the regressor values in deviation from the sample mean, and similarly for $\tilde{y}_i$ and $\tilde{z}_i$, the IV estimator for β_2 can be written as (compare (5.51))

$$\hat{\beta}_{2,IV} = \frac{(1/N)\sum_{i=1}^{N}\tilde{z}_i\tilde{y}_i}{(1/N)\sum_{i=1}^{N}\tilde{z}_i\tilde{x}_i}.$$

If the instrument is valid (and under weak regularity conditions), the estimator is consistent and converges to

$$\beta_2 = \frac{\text{cov}\{z_i, y_i\}}{\text{cov}\{z_i, x_i\}}.$$

However, if the instrument is not correlated with the regressor, the denominator of this expression is zero. In this case, the IV estimator is inconsistent and the asymptotic distribution of $\hat{\beta}_{2,IV}$ deviates substantially from a normal distribution. The instrument is weak if there is some correlation between z_i and x_i, but not enough to make the asymptotic normal distribution provide a good approximation in finite (potentially very large) samples. For example, Bound, Jaeger and Baker (1995) show that part of the results of Angrist and Krueger (1991), who use quarter of birth to instrument for schooling in a wage equation, suffer from the weak instruments problem. Even with samples of more than 300 000 (!) individuals, the IV estimator appeared to be unreliable and misleading.

To figure out whether you have weak instruments, it is useful to examine the reduced form regression and evaluate the explanatory power of the additional instruments that are not included in the equation of interest. Consider the linear model with one endogenous regressor

$$y_i = x_{1i}'\beta_1 + x_{2i}\beta_2 + \varepsilon_i,$$

where $E\{x_{1i}\varepsilon_i\} = 0$ and where additional instruments z_{2i} (for x_{2i}) satisfy $E\{z_{2i}\varepsilon_i\} = 0$. The appropriate reduced form is given by

$$x_{2i} = x_{1i}'\pi_1 + z_{2i}'\pi_2 + v_i.$$

If $\pi_2 = 0$, the instruments in z_{2i} are irrelevant and the IV estimator is inconsistent. If π_2 is 'close to zero', the instruments are weak. The value of the F-statistic for $\pi_2 = 0$ is a measure for the information content contained in the instruments. Staiger and Stock (1997) provide a theoretical analysis of the properties of the IV estimator and provide some guidelines about how large the F-statistic should be for the IV estimator to have good properties. As a simple rule-of-thumb, Stock and Watson (2003, Chapter 10) suggest that you do not have to worry about weak instruments if the F-statistic exceeds 10. In any case, it is a good practice to compute and present the F-statistic of the reduced form in empirical work. If the instruments in z_{2i} are insignificant in the reduced form, you should not put much confidence in the IV results. If you have many instruments available, it may be a good strategy to use the most relevant subset and drop the 'weak' ones.

5.6 The Generalized Method of Moments

The approaches sketched above are special cases of an approach proposed by Hansen (1982), usually referred to as the Generalized Method of Moments (GMM). This approach estimates the model parameters directly from the moment conditions that are imposed by the model. These conditions can be linear in the parameters (as in the above examples) but quite often are nonlinear. To enable identification, the number of moment conditions should be at least as large as the number of unknown parameters. The present section provides a fairly intuitive discussion of the Generalized Method of Moments. First, in the next subsection, we start with a motivating example that illustrates how economic theory can imply nonlinear moment conditions. An extensive, not too technical, overview of GIVE and GMM methodology is given in Hall (1993).

5.6.1 Example

The following example is based on Hansen and Singleton (1982). Consider an individual agent who maximizes the expected utility of current and future consumption by solving

$$\max E_t \left\{ \sum_{s=0}^{S} \delta^s U(C_{t+s}) \right\}, \tag{5.65}$$

where C_{t+s} denotes consumption in period $t+s$, $U(C_{t+s})$ is the utility attached to this consumption level, which is discounted by the discount factor δ $(0 < \delta \leq 1)$, and where E_t is the expectation operator conditional upon all information available at time t. Associated with this problem is a set of intertemporal budget constraints of the form

$$C_{t+s} + q_{t+s} = w_{t+s} + (1 + r_{t+s})q_{t+s-1}, \tag{5.66}$$

where q_{t+s} denotes financial wealth at the end of period $t+s$, r_{t+s} is the return on financial wealth (invested in a portfolio of assets), and w_{t+s} denotes labour income. The budget constraint thus says that labour income plus asset income should be spent on consumption C_{t+s} or saved in q_{t+s}. This maximization problem is hard to solve analytically. Nevertheless, it is still possible to estimate the unknown parameters involved through the first order conditions. The first order conditions of (5.65) subject to (5.66) imply that

$$E_t\{\delta U'(C_{t+1})(1 + r_{t+1})\} = U'(C_t),$$

where U' is the first derivative of U. The right-hand side of this equality denotes the marginal utility of one additional dollar consumed today, while the left-hand side gives the expected marginal utility of saving this dollar until the next period (so that it becomes $1 + r_{t+1}$ dollars) and consuming it then. Optimality thus implies that (expected) marginal utilities are equalized.

As a next step, we can rewrite this equation as

$$E_t \left\{ \frac{\delta U'(C_{t+1})}{U'(C_t)}(1 + r_{t+1}) - 1 \right\} = 0. \tag{5.67}$$

Essentially, this is a (conditional) moment condition which can be exploited to estimate the unknown parameters if we make some assumption about the utility function U. We can do this by transforming (5.67) into a set of unconditional moment conditions.

Suppose z_t is included in the information set. This implies that z_t does not provide any information about the expected value of

$$\frac{\delta U'(C_{t+1})}{U'(C_t)}(1 + r_{t+1}) - 1$$

so that it also holds that[17]

$$E\left\{\left(\frac{\delta U'(C_{t+1})}{U'(C_t)}(1 + r_{t+1}) - 1\right) z_t\right\} = 0. \tag{5.68}$$

Thus we can interpret z_t as a vector of instruments, valid by the assumption of optimal behaviour (rational expectations) of the agent. For simplicity, let us assume that the utility function is of the power form, that is

$$U(C) = \frac{C^{1-\gamma}}{1-\gamma},$$

where γ denotes the (constant) coefficient of relative risk aversion, where higher values of γ correspond to a more risk averse agent. Then we can write (5.68) as

$$E\left\{\left(\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1 + r_{t+1}) - 1\right) z_t\right\} = 0. \tag{5.69}$$

We now have a set of moment conditions which identify the unknown parameters δ and γ, and given observations on $C_{t+1}/C_t, r_{t+1}$ and z_t allow us to estimate them consistently. This requires an extension of the earlier approach to nonlinear functions.

5.6.2 The Generalized Method of Moments

Let us, in general, consider a model that is characterized by a set of R moment conditions as

$$E\{f(w_t, z_t, \theta)\} = 0, \tag{5.70}$$

where f is a vector function with R elements, θ is a K-dimensional vector containing all unknown parameters, w_t is a vector of observable variables that could be endogenous or exogenous, and z_t is the vector of instruments. In the example of the previous subsection $w_t' = (C_{t+1}/C_t, r_{t+1})$; in the linear model of Section 5.5 $w_t' = (y_t, x_t')$.

To estimate θ we take the same approach as before and consider the sample equivalent of (5.70) given by

$$g_T(\theta) \equiv \frac{1}{T}\sum_{t=1}^{T} f(w_t, z_t, \theta). \tag{5.71}$$

If the number of moment conditions R equals the number of unknown parameters K, it would be possible to set the R elements in (5.71) to zero and to solve for θ to obtain

[17] We use the general result that $E\{x_1|x_2\} = 0$ implies that $E\{x_1 g(x_2)\} = 0$ for any function g (see Appendix B).

a unique consistent estimator. If f is nonlinear in θ an analytical solution may not be available. If the number of moment conditions is less than the number of parameters, the parameter vector θ is not identified. If the number of moment conditions is larger, we cannot solve uniquely for the unknown parameters by setting (5.71) to zero. Instead, we choose our estimator for θ such that the vector of sample moments is as close as possible to zero, in the sense that a quadratic form in $g_T(\theta)$ is minimized. That is,

$$\min_\theta Q_T(\theta) = \min_\theta g_T(\theta)' W_T g_T(\theta), \tag{5.72}$$

where, as before, W_T is a positive definite matrix with plim $W_T = W$. The solution to this problem provides the **generalized method of moments** or GMM estimator $\hat{\theta}$. Although we cannot obtain an analytical solution for the GMM estimator in the general case, it can be shown that it is consistent and asymptotically normal under some weak regularity conditions. The heuristic argument presented for the generalized instrumental variables estimator in the linear model extends to this more general setting. Because sample averages converge to population means, which are zero for the true parameter values, an estimator chosen to make these sample moments as close to zero as possible (as defined by (5.72)) will converge to the true value and will thus be consistent. In practice, the GMM estimator is obtained by numerically solving the minimization problem in (5.72), for which a variety of algorithms is available; see Wooldridge (2002, Section 12.7) or Greene (2003, Appendix E) for a general discussion.

As before, different weighting matrices W_T lead to different consistent estimators with different asymptotic covariance matrices. The optimal weighting matrix, which leads to the smallest covariance matrix for the GMM estimator, is the inverse of the covariance matrix of the sample moments. In the absence of autocorrelation it is given by

$$W^{opt} = (E\{f(w_t, z_t, \theta) f(w_t, z_t, \theta)'\})^{-1}.$$

In general this matrix depends upon the unknown parameter vector θ, which presents a problem which we did not encounter in the linear model. The solution is to adopt a multi-step estimation procedure. In the first step we use a suboptimal choice of W_T which does not depend upon θ (for example the identity matrix), to obtain a first consistent estimator $\hat{\theta}_{[1]}$, say. Then, we can consistently estimate the optimal weighting matrix by[18]

$$W_T^{opt} = \left(\frac{1}{T} \sum_{t=1}^{T} f(w_t, z_t, \hat{\theta}_{[1]}) f(w_t, z_t, \hat{\theta}_{[1]})' \right)^{-1}. \tag{5.73}$$

In the second step one obtains the asymptotically efficient (optimal) GMM estimator $\hat{\theta}_{GMM}$. Its asymptotic distribution is given by

$$\sqrt{T}(\hat{\theta}_{GMM} - \theta) \to \mathcal{N}(0, V), \tag{5.74}$$

[18] If there is autocorrelation in $f(w_t, z_t, \theta)$ up to a limited order, the optimal weighting matrix can be estimated using a variant of the Newey-West estimator discussed in Section 5.1; see Greene (2003, Subsection 18.3.4).

where the asymptotic covariance matrix V is given by

$$V = (DW^{opt}D')^{-1}, \tag{5.75}$$

where D is the $K \times R$ derivative matrix

$$D = E\left\{\frac{\partial f(w_t, z_t, \theta)}{\partial \theta'}\right\}. \tag{5.76}$$

Intuitively, the elements in D measure how sensitive a particular moment is with respect to small changes in θ. If the sensitivity with respect to a given element in θ is large, small changes in this element lead to relatively large changes in the objective function $Q_T(\theta)$ and the particular element in θ is relatively accurately estimated. As usual, the covariance matrix in (5.75) can be estimated by replacing the population moments in D and W^{opt} by their sample equivalents, evaluated at $\hat{\theta}_{GMM}$.

The great advantage of the generalized method of moments is that (1) it does not require distributional assumptions, like normality, (2) it can allow for heteroskedasticity of unknown form and (3) it can estimate parameters even if the model cannot be solved analytically from the first order conditions. Unlike most of the cases we discussed before, the validity of the instruments in z_t is beyond doubt if the model leads to a conditional moment restriction (as in (5.67)) and z_t is in the conditioning set. For example, if at time t the agent maximizes expected utility given all publicly available information then any variable that is observed (to the agent) at time t provides a valid instrument.

Unfortunately, there is considerable evidence that the asymptotic distribution in (5.74) often provides a poor approximation to the sampling distribution of the GMM estimator in sample sizes that are typically encountered in empirical work, see, for example, Hansen, Heaton and Yaron (1996). In a recent paper, Stock and Wright (2003) explore the distribution theory for GMM estimators when some or all of the parameters are weakly identified, paying particular attention to variants of the nonlinear model discussed in Subsection 5.6.1. The weak instruments problem, discussed in Subsection 5.5.4, appears also relevant in a general GMM setting.

Finally, we consider the extension of the **overidentifying restrictions test** to nonlinear models. Following the intuition from the linear model, it would be anticipated that if the population moment conditions $E\{f(w_t, z_t, \theta)\} = 0$ are correct then $g_T(\hat{\theta}_{GMM}) \approx 0$. Therefore, the sample moments provide a convenient test of the model specification. Provided that all moment conditions are correct, the test statistic

$$\xi = T g_T(\hat{\theta}_{GMM})' W_T^{opt} g_T(\hat{\theta}_{GMM}),$$

where $\hat{\theta}_{GMM}$ is the optimal GMM estimator and W_T^{opt} is the optimal weighting matrix given in (5.73) (based upon a consistent estimator for θ), is asymptotically Chi-squared distributed with $R - K$ degrees of freedom. Recall that for the exactly identified case, there are zero degrees of freedom and there is nothing that can be tested.

In Section 5.7 we present an empirical illustration using GMM to estimate intertemporal asset pricing models. In Section 10.5 we shall consider another example of GMM, where it is used to estimate a dynamic panel data model. First, we consider a few simple examples.

5.6.3 Some Simple Examples

As a very simple example, assume we are interested in estimating the population mean μ of a variable y_i on the basis of a sample of N observations ($i = 1, 2, \ldots, N$). The moment condition of this 'model' is given by

$$E\{y_i - \mu\} = 0,$$

with sample equivalent

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \mu).$$

By setting this to zero and solving for μ we obtain a method of moments estimator

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i,$$

which is just the sample average.

If we consider the linear model

$$y_i = x_i'\beta + \varepsilon_i$$

again, with instrument vector z_i, the moment conditions are

$$E\{\varepsilon_i z_i\} = E\{(y_i - x_i'\beta)z_i\} = 0.$$

If ε_i is i.i.d. the optimal GMM is the instrumental variables estimator given in (5.43) and (5.56). More generally, the optimal weighting matrix is given by

$$W^{opt} = (E\{\varepsilon_i^2 z_i z_i'\})^{-1},$$

which is estimated unrestrictedly as

$$W_N^{opt} = \left(\frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_i^2 z_i z_i'\right)^{-1},$$

where $\hat{\varepsilon}_i$ is the residual based upon an initial consistent estimator. When it is imposed that ε_i is i.i.d. we can simply use

$$W_N^{opt} = \left(\frac{1}{N}\sum_{i=1}^{N} z_i z_i'\right)^{-1}.$$

The $K \times R$ derivative matrix is given by

$$D = E\{x_i z_i'\},$$

which we can estimate consistently by

$$D_N = \frac{1}{N}\sum_{i=1}^{N} x_i z_i'.$$

In general, the covariance matrix of the *optimal* GMM or GIV estimator $\hat{\beta}$ for β can be estimated as

$$\hat{V}\{\hat{\beta}\} = \left(\sum_{i=1}^{N} x_i z_i'\right)^{-1} \sum_{i=1}^{N} \hat{\varepsilon}_i^2 z_i z_i' \left(\sum_{i=1}^{N} z_i x_i'\right)^{-1}. \tag{5.77}$$

This estimator generalizes (5.62) just as the White heteroskedasticity consistent covariance matrix generalizes the standard OLS expression. Thus, the general GMM set-up allows for heteroskedasticity of ε_i automatically.

5.7 Illustration: Estimating Intertemporal Asset Pricing Models

In the recent finance literature, the GMM framework is frequently used to estimate and test asset pricing models. An asset pricing model, for example the CAPM discussed in Section 2.7, should explain the variation in expected returns for different risky investments. Because some investments are more risky than others, investors may require compensation for bearing this risk by means of a risk premium. This leads to variation in expected returns across different assets. An extensive treatment of asset pricing models and their link with the generalized method of moments is provided in Cochrane (2001).

In this section we consider the consumption-based asset pricing model. This model is derived from the framework sketched in Subsection 5.6.1 by introducing a number of alternative investment opportunities for financial wealth. Assume that there are J alternative risky assets available that the agent can invest in, with returns $r_{j,t+1}$, $j = 1, \ldots, J$, as well as a riskless asset with certain return $r_{f,t+1}$. Assuming that the agent optimally chooses his portfolio of assets, the first order conditions of the problem now imply that

$$E_t\{\delta U'(C_{t+1})(1 + r_{f,t+1})\} = U'(C_t)$$
$$E_t\{\delta U'(C_{t+1})(1 + r_{j,t+1})\} = U'(C_t), \quad j = 1, \ldots, J.$$

This says that the expected marginal utility of investing one additional dollar in asset j is equal for all assets and equal to the marginal utility of consuming this additional dollar today. Assuming power utility, as before, and restricting attention to unconditional expectations[19] the first order conditions can be rewritten as

[19] This means that we restrict attention to moments using instrument $z_t = 1$ only.

$$E\left\{\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1+r_{f,t+1})\right\} = 1 \tag{5.78}$$

$$E\left\{\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(r_{j,t+1}-r_{f,t+1})\right\} = 0, \quad j = 1, \ldots, J, \tag{5.79}$$

where the second set of conditions is written in terms of excess returns, i.e. returns in excess of the riskfree rate.

Let us, for convenience, define the intertemporal marginal rate of substitution

$$m_{t+1}(\theta) \equiv \delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma},$$

where θ contains all unknown parameters. In finance, $m_{t+1}(\theta)$ is often referred to as a stochastic discount factor or a pricing kernel (see Campbell, Lo and MacKinlay, 1997, Chapter 8, or Cochrane, 2001). Alternative asset pricing models are described by alternative specifications for the pricing kernel $m_{t+1}(\theta)$. To see how a choice for $m_{t+1}(\theta)$ provides a model that describes expected returns, we use that for two arbitrary random variables $E\{xy\} = \text{cov}\{x, y\} + E\{x\}E\{y\}$ (see Appendix B), from which it follows that

$$\text{cov}\{m_{t+1}(\theta), r_{j,t+1} - r_{f,t+1}\} + E\{m_{t+1}(\theta)\}E\{r_{j,t+1} - r_{f,t+1}\} = 0.$$

This allows us to write

$$E\{r_{j,t+1} - r_{f,t+1}\} = -\frac{\text{cov}\{m_{t+1}(\theta), r_{j,t+1} - r_{f,t+1}\}}{E\{m_{t+1}(\theta)\}}, \tag{5.80}$$

which says that the expected excess return on any asset j is equal to a risk premium that depends linearly upon the covariance between the asset's excess return and the stochastic discount factor. Knowledge of $m_{t+1}(\theta)$ allows us to describe or explain the cross-sectional variation of expected returns across different assets. In the consumption-based model, this tells us that assets that have a positive covariance with consumption growth (and thus make future consumption more volatile) must promise higher expected returns to induce investors to hold them. Conversely, assets that covary negatively with consumption growth can offer expected returns that are lower than the riskfree rate.[20]

The moment conditions in (5.78)–(5.79) can be used to estimate the unknown parameters δ and γ. In this section we use data[21] that cover monthly returns over the period February 1959–November 1993. The basic assets we consider are ten portfolios of stocks, maintained by the Center for Research in Security Prices at the University of Chicago. These portfolios are size-based, which means that portfolio 1 contains the 10% smallest firms listed at the New York Stock Exchange, while portfolio 10 contains the 10% largest firms that are listed. The riskless return is approximated by the

[20] For example, you may reward a particular asset if it delivers a high return in the situation where you happen to get unemployed.

[21] The data are available in PRICING.

monthly return on a 3 month US Treasury Bill, which does not vary much over time. For consumption we use total US personal consumption expenditures on nondurables and services. It is assumed that the model is valid for a representative agent, whose consumption corresponds to this measure of aggregate per capita consumption. Data on size-based portfolios are used because most asset pricing models tend to underpredict the returns on the stocks of small firms. This is the so-called small firm effect (see Banz, 1981; or Campbell, Lo and MacKinlay, 1997, p. 211).

With one riskless asset and ten risky portfolios, (5.78)–(5.79) provide 11 moment conditions with only two parameters to estimate. These parameters can be estimated using the identity matrix as a suboptimal weighting matrix, using the efficient two-step GMM estimator that was presented above, or using a so-called iterated GMM estimator. This estimator has the same asymptotic properties as the two-step one, but is sometimes argued to have a better small-sample performance. It is obtained by computing a new optimal weighting matrix using the two-step estimator, and using this to obtain a next estimator, $\hat{\theta}_{[3]}$, say, which in turn is used in a weighting matrix to obtain $\hat{\theta}_{[4]}$. This procedure is repeated until convergence.

Table 5.4 presents the estimation results on the basis of the monthly returns from February 1959–November 1993, using one-step GMM (using the identity matrix as weighting matrix) and iterated GMM.[22] The γ estimates are huge and rather imprecise. For the iterated GMM procedure, for example, a 95% confidence interval for γ based upon the approximate normal distribution is as large as (−9.67, 124.47). The estimated risk aversion coefficients of 57.4 and 91.4 are much higher than what is considered economically plausible. This finding illustrates the so-called equity premium puzzle (see Mehra and Prescott, 1985), which reflects that the high risk premia on risky assets (equity) can only be explained in this model if agents are extremely risk averse (compare Campbell, Lo and MacKinlay, 1997, Section 8.2). If we look at the overidentifying restrictions tests, we see, somewhat surprisingly, that they do not reject the joint validity of the imposed moment conditions. This means that the consumption-based asset pricing model is statistically not rejected by the data. This is solely due to the high imprecision of the estimates. Unfortunately this is only a statistical satisfaction and certainly does not imply that the model is economically valuable. The gain in efficiency from the use of the optimal weighting matrix appears to be fairly limited with standard errors that are only up to 20% smaller than for the one-step method.

Table 5.4 GMM estimation results consumption-based asset pricing model

	One-step GMM		Iterated GMM	
	Estimate	s.e.	Estimate	s.e.
δ	0.6996	(0.1436)	0.8273	(0.1162)
γ	91.4097	(38.1178)	57.3992	(34.2203)
ξ ($df = 9$)	4.401	($p = 0.88$)	5.685	($p = 0.77$)

[22] For the one-step GMM estimator the standard errors and the overidentifying restrictions test are computed in a non-standard way. The formulae given in the text do not apply because the optimal weighting matrix is not used. See Cochrane (2001, Chapter 11) for the appropriate expressions. The estimation results in Table 5.4 were obtained using RATS 5.1.

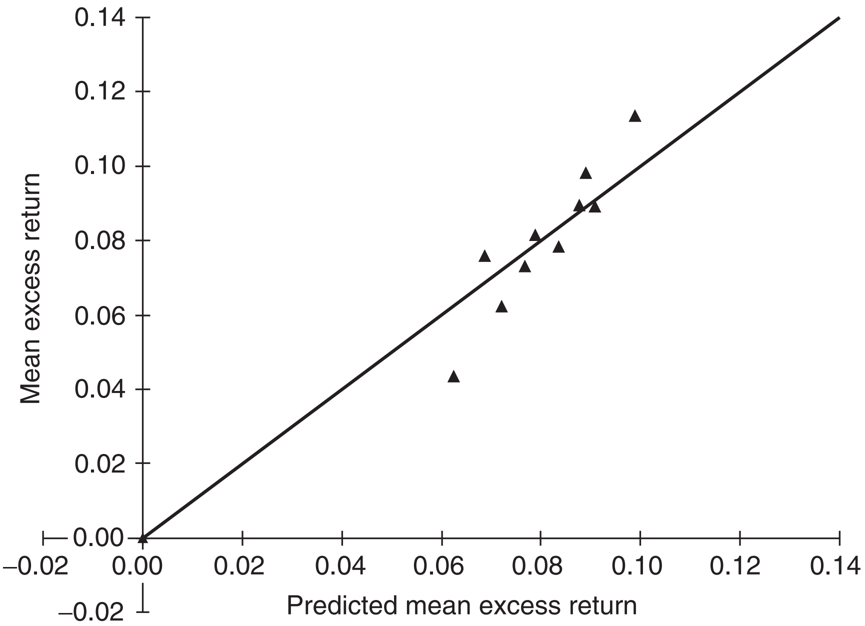

Figure 5.1 Actual versus predicted mean excess returns of size-based portfolios

To investigate the economic value of the above model, it is possible to compute so-called pricing errors (compare Cochrane, 1996). One can directly compute the average expected excess return according to the model, simply by replacing the population moments in (5.80) by the corresponding sample moments and using the estimated values for δ and γ. On the other hand the average excess returns on asset j can be directly computed from the data. In Figure 5.1, we plot the average excess returns against the predicted average excess returns, as well as a 45° line. We do this for the one-step estimator only because, as argued by Cochrane (1996), this estimator minimizes the vector of pricing errors of the 11 assets. Points on the 45° line indicate that the average pricing error is zero. Points above this line indicate that the return of the corresponding asset is underpredicted by the model. The figure confirms our idea that the economic performance of the model is somewhat disappointing. Clearly, the model is unable to fully capture the cross-sectional variation in expected excess returns. The two portfolios with the smallest firms have the highest mean excess return and are both above the 45° line. The model apparently does not solve the small-firm effect as the returns on these portfolios are underpredicted.

Cochrane (1996) also presents a range of alternative asset pricing models that are estimated by GMM, which, in a number of cases, perform much better than the simple consumption-based model discussed here. Marquering and Verbeek (1999) extend the above model by including transaction costs and habit persistence in the utility function.

5.8 Concluding Remarks

This chapter has discussed a variety of models that can be headed under the term ‘stochastic regressors’. Starting from a linear model with an endogenous regressor, we discussed instrumental variables estimation. It was shown how instrumental variables estimation exploits different moment conditions compared to the OLS estimator. If more moment conditions are imposed than unknown parameters, we can use a generalized instrumental variables estimator, which can also be derived in a GMM framework with an optimal weighting matrix. GMM was discussed in detail, with an application to intertemporal asset pricing models. In dynamic models one usually has the advantage that the choice of instruments is less suspect: lagged values can often be assumed to be uncorrelated with current innovations. The big advantage of GMM is that it can estimate the parameters in a model without having to solve the model analytically. That is, there is no need to write the model as $y = \textit{something} + \textit{error term}$. All one needs is conditions in terms of expectations, which are often derived directly from economic theory.

Exercises

Exercise 5.1 (Instrumental Variables)

Consider the following model

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i, \quad i = 1, \ldots, N, \tag{5.81}$$

where (y_i, x_{i2}, x_{i3}) are observed and have finite moments, and ε_i is an unobserved error term. Suppose this model is estimated by ordinary least squares. Denote the OLS estimator by b.

a. What are the *essential* conditions required for unbiasedness of b? What are the *essential* conditions required for consistency of b? Explain the difference between unbiasedness and consistency.

b. Show how the conditions for consistency can be written as moment conditions (if you have not done so already). Explain how a method of moments estimator can be derived from these moment conditions. Is the resulting estimator any different from the OLS one?

Now suppose that $\text{cov}\{\varepsilon_i, x_{i3}\} \neq 0$.

c. Give two examples of cases where one can expect a nonzero correlation between a regressor, x_{i3}, and the error ε_i.

d. In this case, is it possible to still make appropriate inferences based on the OLS estimator, while adjusting the standard errors appropriately?

e. Explain how an instrumental variable, z_i, say, leads to a new moment condition and, consequently, an alternative estimator for β.

f. Why does this alternative estimator lead to a smaller R^2 than the OLS one? What does this say of the R^2 as a measure for the adequacy of the model?

g. Why can we not choose $z_i = x_{i2}$ as an instrument for x_{i3}, even if $E\{x_{i2}\varepsilon_i\} = 0$? Would it be possible to use x_{i2}^2 as an instrument for x_{i3}?

Exercise 5.2 (Returns to Schooling – Empirical)

Consider the data used in Section 5.4, as available in SCHOOLING. The purpose of this exercise is to explore the role of parents' education as instruments to estimate the returns to schooling.

a. Estimate a reduced form for schooling, as reported in Table 5.2, but include mother's and father's education levels, instead of the lived near college dummy. What do these results indicate about the possibility of using parents' education as instruments?

b. Estimate the returns to schooling, on the basis of the same specification as in Section 5.4, using mother's and father's education as instruments (and age and age-squared as instruments for experience and its square).

c. Test the overidentifying restriction.

d. Re-estimate the model using also the lived near college dummy and test the two overidentifying restrictions.

e. Compare and interpret the different estimates on the returns to schooling from Table 5.3, and parts **b** and **d** of this exercise.

Exercise 5.3 (GMM)

An intertemporal utility maximization problem gives the following first order condition

$$E_t\left\{\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1+r_{t+1})\right\}=1,$$

where E_t denotes the expectation operator conditional upon time t information, C_t denotes consumption in period t, r_{t+1} is the return on financial wealth, δ is a discount rate and γ is the coefficient of relative risk aversion. Assume that we have a time series of observations on consumption levels, returns and instrumental variables z_t.

a. Show how the above condition can be written as a set of *unconditional* moment conditions. Explain how we can estimate δ and γ consistently from these moment conditions.

b. What is the minimum number of moment conditions that is required? What do we (potentially) gain by having more moment conditions?

c. How can we improve the efficiency of the estimator for a given set of moment conditions? In which case does this not work?

d. Explain what we mean by 'overidentifying restrictions'. Is this a good or a bad thing?

e. Explain how the overidentifying restrictions test is performed. What is the null hypothesis that is tested? What do you conclude if the test rejects?