The pricing relation in (38) can be tested in much the same manner as the models discussed earlier.[7] The only real difference is that we no longer have to identify the factors. One possible specification is:

$$\varepsilon_t = (\boldsymbol{u}_{1t} \quad \boldsymbol{u}_{2t} \quad \boldsymbol{e}_t)' = \begin{pmatrix} [\boldsymbol{r}_{1t} - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_1]' \\ [\boldsymbol{r}_{2t} - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_2]' \\ [\boldsymbol{Z}_{t-1}\boldsymbol{\delta}_2 - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_1\boldsymbol{\Phi}]' \end{pmatrix} \tag{39}$$

where $\boldsymbol{\Phi} \equiv \boldsymbol{\beta}_1^{-1}\boldsymbol{\beta}_2$. There are k columns in $\boldsymbol{u}_{1t}$, $n-K$ columns in $\boldsymbol{u}_{2t}$ and $n-k$ columns in $\boldsymbol{e}_t$. Thus we have $\ell(2n-k)$ orthogonality conditions and $\ell n + k(n-k)$ parameters. This leaves $(\ell - k)(n-k)$ overidentifying restrictions. Note that both the number of instrumental variables and the total number of assets must be greater than the number of factors.

B. Linear conditional covariance ratios

An important disadvantage of (39) is that the ratio of conditional betas, $\boldsymbol{\Phi} = \boldsymbol{\beta}_1^{-1}\boldsymbol{\beta}_2$, is assumed to be constant. One way to generalize the latent variables model is to assume the elements of $\boldsymbol{\Phi}$ are linear in the instrumental variables.[8] This assumption follows naturally from the previous specifications that imposed the assumption of linear conditional betas. The resulting latent variables system is:

$$\boldsymbol{\eta}_t = (\boldsymbol{u}_{1t} \quad \boldsymbol{u}_{2t} \quad \boldsymbol{\ell}_t)' = \begin{pmatrix} [\boldsymbol{r}_{1t} - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_1]' \\ [\boldsymbol{r}_{2t} - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_2]' \\ [\boldsymbol{Z}_{t-1}\boldsymbol{\delta}_2 - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_1(\boldsymbol{\iota} \otimes \boldsymbol{Z}^*_{t-1})\boldsymbol{\Phi}^*]' \end{pmatrix}, \tag{40}$$

where $\boldsymbol{\iota}$ is a $k \times 1$ vector of ones. With the original set of instruments the dimension of $\boldsymbol{\Phi}^*$ in the final set of moment conditions is $\ell(n-k)$ and the system is not identified. Thus the researcher must specify some subset of the original instruments, $\boldsymbol{Z}^*$, with dimension $\ell^* < \ell$ to be used in the estimation.

Finally, the parameterization in both (39) and (40) can be reduced by substituting the third equation block into the second block. For example,

$$\varepsilon_t = (\boldsymbol{u}_{1t} \quad \boldsymbol{e}_t)' = \begin{pmatrix} [\boldsymbol{r}_{1t} - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_1]' \\ [\boldsymbol{r}_{2t} - \boldsymbol{Z}_{t-1}\boldsymbol{\delta}_1\boldsymbol{\Phi}]' \end{pmatrix}, \tag{41}$$

In this system, it is not necessary to estimate $\boldsymbol{\delta}_2$.

5. Generalized method of moments estimation

Contemporary empirical research in financial economics makes frequent use of a wide variety of econometric techniques. The generalized method of moments has proven to be particularly valuable, however, especially in the area of estimating and testing asset pricing models. This section provides an overview of the gen-

[7] Harvey, Solnik and Zhou (1995) and Zhou (1995) show to construct analytic tests of latent variables models.

[8] See Ferson and Foerster (1994).

eralized method of moments (GMM) procedure. We begin by illustrating the intuition behind GMM using a simple example of classical method of moments estimation. This is followed by brief discussion of the assumptions underlying the GMM approach to estimation and testing along with a review of some of the key distributional results. For detailed proofs of the consistency and asymptotic normality of GMM estimators see Hansen (1982), Gallant and White (1988), and Potscher and Prucha (1991a,b).

A. The Classical method of moments

The easiest way to illustrate the intuition behind the GMM procedure is to consider a simple example of classical method of moments (CMM) estimation. Suppose we observe a random sample $x_1, x_2, \ldots, x_T$ of T observations drawn from a distribution with probability density function $f(x;\boldsymbol{\theta})$, where $\boldsymbol{\theta} \equiv [\theta_1, \theta_2, \ldots, \theta_k]$ denotes a $k \times 1$ vector of unknown parameters. The CMM approach to estimation exploits the fact that in general the j^{th} population moment of x about zero:

$$m_j \equiv \mathrm{E}[x^j] \ , \tag{42}$$

can be written as known function of $\boldsymbol{\theta}$. To implement the CMM procedure we first compute the j^{th} sample moment of x about zero:

$$\hat{m}_j = \frac{1}{T}\sum_{i=1}^{T} x_i^j \ . \tag{43}$$

Then we set the j^{th} sample moment equal to the corresponding population moment for $j = 1, 2, \ldots, k$:

$$\begin{array}{ccc} \hat{m}_1 & = & m_1(\boldsymbol{\theta}) \\ \hat{m}_2 & = & m_2(\boldsymbol{\theta}) \\ \vdots & \vdots & \vdots \\ \hat{m}_k & = & m_k(\boldsymbol{\theta}) \end{array} \tag{44}$$

This yields a set of k equations in k unknowns that can be solved to obtain an estimator for the unknown vector $\boldsymbol{\theta}$. Thus the basic idea behind the CMM procedure is to estimate $\boldsymbol{\theta}$ by replacing population moments with their sample analogues.

Now let's take a more concrete version of the above example. Suppose that $x_1, x_2, \ldots, x_T$ is a random sample of size T drawn from a normal distribution with mean μ and variance σ^2. To obtain the classical method of moments estimators of μ and σ^2 we note that $\sigma^2 = m_2 - (m_1)^2$. This implies that the system of moments equations takes the form:

$$\begin{aligned} &\frac{1}{T}\sum_{i=1}^{T} x_i = \mu \\ &\frac{1}{T}\sum_{i=1}^{T} x_i^2 = \sigma^2 + \mu^2 \ . \end{aligned} \tag{45}$$

Consequently, the CMM estimators for the mean and variance are:

$$\hat{\mu} = \frac{1}{T}\sum_{i=1}^{N} x_i$$
$$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{T} x_i^2 - \left(\frac{1}{T}\sum_{i=1}^{T} x_i\right)^2 \tag{46}$$

Notice that these are also the maximum likelihood estimators of μ and σ^2.

B. The Generalized method of moments

The classical method of moments is just a special case of the generalized method of moments developed by Hansen (1982). This latter procedure provides a general framework for estimation and hypothesis testing that can be used to analyze a wide variety of dynamic economic models. Consider, for example, the class of models that generate conditional moment restrictions of the form:

$$\mathrm{E}_t[\boldsymbol{u}_{t+\tau}] = \mathbf{0} \ , \tag{47}$$

where $\mathrm{E}_t[\cdot]$ is the expectations operator conditional on the information set at time t, $\boldsymbol{u}_{t+\tau} \equiv \boldsymbol{h}(\boldsymbol{X}_{t+\tau}, \boldsymbol{\theta}_0)$ is an $n \times 1$ vector of vector of disturbance terms, $\boldsymbol{X}_{t+\tau}$ is an $s \times 1$ vector of observable random variables, and $\boldsymbol{\theta}_0$ is an $m \times 1$ vector of unknown parameters. The basic idea behind the GMM procedure is to exploit the moment restrictions in (47) to construct a sample objective function whose minimizer is a consistent and asymptotically normal estimate of the unknown vector $\boldsymbol{\theta}_0$.

In order to construct such an objective function, however, we need to make some assumptions about the nature of the data generating process. Let $\boldsymbol{Z}_t$ denote the date t realization of an $\ell \times 1$ vector of observable instrumental variables. We assume, following Hansen (1982), that the vector process $\{\boldsymbol{X}_t, \boldsymbol{Z}_t\}_{t=-\infty}^{\infty}$ is strictly stationary and ergodic. Note that this assumption rules out a number of features sometimes encountered in economic data such as deterministic trends, unit roots, and unconditional heteroskedasticity. It accommodates many common forms of conditional heterogeneity, however, and it does not appear to be overly restrictive in most applications.[9]

With suitable restrictions on the data generating process in place we can proceed to construct the GMM objective function. First we form the Kronecker product:

$$f(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta}_0) \equiv \boldsymbol{u}_{t+\tau} \otimes \boldsymbol{Z}_t \ . \tag{48}$$

Then we note that because $\boldsymbol{Z}_t$ is in the information set at time t, the model in (47) implies that:

[9] Although is possible to establish consistency and asymptotic normality of GMM estimators under weaker assumptions, the associated arguments are too complex for an introductory discussion. The interested reader can consult Potscher and Prucha (1991a,b) for an overview of recent advances in he asymptotic theory of dynamic nonlinear econometric models.

$$\mathrm{E}_t[\boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta}_0)] = \boldsymbol{0} \ . \tag{49}$$

Applying the law of iterated expectations to equation (49) yields the unconditional restriction:

$$\mathrm{E}[\boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta}_0)] = \boldsymbol{0} \ . \tag{50}$$

Equation (50) represents a set of $n\ell$ population orthogonality conditions. The sample analogue of $\mathrm{E}[\boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta})]$:

$$\boldsymbol{g}_T(\boldsymbol{\theta}) \equiv \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta}) \ , \tag{51}$$

forms the basis for the GMM objective function. Note that for any given value of $\boldsymbol{\theta}$ the vector $\boldsymbol{g}_T(\boldsymbol{\theta})$ is just the sample mean of T realizations of the random vector $\boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta})$. Given that $\boldsymbol{f}(\cdot)$ is continuous and $\{\boldsymbol{X}_t, \boldsymbol{Z}_t\}_{t=-\infty}^{\infty}$ is strictly stationary and ergodic we have:

$$\boldsymbol{g}_T(\boldsymbol{\theta}) \xrightarrow{p} \mathrm{E}[\boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta})] \tag{52}$$

by the law of large numbers. Thus if the economic model is valid the vector $\boldsymbol{g}_T(\boldsymbol{\theta}_0)$ should be close to zero when evaluated for a large number of observations. The GMM estimator of $\boldsymbol{\theta}_0$ is obtained by choosing the value of $\boldsymbol{\theta}$ that minimizes the overall deviation of $\boldsymbol{g}_T(\boldsymbol{\theta})$ from zero. As long as $\mathrm{E}[\boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta})]$ is continuous in $\boldsymbol{\theta}$ it follows that this estimator is consistent under fairly general regularity conditions.

If the model is exactly identified ($m = n\ell$), the GMM estimator is the value of $\boldsymbol{\theta}$ that sets the sample moments equal to zero. For the more common situation where the model is overidentified ($m < n\ell$), finding a vector of parameters that sets all of the sample moments equal to zero is not feasible. It is possible, however, to find a value of $\boldsymbol{\theta}$ that sets m linear combinations of the $n\ell$ sample moment conditions equal to zero. We simply let $\boldsymbol{A}_T$ be an $m \times n\ell$ matrix such that $\boldsymbol{A}_T\boldsymbol{g}_T(\boldsymbol{\theta}) = \boldsymbol{0}$ has a well-defined solution. The value of $\boldsymbol{\theta}$ that solves this system of equations is the GMM estimator. Although we have considerable leeway in choosing the weighting matrix $\boldsymbol{A}_T$, Hansen (1982) shows that the variance-covariance matrix of the estimator is minimized by letting $\boldsymbol{A}_T$ equal $\boldsymbol{D}_T'\boldsymbol{S}_T^{-1}$ where $\boldsymbol{D}_T$ and $\boldsymbol{S}_T$ are consistent estimates of:

$$\boldsymbol{D}_0 \equiv \mathrm{E}\left[\left.\frac{\partial \boldsymbol{h}(\boldsymbol{X}_{t+\tau}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_0} \otimes \boldsymbol{Z}_t\right] \quad \text{and} \quad \boldsymbol{S}_0 \equiv \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_0(j) \ , \tag{53}$$

with $\boldsymbol{\Gamma}_0(j) \equiv \mathrm{E}[\boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta}_0)\boldsymbol{f}(\boldsymbol{X}_{t+\tau-j}, \boldsymbol{Z}_{t-j}, \boldsymbol{\theta}_0)']$. Before considering how to derive this result we first have to establish the asymptotic normality of GMM estimators.

C. Asymptotic normality of GMM estimators

We begin by expressing equation (51) as:

$$\sqrt{T}\boldsymbol{g}_T(\boldsymbol{\theta}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta}) \ . \tag{54}$$

The assumption that $\{\boldsymbol{X}_t, \boldsymbol{Z}_t\}_{t=-\infty}^{\infty}$ is stationary and ergodic, along with standard regularity conditions, implies that a version of the central limit theorem holds. In particular we have that:

$$\sqrt{T}\boldsymbol{g}_T(\boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{S}_0) \ , \tag{55}$$

with $\boldsymbol{S}_0$ given by (53). This result allows us to establish the limiting distribution of the GMM estimator $\boldsymbol{\theta}_T$. First we make the following assumptions:

1. The estimator $\boldsymbol{\theta}_T$ converges in probability to $\boldsymbol{\theta}_0$.
2. The weighting matrix $\boldsymbol{A}_T$ converges in probability to $\boldsymbol{A}_0$ where $\boldsymbol{A}_0$ has rank m.
3. Define:

$$\boldsymbol{D}_T \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\left.\frac{\partial \boldsymbol{h}(\boldsymbol{X}_{t+\tau}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_T} \otimes \boldsymbol{Z}_t \right) \ . \tag{56}$$

For any $\boldsymbol{\theta}_T$ such that $\boldsymbol{\theta}_T \xrightarrow{p} \boldsymbol{\theta}_0$ the matrix $\boldsymbol{D}_T$ converges in probability to $\boldsymbol{D}_0$ where $\boldsymbol{D}_0$ has rank m.
Then we apply the mean value theorem to obtain:

$$\boldsymbol{g}_T(\boldsymbol{\theta}_T) = \boldsymbol{g}_T(\boldsymbol{\theta}_0) + \boldsymbol{D}_T^*(\boldsymbol{\theta}_T - \boldsymbol{\theta}_0) \ , \tag{57}$$

where $\boldsymbol{D}_T^*$ is given by (56) with $\boldsymbol{\theta}_T$ replaced by a vector $\boldsymbol{\theta}_T^*$ that lies somewhere within the interval whose endpoints are given by $\boldsymbol{\theta}_T$ and $\boldsymbol{\theta}_0$. Recall that $\boldsymbol{\theta}_T$ is the solution to the system of equations $\boldsymbol{A}_T\boldsymbol{g}_T(\boldsymbol{\theta}) = 0$. So if we premultiply equation (57) by $\boldsymbol{A}_T$ we have:

$$\boldsymbol{A}_T\boldsymbol{g}_T(\boldsymbol{\theta}_0) + \boldsymbol{A}_T\boldsymbol{D}_T^*(\boldsymbol{\theta}_T - \boldsymbol{\theta}_0) = \boldsymbol{0} \ . \tag{58}$$

Solving (58) for $(\boldsymbol{\theta}_T - \boldsymbol{\theta}_0)$ and multiplying by $\sqrt{T}$ gives:

$$\sqrt{T}(\boldsymbol{\theta}_T - \boldsymbol{\theta}_0) = -[\boldsymbol{A}_T\boldsymbol{D}_T^*]^{-1}\boldsymbol{A}_T\sqrt{T}\boldsymbol{g}_T(\boldsymbol{\theta}_0) \ , \tag{59}$$

and by Slutsky's theorem we have:

$$\begin{aligned}\sqrt{T}(\boldsymbol{\theta}_T - \boldsymbol{\theta}_0) &\xrightarrow{d} -[\boldsymbol{A}_0\boldsymbol{D}_0]^{-1}\boldsymbol{A}_0 \\ &\times \{\text{the limiting distribution of } \sqrt{T}\boldsymbol{g}_T(\boldsymbol{\theta}_0)\} \ .\end{aligned} \tag{60}$$

Thus the limiting distribution of the GMM estimator is:

$$\sqrt{T}(\boldsymbol{\theta}_T - \boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0}, (\boldsymbol{A}_0\boldsymbol{D}_0)^{-1}\boldsymbol{A}_0\boldsymbol{S}_0\boldsymbol{A}_0'(\boldsymbol{A}_0\boldsymbol{D}_0)^{-1\prime}) \ . \tag{61}$$

Now that we know the limiting distribution of the generic GMM estimator we can determine the best choice for the weighting matrix $\boldsymbol{A}_T$. The natural metric by which to measure our choice is the variance-covariance matrix of the distribution shown in (61). We want, in other words, to choose the A_T that minimizes the variance-covariance matrix of the limiting distribution of the GMM estimator.

D. The asymptotically efficient weighting matrix

The first step in determining the efficient weighting matrix is to note that $\boldsymbol{S}_0$ is symmetric and positive definite. Thus $\boldsymbol{S}_0$ can be written as $\boldsymbol{S}_0 = \boldsymbol{P}\boldsymbol{P}'$ where $\boldsymbol{P}$ is nonsingular, and we can express the variance-covariance matrix in (61) as:

$$\begin{aligned} \boldsymbol{V} &\equiv (\boldsymbol{A}_0\boldsymbol{D}_0)^{-1}\boldsymbol{A}_0\boldsymbol{S}_0\boldsymbol{A}_0'(\boldsymbol{A}_0\boldsymbol{D}_0)^{-1\prime} \\ &= (\boldsymbol{A}_0\boldsymbol{D}_0)^{-1}\boldsymbol{A}_0\boldsymbol{P}((\boldsymbol{A}_0\boldsymbol{D}_0)^{-1}\boldsymbol{A}_0\boldsymbol{P})' \\ &= (\boldsymbol{H} + (\boldsymbol{D}_0'\boldsymbol{S}_0^{-1}\boldsymbol{D}_0)^{-1}\boldsymbol{D}_0'(\boldsymbol{P}')^{-1})(\boldsymbol{H} + (\boldsymbol{D}_0'\boldsymbol{S}_0^{-1}\boldsymbol{D}_0)^{-1}\boldsymbol{D}_0'(\boldsymbol{P}')^{-1})' \end{aligned} \tag{62}$$

where:

$$\boldsymbol{H} \equiv (\boldsymbol{A}_0\boldsymbol{D}_0)^{-1}\boldsymbol{A}_0\boldsymbol{P} - (\boldsymbol{D}_0'\boldsymbol{S}_0^{-1}\boldsymbol{D}_0)^{-1}\boldsymbol{D}_0'(\boldsymbol{P}')^{-1} \ .$$

At first it may appear a bit odd to define $\boldsymbol{H}$ in this manner, but it simplifies the problem of finding the efficient choice for A_T. To see why this is true note that:

$$\begin{aligned} \boldsymbol{H}\boldsymbol{P}^{-1}\boldsymbol{D}_0 &= (\boldsymbol{A}_0\boldsymbol{D}_0)^{-1}\boldsymbol{A}_0\boldsymbol{P}\boldsymbol{P}^{-1}\boldsymbol{D}_0 - (\boldsymbol{D}_0'\boldsymbol{S}_0^{-1}\boldsymbol{D}_0)^{-1}\boldsymbol{D}_0'(\boldsymbol{P}')^{-1}\boldsymbol{P}^{-1}\boldsymbol{D}_0 \\ &= \boldsymbol{I} - \boldsymbol{I} \\ &= \boldsymbol{0} \end{aligned} \tag{63}$$

As a consequence equation (62) reduces to:

$$\boldsymbol{V} = \boldsymbol{H}\boldsymbol{H}' + (\boldsymbol{D}_0'\boldsymbol{S}_0^{-1}\boldsymbol{D}_0)^{-1} \tag{64}$$

Because $\boldsymbol{H}$ is an $m \times n\ell$ matrix with rank m it follows that $\boldsymbol{H}\boldsymbol{H}'$ is positive definite. Thus $(\boldsymbol{D}_0'\boldsymbol{S}_0^{-1}\boldsymbol{D}_0)^{-1}$ is the lower bound on the asymptotic variance-covariance matrix of the GMM estimator. It is easily verified by direct substitution that choosing $\boldsymbol{A}_0 = \boldsymbol{D}_0'\boldsymbol{S}_0^{-1}$ achieves this lower bound.

This completes our review of the distribution theory for GMM estimators. Next we want to consider some of the practical aspects of GMM estimation and see how we might go about testing the restrictions implied economic models. We begin with a strategy for implementing the GMM procedure.

E. The estimation procedure

To obtain an estimate for the vector of unknown parameters $\boldsymbol{\theta}_0$ we have to solve the system of equations:

$$\boldsymbol{A}_T\boldsymbol{g}_T(\boldsymbol{\theta}) = \boldsymbol{0} \ .$$

Substituting the optimal choice for the weighting matrix into this expression yields:

$$D_T' S_T^{-1} g_T(\theta) = 0 \ , \tag{65}$$

where S_T is a consistent estimate of the matrix S_0. But it is apparent that (65) is just the first-order condition for the problem:

$$\min_{\theta} \quad J_T(\theta) \equiv g_T(\theta)' S_T^{-1} g_T(\theta) \ . \tag{66}$$

So given a consistent estimate of S_0 we can obtain the GMM estimator for θ_0 by minimizing the quadratic form shown in equation (66).

In order to estimate θ_0 we need a consistent estimate of S_0. But, in general, S_0 is a function of θ_0. The solution to this dilemma is to perform a two-step estimation procedure. Initially we set S_T equal to the identify matrix and perform the minimization to get a first-stage estimate for θ_0. Although this estimate is not asymptotically efficient it is still consistent. Thus we can use it to construct a consistent estimate of S_0. Once we have a consistent estimate of S_0 we obtain the second-stage estimate for θ_0 by minimizing the quadratic form shown above.

Let's assume that we have performed the two-step estimation procedure and obtained the efficient GMM estimate of the vector of parameters θ_0. Typically we would like to have some way of evaluating how well the model fits the observed data. One way of obtaining such a goodness-of-fit measure is to construct a test of the overidentifying restrictions.

F. The test for overidentifying restrictions

Suppose the model under consideration is overidentified ($m < n\ell$). Under such circumstances we can develop a test for the overall goodness-of-fit of the model. Recall that by the mean value theorem we can express $g_T(\theta_T)$ as:

$$g_T(\theta_T) = g_T(\theta_0) + D_T^*(\theta_T - \theta_0) \ . \tag{67}$$

If we multiply equation (67) by $\sqrt{T}$ and substitute for $\sqrt{T}(\theta_T - \theta_0)$ from equation (59) we obtain:

$$\sqrt{T} g_T(\theta_T) = (I - D_T^*(A_T D_T^*)^{-1} A_T)\sqrt{T} g_T(\theta_0) \ . \tag{68}$$

Substituting in the optimal choice for A_T yields:

$$\sqrt{T} g_T(\theta_T) = (I - D_T^*(D_T' S_T^{-1} D_T^*)^{-1} D_T' S_T^{-1})\sqrt{T} g_T(\theta_0) \ , \tag{69}$$

so that by Slutsky's theorem:

$$\sqrt{T} g_T(\theta_T) \xrightarrow{d} (I - D_0(D_0' S_0^{-1} D_0)^{-1} D_0' S_0^{-1}) \times N(0, S_0) \ . \tag{70}$$

Because S_0 is symmetric and positive definite it can be factored as $S_0 = PP'$ where P is nonsingular. Thus (70) can be written as:

$$\sqrt{T}\boldsymbol{P}^{-1}\boldsymbol{g}_T(\boldsymbol{\theta}_T)\xrightarrow{d}(\boldsymbol{I}-\boldsymbol{P}^{-1}\boldsymbol{D}_0(\boldsymbol{D}_0'\boldsymbol{S}_0^{-1}\boldsymbol{D}_0)^{-1}\boldsymbol{D}_0'(\boldsymbol{P}')^{-1})\times N(\mathbf{0},\boldsymbol{I}) \ . \tag{71}$$

The matrix premultiplying the normal distribution in (71) is idempotent with rank $n\ell - m$. It follows, therefore, that the overidentifying test statistic:

$$M_T \equiv T\boldsymbol{g}_T(\boldsymbol{\theta}_T)'\boldsymbol{S}_0^{-1}\boldsymbol{g}_T(\boldsymbol{\theta}_T) \tag{72}$$

converges to a central chi-square random variable with $n\ell - m$ degrees of freedom. The limiting distribution of M_T remains the same if we use a consistent estimate $\boldsymbol{S}_T$ in place of $\boldsymbol{S}_0$.

Note that in many respects the test for overidentifying restrictions is analogous to the Lagrange multiplier test in maximum likelihood estimation. The GMM estimator of $\boldsymbol{\theta}_0$ is obtained by setting m linear combinations of the $n\ell$ orthogonality conditions equal to zero. Thus there are $n\ell - m$ linearly independent combinations which have not been set equal to zero. Suppose we took these $n\ell - m$ linear combinations of the moment conditions and set them equal to a $(n\ell - m) \times 1$ vector of unknown parameters α. The system would then be exactly identified and M_T would be identically equal to zero. Imposing the restriction that $\alpha = 0$ yields the efficient GMM estimator along with a quantity $T\boldsymbol{g}_T(\boldsymbol{\theta}_T)'\boldsymbol{S}_T^{-1}\boldsymbol{g}_T(\boldsymbol{\theta}_T)$ that can be viewed as the GMM analogue of the score form of the Lagrange multiplier test statistic.

The test for overidentifying restrictions is appealing because it provides a simple way to gauge how well the model fits the data. It would also be convenient, however, to be able to test restrictions on the vector of parameters for the model. As we shall see, such tests can be constructed in a straightforward manner.

G. Hypothesis testing in GMM

Suppose that we are interested in testing restrictions on the vector of parameters of the form:

$$\boldsymbol{q}(\boldsymbol{\theta}_0) = \mathbf{0} \ , \tag{73}$$

where $\boldsymbol{q}$ is a known $p \times 1$ vector of functions. Let the $p \times m$ matrix $\boldsymbol{Q}_0 \equiv \partial \boldsymbol{q}/\partial \boldsymbol{\theta}'$ denote the Jacobian of $\boldsymbol{q}(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta}_0$. By assumption $\boldsymbol{Q}_0$ has rank p. We know that for the efficient choice of the weighting matrix the limiting distribution of the GMM estimator is:

$$\sqrt{T}(\boldsymbol{\theta}_T - \boldsymbol{\theta}_0)\xrightarrow{d}N(\mathbf{0},(\boldsymbol{D}_0'\boldsymbol{S}_0^{-1}\boldsymbol{D}_0)^{-1}) \ . \tag{74}$$

Thus under fairly general regularity conditions the standard large-sample test criteria are distributed asymptotically as central chi-square random variables with p degrees of freedom when the restrictions hold.

Let $\boldsymbol{\theta}_T^u$ and $\boldsymbol{\theta}_T^r$ denote the unrestricted estimator and the estimator obtained by minimizing $J_T(\boldsymbol{\theta})$ subject to $\boldsymbol{q}(\boldsymbol{\theta}) = 0$. The Wald test statistic is based on the unrestricted estimator. It takes the form:

$$W_T \equiv T\boldsymbol{q}(\boldsymbol{\theta}_T^u)'(\boldsymbol{Q}_T(\boldsymbol{D}_T'\boldsymbol{S}_T^{-1}\boldsymbol{D}_T)^{-1}\boldsymbol{Q}'_T)^{-1}\boldsymbol{q}(\boldsymbol{\theta}_T^u) \ , \tag{75}$$

where $\boldsymbol{Q}_T$, $\boldsymbol{D}_T$ and $\boldsymbol{S}_T$ are consistent estimates of $\boldsymbol{Q}_0$, $\boldsymbol{D}_0$ and $\boldsymbol{S}_0$ computed using $\boldsymbol{\theta}_T^u$. The Lagrange multiplier test statistic is constructed using the gradient of $J_T(\boldsymbol{\theta})$ evaluated at restricted estimator. It is given by:

$$LM_T \equiv T\boldsymbol{g}_T(\boldsymbol{\theta}_T^r)'\boldsymbol{S}_T^{-1}\boldsymbol{D}_T(\boldsymbol{D}_T'\boldsymbol{S}_T^{-1}\boldsymbol{D}_T)^{-1}\boldsymbol{D}_T'\boldsymbol{S}_T^{-1}\boldsymbol{g}_T(\boldsymbol{\theta}_T^r) \ , \tag{76}$$

where $\boldsymbol{D}_T$ and $\boldsymbol{S}_T$ are consistent estimates of $\boldsymbol{D}_0$ and $\boldsymbol{S}_0$ computed from $\boldsymbol{\theta}_T^r$. The likelihood ratio type test statistic is equal to the difference between the over-identifying test statistic for the restricted and unrestricted estimations:

$$LR_T \equiv T(\boldsymbol{g}_T(\boldsymbol{\theta}_T^r)'\boldsymbol{S}_T^{-1}\boldsymbol{g}_T(\boldsymbol{\theta}_T^r) - \boldsymbol{g}_T(\boldsymbol{\theta}_T^u)'\boldsymbol{S}_T^{-1}\boldsymbol{g}_T(\boldsymbol{\theta}_T^u)) \ . \tag{77}$$

The same estimate $\boldsymbol{S}_T$ must be used for both estimations.

It should be clear from the foregoing discussion that a consistent estimate of $\boldsymbol{S}_0$ is one of the key elements of the GMM approach to estimation and testing. In practice there are a number of different methods for estimating $\boldsymbol{S}_0$, and the appropriate method often depends on the specific characteristics of the model under consideration. The discussion below provides an introduction to heteroskedasticity and autocorrelation consistent estimation of the variance-covariance matrix. A more detailed treatment can be found in Andrews (1991).

H. Robust estimation of the variance-covariance matrix

The variance-covariance matrix of $\sqrt{T}\boldsymbol{g}_T(\boldsymbol{\theta}_0)$ is given by:

$$\boldsymbol{S}_0 \equiv \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_0(j) \ , \tag{78}$$

where $\boldsymbol{\Gamma}_0(j) \equiv \mathrm{E}[\boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta}_0)\boldsymbol{f}(\boldsymbol{X}_{t+\tau-j}, \boldsymbol{Z}_{t-j}, \boldsymbol{\theta}_0)']$. Because we have assumed stationarity, this matrix can also be written as:

$$\boldsymbol{S}_0 = \boldsymbol{\Gamma}_0(0) + \sum_{j=1}^{\infty}(\boldsymbol{\Gamma}_0(j) + \boldsymbol{\Gamma}_0(j)') \ , \tag{79}$$

using the relation $\boldsymbol{\Gamma}_0(-j) = \boldsymbol{\Gamma}_0(j)'$. Now we want to consider how we might go about estimating $\boldsymbol{S}_0$ consistently. First take the scenario where the vector $\boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta}_0)$ is serially uncorrelated. Under such circumstances the second term on the right-hand side of equation (79) drops out and

$$\boldsymbol{\Gamma}_T(0) \equiv 1/T\sum_{t=1}^{T}\boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta}_T)\boldsymbol{f}(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta}_T)'$$

provides a consistent estimate for $\boldsymbol{S}_0$.

The case where $\boldsymbol{f}(\cdot)$ exhibits serial correlation is more complicated. Note that the sum in equation (79) contains an infinite number of terms. It is obviously impossible to estimate each of these terms. One way to proceed would be to treat $f(\cdot)$ as if it were serially correlated for a finite number of lags L. Under such circumstances a natural estimator for $\boldsymbol{S}_0$ would be:

$$\boldsymbol{S}_T = \boldsymbol{\Gamma}_T(0) + \sum_{j=1}^{L} (\boldsymbol{\Gamma}_T(j) + \boldsymbol{\Gamma}_T(j)') \ , \tag{80}$$

where $\boldsymbol{\Gamma}_T(j) \equiv 1/T \sum_{t=1+j}^{T} f(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta}_T) f(\boldsymbol{X}_{t+\tau-j}, \boldsymbol{Z}_{t-j}, \boldsymbol{\theta}_T)'$. As long as the individual $\boldsymbol{\Gamma}_T(j)$ in equation (80) are consistent the estimator $\boldsymbol{S}_T$ will be consistent providing that L is allowed to increase at suitable rate as the sample size T increases. But the estimator of $\boldsymbol{S}_0$ in (80) is not guaranteed to be positive semidefinite. This can lead to problems in empirical work.

The solution to this difficulty is to calculate $\boldsymbol{S}_T$ as a weighted sum of the $\boldsymbol{\Gamma}_T(j)$ where the weights gradually decline to zero as j increases. If these weights are chosen appropriately then $\boldsymbol{S}_T$ will be both consistent and positive semidefinite. Suppose we begin by defining the $n\ell(L+1) \times n\ell(L+1)$ partitioned matrix:

$$\boldsymbol{C}_T(L) = \begin{bmatrix} \boldsymbol{\Gamma}_T(0) & \boldsymbol{\Gamma}_T(1)' & \dots & \boldsymbol{\Gamma}_T(L)' \\ \boldsymbol{\Gamma}_T(1) & \boldsymbol{\Gamma}_T(0) & \dots & \boldsymbol{\Gamma}_T(L-1)' \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_T(L) & \boldsymbol{\Gamma}_T(L-1) & \dots & \boldsymbol{\Gamma}_T(0) \end{bmatrix} . \tag{81}$$

The matrix $\boldsymbol{C}_T(L)$ can always be written in the form $\boldsymbol{C}_T(L) = \boldsymbol{Y}'\boldsymbol{Y}$ where $\boldsymbol{Y}$ is an $(T+L) \times n\ell(L+1)$ partitioned matrix. Take $L = 2$ as an example. The matrix $\boldsymbol{Y}$ is given by:

$$\boldsymbol{Y} = \frac{1}{\sqrt{T}} \begin{bmatrix} 0 & 0 & f(\boldsymbol{X}_{1+\tau}, \boldsymbol{Z}_1, \boldsymbol{\theta}_T)' \\ 0 & f(\boldsymbol{X}_{1+\tau}, \boldsymbol{Z}_1, \boldsymbol{\theta}_T)' & \vdots \\ f(\boldsymbol{X}_{1+\tau}, \boldsymbol{Z}_1, \boldsymbol{\theta}_T)' & \vdots & f(\boldsymbol{X}_{T+\tau}, \boldsymbol{Z}_T, \boldsymbol{\theta}_T)' \\ \vdots & f(\boldsymbol{X}_{T+\tau}, \boldsymbol{Z}_T, \boldsymbol{\theta}_T)' & 0 \\ f(\boldsymbol{X}_{T+\tau}, \boldsymbol{Z}_T, \boldsymbol{\theta}_T)' & 0 & 0 \end{bmatrix} . \tag{82}$$

From this result it follows that $\boldsymbol{C}_T(L)$ is a positive semidefinite matrix. Next consider the matrix:

$$\boldsymbol{S}_T(L) = [\alpha_0 \boldsymbol{I} \quad \alpha_1 \boldsymbol{I} \dots \alpha_L \boldsymbol{I}] \begin{bmatrix} \boldsymbol{\Gamma}_T(0) & \dots & \boldsymbol{\Gamma}_T(L)' \\ \boldsymbol{\Gamma}_T(1) & \dots & \boldsymbol{\Gamma}_T(L-1)' \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_T(L) & \dots & \boldsymbol{\Gamma}_T(0) \end{bmatrix} \begin{bmatrix} \alpha_0 \boldsymbol{I} \\ \alpha_1 \boldsymbol{I} \\ \vdots \\ \alpha_L \boldsymbol{I} \end{bmatrix} , \tag{83}$$

where the α_i are scalars. Because $\boldsymbol{S}_T(L)$ is the partitioned-matrix equivalent of a quadratic form in a positive semidefinite matrix it must also be positive semidefinite. Equation (83) can be rearranged to show that:

$$
\begin{aligned}
\boldsymbol{S}_T(L) &= (\alpha_0^2 + \cdots + \alpha_L^2)\boldsymbol{\Gamma}_T(0) \\
&+ \sum_{j=1}^{L}\left(\sum_{i=0}^{L-j}\alpha_i\alpha_{i+j}\right)\left(\boldsymbol{\Gamma}_T(j) + \boldsymbol{\Gamma}_T(j)'\right) .
\end{aligned} \tag{84}
$$

The weighted sum on right-hand side of equation (84) has the general form of an estimator for the variance-covariance matrix $\boldsymbol{S}_0$. Thus if we select the α_i so that the weights in (84) are a decreasing function of L and we allow L to increase with the sample size at an appropriately slow rate we obtain a consistent positive semidefinite estimator for $\boldsymbol{S}_0$.

The modified Bartlett weights proposed by Newey and West (1987) have been used extensively in empirical research. Let w_j be the weight placed on $\boldsymbol{\Gamma}_T(j)$ in the calculation of the variance-covariance matrix. The weighting function for modified Bartlett weights takes the form:

$$
w_j = \begin{cases} 1 - \frac{j}{L+1} & j = 0, 1, 2, \ldots, L \\ 0 & j > L, \end{cases} \tag{85}
$$

where L is the lag truncation parameter. Note that these weights are obtained by setting $\alpha_i = 1/\sqrt{L+1}$ for $i = 0, 1, \ldots, L$. Newey and West (1987) show that if L is allowed to increase at a rate proportional to $T^{1/3}$ then $\boldsymbol{S}_T$ based on these weights will be a consistent estimator of $\boldsymbol{S}_0$. Although the weighting scheme proposed by Newey and West (1987) is popular, recent research has shown that other schemes may be preferable. Andrews (1991) explores both the theoretical and empirical performance of a variety of different weighting functions. Based on his results Parzen weights seem to offer an good combination of analytic tractability and overall performance. The weighting function for Parzen weights is:

$$
w_j = \begin{cases} 1 - \frac{6j^2}{L^2} + \frac{6j^3}{L^3} & 0 \le \frac{j}{L} \le \frac{1}{2} \\ 2\left(1 - \frac{j}{L}\right)^3 & \frac{1}{2} \le \frac{j}{L} \le 1 \\ 0 & \frac{j}{L} > 1 \end{cases} . \tag{86}
$$

The final question we need to address is how choose the lag truncation parameter L in (86). The simplest strategy is to follow the suggestions of Gallant (1987) and set L equal to the integer closest to $T^{1/5}$. The main advantage of this plug-in approach is that it is yields an estimator that depends only on the sample size for the data set in question. An alternative strategy developed by Andrews (1991), however, may lead to better performance in small samples. He suggests the following data-dependent approach: use the first-stage estimate of $\boldsymbol{\theta}_0$ to construct the sample analogue of $f(\boldsymbol{X}_{t+\tau}, \boldsymbol{Z}_t, \boldsymbol{\theta}_0)$. Then estimate a first-order autoregressive model for each element of this vector. The autocorrelation coefficients along with the residual variances can be used to estimate the value of L that minimizes the asymptotic truncated mean-squared-error of the estimator. Andrews (1991) presents Monte Carlo results that suggest that estimators of $\boldsymbol{S}_0$ constructed in this manner perform well under most circumstances.

6. Closing remarks

Asset pricing models often imply that the expected return on an asset can be written as a linear function of one or more beta coefficients that measure the asset's sensitivity to sources of undiversifiable risk in the economy. This linear tradeoff between risk and expected return makes such models both intuitively appealing and analytically tractable. A number of different methods have been proposed for estimating and testing beta pricing models, but the method of instrumental variables is the approach of choice in most situations. The primary advantage of the instrumental variables approach is that it provides a highly tractable way of characterizing time-varying risk and expected returns.

This paper provides an introduction the econometric evaluation of both conditional and unconditional beta pricing models. We present numerous examples of how the instrumental variable methodology can be applied to various models. We began with a discussion of the conditional version of the Sharpe (1964) – Lintner (1965) CAPM and used it to illustrate how the instrumental variables approach could be used to estimate and test single beta models. Then we extended the analysis to models with multiple betas and introduced the concept of latent variables. We also provided an overview of the generalized method of moments approach (GMM) to estimation and testing. All of the techniques developed in this paper have applications in other areas of asset pricing as well.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858.

Bansal, R. and C. R. Harvey (1995). Performance evaluation in the presence of dynamic trading strategies. Working Paper, Duke University, Durham, NC.

Beneish, M. D. and C. R. Harvey (1995). Measurement error and nonlinearity in the earnings-returns relation. Working Paper, Duke University, Durham, NC.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *J. Business* **45**, 444-454.

Blake, I. F. and J. B. Thomas (1968). On a class of processes arising in linear estimation theory. *IEEE Transactions on Information Theory* IT-14, 12–16.

Bollerslev, T., R. F. Engle and J. M. Wooldridge (1988). A capital asset pricing model with time varying covariances. *J. Politic. Econom.* **96**, 116–31.

Breeden, D. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *J. Financ. Econom.* **7**, 265-296.

Campbell, J. Y. (1987). Stock returns and the term structure. *J. Financ. Econom.* **18**, 373–400.

Carhart, M. and R. J. Krail (1994). Testing the conditional CAPM. Working Paper, University of Chicago.

Chu, K. C. (1973). Estimation and decision for linear systems with elliptically random processes. *IEEE Transactions on Automatic Control* AC-**18**, 499–505.

Cochrane, J. (1994). Discrete time empirical finance. Working Paper, University of Chicago.

Devlin, S. J. R. Gnanadesikan and J. R. Kettenring, Some multivariate applications of elliptical distributions. In: S. Ideka et al., eds., *Essays in probability and statistics*, Shinko Tsusho, Tokyo, 365–393.

Dybvig, P. H. and S. A. Ross (1985). Differential information and performance measurement using a security market line. *J. Finance* **40**, 383–400.
Dumas, B. and B. Solnik (1995). The world price of exchange rate risk. *J. Finance* 445–480.
Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *J. Politic. Econom.* **81**, 607–636.
Ferson, W. E. (1990). Are the latent variables in time-varying expected returns compensation for consumption risk. *J. Finance* **45**, 397–430.
Ferson, W. E. (1995). Theory and empirical testing of asset pricing models. In: Robert A. J. W. T. Ziemba and V. Maksimovic, eds. North Holland 145–200
Ferson, W. E., S. R. Foerster and D. B. Keim (1993). General tests of latent variables models and mean-variance spanning. *J. Finance* **48**, 131–156.
Ferson, W. E. and C. R. Harvey (1991). The variation of economic risk premiums. *J. Politic. Econom.* **99**, 285–315.
Ferson, W. E. and C. R. Harvey (1993). The risk and predictability of international equity returns. *Rev. Financ. Stud.* **6**, 527–566.
Ferson, W. E. and C. R. Harvey (1994a). An exploratory investigation of the fundamental determinants of national equity market returns. In: Jeffrey Frankel, ed., *The internationalization of equity markets*, Chicago: University of Chicago Press, 59–138.
Ferson, W. E. and R. A. Korajczyk (1995) Do arbitrage pricing models explain the predictability of stock returns. *J. Business*, 309–350.
Ferson, W. E. and Stephen R. Foerster (1994). Finite sample properties of the Generalized Method of Moments in tests of conditional asset pricing models. *J. Financ. Econom.* **36**, 29–56.
Gallant, A. R. (1981). On the bias in flexible functional forms and an essentially unbiased form: The Fourier flexible form. *J. Econometrics* **15**, 211-224.
Gallant, A. R. (1987). Nonlinear statistical models. John Wiley and Sons, NY.
Gallant, A. R. and G. E. Tauchen (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes. *Econometrica* **57**, 1091–1120.
Gallant, A. R. and H. White (1988). A unified theory of estimation and inference for nonlinear dynamic models. Basil Blackwell, NY.
Gallant, A. R. and H. White (1990). On learning the derivatives of an unknown mapping with multilayer feedforward networks. University of California at San Diego.
Gibbons, M. R. and W. E. Ferson (1985). Tests of asset pricing models with changing expectations and an unobservable market portfolio. *J. Financ. Econom.* **14**, 217–236.
Glodjo, A. and C. R. Harvey (1995). Forecasting foreign exchange market returns via entropy coding. Working Paper, Duke University, Durham NC.
Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–1054.
Hansen, L. P. and R. J. Hodrick (1983). Risk averse speculation in the forward foreign exchange market: An econometric analysis of linear models. In: Jacob A. Frenkel, ed., Exchange rates and international macroeconomics, University of Chicago Press, Chicago, IL.
Hansen, L. P. and R. Jagannathan (1991). Implications of security market data for models of dynamic economies. *J. Politic. Econom.* **99**, 225-262.
Hansen, L. P. and R. Jagannathan (1994). Assessing specification errors in stochastic discount factor models. Unpublished working paper, University of Chicago, Chicago, IL.
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**, 1269–1285.
Harvey, C. R. (1989). Time-varying conditional covariances in tests of asset pricing models. *J. Financ. Econom.* **24**, 289–317.
Harvey, C. R. (1991a). The world price of covariance risk. *J. Finance* **46**, 111–157.
Harvey, C. R. (1991b). The specification of conditional expectations. Working Paper, Duke University.

Harvey, C. R. (1995), Predictable Risk and returns in emerging markets, Rev. Financ. Stud. 773–816.
Harvey, C. R. and C. Kirby (1995). Analytic tests of factor pricing models. Working Paper, Duke University, Durham, NC.
Harvey, C. R., B. H. Solnik and G. Zhou (1995). What determines expected international asset returns? Working Paper, Duke University, Durham, NC.
Huang, R. D. (1989). Tests of the conditional asset pricing model with changing expectations. Unpublished working Paper, Vanderbilt University, Nashville, TN.
Jagannathan, R. and Z. Wang (1996). The CAPM is alive and well. *J. Finance* **51**, 3–53.
Kan, R. and C. Zhang (1995). A test of conditional asset pricing models. Working Paper, University of Alberta, Edmonton, Canada.
Keim, D. B. and R. F. Stambaugh (1986). Predicting returns in the bond and stock market. *J. Financ. Econom.* **17**, 357–390.
Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā*, series *A*, 419–430.
Kirby, C (1995). Measuring the predictable variation in stock and bond returns. Working Paper, Rice University, Houston, Tx.
Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Rev. Econom. Statist.* **47**, 13–37.
Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica* **41**, 867–887.
Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity-consistent covariance matrix. *Econometrica* **55**, 703–708.
Potscher, B. M. and I. R. Prucha (1991a). Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part I: Consistency and approximation concepts. *Econometric Rev.* **10**, 125–216.
Potscher, B. M. and I. R. Prucha (1991b). Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part II: Asymptotic normality. *Econometric Rev.* **10**, 253–325.
Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *J. Econom. Theory* **13**, 341-360.
Shanken, J. (1990). Intertemporal asset pricing: An empirical investigation. *J. Econometrics* **45**, 99–120.
Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *J. Finance* **19**, 425–442.
Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* London: Chapman and Hall.
Solnik, B. (1991). The economic significance of the predictability of international asset returns. Working Paper, HEC-School of Management.
Vershik, A. M. (1964). Some characteristics properties of Gaussian stochastic processes. *Theory Probab. Appl.* **9**, 353–356.
White, H. (1980). A heteroskedasticity consistent covariance matrix estimator and a direct test of heteroskedasticity. *Econometrica* **48**, 817–838.
Zhou, G. (1995). Small sample rank tests with applications to asset pricing. *J. Empirical Finance* **2**, 71–94.