CHAPTER 11
Simultaneous Equations Model

11.1 Introduction

Economists formulate models for consumption, production, investment, money demand and money supply, labor demand and labor supply to attempt to explain the workings of the economy. These behavioral equations are estimated equation by equation or jointly as a system of equations. These are known as *simultaneous equations models*. Much of today's econometrics have been influenced and shaped by a group of economists and econometricians known as the Cowles Commission who worked together at the University of Chicago in the late 1940's, see Chapter 1. Simultaneous equations models had their genesis in economics during that period. Haavelmo's (1944) work emphasized the use of the probability approach to formulating econometric models. Koopmans and Marschak (1950) and Koopmans and Hood (1953) in two influential Cowles Commission monographs provided the appropriate statistical procedures for handling simultaneous equations models. In this chapter, we first give simple examples of simultaneous equations models and show why the least squares estimator is no longer appropriate. Next, we discuss the important problem of identification and give a simple necessary but not sufficient condition that helps check whether a specific equation is identified. Sections 11.2 and 11.3 give the estimation of a single and a system of equations using instrumental variable procedures. Section 11.4 gives a test of over-identification restrictions whereas, section 11.5 gives a Hausman specification test. Section 11.6 concludes with an empirical example. The Appendix revisits the identification problem and gives a necessary and sufficient condition for identification.

11.1.1 Simultaneous Bias

Example 1: Consider a simple Keynesian model with no government

$$\begin{aligned} C_t &= \alpha + \beta Y_t + u_t \qquad t = 1, 2, \ldots, T && (11.1)\\ Y_t &= C_t + I_t && (11.2) \end{aligned}$$

where C_t denotes consumption, Y_t denotes disposable income, and I_t denotes autonomous investment. This is a system of two simultaneous equations, also known as *structural equations* with the second equation being an identity. The first equation can be estimated by OLS giving

$$\widehat{\beta}_{OLS} = \sum_{t=1}^{T} y_t c_t / \sum_{t=1}^{T} y_t^2 \quad \text{and} \quad \widehat{\alpha}_{OLS} = \bar{C} - \widehat{\beta}_{OLS}\bar{Y} \tag{11.3}$$

with y_t and c_t denoting Y_t and C_t in deviation form, i.e., $y_t = Y_t - \bar{Y}$, and $\bar{Y} = \sum_{t=1}^{T} Y_t/T$. Since I_t is autonomous, it is an *exogenous* variable determined outside the system, whereas C_t and Y_t are *endogenous* variables determined by the system. Let us solve for Y_t and C_t in terms of the constant and I_t. The resulting two equations are known as the *reduced form* equations

$$\begin{aligned} C_t &= \alpha/(1-\beta) + \beta I_t(1-\beta) + u_t/(1-\beta) && (11.4)\\ Y_t &= \alpha/(1-\beta) + I_t/(1-\beta) + u_t/(1-\beta) && (11.5) \end{aligned}$$

B.H. Baltagi, *Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-20059-5_11,

These equations express each endogenous variable in terms of exogenous variables and the error terms. Note that both Y_t and C_t are a function of u_t, and hence both are correlated with u_t. In fact, $Y_t - E(Y_t) = u_t/(1-\beta)$, and

$$\text{cov}(Y_t, u_t) = E[(Y_t - E(Y_t))u_t] = \sigma_u^2/(1-\beta) \geq 0 \quad \text{if} \quad 0 \leq \beta \leq 1 \tag{11.6}$$

This holds because $u_t \sim (0, \sigma_u^2)$ and I_t is exogenous and independent of the error term. Equation (11.6) shows that the right hand side regressor in (11.1) is correlated with the error term. This causes the OLS estimates to be *biased* and *inconsistent*. In fact, from (11.1),

$$c_t = C_t - \bar{C} = \beta y_t + (u_t - \bar{u})$$

and substituting this expression in (11.3), we get

$$\widehat{\beta}_{OLS} = \beta + \textstyle\sum_{t=1}^T y_t u_t / \sum_{t=1}^T y_t^2 \tag{11.7}$$

From (11.7), it is clear that $E(\widehat{\beta}_{OLS}) \neq \beta$, since the expected value of the second term is not necessarily zero. Also, using (11.5) one gets

$$y_t = Y_t - \bar{Y} = [i_t + (u_t - \bar{u})]/(1-\beta)$$

where $i_t = I_t - \bar{I}$ and $\bar{I} = \sum_{t=1}^T I_t/T$. Defining $m_{yy} = \sum_{t=1}^T y_t^2/T$, we get

$$m_{yy} = (m_{ii} + 2m_{iu} + m_{uu})/(1-\beta)^2 \tag{11.8}$$

where $m_{ii} = \sum_{t=1}^T i_t^2/T$, $m_{iu} = \sum_{t=1}^T i_t(u_t - \bar{u})/T$ and $m_{uu} = \sum_{t=1}^T (u_t - \bar{u})^2/T$. Also,

$$m_{yu} = (m_{iu} + m_{uu})/(1-\beta) \tag{11.9}$$

Using the fact that plim $m_{iu} = 0$ and plim $m_{uu} = \sigma_u^2$, we get

$$\text{plim } \widehat{\beta}_{OLS} = \beta + \text{plim } (m_{yu}/m_{yy}) = \beta + [\sigma_u^2(1-\beta)/(\text{plim } m_{ii} + \sigma_u^2)]$$

which shows that $\widehat{\beta}_{OLS}$ overstates β if $0 \leq \beta \leq 1$.

Example 2: Consider a simple demand and supply model

$$\begin{aligned} Q_t^d &= \alpha + \beta P_t + u_{1t} && (11.10)\\ Q_t^s &= \gamma + \delta P_t + u_{2t} && (11.11)\\ Q_t^d &= Q_t^s = Q_t \quad t = 1, 2, \ldots, T && (11.12) \end{aligned}$$

Substituting the equilibrium condition (11.12) in (11.10) and (11.11), we get

$$\begin{aligned} Q_t &= \alpha + \beta P_t + u_{1t} && (11.13)\\ Q_t &= \gamma + \delta P_t + u_{2t} \quad t = 1, 2, \ldots, T && (11.14) \end{aligned}$$

For the demand equation (11.13), the sign of β is expected to be negative, while for the supply equation (11.14), the sign of δ is expected to be positive. However, we only observe one equilibrium pair (Q_t, P_t) and these are not labeled demand or supply quantities and prices. When we run the OLS regression of Q_t on P_t we do not know what we are estimating, demand or supply? In fact, any linear combination of (11.13) and (11.14) looks exactly like (11.13) or (11.14). It will have a constant, Price, and a disturbance term in it. Since demand or supply cannot be distinguished from this 'mongrel' we have what is known as an *identification problem*. If the demand equation (or the supply equation) looked different from this mongrel, then this particular equation would be identified. More on this later. For now let us examine the properties of the OLS estimates of the demand equation. It is well known that

$$\widehat{\beta}_{OLS} = \sum_{t=1}^{T} q_t p_t / \sum_{t=1}^{T} p_t^2 = \beta + \sum_{t=1}^{T} p_t(u_{1t} - \bar{u}_1) / \sum_{t=1}^{T} p_t^2 \tag{11.15}$$

where q_t and p_t denote Q_t and P_t in deviation form, i.e., $q_t = Q_t - \bar{Q}$. This estimator is unbiased depending on whether the last term in (11.15) has zero expectations. In order to find this expectation we solve the structural equations in (11.13) and (11.14) for Q_t and P_t

$$Q_t = (\alpha\delta - \gamma\beta)/(\delta - \beta) + (\delta u_{1t} - \beta u_{2t})/(\delta - \beta) \tag{11.16}$$
$$P_t = (\alpha - \gamma)/(\delta - \beta) + (u_{1t} - u_{2t})/(\delta - \beta) \tag{11.17}$$

(11.16) and (11.17) are known as the reduced form equations. Note that both Q_t and P_t are functions of both errors u_1 and u_2. Hence, P_t is correlated with u_{1t}. In fact,

$$p_t = (u_{1t} - \bar{u}_1)/(\delta - \beta) - (u_{2t} - \bar{u}_2)/(\delta - \beta) \tag{11.18}$$

and

$$\text{plim} \sum_{t=1}^{T} p_t(u_{1t} - \bar{u}_1)/T = (\sigma_{11} - \sigma_{12})/(\delta - \beta) \tag{11.19}$$

$$\text{plim} \sum_{t=1}^{T} p_t^2/T = (\sigma_{11} + \sigma_{22} - 2\sigma_{12})/(\delta - \beta)^2 \tag{11.20}$$

where $\sigma_{ij} = \text{cov}(u_{it}, u_{jt})$ for $i, j = 1, 2$; and $t = 1, \ldots, T$. Hence, from (11.15)

$$\text{plim}\ \widehat{\beta}_{OLS} = \beta + (\sigma_{11} - \sigma_{12})(\delta - \beta)/(\sigma_{11} + \sigma_{22} - 2\sigma_{12}) \tag{11.21}$$

and the last term is not necessarily zero, implying that $\widehat{\beta}_{OLS}$ is *not consistent* for β. Similarly, one can show that the OLS estimator for δ is not consistent, see problem 1. This *simultaneous bias* is once again due to the correlation of the right hand side variable (price) with the error term u_1. This correlation could be due to the fact that P_t is a function of u_{2t}, from (11.17), and u_{2t} and u_{1t} are correlated, making P_t correlated with u_{1t}. Alternatively, P_t is a function of Q_t, from (11.13) or (11.14), and Q_t is a function of u_{1t}, from (11.13), making P_t a function of u_{1t}. Intuitively, if a shock in demand (i.e., a change in u_{1t}) shifts the demand curve, the new intersection of demand and supply determines a new equilibrium price and quantity. This new price is therefore, affected by the change in u_{1t}, and is correlated with it.

In general, whenever a right hand side variable is correlated with the error term, the OLS estimates are biased and inconsistent. We refer to this as an *endogeneity problem*. Recall, Figure 3 of Chapter 3 with $\text{cov}(P_t, u_{1t}) > 0$. This shows that P_t's above their mean are on the average associated with u_{1t}'s above their mean, (i.e., $u_{1t} > 0$). This implies that the quantity Q_t associated with this particular P_t is on the average above the true line $(\alpha + \beta P_t)$. This is true for all observations to the right of $E(P_t)$. Similarly, any P_t to the left of $E(P_t)$ is on the average associated with a u_{1t} below its mean, (i.e., $u_{1t} < 0$). This implies that quantities associated with prices below their mean $E(P_t)$ are on the average data points that lie below the true line. With this observed data, the estimated line using OLS will always be biased. In this case, the intercept estimate is biased downwards, whereas the slope estimate is biased upwards. This bias does not disappear with more data, as any new observation will on the average be either above the true line if $P_t > E(P_t)$ or below the line if $P_t < E(P_t)$. Hence, these OLS estimates are inconsistent.

Deaton (1997, p. 95) has a nice discussion of endogeneity problems in development economics. One important example pertains to farm size and farm productivity. Empirical studies using OLS have found an inverse relationship between productivity as measured by log(Output/Acre) and farm size as measured by (Acreage). This seems counter-intuitive as it suggests that smaller farms are more productive than larger farms. Economic explanations of this phenomenon include the observation that hired labor (which is typically used on large farms) is of lower quality than family labor (which is typically used on small farms). The latter needs less monitoring and can be entrusted with valuable animals and machinery. Another explanation is that this phenomenon is an optimal response by small farmers to uncertainty. It could also be a sign of inefficiency as farmers work too much on their own farms pushing their marginal productivity below market wage. How could this be an endogeneity problem? After all, the amount of land is outside the control of the farmer. This is true, but that does not mean that acreage is uncorrelated with the disturbance term. After all, size is unlikely to be independent of the *quality* of land. "Desert farms that are used for low-intensity animal grazing are typically larger than garden farms, where the land is rich and output/acre is high." In this case, land quality is negatively correlated with land size. It takes more acres to sustain a cow in West Texas than in less arid areas. This negative correlation between acres, the explanatory variable and quality of land which is an omitted variable included in the error term introduces endogeneity. This in turn results in downward bias of the OLS estimate of acreage on productivity.

Endogeneity can also be caused by *sample selection*. Gronau (1973) observed that women with small children had higher wages than women with no children. An economic explanation is that women with children have higher reservation wages and as a result fewer of them work. Of those that work, their observed wages are higher than those without children. The endogeneity works through the unobserved component in the working women's wage that induces her to work. This is positively correlated with the number of children she has and therefore introduces upward biases in the OLS estimate of the effect of the number of children on wages.

11.1.2 The Identification Problem

In general, we can think of any structural equation, say the first, as having one left hand side endogenous variable y_1, g_1 right hand side endogenous variables, and k_1 right hand side exogenous variables. The right hand side endogenous variables are correlated with the error term rendering OLS on this equation biased and inconsistent. Normally, for each endogenous variable, there exists a corresponding structural equation explaining its behavior in the model. We say that a system of simultaneous equations is *complete* if there are as many endogenous variables as there are equations. To correct for the simultaneous bias we need to replace the right hand side endogenous variables in this equation by variables which are highly correlated with the ones they are replacing but not correlated with the error term. Using the method of instrumental variable estimation, discussed below, we will see that these variables turn out to be the predictors obtained by regressing each right hand side endogenous variable on a subset of all the exogenous variables in the system. Let us assume that there are K exogenous variables in the simultaneous system. What set of exogenous variables should we use that would lead to consistent estimates of this structural equation? A search for the minimum set needed for consistency leads us to the *order condition* for identification.

The Order Condition for Identification: A *necessary* condition for identification of any structural equation is that the number of excluded exogenous variables from this equation are greater than or equal to the number of right hand side included endogenous variables. Let K be the number of exogenous variables in the system, then this condition requires $k_2 \geq g_1$, where $k_2 = K - k_1$.

Let us consider the demand and supply equations given in (11.13) and (11.14) but assume that the supply equation has in it an extra variable W_t denoting weather conditions. In this case the demand equation has one right hand side endogenous variable P_t, i.e., $g_1 = 1$ and one excluded exogenous variable W_t, making $k_2 = 1$. Since $k_2 \geq g_1$, this *order condition* is satisfied, in other words, based on the order condition alone we cannot conclude that the demand equation is unidentified. The supply equation, however, has $g_1 = 1$ and $k_2 = 0$, making this equation unidentified, since it does not satisfy the order condition for identification. Note that this condition is only *necessary* but not sufficient for identification. In other words, it is useful only if it is not satisfied, in which case the equation in question is not identified. Note that any linear combination of the new supply and demand equations would have a constant, price and weather. This looks like the supply equation but not like demand. This is why the supply equation is not identified. In order to prove once and for all whether the demand equation is identified, we need the *rank condition* for identification and this will be discussed in details in the Appendix to this chapter. Adding a third variable to the supply equation like the amount of fertilizer used F_t will not help the supply equation any, since a linear combination of supply and demand will still look like supply. However, it does help the identification of the demand equation. Denote by $\ell = k_2 - g_1$, the *degree of over-identification*. In (11.13) and (11.14) both equations are unidentified (or *under-identified*) with $\ell = -1$. When W_t is added to the supply equation, $\ell = 0$ for the demand equation, and it is just-identified. When both W_t and F_t are included in the supply equation, $\ell = 1$ and the demand equation is *over-identified*.

Without the use of matrices, we can describe a two-stage least squares method that will estimate the demand equation consistently. First, we run the right hand side endogenous variable P_t on a constant and W_t and get $\widehat{P}_t$, then replace P_t in the demand equation with $\widehat{P}_t$ and perform this second stage regression. In other words, the first step regression is

$$P_t = \pi_{11} + \pi_{12} W_t + v_t \tag{11.22}$$

with $\widehat{v}_t = P_t - \widehat{P}_t$ satisfying the OLS normal equations $\sum_{t=1}^T \widehat{v}_t = \sum_{t=1}^T \widehat{v}_t W_t = 0$. The second stage regression is

$$Q_t = \alpha + \beta \widehat{P}_t + \epsilon_t \tag{11.23}$$

with $\sum_{t=1}^T \widehat{\epsilon}_t = \sum_{t=1}^T \widehat{\epsilon}_t \widehat{P}_t = 0$. Using (11.13) and (11.23), we can write

$$\epsilon_t = \beta(P_t - \widehat{P}_t) + u_{1t} = \beta \widehat{v}_t + u_{1t} \tag{11.24}$$

so that $\sum_{t=1}^T \epsilon_t = \sum_{t=1}^T u_{1t}$ and $\sum_{t=1}^T \epsilon_t \widehat{P}_t = \sum_{t=1}^T u_{1t} \widehat{P}_t$ using the fact that $\sum_{t=1}^T \widehat{v}_t = \sum_{t=1}^T \widehat{v}_t \widehat{P}_t = 0$. So the new error ϵ_t behaves as the original disturbance u_{1t}. However, our right hand side variable is now $\widehat{P}_t$ which is independent of u_{1t} since it is a linear combination of exogenous variables only. We essentially decomposed P_t into two parts, the first part $\widehat{P}_t$ is a linear combination of exogenous variables and therefore, independent of the u_{1t}'s. The second part is $\widehat{v}_t$ which is correlated with u_{1t}. In fact, this is the source of simultaneous bias. The two parts $\widehat{P}_t$ and $\widehat{v}_t$ are orthogonal to each other by construction. Hence when the $\widehat{v}_t$'s become part of the new error ϵ_t, they are orthogonal to the new regressor $\widehat{P}_t$. Furthermore, $\widehat{P}_t$ is also independent of u_{1t}.

Why would this procedure not work on the estimation of (11.13) if the model is given by equations (11.13) and (11.14). The answer is that in (11.22) we will only have a constant, and no W_t. When we try to run the second-stage regression in (11.23) the regression will fail because of perfect multicollinearity between the constant and $\widehat{P}_t$. This will happen whenever the *order condition* is not satisfied and the equation is not identified, see Kelejian and Oates (1989). Hence, in order for it to succeed in the second stage we need at least one excluded exogenous variable from the demand equation that is in the supply equation, i.e., variables like W_t or F_t. Therefore, whenever the second-stage regression fails because of perfect multicollinearity between the right hand side regressors, this implies that the *order condition* of identification is not satisfied.

In general, if we are given an equation like

$$y_1 = \alpha_{12}y_2 + \beta_{11}X_1 + \beta_{12}X_2 + u_1 \tag{11.25}$$

the order condition requires the existence of at least one exogenous variable excluded from (11.25), say X_3. These extra exogenous variables like X_3 usually appear in other equations of our simultaneous equation model. In the first step regression we run

$$y_2 = \pi_{21}X_1 + \pi_{22}X_2 + \pi_{23}X_3 + v_2 \tag{11.26}$$

with the OLS residuals $\widehat{v}_2$ satisfying

$$\textstyle\sum_{t=1}^T \widehat{v}_{2t}X_{1t} = 0; \quad \sum_{t=1}^T \widehat{v}_{2t}X_{2t} = 0; \quad \sum_{t=1}^T \widehat{v}_{2t}X_{3t} = 0 \tag{11.27}$$

and in the second step, we run the regression

$$y_1 = \alpha_{12}\widehat{y}_2 + \beta_{11}X_1 + \beta_{12}X_2 + \epsilon_1 \tag{11.28}$$

where $\epsilon_1 = \alpha_{12}(y_2 - \widehat{y}_2) + u_1 = \alpha_{12}\widehat{v}_2 + u_1$. This regression will lead to consistent estimates, because

$$\begin{aligned}&\textstyle\sum_{t=1}^T \widehat{y}_{2t}\epsilon_{1t} = \sum_{t=1}^T \widehat{y}_{2t}u_{1t}; \quad \sum_{t=1}^T X_{1t}\epsilon_{1t} = \sum_{t=1}^T X_{1t}u_{1t};\\ &\textstyle\sum_{t=1}^T X_{2t}\epsilon_{1t} = \sum_{t=1}^T X_{2t}u_{1t}\end{aligned} \tag{11.29}$$

and u_{1t} is independent of the exogenous variables. In order to solve for 3 structural parameters α_{12}, β_{11} and β_{12} one needs three linearly independent OLS normal equations. $\sum_{t=1}^T \widehat{y}_{2t}\widehat{\epsilon}_{1t} = 0$ is a new piece of information provided y_2 is regressed on at least one extra variable besides X_1 and X_2. Otherwise, $\sum_{t=1}^T X_{1t}\widehat{\epsilon}_{1t} = \sum_{t=1}^T X_{2t}\widehat{\epsilon}_{1t} = 0$ are the only two linearly independent normal equations in three structural parameters.

What happens if there is another right hand side endogenous variable, say y_3? In that case (11.25) becomes

$$y_1 = \alpha_{12}y_2 + \alpha_{13}y_3 + \beta_{11}X_1 + \beta_{12}X_2 + u_1 \tag{11.30}$$

Now we need at least two exogenous variables that are excluded from (11.30) for the order condition to be satisfied, and the second stage regression to run. Otherwise, we will have less linearly independent equations than there are structural parameters to estimate, and the second stage regression will fail. Also, y_2 and y_3 should be regressed on the *same set* of exogenous variables. Furthermore, this set of second-stage regressors *should always include the right hand side* exogenous variables of (11.30). These two conditions will ensure consistency of the estimates. Let X_3 and X_4 be the excluded exogenous variables from (11.30). Our first step regression would regress y_2 and y_3 on X_1, X_2, X_3 and X_4 to get $\widehat{y}_2$ and $\widehat{y}_3$, respectively. The second stage regression would regress y_1 on $\widehat{y}_2$, $\widehat{y}_3$, X_1 and X_2. From the first step regressions we have

$$y_2 = \widehat{y}_2 + \widehat{v}_2 \quad \text{and} \quad y_3 = \widehat{y}_3 + \widehat{v}_3 \tag{11.31}$$

where $\widehat{y}_2$ and $\widehat{y}_3$ are linear combinations of the X's, and $\widehat{v}_2$ and $\widehat{v}_3$ are the residuals. The second stage regression has the following normal equations

$$\textstyle\sum_{t=1}^{T} \widehat{y}_{2t}\widehat{\epsilon}_{1t} = \sum_{t=1}^{T} \widehat{y}_{3t}\widehat{\epsilon}_{1t} = \sum_{t=1}^{T} X_{1t}\widehat{\epsilon}_{1t} = \sum_{t=1}^{T} X_{2t}\widehat{\epsilon}_{1t} = 0 \tag{11.32}$$

where $\widehat{\epsilon}_1$ denotes the residuals from the second stage regression. In fact

$$\epsilon_1 = \alpha_{12}\widehat{v}_2 + \alpha_{13}\widehat{v}_3 + u_1 \tag{11.33}$$

Now $\sum_{t=1}^{T} \epsilon_{1t}\widehat{y}_{2t} = \sum_{t=1}^{T} u_{1t}\widehat{y}_{2t}$ because $\sum_{t=1}^{T} \widehat{v}_{2t}\widehat{y}_{2t} = \sum_{t=1}^{T} \widehat{v}_{3t}\widehat{y}_{2t} = 0$. The latter holds because $\widehat{y}_2$, the predictor, is orthogonal to $\widehat{v}_2$, the residual. Also, $\widehat{y}_2$ is orthogonal to $\widehat{v}_3$ if y_2 is regressed on a set of X's that are a subset of the regressors included in the first step regression of y_3. Similarly, $\sum_{t=1}^{T} \epsilon_{1t}\widehat{y}_{3t} = \sum_{t=1}^{T} u_{1t}\widehat{y}_{3t}$ if y_3 is regressed on a set of exogenous variables that are a subset of the X's included in the first step regression of y_2. Combining these two conditions leads to the following fact: y_2 and y_3 have to be regressed on the *same set* of exogenous variables for the composite error term to behave like the original error. Furthermore these exogenous variables *should include the included X's* on the right hand side of the equation to be estimated, i.e., X_1 and X_2, otherwise, $\sum_{t=1}^{T} \epsilon_{1t}X_{1t}$ is not necessarily equal to $\sum_{t=1}^{T} u_{1t}X_{1t}$, because $\sum_{t=1}^{T} \widehat{v}_{2t}X_{1t}$ or $\sum_{t=1}^{T} \widehat{v}_{3t}X_{1t}$ are not necessarily zero. For further analysis along these lines, see problem 2.

11.2 Single Equation Estimation: Two-Stage Least Squares

In matrix form, we can write the first structural equation as

$$y_1 = Y_1\alpha_1 + X_1\beta_1 + u_1 = Z_1\delta_1 + u_1 \tag{11.34}$$

where y_1 and u_1 are $(T \times 1)$, Y_1 denotes the right hand side endogenous variables which is $(T \times g_1)$ and X_1 is the set of right hand side included exogenous variables which is $(T \times k_1)$, α_1 is of dimension g_1 and β_1 is of dimension k_1. $Z_1 = [Y_1, X_1]$ and $\delta_1' = (\alpha_1', \beta_1')$. We require the existence of excluded exogenous variables, from (11.34), call them X_2, enough to identify this equation. These excluded exogenous variables appear in the other equations in the simultaneous model. Let the set of all exogenous variables be $X = [X_1, X_2]$ where X is of dimension $(T \times k)$. For the order condition to be satisfied for equation (11.34) we must have $(k - k_1) \geq g_1$. If *all* the exogenous variables in the system are included in the first step regression, i.e., Y_1 is regressed on X to get $\widehat{Y}_1$, the resulting second stage least squares estimator obtained from regressing y_1 on $\widehat{Y}_1$ and X_1 is called two-stage least squares (2SLS). This method was proposed independently by Basmann (1957) and Theil (1953). In matrix form $\widehat{Y}_1 = P_X Y_1$ is the predictor of the right hand side endogenous variables, where P_X is the projection matrix $X(X'X)^{-1}X'$. Replacing Y_1 by $\widehat{Y}_1$ in (11.34), we get

$$y_1 = \widehat{Y}_1\alpha_1 + X_1\beta_1 + w_1 = \widehat{Z}_1\delta_1 + w_1 \tag{11.35}$$

where $\widehat{Z}_1 = [\widehat{Y}_1, X_1]$ and $w_1 = u_1 + (Y_1 - \widehat{Y}_1)\alpha_1$. Running OLS on (11.35) one gets

$$\widehat{\delta}_{1,2SLS} = (\widehat{Z}_1'\widehat{Z}_1)^{-1}\widehat{Z}_1'y_1 = (Z_1'P_XZ_1)^{-1}Z_1'P_Xy_1 \tag{11.36}$$

where the second equality follows from the fact that $\widehat{Z}_1 = P_XZ_1$ and the fact that P_X is idempotent. The former equality holds because $P_XX = X$, hence $P_XX_1 = X_1$, and $P_XY_1 = \widehat{Y}_1$. If there is only one right hand side endogenous variable, running the first-stage regression y_2 on X_1 and X_2 and testing that the coefficients of X_2 are *all* zero against the hypothesis that at least one of these coefficients is different from zero is a test for rank identification. In case of several right hand side endogenous variables, things get complicated, see Cragg and Donald (1996), but one can still run the first-stage regressions for each right hand side endogenous variable to make sure that at least one element of X_2 is significantly different from zero.[1] This is not sufficient for the rank condition but it is a good diagnostic for whether the rank condition fails. If we fail to meet this requirement we should question our 2SLS estimator.

Two-stage least squares can also be thought of as a simple instrumental variables estimator with the set of instruments $W = \widehat{Z}_1 = [\widehat{Y}_1, X_1]$. Recall that Y_1 is correlated with u_1, rendering OLS inconsistent. The idea of simple instrumental variables is to find a set of instruments, say W for Z_1 with the following properties: (1) plim $W'u_1/T = 0$, the instruments have to be *exogenous*, i.e., uncorrelated with the error term, otherwise this defeats the purpose of the instruments and result in inconsistent estimates. (2) plim $W'W/T = Q_w \neq 0$, where Q_w is finite and positive definite, the W's should not be perfectly multicollinear. (3) W should be highly correlated with Z_1, i.e., the instruments should be *highly relevant*, not *weak instruments* as we will explain shortly. In fact, plim $W'Z_1/T$ should be finite and of full rank $(k_1 + g_1)$. Premultiplying (11.34) by W', we get

$$W'y_1 = W'Z_1\delta_1 + W'u_1 \tag{11.37}$$

In this case, $W = \widehat{Z}_1$ is of the same dimension as Z_1, and since plim $W'Z_1/T$ is square and of full rank $(k_1 + g_1)$, the *simple instrumental variable* (IV) estimator of δ_1 becomes

$$\widehat{\delta}_{1,IV} = (W'Z_1)^{-1}W'y_1 = \delta_1 + (W'Z_1)^{-1}W'u_1 \tag{11.38}$$

with plim $\widehat{\delta}_{1,IV} = \delta_1$ which follows from (11.37) and the fact that plim $W'u_1/T = 0$.

Digression: In the general linear model, $y = X\beta + u$, X is the set of instruments for X. Premultiplying by X' we get $X'y = X'X\beta + X'u$ and using the fact that plim $X'u/T = 0$, one gets

$$\widehat{\beta}_{IV} = (X'X)^{-1}X'y = \widehat{\beta}_{OLS}.$$

This estimator is consistent as long as X and u are uncorrelated. In the simultaneous equation model for the first structural equation given in (11.34), the right hand side regressors Z_1 include endogenous variables Y_1 that are correlated with u_1. Therefore OLS on (11.34) will lead to inconsistent estimates, since the matrix of instruments $W = Z_1$, and Z_1 is correlated with u_1. In fact,

$$\widehat{\delta}_{1,OLS} = (Z_1'Z_1)^{-1}Z_1'y_1 = \delta_1 + (Z_1'Z_1)^{-1}Z_1'u_1$$

with plim $\widehat{\delta}_{1,OLS} \neq \delta_1$ since plim $Z_1'u_1/T \neq 0$.

Denote by $e_{1,OLS} = y_1 - Z_1\widehat{\delta}_{1,OLS}$ as the OLS residuals on the first structural equation, then

$$\text{plim } s_1^2 = \frac{e_{1,OLS}'e_{1,OLS}}{T-(g_1+k_1)} = \text{plim}\frac{u_1'\bar{P}_{Z_1}u_1}{T-(g_1+k_1)} = \sigma_{11} - \text{plim}\frac{u_1'Z_1(Z_1'Z_1)^{-1}Z_1'u_1}{T-(g_1+k_1)} \leq \sigma_{11},$$

since the last term is positive. Only if plim $Z_1'u_1/T$ is zero will plim $s_1^2 = \sigma_{11}$, otherwise it is smaller. OLS fits very well, it minimizes $(y_1 - Z_1\delta_1)'(y_1 - Z_1\delta_1)$. Since Z_1 and u_1 are correlated, OLS attributes part of the variation in y_1 that is due to u_1 incorrectly to the regressor Z_1.

Both the simple IV and OLS estimators can be interpreted as method of moments estimators. These were discussed in Chapter 2. For OLS, the population moment conditions are given by $E(X'u) = 0$ and the corresponding sample moment conditions yield $X'(y - X\widehat{\beta})/T = 0$. Solving for $\widehat{\beta}$ results in $\widehat{\beta}_{OLS}$. Similarly, the population moment conditions for the simple IV estimator in (11.37) are $E(W'u_1) = 0$ and the corresponding sample moment conditions yield $W'(y_1 - Z_1\widehat{\delta}_1)/T = 0$. Solving for $\widehat{\delta}_1$ results in $\widehat{\delta}_{1,IV}$ given in (11.38).

If $W = [\widehat{Y}_1, X_1]$, then (11.38) results in

$$\widehat{\delta}_{1,IV} = \begin{bmatrix} \widehat{Y}_1'Y_1 & \widehat{Y}_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{Y}_1'y_1 \\ X_1'y_1 \end{bmatrix} \tag{11.39}$$

which is the same as (11.36)

$$\widehat{\delta}_{1,2SLS} = \begin{bmatrix} \widehat{Y}_1'\widehat{Y}_1 & \widehat{Y}_1'X_1 \\ X_1'\widehat{Y}_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{Y}_1'y_1 \\ X_1'y_1 \end{bmatrix} \tag{11.40}$$

provided $\widehat{Y}_1'\widehat{Y}_1 = \widehat{Y}_1'Y_1$, and $X_1'Y_1 = X_1'\widehat{Y}_1$. The latter conditions hold because $\widehat{Y}_1 = P_XY_1$, and $P_XX_1 = X_1$.

In general, let X^* be our set of first stage regressors. An IV estimator with $\widehat{Y}_1^* = P_{X^*}Y_1$, i.e., with every right hand side y regressed on the *same set of regressors* X^*, will satisfy

$$\widehat{Y}_1^{*\prime}\widehat{Y}_1^* = \widehat{Y}_1^{*\prime}P_{X^*}Y_1 = \widehat{Y}_1^{*\prime}Y_1$$

In addition, for $X_1'\widehat{Y}_1^*$ to equal $X_1'Y_1$, X_1 has to be a subset of the regressors in X^*. Therefore X^* *should include* X_1 *and at least as many* X*'s from* X_2 *as is required for identification*, i.e., (at least g_1 of the X's from X_2). In this case, the IV estimator using $W^* = [\widehat{Y}_1^*, X_1]$ will result in the same estimator as that obtained by a two stage regression where in the first step $\widehat{Y}_1^*$ is obtained by regressing Y_1 on X^*, and in the second step y_1 is regressed on W^*. Note that these are the same conditions required for consistency of an IV estimator. Note also, that if this equation is just-identified, then there is exactly g_1 of the X's excluded from that equation. In other words, X_2 is of dimension $(T \times g_1)$, and $X^* = X$ is of dimension $T \times (g_1 + k_1)$. Problem 3 shows that 2SLS in this case reduces to an IV estimator with $W = X$, i.e.

$$\widehat{\delta}_{1,2SLS} = \widehat{\delta}_{1,IV} = (X'Z_1)^{-1}X'y_1 \tag{11.41}$$

Note that if the first equation is over-identified, then $X'Z_1$ is not square and (11.41) cannot be computed.

Rather than having W, the matrix of instruments, be of exactly the same dimension as Z_1 which is required for the expression in (11.38), one can define *a generalized instrumental variable* in terms of a general matrix W of dimension $T \times \ell$ where $\ell \geq g_1 + k_1$. The latter condition is the *order condition* for identification. In this case, $\widehat{\delta}_{1,IV}$ is obtained as GLS on (11.37). Using the fact that

$$\text{plim } W'u_1u_1'W/T = \sigma_{11} \text{ plim } W'W/T,$$

one gets

$$\widehat{\delta}_{1,IV} = (Z_1'P_WZ_1)^{-1}Z_1'P_Wy_1 = \delta_1 + (Z_1'P_WZ_1)^{-1}Z_1'P_Wu_1$$

with plim $\widehat{\delta}_{1,IV} = \delta_1$ and limiting covariance matrix σ_{11} plim $(Z_1'P_WZ_1/T)^{-1}$. Therefore, 2SLS can be obtained as a generalized instrumental variable estimator with $W = X$. This also means that 2SLS of δ_1 can be obtained as GLS on (11.34) after premultiplication by X', see problem 4. Note that GLS on (11.37) minimizes $(y_1 - Z_1\delta_1)'P_W(y_1 - Z_1\delta_1)$ which yields the first-order conditions

$$Z_1'P_W(y_1 - Z_1\widehat{\delta}_{1,IV}) = 0$$

the solution of which is $\widehat{\delta}_{1,IV} = (Z_1'P_WZ_1)^{-1}Z_1'P_Wy_1$. It can also be shown that 2SLS and the generalized instrumental variables estimators are special cases of a Generalized Method of Moments (GMM) estimator considered by Hansen (1982). See Davidson and MacKinnon (1993) and Hall (1993) for an introduction to GMM.

For the matrix $Z_1'P_WZ_1$ to be of full rank and invertible, a *necessary* condition is that W must be of full rank $\ell \geq (g_1 + k_1)$. This is in fact, the *order condition* of identification. If $\ell = g_1 + k_1$, then this equation is just-identified. Also, $W'Z_1$ is square and nonsingular. Problem 10 asks the reader to verify that the generalized instrumental variable estimator reduces to the simple instrumental variable estimator given in (11.38). Also, under just-identification the minimized value of the criterion function is zero.

One of the biggest problems with IV estimation is the choice of the instrumental variables W. We have listed some necessary conditions for this set of instruments to yield consistent estimators of the structural coefficients. However, different choices by different researchers may yield different estimates in finite samples. Using more instruments will yield more efficient IV estimation. Let W_1 and W_2 be two sets of IV's with W_1 being spanned by the space of W_2. In this case, $P_{W_2}W_1 = W_1$ and therefore, $P_{W_2}P_{W_1} = P_{W_1}$. For the corresponding IV estimators

$$\widehat{\delta}_{1,W_i} = (Z_1'P_{W_i}Z_1)^{-1}Z_1'P_{W_i}y_1 \quad \text{for } i = 1, 2$$

are both consistent for δ_1 as long as plim $W_i'u_1/T = 0$ and have asymptotic covariance matrices

$$\sigma_{11} \text{ plim } (Z_1'P_{W_i}Z_1/T)^{-1}$$

Note that $\widehat{\delta}_{1,W_2}$ is at least as efficient as $\widehat{\delta}_{1,W_1}$, if the difference in their asymptotic covariance matrices is positive semi-definite, i.e., if

$$\sigma_{11}\left[\text{plim}\frac{Z_1'P_{W_1}Z_1}{T}\right]^{-1} - \sigma_{11}\left[\text{plim}\frac{Z_1'P_{W_2}Z_1}{T}\right]^{-1}$$

is p.s.d. This holds, if $Z_1'P_{W_2}Z_1 - Z_1'P_{W_1}Z_1$ is p.s.d. This last condition holds since $P_{W_2} - P_{W_1}$ is idempotent. Problem 11 asks the reader to verify this result. $\widehat{\delta}_{1,W_2}$ is more efficient than $\widehat{\delta}_{1,W_1}$ since W_2 explains Z_1 at least as well as W_1. This seems to suggest that one should use as many instruments as possible. If T is large this is a good strategy. But, if T is finite, there will be a trade-off between this gain in asymptotic efficiency and the introduction of more finite sample bias in our IV estimator.

In fact, the more instruments we use, the more will $\widehat{Y}_1$ resemble Y_1 and the more bias is introduced in this second stage regression. The extreme case where Y_1 is perfectly predicted by $\widehat{Y}_1$ returns us to OLS which we know is biased. On the other hand, if our set of instruments have little ability in predicting Y_1, then the resulting instrumental variable estimator will be inefficient and its asymptotic distribution will not resemble its finite sample distribution, see Nelson and Startz (1990). If the number of instruments is fixed and the coefficients of the instruments in the first stage regression go to zero at the rate $1/\sqrt{T}$, indicating weak correlation, Staiger and Stock (1997) find that even as T increases, IV estimation is not consistent and has a nonstandard asymptotic distribution. Bound et al. (1995) recommend reporting the R^2 or the F-statistic of the first stage regression as a useful indicator of the quality of IV estimates.

Instrumental variables are important for obtaining consistent estimates when endogeneity is suspected. However, invalid instruments can produce meaningless results. How do we know whether our instruments are valid? Stock and Watson (2003) draw an analogy between a relevant instrument and a large sample. The more relevant the instrument, i.e., the more the variation in the right hand side endogenous variable that is explained by this instrument, the more accurate the resulting estimator. This is similar to the observation that the larger the sample size, the more accurate the estimator. They argue that the instruments should not just be relevant, but highly relevant if the normal distribution is to provide a good approximation to the sampling distribution of 2SLS. Weak instruments explain little of the variation in the right hand side endogenous variable they are instrumenting. This renders the normal distribution as a poor approximation to the sampling distribution of 2SLS, even if the sample size is large. Stock and Watson (2003) suggest a simple rule of thumb to check for weak instruments. If there is one right hand side endogenous variable, the first-stage regression can test for the significance of the excluded exogenous variables (or instruments) using an F-statistic. This first-stage F-statistic should be larger than 10.[2] Stock and Watson (2003) suggest that a first-stage F-statistic less than 10 indicates weak instruments which casts doubt on the validity of 2SLS, since with weak instruments, 2SLS will be biased even in large samples and the corresponding t-statistics and confidence intervals will be unreliable. Finding weak instruments, one can search for additional stronger instruments, or use alternative estimators than 2SLS which are less sensitive to weak instruments like LIML. Deaton (1997, p. 112) argues that it is difficult to find instruments that are exogenous while at the same time highly correlated with the endogenous variables they are instrumenting. He argues that it is easy to generate 2SLS estimates that are different from OLS but much harder to make the case that these 2SLS estimates are necessarily better than OLS. "Credible identification and estimation of structural equations almost always requires real creativity, and creativity cannot be reduced to a formula." Stock and Watson (2003, p. 371) show that for the case of a single right hand side endogenous variable with no included exogenous variables and one weak instrument, the distribution of the 2SLS estimators is non-normal even for large samples, with the mean of the sampling distribution of the 2SLS estimator approximately equal to the true coefficient plus the asymptotic bias of the OLS estimator divided by $(E(F)-1)$ where F is the first-stage F-statistic. If $E(F)=10$, then the large sample bias of 2SLS is (1/9) that of the large sample bias of OLS. They argue that this rule of thumb is an acceptable cutoff for most empirical applications.

2SLS is a single equation estimator. The focus is on a particular equation. $[y_1, Y_1, X_1]$ is specified and therefore all that is needed to perform 2SLS is the matrix X of all exogenous variables in the system. If a researcher is interested in a particular behavioral economic relationship which may be a part of a big model consisting of several equations, one need not specify the whole model to perform 2SLS on that equation, all that is needed is the matrix of all exogenous variables in that system. Empirical studies involving one structural equation, specify which right hand side variables are endogenous and proceed by estimating this equation via an IV procedure that usually includes all the feasible exogenous variables available to the researcher. If this set of exogenous variables does not include all the X's in the system, this estimation method is not 2SLS. However, it is a consistent IV method which we will call feasible 2SLS.

Substituting (11.34) in (11.36), we get

$$\widehat{\delta}_{1,2SLS} = \delta_1 + (Z_1'P_XZ_1)^{-1}Z_1'P_Xu_1 \tag{11.42}$$

with plim $\widehat{\delta}_{1,2SLS} = \delta_1$ and an asymptotic variance covariance matrix given by σ_{11} plim $(Z_1'P_XZ_1/T)^{-1}$. σ_{11} is estimated from the 2SLS residuals $\widehat{u}_1 = y_1 - Z_1\widehat{\delta}_{1,2SLS}$, by computing $s_{11} = \widehat{u}_1'\widehat{u}_1/(T-g_1-k_1)$. It is important to emphasize that s_{11} is obtained from the 2SLS residuals of the original equation (11.34), not (11.35). In other words, s_{11} is not the mean squared error (i.e., s^2) of the second stage regression given in (11.35). The latter regression has $\widehat{Y}_1$ in it and not Y_1. Therefore, the asymptotic variance covariance matrix of 2SLS can be estimated by $s_{11}(Z_1'P_XZ_1)^{-1} = s_{11}(\widehat{Z}_1'\widehat{Z}_1)^{-1}$. The t-statistics reported by 2SLS packages are based on the standard errors obtained from the square root of the diagonal elements of this matrix. These standard errors and t-statistics can be made robust for heteroskedasticity by computing $(\widehat{Z}_1'\widehat{Z}_1)^{-1}(\widehat{Z}_1'diag[\widehat{u}_i^2]\widehat{Z}_1)(\widehat{Z}_1'\widehat{Z}_1)^{-1}$ where $\widehat{u}_i$ denotes the i-th 2SLS residual. Wald type statistics for $H_o; R\delta_1 = r$ based on 2SLS estimates of δ_1 can be obtained as in equation (7.41) with $\widehat{\delta}_{1,2SLS}$ replacing $\widehat{\beta}_{OLS}$ and $var(\widehat{\delta}_{1,2SLS}) = s_{11}(\widehat{Z}_1'\widehat{Z}_1)^{-1}$ replacing $var(\widehat{\beta}_{OLS}) = s_{11}(X'X)^{-1}$. This can be made robust for heteroskedasticity by using the robust variance covariance matrix of $\widehat{\delta}_{1,2SLS}$ described above. The resulting Wald statistic is asymptotically distributed as χ_q^2 under the null hypothesis, with q being the number of restrictions imposed by $R\delta_1 = r$.

LM type tests for exclusion restrictions, like a subset of δ_1 set equal to zero can be performed by running the *restricted* 2SLS residuals on the matrix of *unrestricted* second stage regressors $\widehat{Z}_1$. The test statistic is given by TR_u^2 where R_u^2 denotes the uncentered R^2. This is asymptotically distributed as χ_q^2 under the null hypothesis, where q is the number of coefficients in δ_1 set equal to zero. Note that it does not matter whether the exclusion restrictions are imposed on β_1 or α_1, i.e., whether the excluded variables to be tested are endogenous or exogenous. An F-test for these exclusion restrictions can be constructed based on the restricted and unrestricted residual sums of squares from the second stage regression. The denominator of this F-statistic, however, is based on the unrestricted 2SLS residual sum of squares as reported by the 2SLS package. Of course, one has to adjust the numerator and denominator by the appropriate degrees of freedom. Under the null, this is asymptotically distributed as $F(q, T-(g_1+k_1))$. See Wooldridge (1990) for details. Also, see the over-identification test in Section 11.5.

Finite sample properties of 2SLS are model specific, see Mariano (2001) for a useful summary. One important result is that the absolute moments of positive order for 2SLS are finite up to the order of over-identification. So, for the 2SLS estimator to have a mean and variance, we need the degree of over-identification to be at least 2. This also means that for a just-identified model, no moments for 2SLS exist. For 2SLS, the absolute bias is an increasing function of the degree of over-identification. For the case of one right hand side included endogenous regressor, like equation (11.25), the size of OLS bias relative to 2SLS gets larger, the lower the degree of over-identification, the bigger the sample size, the higher the absolute value of the correlation between the disturbances and the endogenous regressor y_2 and the higher the *concentration parameter* μ^2. The latter is defined as $\mu^2 = E(y_2)'(P_X - P_{X_1})E(y_2)/\omega^2$ and $\omega^2 = \text{var}(y_{2t})$. In terms of MSE, larger values of μ^2 and large sample size favor 2SLS over OLS.

Another important single equation estimator is the *Limited Information Maximum Likelihood* (LIML) estimator which as the name suggests maximizes the likelihood function pertaining to the endogenous variables appearing in the estimated equation only. Excluded exogenous variables from this equation as well as the identifiability restrictions on other equations in the system are disregarded in the likelihood maximization. For details, see Anderson and Rubin (1950). LIML is invariant to the normalization choice of the dependent variable whereas 2SLS is not. This invariancy of LIML is in the spirit of a simultaneous equation model where normalization should not matter. Under just-identification 2SLS and LIML are equivalent. LIML is also known as the *Least Variance Ratio* (LVR) method, since the LIML estimates can be obtained by minimizing a ratio of two variances or equivalently the ratio of two residual sum of squares. Using equation (11.34), one can write

$$y_1^* = y_1 - Y_1\alpha = X_1\beta_1 + u_1$$

For a choice of α_1 one can compute y_1^* and regress it on X_1 to get the residual sum of squares RSS_1. Now regress y_1^* on X_1 and X_2 and compute the residual sum of squares RSS_2. Equation (11.34) states that X_2 does not enter the specification of that equation. In fact, this is where our identifying restrictions come from and the excluded exogenous variables that are used as instrumental variables. If these identifying restrictions are true, adding X_2 to the regression of y_1^* and X_1 should lead to minimal reduction in RSS_1. Therefore, the LVR method finds the α_1 that will minimize the ratio (RSS_1/RSS_2). After α_1 is estimated, β_1 is obtained from regressing y_1^* on X_1. In contrast, it can be shown that 2SLS minimizes $RSS_1 - RSS_2$. For details, see Johnston (1984) or Mariano (2001). Estimator bias is less of a problem for LIML than 2SLS. In fact as the number of instruments increase with the sample size such that their ratio is a constant, Bekker (1994) shows that 2SLS becomes *inconsistent* while LIML remains consistent. Both estimators are special cases of the following estimator:

$$\widehat{\delta}_1 = (Z_1'P_XZ_1 - \widehat{\theta}Z_1'Z_1)^{-1}(Z_1'P_Xy_1 - \widehat{\theta}Z_1'y_1)$$

with $\widehat{\theta} = 0$ yielding 2SLS, and $\widehat{\theta} =$ the *smallest eigenvalue* of $\{(D_1'D_1)^{-1}D_1'P_XD_1\}$ yielding LIML, where $D_1 = [y_1, Z_1]$.

Example 3: Simple Keynesian Model

For the data from the Economic Report of the President, given in Table 5.3, consider the simple Keynesian model with no government

$$C_t = \alpha + \beta Y_t + u_t \quad t = 1, 2, \ldots, T$$

with $Y_t = C_t + I_t$.

Table 11.1 Two-Stage Least Squares

Dependent Variable: CONSUMP
Method: Two-Stage Least Squares
Sample: 1959 2007
Included observations: 49
Instrument specification: INV
Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4081.653	3194.839	1.277577	0.2077
Y	0.685609	0.172415	3.976513	0.0002
R-squared	0.904339	Mean dependent var		16749.10
Adjusted R-squared	0.902304	S.D. dependent var		5447.060
S.E. of regression	1702.552	Sum squared resid		1.36E+08
F-statistic	15.81265	Durbin-Watson stat		0.014554
Prob (F-statistic)	0.000240	Second-Stage SSR		1.38E+09
J-statistic	0.000000	Instrument rank		2

The OLS estimates of the consumption function yield:

$$\begin{aligned} C_t = &-1343.31 + \ 0.979 \ \ Y_t + \text{residuals} \\ &\ (219.56) \quad\ (0.011) \end{aligned}$$

The 2SLS estimates assuming that I_t is exogenous and is the only instrument available, yield

$$\begin{aligned} C_t = &\ 4081.65 + \ 0.686 \ \ Y_t + \text{residuals} \\ &\ (3194.8) \quad (0.172) \end{aligned}$$

Table 11.1 reports these 2SLS results using EViews. Note that the OLS estimate of the intercept is understated, while that of the slope estimate is overstated indicating positive correlation between Y_t and the error as described in (11.6). The standard errors of 2SLS are bigger than those of OLS. This is always the case for an instrumental variable estimator as will be shown analytically for a simple regression in Example 4 below.

OLS on the reduced form equations yield

$$\begin{aligned} C_t = &\ 12982.72 + \ 2.18 \ \ I_t + \text{residuals} \quad \text{and} \quad Y_t = 12982.72 + \ 3.18 \ \ I_t + \text{residuals} \\ &\ (3110.4) \quad (1.74) \qquad\qquad\qquad\qquad\qquad\quad\ (3110.4) \quad (1.74) \end{aligned}$$

From example (A.5) in the Appendix, we see that $\widehat{\beta} = \widehat{\pi}_{12}/\widehat{\pi}_{22} = 2.18/3.18 = 0.686$ as described in (A.24). Also, $\widehat{\beta} = (\widehat{\pi}_{22} - 1)/\widehat{\pi}_{22} = (3.18 - 1)/3.18 = 2.18/3.18 = 0.686$ as described in (A.25). Similarly, $\widehat{\alpha} = \widehat{\pi}_{11}/\widehat{\pi}_{22} = \widehat{\pi}_{21}/\widehat{\pi}_{22} = 12982.72/3.18 = 4081.65$ as described in (A.22).

This confirms that under just-identification, the 2SLS estimates of the structural coefficients are identical to the *Indirect Least Squares* (ILS) estimates. The latter estimates uniquely solve for the structural parameter estimates from the reduced form estimates under just-identification. Note that in this case both 2SLS and ILS estimates of the consumption equation are identical to the simple IV estimator using I_t as an instrument for Y_t; i.e., $\widehat{\beta}_{IV} = m_{ci}/m_{yi}$ as shown in (A.24).

11.2.1 Spatial Lag Dependence

An alternative popular model for spatial lag dependence considered in Section 9.9 is given by:

$$y = \rho Wy + X\beta + \epsilon$$

where $\epsilon \sim \text{IIN}(0, \sigma^2)$, see Anselin (1988). Here y_i may denote output in region i which is affected by output of its neighbors through the spatial coefficient ρ and the weight matrix W. Recall from section 9.9, W is a *known* weight matrix with zero elements along its diagonal. It could be a contiguity matrix having elements 1 if its a neighboring region and zero otherwise. Usually this is normalized such that each row sums to 1. Alternatively, W could be based on distances from neighbors again normalized such that each row sums to 1. It is clear that the presence of Wy as a regressor introduces *endogeneity*. Assuming $(I_n - \rho W)$ nonsingular, one can solve for the reduced form model:

$$y = (I_n - \rho W)^{-1} X\beta + \epsilon^*$$

where $\epsilon^* = (I_n - \rho W)^{-1}\epsilon$ has mean zero and variance covariance matrix which has the same form as (9.38), i.e.,

$$\Sigma = E(\epsilon^* \epsilon^{*\prime}) = \sigma^2\Omega = \sigma^2(I_n - \rho W)^{-1}(I_n - \rho W')^{-1}$$

For $|\rho| < 1$, one obtains

$$(I_n - \rho W)^{-1} = I_n + \ \rho W + \rho^2 W^2 + \rho^3 W^3 + ...$$

Hence

$$E(y/X) = (I_n - \rho W)^{-1} X\beta = X\beta + \ \rho WX\beta + \rho^2 W^2 X\beta + \rho^3 W^3 X\beta + ...$$

This also means that

$$E(Wy/X) = W(I_n - \rho W)^{-1} X\beta = WX\beta + \ \rho W^2 X\beta + \rho^2 W^3 X\beta + \rho^3 W^4 X\beta + ...$$

Based on this last expression, Kelejian and Robinson (1993) and Kelejian and Prucha (1998) suggest the use of a subset of the following instrumental variables:

$$\{X, WX, W^2X, W^3X, W^4X, ...\}$$

Lee (2003) suggested using the optimal instrument matrix:

$$\{X, W(I_n - \widehat{\rho}W)^{-1} X\widehat{\beta}\}$$

where the values for $\widehat{\rho}$ and $\widehat{\beta}$ are obtained from a first stage IV estimator, using $\{X, WX\}$ as instruments, possibly augmented with W^2X. Note that Lee's (2003) instruments involve inverting a matrix of dimension n. Kelejian, et al. (2004) suggest an approximation based upon:

$$\{X, \sum_{s=0}^{r} \widehat{\rho}^s W^{s+1} X\widehat{\beta}\}$$

where r, the highest order of this approximation depends upon the sample size, with $r = o(n^{1/2})$. In their Monte Carlo experiments, they set $r = n^c$ where $c = 0.25, 0.35$, and 0.45. This is a natural application of 2SLS to deal with the problem of spatial lag dependence.

11.3 System Estimation: Three-Stage Least Squares

If the entire simultaneous equations model is to be estimated, then one should consider system estimators rather than single equation estimators. System estimators take into account the zero restrictions in every equation as well as the variance-covariance matrix of the disturbances of the whole system. One such system estimator is *Three-Stage Least Squares* (3SLS) where the structural equations are stacked on top of each other, just like a set of SUR equations,

$$y = Z\delta + u \tag{11.43}$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_G \end{bmatrix}; \ Z = \begin{bmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & \dots & 0 \\ \vdots & \dots & \vdots & \dots \\ 0 & \dots & & Z_G \end{bmatrix}; \ \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_G \end{bmatrix}; \ u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_G \end{bmatrix}$$

and u has zero mean and variance-covariance matrix $\Sigma \otimes I_T$, indicating the possible correlation among the disturbances of the different structural equations. $\Sigma = [\sigma_{ij}]$, with $E(u_i u_j') = \sigma_{ij} I_T$, for $i, j = 1, 2, \dots, G$. This $\otimes$ notation was used in Chapter 9 and defined in the Appendix to Chapter 7. Problem 4 shows that premultiplying the i-th structural equation by X' and performing GLS on the transformed equation results in 2SLS. For the system given in (11.43), the analogy is obtained by premultiplying by $(I_G \otimes X')$, i.e., each equation by X', and performing GLS on the whole system. The transformed error $(I_G \otimes X')u$ has a zero mean and variance-covariance matrix $\Sigma \otimes (X'X)$. Hence, GLS on the entire system obtains

$$\begin{aligned} \widehat{\delta}_{GLS} &= \{Z'(I_G \otimes X)[\Sigma^{-1} \otimes (X'X)^{-1}](I_G \otimes X')Z\}^{-1} \\ &\quad \{Z'(I_G \otimes X)[\Sigma^{-1} \otimes (X'X)^{-1}](I_G \otimes X')y \end{aligned} \tag{11.44}$$

which upon simplifying yields

$$\widehat{\delta}_{GLS} = \{Z'[\Sigma^{-1} \otimes P_X]Z\}^{-1}\{Z'[\Sigma^{-1} \otimes P_X]y\} \tag{11.45}$$

Σ has to be estimated to make this estimator operational. Zellner and Theil (1962), suggest getting the 2SLS residuals for the i-th equation, say $\widehat{u}_i = y_i - Z_i\widehat{\delta}_{i,2SLS}$ and estimating Σ by $\widehat{\Sigma} = [\widehat{\sigma}_{ij}]$ where

$$\widehat{\sigma}_{ij} = [\widehat{u}_i'\widehat{u}_j/(T - g_i - k_i)^{1/2}(T - g_j - k_j)^{1/2}] \quad \text{for} \quad i, j = 1, 2, \dots, G.$$

If $\widehat{\Sigma}$ is substituted for Σ in (11.45), the resulting estimator is called 3SLS:

$$\widehat{\delta}_{3SLS} = \{Z'[\widehat{\Sigma}^{-1} \otimes P_X]Z\}^{-1}\{Z'[\widehat{\Sigma}^{-1} \otimes P_X]y\} \tag{11.46}$$

The asymptotic variance-covariance matrix of $\widehat{\delta}_{3SLS}$ can be estimated by $\{Z'[\widehat{\Sigma}^{-1} \otimes P_X]Z\}^{-1}$. If the system of equations (11.43) is properly specified, 3SLS is more efficient than 2SLS. But if say, the second equation is improperly specified while the first equation is properly specified, then a system estimator like 3SLS will be contaminated by this misspecification whereas a single equation estimator like 2SLS on the first equation is not. So, if the first equation is of interest it does not pay to go to a system estimator in this case.

Two sufficient conditions exist for the equivalence of 2SLS and 3SLS, these are the following: (i) Σ is diagonal, and (ii) every equation is just identified. Problem 5 leads you step by step through these results. It is also easy to show, see problem 5, that a necessary and sufficient condition for 3SLS to be equivalent to 2SLS on each equation is given by

$$\sigma^{ij}\widehat{Z}_i'\bar{P}_{\widehat{Z}_j} = 0 \quad \text{for} \quad i,j = 1,2,\ldots,G$$

where $\widehat{Z}_i = P_X Z_i$, see Baltagi (1989). This is similar to the condition derived in the seemingly unrelated regressions case except it involves the set of second stage regressors of 2SLS. One can easily see that besides the two sufficient conditions given above, $\widehat{Z}_i'\bar{P}_{\widehat{Z}_j} = 0$ states that the set of second stage regressors of the i-th equation have to be a perfect linear combination of those in the j-th equation and vice versa. A similar condition was derived by Kapteyn and Fiebig (1981). If some equations in the system are over-identified while others are just-identified, the 3SLS estimates of the over-identified equations can be obtained by running 3SLS ignoring the just-identified equations. The 3SLS estimates of each just-identified equation differ from those of 2SLS by a vector which is a linear function of the 3SLS residuals of the over-identified equations, see Theil (1971) and problem 17.

11.4 Test for Over-Identification Restrictions

We emphasized *instrument relevance*, now we turn to *instrument exogeneity*. Under just-identification, one cannot statistically test instruments for exogeneity. This choice of exogenous instruments requires making an expert judgement based on knowledge of the empirical application. However, if the first structural equation is over-identified, i.e., the number of instruments ℓ is larger than the number of right hand side variables $(g_1 + k_1)$, then one can test these over-identifying restrictions. A likelihood ratio test for this over-identification condition based on maximum likelihood procedures was given by Anderson and Rubin (1950). This version of the test requires the computation of LIML. This was later modified by Basmann (1960) so that it could be based on the 2SLS procedure. Here we present a simpler alternative based on Davidson and MacKinnon (1993) and Hausman (1983). In essence, one is testing

$$H_o;\, y_1 = Z_1\delta_1 + u_1 \quad \text{versus} \quad H_1;\, y_1 = Z_1\delta_1 + W^*\gamma + u_1 \tag{11.47}$$

where $u_1 \sim \text{IID}(0, \sigma_{11} I_T)$. Let W be the matrix of instruments of full rank ℓ. Also, let W^* be a subset of instruments W, of dimension $(\ell - k_1 - g_1)$, that are linearly independent of $\widehat{Z}_1 = P_W Z_1$. In this case, the matrix $[\widehat{Z}_1, W^*]$ has full rank ℓ and therefore, spans the same space as W. A test for over-identification is a test for $\gamma = 0$. In other words, W^* has no ability to explain any variation in y_1 that is not explained by Z_1 using the matrix of instruments W.

If W^* is correlated with u_1 or the first structural equation (11.34) is misspecified, say by Z_1 not including some variables in W^*, then $\gamma \neq 0$. Hence, testing $\gamma = 0$ should be interpreted as a joint test for the validity of the matrix of instruments W *and* the proper specification of (11.34) see, Davidson and MacKinnon (1993). Testing H_o; $\gamma = 0$ can be obtained as an asymptotic F-test as follows:

$$\frac{(RRSS^* - URSS^*)/(\ell - (g_1 + k_1))}{URSS/(T - \ell)} \tag{11.48}$$

This is asymptotically distributed as $F(\ell-(g_1+k_1), T-\ell)$ under H_o. Using instruments W, we regress Z_1 on W and get $\widehat{Z}_1$, then obtain the restricted 2SLS estimate $\widetilde{\delta}_{1,2SLS}$ by regressing y_1 on $\widehat{Z}_1$. The restricted residual sum of squares from the *second stage regression* is $RRSS^* = (y_1-\widehat{Z}_1\widetilde{\delta}_{1,2SLS})'(y_1-\widehat{Z}_1\widetilde{\delta}_{1,2SLS})$. Next we regress y_1 on $\widehat{Z}_1$ and W^* to get the unrestricted 2SLS estimates $\widehat{\delta}_{1,2SLS}$ and $\widehat{\gamma}_{2SLS}$. The unrestricted residual sum of squares from the *second stage regression* is $URSS^* = (y_1-\widehat{Z}_1\widehat{\delta}_{1,2SLS}-W^*\widehat{\gamma}_{2SLS})'(y_1-\widehat{Z}_1\widehat{\delta}_{1,2SLS}-W^*\widehat{\gamma}_{2SLS})$. The URSS in (11.48) is the 2SLS residuals sum of squares from the unrestricted model which is obtained as follows $(y_1-Z_1\widehat{\delta}_{1,2SLS}-W^*\widehat{\gamma}_{2SLS})'(y_1-Z_1\widehat{\delta}_{1,2SLS}-W^*\widehat{\gamma}_{2SLS})$. URSS differs from $URSS^*$ in that Z_1 rather than $\widehat{Z}_1$ is used in obtaining the residuals. Note that this differs from the Chow-test in that the denominator is not based on $URSS^*$, see Wooldridge (1990).

This test does not require the construction of W^* for its implementation. This is because the model under H_1 is just-identified with as many regressor as there are instruments. This means that its

$$URSS^* = y_1'\bar{P}_W y_1 = y_1'y_1 - y_1'P_W y_1$$

see problem 10. It is easy to show, see problem 12, that

$$RRSS^* = y_1'\bar{P}_{\widehat{Z}_1} y_1 = y_1'y_1 - y_1'P_{\widehat{Z}_1} y_1$$

where $\widehat{Z}_1 = P_W Z_1$. Hence,

$$RRSS^* - URSS^* = y_1'P_W y_1 - y_1'P_{\widehat{Z}_1} y_1 \tag{11.49}$$

The test for over-identification can therefore be based on $RRSS^* - URSS^*$ divided by a *consistent* estimate of σ_{11}, say,

$$\widetilde{\sigma}_{11} = (y_1 - Z_1\widetilde{\delta}_{1,2SLS})'(y_1 - Z_1\widetilde{\delta}_{1,2SLS})/T \tag{11.50}$$

Problem 12 shows that the resulting test statistic is exactly that proposed by Hausman (1983). In a nutshell, the Hausman over-identification test regresses the 2SLS residuals $y_1 - Z_1\widetilde{\delta}_{1,2SLS}$ on the matrix W of *all* pre-determined variables in the model. The test statistic is T times the *uncentered* R^2 of this regression. See the Appendix to Chapter 3 for a definition of *uncentered* R^2. This test statistic is asymptotically distributed as χ^2 with $\ell-(g_1+k_1)$ degrees of freedom. Large values of this statistic reject the null hypothesis.

Alternatively, one can get this test statistic as a Gauss-Newton Regression (GNR) on the unrestricted model in (11.47). To see this, recall from section 8.4 that the GNR applies to a general nonlinear model $y_t = x_t(\beta) + u_t$. Using the set of instruments W, the GNR becomes

$$y - x(\widetilde{\beta}) = P_W X(\widetilde{\beta})b + \text{residuals}$$

where $\widetilde{\beta}$ denotes the restricted instrumental variable estimate of β under the null hypothesis and $X(\beta)$ is the matrix of derivatives with typical elements $X_{ij}(\beta) = \partial x_i(\beta)/\partial\beta_j$ for $j = 1,\dots,k$. Thus, the only difference between this GNR and that in Chapter 8 is that the regressors are multiplied by P_W, see Davidson and MacKinnon (1993, p. 226). Therefore, the GNR for (11.47) yields

$$y_1 - Z_1\widetilde{\delta}_{1,2SLS} = \widehat{Z}_1 b_1 + W^* b_2 + \text{residuals} \tag{11.51}$$

since $P_W[Z_1, W^*] = [\widehat{Z}_1, W^*]$ and $\widetilde{\delta}_{1,2SLS}$ is the restricted estimator under H_o; $\gamma = 0$. But, $[\widehat{Z}_1, W^*]$ spans the same space as W, see problem 12. Hence, the GNR in (11.51) is equivalent to running the 2SLS residuals on W and computing T times the *uncentered* R^2 as described above. Once again, it is clear that W^* need not be constructed.

The basic intuition behind the test for over-identification restriction rests on the fact that one can compute several legitimate IV estimators if all these instruments are relevant and exogenous. For example, suppose there are two instruments and one right hand side endogenous variable. Then one can compute two IV estimators using each instrument separately. If these IV estimators produce very different estimates, then may be one instrument or the other or both are not exogenous. The over-identification test we just described implicitly makes this comparison without actually computing all possible IV estimates. Exogenous instruments have to be uncorrelated with the disturbances. This suggests that the 2SLS residuals have to be uncorrelated with the instruments. This is the basis for the TR_u^2 test statistic. If all the instruments are exogenous, the regression coefficient estimates should all be not significantly different from zero and the R_u^2 should be low.

11.5 Hausman's Specification Test[3]

A critical assumption for the linear regression model $y = X\beta + u$ is that the set of regressors X are uncorrelated with the error term u. Otherwise, we have simultaneous bias and OLS is inconsistent. Hausman (1978) proposed a general specification test for H_o; $E(u/X) = 0$ versus H_1; $E(u/X) \neq 0$. Two estimators are needed to implement this test. The first estimator must be a consistent and efficient estimator of β under H_o which becomes inconsistent under H_1. Let us denote this efficient estimator under H_o by $\widehat{\beta}_o$. The second estimator, denoted by $\widehat{\beta}_1$, must be consistent for β under both H_o and H_1, but inefficient under H_o. Hausman's test is based on the difference between these two estimators $\widehat{q} = \widehat{\beta}_1 - \widehat{\beta}_o$. Under H_o; plim $\widehat{q}$ is zero, while under H_1; plim $\widehat{q} \neq 0$. Hausman (1978) shows that var$(\widehat{q})$ = var$(\widehat{\beta}_1)-$ var$(\widehat{\beta}_o)$ and Hausman's test becomes

$$m = \widehat{q}'[\text{var}(\widehat{q})]^{-1}\widehat{q} \tag{11.52}$$

which is asymptotically distributed under H_o as χ_k^2 where k is the dimension of β.

It remains to show that var$(\widehat{q})$ is the difference between the two variances. This can be illustrated for a single regressor case without matrix algebra, see Maddala (1992, page 507). First, one shows that cov$(\widehat{\beta}_o, \widehat{q}) = 0$. To prove this, consider a new estimator of β defined as $\widehat{\widehat{\beta}} = \widehat{\beta}_o + \lambda\widehat{q}$ where λ is an arbitrary constant. Under H_o, plim $\widehat{\widehat{\beta}} = \beta$ for every λ and

$$\text{var}(\widehat{\widehat{\beta}}) = \text{var}(\widehat{\beta}_o) + \lambda^2\text{var}(\widehat{q}) + 2\lambda\text{cov}(\widehat{\beta}_o, \widehat{q})$$

Since $\widehat{\beta}_o$ is efficient, var$(\widehat{\widehat{\beta}}) \geq$ var$(\widehat{\beta}_o)$ which means that λ^2 var$(\widehat{q})+2\lambda$ cov$(\widehat{\beta}_o, \widehat{q}) \geq 0$ for every λ. If cov$(\widehat{\beta}_o, \widehat{q}) > 0$, then for $\lambda = -$cov$(\widehat{\beta}_o, \widehat{q})/var(\widehat{q})$ the above inequality is violated. Similarly, if cov$(\widehat{\beta}_o, \widehat{q}) < 0$, then for λ=-cov$(\widehat{\beta}_o, \widehat{q})/var(\widehat{q})$ the above inequality is violated. Therefore, under H_o, for the above inequality to be satisfied for every λ, it must be the case that cov$(\widehat{\beta}_o, \widehat{q}) = 0$.

Now, $\widehat{q} = \widehat{\beta}_1 - \widehat{\beta}_o$ can be rewritten as $\widehat{\beta}_1 = \widehat{q} + \widehat{\beta}_o$ with var$(\widehat{\beta}_1)$ = var$(\widehat{q})+$ var$(\widehat{\beta}_o)+2$cov$(\widehat{q}, \widehat{\beta}_o)$. Using the fact that the last term is zero, we get the required result: var$(\widehat{q})$ = var$(\widehat{\beta}_1)-$ var$(\widehat{\beta}_o)$.

Example 4: Consider the simple regression without a constant

$$y_t = \beta x_t + u_t \qquad t = 1, 2, \ldots, T \quad \text{with} \quad u_t \sim \text{IIN}(0, \sigma^2)$$

and where β is a scalar. Under H_o; $E(u_t/x_t) = 0$ and OLS is efficient and consistent. Under H_1; OLS is not consistent. Using w_t as an instrumental variable, with w_t uncorrelated with u_t and preferably highly correlated with x_t, yields the following IV estimator of β:

$$\widehat{\beta}_{IV} = \sum_{t=1}^T y_t w_t / \sum_{t=1}^T x_t w_t = \beta + \sum_{t=1}^T w_t u_t / \sum_{t=1}^T x_t w_t$$

with plim $\widehat{\beta}_{IV} = \beta$ under H_o and H_1 and

$$var(\widehat{\beta}_{IV}) = \sigma^2 \sum_{t=1}^T w_t^2 / (\sum_{t=1}^T x_t w_t)^2 = \frac{\sigma^2}{\sum_{t=1}^T x_t^2}\left[\frac{1}{r_{xw}^2}\right]$$

where $r_{xw}^2 = (\sum_{t=1}^T x_t w_t)^2 / \sum_{t=1}^T w_t^2 \sum_{t=1}^T x_t^2$. Also,

$$\widehat{\beta}_{OLS} = \sum_{t=1}^T x_t y_t / \sum_{t=1}^T x_t^2 = \beta + \sum_{t=1}^T x_t u_t / \sum_{t=1}^T x_t^2$$

with plim $\widehat{\beta}_{OLS} = \beta$ under H_o but plim $\widehat{\beta}_{OLS} \neq \beta$ under H_1, with var$(\widehat{\beta}_{OLS}) = \sigma^2 / \sum_{t=1}^T x_t^2$. One can see that var$(\widehat{\beta}_{OLS}) \leq$ var$(\widehat{\beta}_{IV})$ since $0 \leq r_{xw}^2 \leq 1$. In fact for a weak IV, r_{xw}^2 is small and close to zero and the $var(\widehat{\beta}_{IV})$ blows up. Strong IV is where r_{xw}^2 is close to 1 and the closer is the $var(\widehat{\beta}_{IV})$ to var$(\widehat{\beta}_{OLS})$. In this case, $\widehat{q} = \widehat{\beta}_{IV} - \widehat{\beta}_{OLS}$ and plim $\widehat{q} \neq 0$ under H_1, while plim $\widehat{q} = 0$ under H_o, with

$$\text{var}(\widehat{q}) = \text{var}(\widehat{\beta}_{IV}) - \text{var}(\widehat{\beta}_{OLS}) = \frac{\sigma^2}{\sum_{t=1}^T x_t^2}\left[\frac{1}{r_{xw}^2} - 1\right] = \text{var}(\widehat{\beta}_{OLS})\left[\frac{1 - r_{xw}^2}{r_{xw}^2}\right]$$

Therefore, Hausman's test statistic is

$$m = \widehat{q}^2 r_{xw}^2 / \text{var}(\widehat{\beta}_{OLS})(1 - r_{xw}^2)$$

which is asymptotically distributed as χ_1^2 under H_o. Note that the same estimator of σ^2 is used for var$(\widehat{\beta}_{IV})$ and var$(\widehat{\beta}_{OLS})$. This is the estimator of σ^2 obtained under H_o.

The Hausman-test can also be obtained from the following *augmented regression*:

$$y_t = \beta x_t + \gamma \widehat{x}_t + \epsilon_t$$

where $\widehat{x}_t$ is the predicted value of x_t from regressing it on the instrumental variable w_t. Problem 13 asks the reader to show that Hausman's test statistic can be obtained by testing $\gamma = 0$.

The IV estimator assumes that w_t and u_t are uncorrelated. If this is violated in practice, then the IV estimator is *inconsistent* and the asymptotic bias can be aggravated if in addition this is a weak instrument. To see this,

$$\text{plim}\widehat{\beta}_{IV} = \beta + \text{plim}(\sum_{t=1}^T w_t u_t / \sum_{t=1}^T x_t w_t) = \beta + \frac{cov(w_t, u_t)}{cov(x_t, w_t)} = \beta + \frac{corr(w_t, u_t)}{corr(x_t, w_t)}\frac{\sigma_x}{\sigma_u}$$

where $cov(w_t, u_t)$ and $corr(x_t, w_t)$ denote the population covariance and correlation, respectively. Also, σ_x and σ_u denote the population standard deviations. If $corr(w_t, u_t) \neq 0$, there is an asymptotic bias in $\widehat{\beta}_{IV}$. If $corr(w_t, u_t)$ is small and the instrument is strong (with a large $corr(x_t, w_t)$), this bias could be small. However, this bias could be large, even if $corr(w_t, u_t)$ is small, in the case of weak instruments (small $corr(x_t, w_t)$). This warns researchers of using an instrument which they deem much better than x_t because it is less correlated with u_t. If this instrument is additionally weak, the bias of the resulting IV estimator will be enlarged due to the smaller $corr(x_t, w_t)$. In sum, *weak* instruments may have *large* asymptotic bias in practice even if they are *slightly* correlated with the error term.

In matrix form, the Durbin-Wu-Hausman test for the first structural equation is based upon the difference between OLS and IV estimation of (11.34) using the matrix of instruments W. In particular, the vector of contrasts is given by

$$\begin{aligned} \widehat{q} &= \widehat{\delta}_{1,IV} - \widehat{\delta}_{1,OLS} = (Z_1'P_W Z_1)^{-1}[Z_1'P_W y_1 - (Z_1'P_W Z_1)(Z_1'Z_1)^{-1}Z_1'y_1] \\ &= (Z_1'P_W Z_1)^{-1}[Z_1'P_W \bar{P}_{Z_1} y_1] \end{aligned} \tag{11.53}$$

Under the null hypothesis, $\widehat{q} = (Z_1'P_W Z_1)^{-1}Z_1'P_W \bar{P}_{Z_1}u_1$. The test for $\widehat{q} = 0$ can be based on the test for $Z_1'P_W\bar{P}_{Z_1}u_1$ having mean zero asymptotically. This last vector is of dimension $(g_1 + k_1)$. However, not all of its elements are necessarily random variables since $\bar{P}_{Z_1}$ may annihilate some columns of the second stage regressors $\widehat{Z}_1 = P_W Z_1$. In fact, all the included X's which are part of W, i.e., X_1, will be annihilated by $\bar{P}_{Z_1}$. Only the g_1 linearly independent variables $\hat{Y}_1 = P_W Y_1$ are not annihilated by $\bar{P}_{Z_1}$.

Our test focuses on the vector $\hat{Y}_1'\bar{P}_{Z_1}u_1$ having mean zero asymptotically. Now consider the artificial regression

$$y_1 = Z_1\delta_1 + \hat{Y}_1\gamma + \text{residuals} \tag{11.54}$$

Since $[Z_1, \hat{Y}_1]$, $[Z_1, \widehat{Z}_1]$, $[Z_1, Z_1 - \widehat{Z}_1]$ and $[Z_1, Y_1 - \widehat{Y}_1]$ all span the same column space, this regression has the same sum of squares residuals as

$$y_1 = Z_1\delta_1 + (Y_1 - \hat{Y}_1)\eta + \text{residuals} \tag{11.55}$$

The DWH-test may be based on either of these regressions. It is equivalent to testing $\gamma = 0$ in (11.54) or $\eta = 0$ in (11.55) using an F-test. This is asymptotically distributed as $F(g_1, T - 2g_1 - k_1)$. Davidson and MacKinnon (1993, p. 239) warn about interpreting this test as one of exogeneity of Y_1 (the variables in Z_1 not in the space spanned by W). They argue that what is being tested is the consistency of the OLS estimates of δ_1, not that every column of Z_1 is independent of u_1.

In practice, one may be sure about using W_2 as a set of IV's but is not sure whether some r additional variables in Z_1 are legitimate as instruments. The DWH-test in this case will be based upon the difference between two IV estimators for δ_1. The first is $\widehat{\delta}_{2,IV}$ based on W_2 and the second is $\widehat{\delta}_{1,IV}$ based on W_1. The latter set includes W_2 and the additional r variables in Z_1.

$$\begin{aligned} \widehat{\delta}_{2,IV} - \widehat{\delta}_{1,IV} &= (Z_1'P_{W_2}Z_1)^{-1}[Z_1'P_{W_2}y_1 - (Z_1'P_{W_2}Z_1)(Z_1'P_{W_1}Z_1)^{-1}Z_1'P_{W_1}y_1] \\ &= (Z_1'P_{W_2}Z_1)^{-1}Z_1'P_{W_2}(\bar{P}_{P_{W_1}Z_1})y_1 \end{aligned} \tag{11.56}$$

since $P_{W_2}P_{W_1} = P_{W_2}$. The DWH-test is based on this contrast having mean zero asymptotically. Once again this last vector has dimension $g_1 + k_1$ and not all its elements are necessarily random variables since $\bar{P}_{P_{W_1}Z_1}$ annihilates some columns of $P_{W_2}Z_1$. This test can be based on the following artificial regression:

$$y_1 = Z_1\delta_1 + P_{W_2}Z_1^*\gamma + \text{residuals} \tag{11.57}$$

where $P_{W_2}Z_1^*$ consists of the r columns of $P_{W_2}Z_1$ that are not annihilated by $\bar{P}_{W_1}$. Regression (11.57) is performed with W_1 as the set of IV's and $\gamma = 0$ is tested using an F-test.

11.6 Empirical Examples

Example 1: Crime in North Carolina. Cornwell and Trumbull (1994) estimated an economic model of crime using data on 90 counties in North Carolina observed over the years 1981–87. This data set is available on the Springer web site as CRIME.DAT. Here, we consider cross-section data for 1987 and reconsider the full panel data set in Chapter 12. Table 11.2 gives the OLS estimates relating the crime rate (which is an FBI index measuring the number of crimes divided by the county population) to a set of explanatory variables. This was done using Stata. All variables are in logs except for the regional dummies. The explanatory variables consist of the probability of arrest (which is measured by ratio of arrests to offenses), probability of conviction given arrest (which is measured by the ratio of convictions to arrests), probability of a prison sentence given a conviction (measured by the proportion of total convictions resulting in prison sentences); average prison sentence in days as a proxy for sanction severity. The number of police per capita as a measure of the county's ability to detect crime, the population density which is the county population divided by county land area, a dummy variable indicating whether the county is in the SMSA with population larger than 50,000. Percent minority, which is the

Table 11.2 Least Squares Estimates: Crime in North Carolina

Source	SS	df	MS		
Model	22.8072483	20	1.14036241	Number of obs = 90	F(20,69) = 19.71
Residual	3.99245334	69	.057861643	Prob > F = 0.0000	R-squared = 0.8510
Total	26.7997016	89	.301120243	Adj R-squared = 0.8078	Root MSE = .24054

lcrmrte	Coef.	Std. Err.	t	$P > \|t\|$	[95% Conf.	Interval]
lprbarr	−.4522907	.0816261	−5.54	0.000	−.6151303	−.2894511
lprbconv	−.3003044	.0600259	−5.00	0.000	−.4200527	−.180556
lprbpris	−.0340435	.1251096	−0.27	0.786	−.2836303	.2155433
lavgsen	−.2134467	.1167513	−1.83	0.072	−.4463592	.0194659
lpolpc	.3610463	.0909534	3.97	0.000	.1795993	.5424934
ldensity	.3149706	.0698265	4.51	0.000	.1756705	.4542707
lwcon	.2727634	.2198714	1.24	0.219	−.165868	.7113949
lwtuc	.1603777	.1666014	0.96	0.339	−.171983	.4927385
lwtrd	.1325719	.3005086	0.44	0.660	−.4669263	.7320702
lwfir	−.3205858	.251185	−1.28	0.206	−.8216861	.1805146
lwser	−.2694193	.1039842	−2.59	0.012	−.4768622	−.0619765
lwmfg	.1029571	.1524804	0.68	0.502	−.2012331	.4071472
lwfed	.3856593	.3215442	1.20	0.234	−.2558039	1.027123
lwsta	−.078239	.2701264	−0.29	0.773	−.6171264	.4606485
lwloc	−.1774064	.4251793	−0.42	0.678	−1.025616	.670803
lpctymle	.0326912	.1580377	0.21	0.837	−.2825855	.3479678
lpctmin	.2245975	.0519005	4.33	0.000	.1210589	.3281361
west	−.087998	.1243235	−0.71	0.481	−.3360167	.1600207
central	−.1771378	.0739535	−2.40	0.019	−.3246709	−.0296046
urban	−.0896129	.1375084	−0.65	0.517	−.3639347	.184709
_cons	−3.395919	3.020674	−1.12	0.265	−9.421998	2.630159

Table 11.3 Instrumental Variables (2SLS) Regression: Crime in North Carolina

Source	SS	df	MS		
Model	22.6350465	20	1.13175232	Number of obs	= 90
Residual	4.16465515	69	.060357321	F(20,69)	= 17.35
				Prob > F	= 0.0000
Total	26.7997016	89	.301120243	R-squared	= 0.8446
				Adj R-squared	= 0.7996
				Root MSE	= .24568

lcrmrte	Coef.	Std. Err.	t	$P > \|t\|$	[95% Conf.	Interval]
lprbarr	−.4393081	.2267579	−1.94	0.057	−.8916777	.0130615
lpolpc	.5136133	.1976888	2.60	0.011	.1192349	.9079918
lprbconv	−.2713278	.0847024	−3.20	0.002	−.4403044	−.1023512
lprbpris	−.0278416	.1283276	−0.22	0.829	−.2838482	.2281651
lavgsen	−.280122	.1387228	−2.02	0.047	−.5568663	−.0033776
ldensity	.3273521	.0893292	3.66	0.000	.1491452	.505559
lwcon	.3456183	.2419206	1.43	0.158	−.137	.8282366
lwtuc	.1773533	.1718849	1.03	0.306	−.1655477	.5202542
lwtrd	.212578	.3239984	0.66	0.514	−.433781	.8589371
lwfir	−.3540903	.2612516	−1.36	0.180	−.8752731	.1670925
lwser	−.2911556	.1122454	−2.59	0.012	−.5150789	−.0672322
lwmfg	.0642196	.1644108	0.39	0.697	−.263771	.3922102
lwfed	.2974661	.3425026	0.87	0.388	−.3858079	.9807402
lwsta	.0037846	.3102383	0.01	0.990	−.615124	.6226931
lwloc	−.4336541	.5166733	−0.84	0.404	−1.464389	.597081
lpctymle	.0095115	.1869867	0.05	0.960	−.3635166	.3825397
lpctmin	.2285766	.0543079	4.21	0.000	.1202354	.3369179
west	−.0952899	.1301449	−0.73	0.467	−.3549219	.1643422
central	−.1792662	.0762815	−2.35	0.022	−.3314437	−.0270888
urban	−.1139416	.143354	−0.79	0.429	−.3999251	.1720419
_cons	−1.159015	3.898202	−0.30	0.767	−8.935716	6.617686

Instrumented: lprbarr lpolpc
Instruments: lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd lwfir lwser lwmfg lwfed lwsta lwloc lpctymle lpctmin west central ltaxpc lmix

proportion of the county's population that is minority or non-white. Percent young male which is the proportion of the county's population that is males and between the ages of 15 and 24. Regional dummies for western and central counties. Opportunities in the legal sector captured by the average weekly wage in the county by industry. These industries are: construction; transportation, utilities and communication; wholesale and retail trade; finance, insurance and real estate; services; manufacturing; and federal, state and local government.

Results show that the probability of arrest as well as conviction given arrest have a negative and significant effect on the crime rate with estimated elasticities of −0.45 and −0.30 respectively. The probability of imprisonment given conviction as well as the sentence severity have a negative but insignificant effect on the crime rate. The greater the number of police per capita, the greater the number of reported crimes per capita. The estimated elasticity is 0.36 and it is significant. This could be explained by the fact that the larger the police force, the larger the reported crime. Alternatively, this could be an endogeneity problem with more crime resulting

Table 11.4 Hausman's Test: Crime in North Carolina

	Coefficients			
	(b) b2sls	(B) bols	(b–B) Difference	sqrt(diag(V_b–V_B)) S.E.
lprbarr	−.4393081	−.4522907	.0129826	.2115569
lpolpc	.5136133	.3610463	.152567	.1755231
lprbconv	−.2713278	−.3003044	.0289765	.0597611
lprbpris	−.0278416	−.0340435	.0062019	.0285582
lavgsen	−.280122	−.2134467	−.0666753	.0749208
ldensity	.3273521	.3149706	.0123815	.0557132
lwcon	.3456183	.2727634	.0728548	.1009065
lwtuc	.1773533	.1603777	.0169755	.0422893
lwtrd	.212578	.1325719	.0800061	.1211178
lwfir	−.3540903	−.3205858	−.0335045	.0718228
lwser	−.2911556	−.2694193	−.0217362	.0422646
lwmfg	.0642196	.1029571	−.0387375	.0614869
lwfed	.2974661	.3856593	−.0881932	.1179718
lwsta	.0037846	−.078239	.0820236	.1525764
lwloc	−.4336541	−.1774064	−.2562477	.293554
lpctymle	.0095115	.0326912	−.0231796	.0999404
lpctmin	.2285766	.2245975	.0039792	.0159902
west	−.0952899	−.087998	−.0072919	.0384885
central	−.1792662	−.1771378	−.0021284	.0187016
urban	−.1139416	−.0896129	−.0243287	.0405192

b = consistent under Ho and Ha; obtained from ivreg

B = inconsistent under Ha, efficient under Ho; obtained from regress

Test: Ho: difference in coefficients not systematic

chi2(20) = (b−B)'[(V_b−V_B)^(−1)](b−B)
= 0.87
Prob > chi2 = 1.0000

in the hiring of more police. The higher the density of the population the higher the crime rate. The estimated elasticity is 0.31 and it is significant. Returns to legal activity are insignificant except for wages in the service sector. This has a negative and significant effect on crime with an estimated elasticity of −0.27. Percent young male is insignificant, while percent minority is positive and significant with an estimated elasticity of 0.22. The central dummy variable is negative and significant while the western dummy variable is not significant. Also, the urban dummy variable is insignificant. Cornwell and Trumbull (1994) worried about the endogeneity of police per capita and the probability of arrest. They used as instruments two additional variables. Offense mix which is the ratio of crimes involving face to face contact (such as robbery, assault and rape) to those that do not. The rationale for using this variable is that arrest is facilitated by positive identification of the offender. The second instrument is per capita tax revenue. This is justified on the basis that counties with preferences for law enforcement will vote for higher taxes to fund a larger police force.

The 2SLS estimates are reported in Table 11.3. The probability of arrest has an estimated elasticity of -0.44 but now with a p-value of 0.057. The probability of conviction given arrest has an estimated elasticity of -0.27 still significant. The probability of imprisonment given conviction is still insignificant while the sentence severity is now negative and significant with an estimated elasticity of -0.28. Police per capita has a higher elasticity of 0.51 still significant. The remaining estimates are slightly affected. In fact, the Hausman test based on the difference between the OLS and 2SLS estimates is shown in Table 11.4. This is computed using Stata and it contrasts 20 slope coefficient estimates. The Hausman test statistic is 0.87 and is asymptotically distributed as χ^2_{20}. This is insignificant, and shows that the 2SLS and OLS estimates are not significantly different given this model specification and the specific choice of instruments. Note that this is a just-identified equation and one cannot test for over-identification.

Note that the 2sls estimates and the Hausman test are *not* robust to heteroskedasticity. Table 11.5 gives the 2sls estimates by running *ivregress* in Stata with the *robust* variance-covariance matrix option.

Note that the estimates remain the same as Table 11.3 but the standard errors for the right hand side endogenous variables Y_1 are now larger. The option (*estat endogenous*) generates the Hausman test which tests whether 2sls is different from OLS based now on the robust variance-covariance estimate. The F(2,67) statistic observed is 0.455 and has a p-value 0.636 which is not significant. This F-statistic could have been generated by an artificial regression as follows: Obtain the residuals from the first stage regressions, see Tables 11.6 and 11.7 below. Call these residuals *v1hat* and *v2hat*. Include them as additional variables in the original equation and run robust least squares. The robust Hausman test is equivalent to testing that the coefficients of these two residuals are jointly zero. The Stata commands (without showing the output of this artificial regression) and the resulting test statistic are shown below:

```
. quietly regress lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc ldensity
lwcon lwtuc lwtrd lwfir lwser lwmfg lwfed lwsta lwloc lpctymle lpctmin west
central urban v1hat v2hat if year==87, vce(robust)

. test v1hat v2hat

(1) v1hat = 0
(2) v2hat = 0

     F(2,67)  = 0.46
     Prob > F = 0.6361
```

This is the same statistic obtained above with *estat endogenous*.

Tables 11.6 and 11.7 give the first-stage regressions for the probability of arrest and police per capita. The R^2 of these regressions are 0.47 and 0.56, respectively. The F-statistics for the significance of all slope coefficients are 3.11 and 4.42, respectively. The additional instruments (offense mix and per capita tax revenue) are jointly significant in both regressions yielding F-statistics of 5.78 and 10.56 with p-values of 0.0048 and 0.0001, respectively. Although there are two right hand side endogenous regressors in the crime equation rather than one, the Stock and Watson 'rule of thumb' suggest that these instruments may be weak.

Table 11.5 Robust Variance Covariance (2SLS) Regression: Crime in North Carolina

Instrumental variables (2SLS) regression

Number of obs	=	90
Wald chi2(20)	=	1094.07
Prob > chi2	=	0.0000
R-squared	=	0.8446
Root MSE	=	.21511

lcrmrte	Coef.	Robust Std. Err.	z	$P>\|z\|$	[95% Conf.	Interval]
lprbarr	−.4393081	.311466	−1.41	0.158	−1.04977	.1711541
lpolpc	.5136133	.2483426	−2.07	0.039	.0268707	1.000356
lprbconv	−.2713278	.1138502	−2.38	0.017	−.4944701	−.0481855
lprbpris	−.0278416	.1339361	−0.21	0.835	−.2903516	.2346685
lavgsen	−.280122	.1204801	−2.33	0.020	−.5162587	−.0439852
ldensity	.3273521	.0983388	3.33	0.001	.1346116	.5200926
lwcon	.3456183	.1961291	1.76	0.078	−.0387877	.7300243
lwtuc	.1773533	.1942597	0.91	0.361	−.2033887	.5580952
lwtrd	.212578	.2297782	0.93	0.355	−.2377789	.6629349
lwfir	−.3540903	.2299624	−1.54	0.124	−.8048082	.0966276
lwser	−.2911556	.0865243	−3.37	0.001	−.4607401	−.121571
lwmfg	.0642196	.1459929	0.44	0.660	−.2219213	.3503605
lwfed	.2974661	.3089013	0.96	0.336	−.3079692	.9029015
lwsta	.0037846	.2861629	0.01	0.989	−.5570843	.5646535
lwloc	−.4336541	.4840087	−0.90	0.370	−1.382294	.5149856
lpctymle	.0095115	.2232672	0.04	0.966	−.4280842	.4471073
lpctmin	.2285766	.0531983	4.30	0.000	.1243099	.3328434
west	−.0952899	.1293715	−0.74	0.461	−.3488534	.1582736
central	−.1792662	.0651109	−2.75	0.006	−.3068813	−.0516512
urban	−.1139416	.1065919	−1.07	0.285	−.3228579	.0949747
_cons	−1.159015	3.791608	−0.31	0.760	−8.59043	6.2724

Instrumented: lprbarr lpolpc
Instruments: lprbconv lprbpris lavgsen ldensity lwcon lwtuc lwtrd lwfir lwser lwmfg lwfed lwsta lwloc lpctymle lpctmin west central urban ltaxpc lmix

One could obtain a lot of diagnostics on weak instruments after (*ivregress 2sls*) in Stata by issuing the command (*estat firststage*). This is done in Table 11.8. The option *forcenonrobust* is forcing these diagnostics to be done for a robust regression where the econometric theory behind their derivation need not apply.

This is a just-identified equation, so we cannot test over-identification. We have already seen the R-squared of the first stage regressions (0.47 and 0.56). These are not low enough to flag possible weak instruments. But these R-squared measures may be due mostly to the inclusion of the right hand side exogenous variables X_1. We want to know the additional contribution of the instruments $X_2 =$ (ltaxpc lmix) over and above X_1. The partial R-squared provide such measures and yield lower numbers. For lprbarr, this is 0.32. This is the correlation between lprbarr and the instruments $X_2 =$ (ltaxpc lmix) *after* including the right hand side exogenous variables X_1. The F(2,69) statistic tests the joint significance of the excluded instruments X_2 in the first stage regressions. These F-statistics of 6.58 and 6.68 are certainly less than 10.

Table 11.6 First Stage Regression: Probability of Arrest

Source	SS	df	MS			
Model	6.84874028	20	0.342437014			
Residual	7.59345096	69	0.110050014			
Total	14.4421912	89	0.162271812			

Number of obs	=	90
F(20,69)	=	3.11
Prob > F	=	0.0002
R-squared	=	0.4742
Adj R-squared	=	0.3218
Root MSE	=	0.33174

lprbarr	Coef.	Std. Err.	t	$P > \|t\|$	[95% Conf.	Interval]
lprbconv	–.1946392	.0877581	–2.22	0.030	–.3697119	–.0195665
lprbpris	–.0240173	.1732583	–0.14	0.890	–.3696581	.3216236
lavgsen	.1565061	.1527134	1.02	0.309	–.1481488	.4611611
ldensity	–.2211654	.0941026	–2.35	0.022	–.408895	–.0334357
lwcon	–.2024569	.3020226	–0.67	0.505	–.8049755	.4000616
lwtuc	–.0461931	.230479	–0.20	0.842	–.5059861	.4135999
lwtrd	.0494793	.4105612	0.12	0.904	–.769568	.8685266
lwfir	.050559	.3507405	0.14	0.886	–.6491492	.7502671
lwser	.0551851	.1500094	0.37	0.714	–.2440754	.3544456
lwmfg	.0550689	.2138375	0.26	0.798	–.3715252	.481663
lwfed	.2622408	.4454479	0.59	0.558	–.6264035	1.150885
lwsta	–.4843599	.3749414	–1.29	0.201	–1.232347	.2636277
lwloc	.7739819	.5511607	1.40	0.165	–.3255536	1.873517
lpctymle	–.3373594	.2203286	–1.53	–0.130	.776903	.1021842
lpctmin	–.0096724	.0729716	–0.13	–0.895	.1552467	.1359019
west	.0701236	.1756211	0.40	0.691	–.280231	.4204782
central	.0112086	.1034557	0.11	0.914	–.1951798	.217597
urban	–.0150372	.2026425	–0.07	0.941	–.4192979	.3892234
ltaxpc	–.1938134	.1755345	–1.10	0.273	–.5439952	.1563684
lmix	.2682143	.0864373	3.10	0.003	.0957766	.4406519
_cons	–4.319234	3.797113	–1.14	0.259	–11.89427	3.255799

This is the rule of thumb suggested by Stock and Watson for the case of one right hand side endogenous variable. Note that we computed these statistics above but without the robust variance-covariance matrix option. Shea's partial R-squared of 0.135 for lprbarr is the R-squared from running a regression of residuals on residuals. The first residuals come from regressing lprbarr on the right hand side included exogenous variables X_1. The second set of residuals come from regressing the right hand side included exogenous variables X_1 on the set of instruments $X_2 =$ (ltaxpc lmix).

Because we have more than one right hand side endogenous variable, Stock and Yogo (2005) suggest using the minimum eigenvalue of a matrix analog of the F-statistic originally proposed by Cragg and Donald (1993) to test for identification. A low minimum eigenvalue statistic indicate weak instruments. If there is only *one* right hand side endogenous variable, this reverts back to the F-statistic which is compared to 10 by the ad hoc rule of Stock and Watson. The critical values for this minimum eigenvalue statistic are dependent on how much relative bias we are willing to tolerate relative to OLS in case of weak instruments. This is available only when the degree of over-identification is *two or more*. This is why Stata does not report it in this just-identified example. However, Stata does report a second test which applies even for the

Table 11.7 First Stage Regression: Police per Capita

Source	SS	df	MS			
Model	6.99830344	20	0.349915172	Number of obs	=	90
Residual	5.46683312	69	0.079229465	F(20,69)	=	4.42
Total	12.4651366	89	0.140057714	Prob > F	=	0.0000
				R-squared	=	0.5614
				Adj R-squared	=	0.4343
				Root MSE	=	0.28148

lpolpc	Coef.	Std. Err.	t	$P > \|t\|$	[95% Conf.	Interval]
lprbconv	.0037744	.0744622	0.05	0.960	–.1447736	.1523223
lprbpris	–.0487064	.1470085	–0.33	0.741	–.3419802	.2445675
lavgsen	.3958972	.1295763	3.06	0.003	.1373996	.6543948
ldensity	.0201292	.0798454	0.25	0.802	–.1391581	.1794165
lwcon	–.5368469	.2562641	–2.09	0.040	–1.04808	–.025614
lwtuc	–.0216638	.1955598	–0.11	0.912	–.411795	.3684674
lwtrd	–.4207274	.3483584	–1.21	0.231	–1.115683	.2742286
lwfir	.0001257	.2976009	0.00	1.000	–.5935718	.5938232
lwser	.0973089	.1272819	0.76	0.447	–.1566116	.3512293
lwmfg	.1710295	.1814396	0.94	0.349	–.1909327	.5329916
lwfed	.8555422	.3779595	2.26	0.027	.1015338	1.609551
lwsta	–.1118764	.3181352	–0.35	0.726	–.7465387	.5227859
lwloc	1.375102	.4676561	2.94	0.004	.4421535	2.30805
lpctymle	.4186939	.1869473	2.24	0.028	.0457442	.7916436
lpctmin	–.0517966	.0619159	–0.84	0.406	–.1753154	.0717222
west	.1458865	.1490133	0.98	0.331	–.151387	.4431599
central	.0477227	.0877814	0.54	0.588	–.1273964	.2228419
urban	–.1192027	.1719407	–0.69	0.490	–.4622151	.2238097
ltaxpc	.5601989	.1489398	3.76	0.000	.2630721	.8573258
lmix	.2177256	.0733414	2.97	0.004	.0714135	.3640378
_cons	–16.33148	3.221824	–5.07	0.000	–22.75884	–9.904113

just-identified case. This is based on size distortions of the Wald test for the joint significance of the right hand side endogenous variables Y_1 at the 5% level. The observed minimum eigenvalue statistic of 5.31 is between the critical values of 7.03 and 4.58 and indicates a 2SLS relative bias of more than 10% and less than 15% when it should be 5%.

Example 2: Growth and Inequality Reconsidered. Lundberg and Squire (2003) estimate a two equation model of growth and inequality using 3SLS, see section 10.5 for the SUR specification where all the explanatory variables were assumed to be exogenous. The first equation relates Growth (dly) to education (adult years schooling: yrt), the share of government consumption in GDP (gov), M2/GDP (m2y), Inflation (inf), Sachs-Warner measure of openness (swo), changes in the terms of trade (dtot), initial income (f_pcy), dummy for 1980s (d80) and dummy for 1990s (d90). The second equation relates the Gini coefficient (gih) to education, M2/GDP, civil liberties index (civ), mean land Gini (mlg), mean land Gini interacted with a dummy for developing countries (mlgldc). The data contains 119 observations for 38 countries over the period 1965-1990, and can be obtained from

http://www.res.org.uk/economic/datasets/datasetlist.asp.

Table 11.8 Weak IV Diagnostics: The Crime Example

. estat firststage, forcenonrobust all
First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust $F(2, 69)$	Prob > F
lprbarr	0.4742	0.3218	0.1435	6.57801	0.0024
lpolpc	0.5614	0.4343	0.2344	6.68168	0.0022

Shea's partial R-squared

Variable	Shea's Partial R-sq.	Shea's Adj. Partial R-sq.
lprbarr	0.1352	–0.0996
lpolpc	0.2208	0.0093

Minimum eigenvalue statistic = 5.31166

Critical Values	# of endogenous regressors:			2
Ho: Instruments are weak	# of excluded instruments:			2
	5%	10%	20%	30%
2SLS relative bias		(not available)		
	10%	15%	20%	25%
2SLS Size of nominal 5% Wald test	7.03	4.58	3.95	3.63
LIML Size of nominal 5% Wald test	7.03	4.58	3.95	3.63

Education, government, M2/GDP, inflation, Sachs-Warner measure of openness, civil liberties index, mean land Gini, mean land Gini interacted with a dummy for developing countries(ldc) are assumed to be *endogenous*. Instruments include initial values of all variables (except land Gini and income), population, urban share, life expectancy, fertility, initial female literacy and democracy, arable area, dummies for oil and non-oil commodity exporters, and legal origin. Table 11.9 reports the 3SLS estimates of these two equations using the *reg3* command in Stata. The results replicate those reported in Table 1 of Lundberg and Squire (2003, p. 334). Allowing for endogeneity, these results still show that openness enhances growth and education reduces inequality.

Notes

1. A heteroskedasticity-robust statistic is recommended especially if y_2 has discrete characteristics.

2. Why 10? See the proof in Appendix 10.4 of Stock and Watson (2003).

3. This test is also known as the Durbin-Wu-Hausman test, following the work of Durbin (1954), Wu (1973) and Hausman (1978).

Table 11.9 Growth and Inequality: 3SLS Estimates

reg3 (Growth: dly = yrt gov m2y inf swo dtot f_pcy d80 d90) (Inequality: gih = yrt m2y civ mlg mlgldc), exog(commod f_civ f_dem f_dtot f_flit f_gov f_inf f_m2y f_swo f_yrt pop urb lex lfr marea oil legor_fr legor_ge legor_mx legor_sc legor_uk) endog(yrt gov m2y inf swo civ mlg mlgldc)

Three-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
Growth	119	9	2.34138	0.3905	65.55	0.0000
Inequality	119	5	7.032975	0.4368	94.28	0.0000

	Coef.	Std. Err.	z	$P>\|z\|$	[95% Conf.	Interval]
Growth						
yrt	–.0280625	.1827206	–0.15	0.878	–.3861882	.3300632
gov	–.0533221	.0447711	–1.19	0.234	–.1410718	.0344276
m2y	.0085368	.0199759	0.43	0.669	–.0306152	.0476889
inf	–.0008174	.0025729	–0.32	0.751	–.0058602	.0042254
swo	4.162776	.9499015	4.38	0.000	2.301003	6.024548
dtot	26.03736	23.05123	1.13	0.259	–19.14221	71.21694
f_pcy	–1.38017	.5488437	–2.51	0.012	–2.455884	–.3044564
d80	–1.560392	.545112	–2.86	0.004	–2.628792	–.4919922
d90	–3.413661	.6539689	–5.22	0.000	–4.695417	–2.131906
_cons	13.00837	3.968276	3.28	0.001	5.230693	20.78605
Inequality						
yrt	–1.244464	.4153602	–3.00	0.003	–2.058555	–.4303731
m2y	–.120124	.0581515	–2.07	0.039	–.2340989	–.0061492
civ	.2531189	.7277433	0.35	0.728	–1.173232	1.67947
mlg	.292672	.0873336	3.35	0.001	.1215012	.4638428
mlgldc	–.0547843	.0576727	–0.95	0.342	–.1678207	.0582522
_cons	33.13231	5.517136	6.01	0.000	22.31893	43.9457

Endogenous variables:	dly gih yrt gov m2y inf swo civ mlg mlgldc
Exogenous variables:	dtot f_pcy d80 d90 commod f_civ f_dem f_dtot f_flit f_gov f_inf f_m2y f_swo f_yrt pop urb lex lfr marea oil legor_fr legor_ge legor_mx legor_sc legor_uk

Problems

1. *The Inconsistency of OLS.* Show that the OLS estimator of δ in (11.14), which can be written as

$$\widehat{\delta}_{OLS} = \sum_{t=1}^{T} p_t q_t / \sum_{t=1}^{T} p_t^2$$

is not consistent for δ. **Hint:** Write $\widehat{\delta}_{OLS} = \delta + \sum_{t=1}^{T} p_t(u_{2t} - \bar{u}_2)/\sum_{t=1}^{T} p_t^2$, and use (11.18) to show that

$$\text{plim } \widehat{\delta}_{OLS} = \delta + (\sigma_{12} - \sigma_{22})(\delta - \beta)/[\sigma_{11} + \sigma_{22} - 2\sigma_{12}].$$

2. *When Is the IV Estimator Consistent?* Consider equation (11.30) and let X_3 and X_4 be the only two other exogenous variables in this system.

(a) Show that a two-stage estimator which regresses y_2 on X_1, X_2 and X_3 to get $y_2 = \hat{y}_2 + \hat{v}_2$, and y_3 on X_1, X_2 and X_4 to get $y_3 = \hat{y}_3 + \hat{v}_3$, and then regresses y_1 on $\hat{y}_2$, $\hat{y}_3$ and X_1 and

X_2 does not necessarily yield consistent estimators. **Hint:** Show that the composite error is $\epsilon_1 = (u_1 + \alpha_{12}\hat{v}_2 + \alpha_{13}\hat{v}_3)$ and $\sum_{t=1}^T \hat{\epsilon}_{1t}\hat{y}_{2t} \neq 0$, because $\sum_{t=1}^T \hat{y}_{2t}\hat{v}_{3t} \neq 0$. The latter does not hold because $\sum_{t=1}^T X_{3t}\hat{v}_{3t} \neq 0$. (This shows that if both y's are not regressed on the *same set* of X's, the resulting two stage regression estimates are not consistent).

(b) Show that the two-stage estimator which regresses y_2 and y_3 on X_2, X_3 and X_4 to get $y_2 = \hat{y}_2 + \hat{v}_2$ and $y_3 = \hat{y}_3 + \hat{v}_3$ and then regresses y_1 on $\hat{y}_2$, $\hat{y}_3$ and X_1 and X_2 is not necessarily consistent. **Hint:** Show that the composite error term $\epsilon_1 = u_1 + \alpha_{12}\hat{v}_2 + \alpha_{13}\hat{v}_3$ does not satisfy $\sum_{t=1}^T \hat{\epsilon}_{1t}X_{1t} = 0$, since $\sum_{t=1}^T \hat{v}_{2t}X_{1t} \neq 0$ and $\sum_{t=1}^T \hat{v}_{3t}X_{1t} \neq 0$. (This shows that if one of the *included* X's is not included in the first stage regression, then the resulting two-stage regression estimates are not consistent).

3. *Just-Identification and Simple IV.* If equation (11.34) is just-identified, then X_2 is of the same dimension as Y_1, i.e., both are $T \times g_1$. Hence, Z_1 is of the same dimension as X, both of dimension $T \times (g_1 + k_1)$. Therefore, $X'Z_1$ is a square nonsingular matrix of dimension $(g_1 + k_1)$. Hence, $(Z_1'X)^{-1}$ exists. Using this fact, show that $\widehat{\delta}_{1,2SLS}$ given by (11.36) reduces to $(X'Z_1)^{-1}X'y_1$. This is exactly the IV estimator with $W = X$, given in (11.41). Note that this is only feasible if $X'Z_1$ is square and nonsingular.

4. *2SLS Can Be Obtained as GLS.* Premultiply equation (11.34) by X' and show that the transformed disturbances $X'u_1 \sim (0, \sigma_{11}(X'X))$. Perform GLS on the resulting transformed equation and show that $\widehat{\delta}_{1,GLS}$ is $\widehat{\delta}_{1,2SLS}$, given by (11.36).

5. *The Equivalence of 3SLS and 2SLS.*

(a) Show that $\widehat{\delta}_{3SLS}$ given in (11.46) reduces to $\widehat{\delta}_{2SLS}$, when (i) Σ is diagonal, or (ii) every equation in the system is just-identified. **Hint:** For (i); show that $\widehat{\Sigma}^{-1} \otimes P_X$ is block-diagonal with the i-th block consisting of $P_X/\widehat{\sigma}_{ii}$. Also, Z is block-diagonal, therefore, $\{Z'[\widehat{\Sigma}^{-1} \otimes P_X]Z\}^{-1}$ is block-diagonal with the i-th block consisting of $\widehat{\sigma}_{ii}(Z_i'P_XZ_i)^{-1}$. Similarly, computing $Z'[\widehat{\Sigma}^{-1} \otimes P_X]y$, one can show that the i-th element of $\widehat{\delta}_{3SLS}$ is $(Z_i'P_XZ_i)^{-1}Z_i'P_Xy_i = \widehat{\delta}_{i,2SLS}$. For (ii); show that $Z_i'X$ is square and nonsingular under just-identification. Therefore, $\widehat{\delta}_{i,2SLS} = (X'Z_i)^{-1}X'y_i$ from problem 3. Also, from (11.44), we get

$$\begin{aligned}\widehat{\delta}_{3SLS} &= \{\text{diag}[Z_i'X](\widehat{\Sigma}^{-1} \otimes (X'X)^{-1})\text{diag}[X'Z_i]\}^{-1} \\ &\quad \{\text{diag}[Z_i'X](\widehat{\Sigma}^{-1} \otimes (X'X)^{-1})(I_G \otimes X')y\}.\end{aligned}$$

Using the fact that $Z_i'X$ is square, one can show that $\widehat{\delta}_{i,3SLS} = (X'Z_i)^{-1}X'y_i$.

(b) Premultiply the system of equations in (11.43) by $(I_G \otimes P_X)$ and let $y^* = (I_G \otimes P_X)y$, $Z^* = (I_G \otimes P_X)Z$ and $u^* = (I_G \otimes P_X)u$, then $y^* = Z^*\delta + u^*$.Show that OLS on this transformed model yields 2SLS on each equation in (11.43). Show that GLS on this model yields 3SLS (knowing the true Σ) given in (11.45). Note that $\text{var}(u^*) = \Sigma \otimes P_X$ and its generalized inverse is $\Sigma^{-1} \otimes P_X$. Use the Milliken and Albohali condition for the equivalence of OLS and GLS given in equation (9.7) of Chapter 9 to deduce that 3SLS is equivalent to 2SLS if $Z^{*\prime}(\Sigma^{-1} \otimes P_X)\bar{P}_{Z^*} = 0$. Show that this reduces to the following necessary and sufficient condition $\sigma^{ij}\widehat{Z}_i'\bar{P}_{\widehat{Z}_j} = 0$ for $i \neq j$, see Baltagi (1989). **Hint:** Use the fact that

$$Z^* = \text{diag}[P_XZ_i] = \text{diag}[\widehat{Z}_i] \quad \text{and} \quad \bar{P}_{Z^*} = \text{diag}[\bar{P}_{\widehat{Z}_i}].$$

Verify that the two sufficient conditions given in part (a) satisfy this necessary and sufficient condition.

6. Consider the following demand and supply equations:

$$\begin{aligned} Q &= a - bP + u_1 \\ Q &= c + dP + eW + fL + u_2 \end{aligned}$$

where W denotes weather conditions affecting supply, and L denotes the supply of immigrant workers available at harvest time.

   (a) Write this system in the matrix form given by equation (A.1) in the Appendix.

   (b) What does the *order-condition* for identification say about these two equations?

   (c) Premultiply this system by a nonsingular matrix $F = [f_{ij}]$, for $i, j = 1, 2$. What restrictions must the matrix F satisfy if the transformed model is to satisfy the same restrictions of the original model? Show that the first row of F is in fact the first row of an identity matrix, but the second row of F is not the second row of an identity matrix. What do you conclude?

7. Answer the same questions in problem 6 for the following model:

$$\begin{aligned} Q &= a - bP + cY + dA + u_1 \\ Q &= e + fP + gW + hL + u_2 \end{aligned}$$

where Y is real income and A is real assets.

8. Consider example (A.1) in the Appendix. Recall, that system of equations (A.3) and (A.4) are just-identified.

   (a) Construct ϕ for the demand equation (A.3) and show that $A\phi = (0, -f)'$ which is of rank 1 as long as $f \neq 0$. Similarly, construct ϕ for the supply equation (A.4) and show that $A\phi = (-c, 0)'$ which is of rank 1 as long as $c \neq 0$.

   (b) Using equation (A.17), show how the structural parameters can be retrieved from the reduced form parameters. Derive the reduced form equations for this system and verify the above relationships relating the reduced form and structural form parameters.

9. Derive the reduced form equations for the model given in problem 6, and show that the structural parameters of the second equation cannot be derived from the reduced form parameters. Also, show that there are more than one way of expressing the structural parameters of the first equation in terms of the reduced form parameters.

10. *Just-Identified Model.* Consider the just-identified equation

$$y_1 = Z_1\delta_1 + u_1$$

with W, the matrix of instruments for this equation of dimension $T \times \ell$ where $\ell = g_1 + k_1$ the dimension of Z_1. In this case, $W'Z_1$ is square and nonsingular.

   (a) Show that the *generalized instrumental variable* estimator given below (11.41) reduces to the *simple instrumental variable* estimator given in (11.38).

   (b) Show that the minimized value of the criterion function for this just-identified model is zero, i.e., show that $(y_1 - Z_1\widehat{\delta}_{1,IV})'P_W(y_1 - Z_1\widehat{\delta}_{1,IV}) = 0$.

   (c) Conclude that the residual sum of squares of the second stage regression of this just-identified model is the same as that obtained by regressing y_1 on the matrix of instruments W, i.e., show that $(y_1 - \widehat{Z}_1\widehat{\delta}_{1,IV})'(y_1 - \widehat{Z}_1\widehat{\delta}_{1,IV}) = y_1'\bar{P}_W y_1$ where $\widehat{Z}_1 = P_W Z_1$. **Hint:** Show that $P_{\widehat{Z}_1} = P_{P_W Z_1} = P_W$, under just-identification.

11. *The More Valid Instruments the Better.* Let W_1 and W_2 be two sets of instrumental variables for the first structural equation given in (11.34). Suppose that W_1 is spanned by the space of W_2. Verify that the resulting IV estimator of δ_1 based on W_2 is at least as efficient as that based on W_1. **Hint:** Show that $P_{W_2}W_1 = W_1$ and that $P_{W_2} - P_{W_1}$ is idempotent. Conclude that the difference in the corresponding asymptotic covariances of these IV estimators is positive semi-definite. (This shows that increasing the number of legitimate instruments should improve the asymptotic efficiency of an IV estimator).

12. *Testing for Over-Identification.* In testing $H_o;\ \gamma = 0$ versus $H_1;\ \gamma \neq 0$ in section 11.4, equation (11.47):

    (a) Show that the second stage regression of 2SLS on the unrestricted model $y_1 = Z_1\delta_1 + W^*\gamma + u_1$ with the matrix of instruments W yields the following residual sum of squares:

    $$URSS^* = y_1'\bar{P}_W y_1 = y_1'y_1 - y_1'P_W y_1$$

    Hint: Use the results of problem 10 for the just-identified case.

    (b) Show that the second stage regression of 2SLS on the restricted model $y_1 = Z_1\delta_1 + u_1$ with the matrix of instruments W yields the following residual sum of squares:

    $$RRSS^* = y_1'\bar{P}_{\widehat{Z}_1} y_1 = y_1'y_1 - y_1'P_{\widehat{Z}_1} y_1$$

    where $\widehat{Z}_1 = P_W Z_1$ and $P_{\widehat{Z}_1} = P_W Z_1(Z_1'P_W Z_1)^{-1}Z_1'P_W$. Conclude that $RRSS^* - URSS^*$ yields (11.49).

    (c) Consider the test statistic $(RRSS^* - URSS^*)/\widetilde{\sigma}_{11}$ where $\widetilde{\sigma}_{11}$ is given by (11.50) as the usual 2SLS residual sum of squares under H_o divided by T. Show that it can be written as Hausman's (1983) test statistic, i.e., TR_u^2 where R_u^2 is the *uncentered* R^2 of the regression of 2SLS residuals $(y_1 - Z_1\widetilde{\delta}_{1,2SLS})$ on the matrix of all pre-determined variables W. **Hint:** Show that the regression sum of squares $(y_1 - Z_1\widetilde{\delta}_{1,2SLS})'P_W(y_1 - Z_1\widetilde{\delta}_{1,2SLS}) = (RRSS^* - URSS^*)$ given in (11.49).

    (d) Verify that the test for H_o based on the GNR for the model given in part (a) yields the same TR_u^2 test statistic described in part (c).

13. *Hausman's Specification Test:* OLS versus 2SLS. This is based on Maddala (1992, page 511). For the simple regression

    $$y_t = \beta x_t + u_t \quad t = 1, 2 \ldots, T$$

    where β is scalar and $u_t \sim \text{IIN}(0, \sigma^2)$. Let w_t be an instrumental variable for x_t. Run x_t on w_t and get $x_t = \widehat{\pi}w_t + \widehat{\nu}_t$ or $x_t = \hat{x}_t + \widehat{\nu}_t$ where $\hat{x}_t = \widehat{\pi}w_t$.

    (a) Show that in the *augmented regression* $y_t = \beta x_t + \gamma\hat{x}_t + \epsilon_t$ a test for $\gamma = 0$ based on OLS from this regression yields Hausman's test-statistic. **Hint:** Show that $\widehat{\gamma}_{OLS} = \widehat{q}/(1 - r_{xw}^2)$ where

    $$r_{xw}^2 = \left(\sum_{t=1}^T x_t w_t\right)^2 / \sum_{t=1}^T w_t^2 \sum_{t=1}^T x_t^2.$$

    Next, show that $\text{var}(\widehat{\gamma}_{OLS}) = \text{var}(\widehat{\beta}_{OLS})/r_{xw}^2(1 - r_{xw}^2)$. Conclude that

    $$\widehat{\gamma}_{OLS}^2/\text{var}(\widehat{\gamma}_{OLS}) = \widehat{q}^2 r_{xw}^2/[\text{var}(\widehat{\beta}_{OLS})(1 - r_{xw}^2)]$$

    is the Hausman (1978) test statistic m given in section 11.5.

(b) Show that the same result in part (a) could have been obtained from the *augmented regression*

$$y_t = \beta x_t + \gamma \widehat{\nu}_t + \eta_t$$

where $\widehat{\nu}_t$ is the residual from the regression of x_t on w_t.

14. Consider the following structural equation: $y_1 = \alpha_{12}y_2 + \alpha_{13}y_3 + \beta_{11}X_1 + \beta_{12}X_2 + u_1$ where y_2 and y_3 are endogenous and X_1 and X_2 are exogenous. Also, suppose that the excluded exogenous variables include X_3 and X_4.

(a) Show that Hausman's test statistic can be obtained from the augmented regression:

$$y_1 = \alpha_{12}y_2 + \alpha_{13}y_3 + \beta_{11}X_1 + \beta_{12}X_2 + \gamma_2\hat{y}_2 + \gamma_3\hat{y}_3 + \epsilon_1$$

where $\hat{y}_2$ and $\hat{y}_3$ are predicted values from regressing y_2 and y_3 on $X = [X_1,X_2,X_3,X_4]$. Hausman's test is equivalent to testing H_o; $\gamma_2 = \gamma_3 = 0$. See equation (11.54).

(b) Show that the same results in part (a) hold if we had used the following augmented regression:

$$y_1 = \alpha_{12}y_2 + \alpha_{13}y_3 + \beta_{11}X_1 + \beta_{12}X_2 + \gamma_2\widehat{\nu}_2 + \gamma_3\widehat{\nu}_3 + \eta_1$$

where $\widehat{\nu}_2$ and $\widehat{\nu}_3$ are the residuals from running y_2 and y_3 on $X = [X_1, X_2, X_3, X_4]$. See equation (11.55). **Hint:** Show that the regressions in (a) and (b) have the same residual sum of squares.

15. For the artificial regression given in (11.55):

(a) Show that OLS on this model yields $\widehat{\delta}_{1,OLS} = \widehat{\delta}_{1,IV} = (Z_1'P_WZ_1)^{-1}Z_1'P_Wy_1$. **Hint:** $Y_1 - \hat{Y}_1 = \bar{P}_WY_1$. Use the FWL Theorem to residual out these variables in (11.55) and use the fact that $\bar{P}_WZ_1 = [\bar{P}_WY_1, 0]$.

(b) Show that the $\text{var}(\widehat{\delta}_{1,OLS}) = \widetilde{s}_{11}(Z_1'P_WZ_1)^{-1}$ where $\widetilde{s}_{11}$ is the mean squared error of the OLS regression in (11.55). Note that when $\eta \neq 0$ in (11.55), IV estimation is necessary and $\widetilde{s}_{11}$ underestimates σ_{11} and will have to be replaced by $(y_1 - Z_1\widehat{\delta}_{1,IV})'(y_1 - Z_1\widehat{\delta}_{1,IV})/T$.

16. *Recursive Systems.* A recursive system has two crucial features: B is a triangular matrix and Σ is a diagonal matrix. For this special case of the simultaneous equations model, OLS is still consistent, and under normality of the disturbances still maximum likelihood. Let us consider a specific example:

$$\begin{aligned} y_{1t} + \gamma_{11}x_{1t} + \gamma_{12}x_{2t} &= u_{1t} \\ \beta_{21}y_{1t} + y_{2t} + \gamma_{23}x_{3t} &= u_{2t} \end{aligned}$$

In this case, $B = \begin{bmatrix} 1 & 0 \\ \beta_{21} & 1 \end{bmatrix}$ is triangular and $\sum = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix}$ is assumed diagonal.

(a) Check the identifiability conditions of this recursive system.

(b) Solve for the reduced form and show that y_{1t} is only a function of the x_t's and u_{1t}, while y_{2t} is a function of the x_t's and a linear combination of u_{1t} and u_{2t}.

(c) Show that OLS on the first structural equation yields consistent estimates. **Hint:** There are no right hand side y's for the first equation. Show that despite the presence of y_1 in the second equation, OLS of y_2 on y_1 and x_3 yields consistent estimates. Note that y_1 is a function of u_1 only and u_1 and u_2 are not correlated.

(d) Under the normality assumption on the disturbances, the likelihood function conditional on the x's is given by

$$L(B, \Gamma, \Sigma) = (2\pi)^{-T/2}|B|^T|\Sigma|^{-T/2}\exp(-\frac{1}{2}\sum_{t=1}^T u_t'\Sigma^{-1}u_t)$$

where in this two equation case $u_t' = (u_{1t}, u_{2t})$. Since B is triangular, $|B| = 1$. Show that maximizing L with respect to B and Γ is equivalent to minimizing $Q = \sum_{t=1}^T u_t'\Sigma^{-1}u_t$. Conclude that when Σ is diagonal, Σ^{-1} is diagonal and $Q = \sum_{t=1}^T u_{1t}^2/\sigma_{11} + \sum_{t=1}^T u_{2t}^2/\sigma_{22}$. Hence, maximizing the likelihood with respect to B and Γ is equivalent to running OLS on each equation *separately*.

17. *Hausman's Specification Test: 2SLS Versus 3SLS.* This is based on Holly (1988). Consider the two-equations model,

$$\begin{aligned} y_1 &= \alpha y_2 + \beta_1 x_1 + \beta_2 x_2 + u_1 \\ y_2 &= \gamma y_1 + \beta_3 x_3 + u_2 \end{aligned}$$

where y_1 and y_2 are endogenous; x_1, x_2 and x_3 are exogenous (the y's and the x's are $n \times 1$ vectors). The standard assumptions are made on the disturbance vectors u_1 and u_2. With the usual notation, the model can also be written as

$$\begin{aligned} y_1 &= Z_1\delta_1 + u_1 \\ y_2 &= Z_2\delta_2 + u_2 \end{aligned}$$

The following notation will be used: $\widetilde{\delta} = 2SLS$, $\widetilde{\widetilde{\delta}} = 3SLS$, and the corresponding residuals will be denoted as $\widetilde{u}$ and $\widetilde{\widetilde{u}}$, respectively.

(a) Assume that $\alpha\gamma \neq 1$. Show that the 3SLS estimating equations reduce to

$$\begin{aligned} &\widetilde{\sigma}^{11}X'\widetilde{\widetilde{u}}_1 + \widetilde{\sigma}^{12}X'\widetilde{\widetilde{u}}_2 = 0 \\ &\widetilde{\sigma}^{12}Z_2'P_X\widetilde{\widetilde{u}}_1 + \widetilde{\sigma}^{22}Z_2'P_X\widetilde{\widetilde{u}}_2 = 0 \end{aligned}$$

where $X = (x_1, x_2, x_3)$, $\Sigma = [\sigma_{ij}]$ is the structural form covariance matrix, and $\Sigma^{-1} = [\sigma^{ij}]$ for $i, j = 1, 2$.

(b) Deduce that $\widetilde{\widetilde{\delta}}_2 = \widetilde{\delta}_2$ and $\widetilde{\widetilde{\delta}}_1 = \widetilde{\delta}_1 - (\widetilde{\sigma}_{12}/\widetilde{\sigma}_{22})(Z_1'P_XZ_1)^{-1}Z_1'P_X\widetilde{u}_2$. This proves that the 3SLS estimator of the over-identified second equation is equal to its 2SLS counterpart. Also, the 3SLS estimator of the just-identified first equation differs from its 2SLS (or indirect least squares) counterpart by a linear combination of the 2SLS (or 3SLS) residuals of the over-identified equation, see Theil (1971).

(c) How would you interpret a Hausman-type test where you compare $\widetilde{\widetilde{\delta}}_1$ and $\widetilde{\delta}_1$? Show that it is nR^2 where R^2 is the R-squared of the regression of $\widetilde{u}_2$ on the set of second stage regressors of both equations $\widehat{Z}_1$ and $\widehat{Z}_2$. **Hint:** See the solution by Baltagi (1989).

18. For the two-equation simultaneous model

$$\begin{aligned} &y_{1t} = \beta_{12}y_{2t} + \gamma_{11}x_{1t} + u_{1t} \\ &y_{2t} = \beta_{21}y_{1t} + \gamma_{22}x_{2t} + \gamma_{23}x_{3t} + u_{2t} \end{aligned}$$

With

$$X'X = \begin{bmatrix} 20 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 10 \end{bmatrix} X'Y = \begin{bmatrix} 5 & 10 \\ 40 & 20 \\ 20 & 30 \end{bmatrix} Y'Y = \begin{bmatrix} 3 & 4 \\ 4 & 8 \end{bmatrix}$$

(a) Determine the identifiability of each equation with the aid of the order and rank conditions for identification.

(b) Obtain the OLS normal equations for both equations. Solve for the OLS estimates.

(c) Obtain the 2SLS normal equations for both equations. Solve for the 2SLS estimates.

(d) Can you estimate these equations using Indirect Least Squares? Explain.

19. Laffer (1970) considered the following supply and demand equations for Traded Money:

$$\begin{aligned}\log(TM/P) &= \alpha_o + \alpha_1 \log(RM/P) + \alpha_2 \log\, i + u_1 \\ \log(TM/P) &= \beta_o + \beta_1 \log(Y/P) + \beta_2 \log\, i + \beta_3 \log(S1) + \beta_4 \log(S2) + u_2\end{aligned}$$

where

$$\begin{aligned}TM &= \text{Nominal total trade money} \\ RM &= \text{Nominal effective reserve money} \\ Y &= \text{GNP in current dollars} \\ S2 &= \text{Degree of market utilization} \\ i &= \text{short-term rate of interest} \\ S1 &= \text{Mean real size of the representative economic unit } (1939 = 100) \\ P &= \text{GNP price deflator } (1958 = 100)\end{aligned}$$

The basic idea is that trade credit is a line of credit and the unused portion represents purchasing power which can be used as a medium of exchange for goods and services. Hence, Laffer (1970) suggests that trade credit should be counted as part of the money supply. Besides real income and the short-term interest rate, the demand for real traded money includes $\log(S1)$ and $\log(S2)$. $S1$ is included to capture economies of scale. As $S1$ increases, holding everything else constant, the presence of economies of scale would mean that the demand for traded money would decrease. Also, the larger $S2$, the larger the degree of market utilization and the more money is needed for transaction purposes.

The data are provided on the Springer web site as LAFFER.ASC. This data covers 21 annual observations over the period 1946–1966. This was obtained from Lott and Ray (1992). Assume that (TM/P) and i are endogenous and the rest of the variables in this model are exogenous.

(a) Using the *order condition* for identification, determine whether the demand and supply equations are identified? What happens if you used the *rank condition* of identification?

(b) Estimate this model using OLS.

(c) Estimate this model using 2SLS.

(d) Estimate this model using 3SLS. Compare the estimates and their standard errors for parts (b), (c) and (d).

(e) Test the over-identification restriction of each equation.

(f) Run Hausman's specification test on each equation basing it on OLS and 2SLS.

(g) Run Hausman's specification test on each equation basing it on 2SLS and 3SLS.

20. The market for a certain good is expressed by the following equations:

$$\begin{aligned}D_t &= \alpha_0 - \alpha_1 P_t + \alpha_2 X_t + u_{1t} \qquad (\alpha_1, \alpha_2 > 0) \\ S_t &= \beta_0 + \beta_1 P_t + u_{2t} \qquad (\beta_1 > 0) \\ D_t &= S_t = Q_t\end{aligned}$$

where D_t is the quantity demanded, S_t is the quantity supplied, X_t is an exogenous demand shift variable. (u_{1t}, u_{2t}) is an IID random vector with zero mean and covariance matrix $\Sigma = [\sigma_{ij}]$ for $i, j = 1, 2$.

(a) Examine the identifiability of the model under the assumptions given above using the order and rank conditions of identification.

(b) Assuming the moment matrix of exogenous variables converge to a finite non-zero matrix, derive the simultaneous equation bias in the OLS estimator of β_1.

(c) If $\sigma_{12} = 0$ would you expect this bias to be positive or negative? Explain.

21. Consider the following three equations simultaneous model

$$\begin{aligned} y_1 &= \alpha_1 + \beta_2 y_2 + \gamma_1 X_1 + u_1 && (1) \\ y_2 &= \alpha_2 + \beta_1 y_1 + \beta_3 y_3 + \gamma_2 X_2 + u_2 && (2) \\ y_3 &= \alpha_3 + \gamma_3 X_3 + \gamma_4 X_4 + \gamma_5 X_5 + u_3 && (3) \end{aligned}$$

where the X's are exogenous and the y's are endogenous.

(a) Examine the identifiability of this system using the order and rank conditions.

(b) How would you estimate equation (2) by 2SLS? Describe your procedure step by step.

(c) Suppose that equation (1) was estimated by running y_2 on a constant X_2 and X_3 and the resulting predicted $\hat{y}_2$ was substituted in (1), and OLS performed on the resulting model. Would this estimating procedure yield consistent estimates of α_1, β_2 and γ_1? Explain your answer.

(d) How would you test for the over-identification restrictions in equation (1)?

22. *Equivariance of Instrumental Variables Estimators.* This is based on Sapra (1997). For the structural equation given in (11.34), let the matrix of instruments W be of dimension $T \times \ell$ where $\ell \geq g_1 + k_1$ as described below (11.41). Then the corresponding instrumental variable estimator of δ_1 given below (11.41) is $\widehat{\delta}_{1,IV}(y_1) = (Z_1' P_W Z_1)^{-1} Z_1' P_W y_1$.

(a) Show that this IV estimator is an equivariant estimator of δ_1, i.e., show that for any linear transformation $y_1^* = a y_1 + Z_1 b$ where a is a positive scalar and b is an $(\ell \times 1)$ real vector, the following relationship holds:

$$\widehat{\delta}_{1,IV}(y_1^*) = a\widehat{\delta}_{1,IV}(y_1) + b.$$

(b) Show that the variance estimator

$$\widehat{\sigma}^2(y_1) = (y_1 - Z_1\widehat{\delta}_{1,IV}(y_1))'(y_1 - Z_1\widehat{\delta}_{1,IV}(y_1))/T$$

is equivariant for σ^2, i.e., show that $\widehat{\sigma}^2(y_1^*) = a^2\widehat{\sigma}^2(y_1)$.

23. *Identification and Estimation of a Simple Two-Equation Model.* This is based on Holly (1987). Consider the following two equation model

$$\begin{aligned} y_{t1} &= \alpha + \beta y_{t2} + u_{t1} \\ y_{t2} &= \gamma + y_{t1} + u_{t2} \end{aligned}$$

where the y's are endogenous variables and the u's are serially independent disturbances that are identically distributed with zero means and nonsingular covariance matrix $\Sigma = [\sigma_{ij}]$ where $E(u_{ti}u_{tj}) = \sigma_{ij}$ for $i, j = 1, 2$, and all $t = 1, 2, \ldots, T$. The reduced form equations are given by

$$y_{t1} = \pi_{11} + \nu_{t1} \quad \text{and} \quad y_{t2} = \pi_{21} + \nu_{t2}$$

with $\Omega = [\omega_{ij}]$ where $E(\nu_{ti}\nu_{tj}) = \omega_{ij}$ for $i, j = 1, 2$ and all $t = 1, 2, \ldots, T$.

(a) Examine the identification of this system of two equations when no further information is available.

(b) Repeat part (a) when $\sigma_{12} = 0$.

(c) Assuming $\sigma_{12} = 0$, show that $\widehat{\beta}_{OLS}$, the OLS estimator of β in the first equation is not consistent.

(d) Assuming $\sigma_{12} = 0$, show that an alternative consistent estimator of β is an IV estimator using $z_t = [(y_{t2} - \bar{y}_2) - (y_{t1} - \bar{y}_1)]$ as an instrument for y_{t2}.

(e) Show that the IV estimator of β obtained from part (d) is also an indirect least squares estimator of β. **Hint:** See the solution by Singh and Bhat (1988).

24. *Errors in Measurement and the Wald (1940) Estimator.* This is based on Farebrother (1985). Let y_i^* be permanent consumption and X^* be permanent income, both are measured with error:

$$y_i^* = \beta x_i^* \quad \text{where} \quad y_i = y_i^* + \epsilon_i \quad \text{and} \quad x_i = x_i^* + u_i \quad \text{for} \quad i = 1, 2, \ldots, n.$$

Let x_i^*, ϵ_i and u_i be independent normal random variables with zero means and variances σ_*^2, σ_ϵ^2 and σ_u^2, respectively. Wald (1940) suggested the following estimator of β: Order the sample by the x_i's and split the sample into two. Let $(\bar{y}_1, \bar{x}_1)$ be the sample mean of the first half of the sample and $(\bar{y}_2, \bar{x}_2)$ be the sample mean of the second half of this sample. Wald's estimator of β is $\widehat{\beta}_W = (\bar{y}_2 - \bar{y}_1)/(\bar{x}_2 - \bar{x}_1)$. It is the slope of the line joining these two sample mean observations.

(a) Show that $\widehat{\beta}_W$ can be interpreted as a simple IV estimator with instrument

$$\begin{aligned} z_i &= 1 \quad \text{for} \quad x_i \geq \text{median}(x) \\ &= -1 \quad \text{for} \quad x_i < \text{median}(x) \end{aligned}$$

where median(x) is the sample median of $x_1, x_2, \ldots, x_n$.

(b) Define $w_i = \rho^2 x_i^* - \tau^2 u_i$ where $\rho^2 = \sigma_u^2/(\sigma_u^2 + \sigma_*^2)$ and $\tau^2 = \sigma_*^2/(\sigma_u^2 + \sigma_*^2)$. Show that $E(x_i w_i) = 0$ and that $w_i \sim N(0, \sigma_*^2 \sigma_u^2/(\sigma_*^2 + \sigma_u^2))$.

(c) Show that $x_i^* = \tau^2 x_i + w_i$ and use it to show that

$$E(\widehat{\beta}_W / x_1, \ldots, x_n) = E(\widehat{\beta}_{OLS} / x_1, \ldots, x_n) = \beta \tau^2.$$

Conclude that the exact small sample bias of $\widehat{\beta}_{OLS}$ and $\widehat{\beta}_W$ are the same.

25. *Comparison of t-ratios.* This is based on Holly (1990). Consider the two equations model

$$y_1 = \alpha y_2 + X\beta + u_1 \quad \text{and} \quad y_2 = \gamma y_1 + X\beta + u_2$$

where α and γ are scalars, y_1 and y_2 are $T \times 1$ and X is a $T \times (K-1)$ matrix of exogenous variables. Assume that $u_i \sim N(0, \sigma_i^2 I_T)$ for $i = 1, 2$. Show that the t-ratios for H_o^a; $\alpha = 0$ and H_o^b; $\gamma = 0$ using $\widehat{\alpha}_{OLS}$ and $\widehat{\gamma}_{OLS}$ are the same. Comment on this result. **Hint:** See the solution by Farebrother (1991).

26. *Degeneration of Feasible GLS to 2SLS in a Limited Information Simultaneous Equations Model.* This is based on Gao and Lahiri (2000). Consider a simple limited information simultaneous equations model,

$$\begin{aligned} y_1 &= \gamma y_2 + u, && (1) \\ y_2 &= X\beta + v, && (2) \end{aligned}$$

where y_1, y_2 are $N \times 1$ vectors of observations on two endogenous variables. X is $N \times K$ matrix of predetermined variables of the system, and $K \geq 1$ such that (1) is identified. Each row of

(u, v) is assumed to be i.i.d. $(0, \Sigma)$, and Σ is p.d. In this case, $\widehat{\gamma}_{2SLS} = (y_2' P_X y_2)^{-1} y_2' P_X y_1$, where $P_X = X(X'X)^{-1}X'$. The residuals $\widehat{u} = y_1 - \widehat{\gamma}_{2SLS} y_2$ and $\widehat{v} = My_2$, where $M = I_N - P_X$ are used to generate a consistent estimate for Σ

$$\widehat{\Sigma} = \frac{1}{N} \begin{bmatrix} \widehat{u}'\widehat{u} & \widehat{u}'\widehat{v} \\ \widehat{v}'\widehat{u} & \widehat{v}'\widehat{v} \end{bmatrix}.$$

Show that a feasible GLS estimate of γ using $\widehat{\Sigma}$ degenerates to $\widehat{\gamma}_{2SLS}$.

27. *Equality of Two IV Estimators.* This is based on Qian (1999). Consider the following linear regression model:

$$y_i = x_i'\beta + \epsilon = x_{1i}'\beta_1 + x_{2i}'\beta_2 + \epsilon_i, \qquad i = 1, 2, \ldots, N, \tag{1}$$

where the dimensions of x_i', x_{1i}' and x_{2i}' are $1 \times K$, $1 \times K_1$ and $1 \times K_2$, respectively, with $K = K_1 + K_2$. x_i may be correlated with ϵ_i, but we have instruments z_i such that $E(\epsilon_i | z_i') = 0$ and $E(\epsilon_i^2 | z_i') = \sigma^2$. Partition the instruments into two subsets: $z_i' = (z_{1i}', z_{2i}')$, where the dimensions of z_i, z_{1i}, and z_{2i} are L, L_1 and L_2, with $L = L_1 + L_2$. Assume that $E(z_{1i}x_i')$ has full column rank (so $L_1 \geq K$); this ensures that β can be estimated consistently using the subset of instruments z_{1i} only or using the entire set $z_i' = (z_{1i}', z_{2i}')$. We also assume that (y_i, x_i', z_i') is covariance stationary.

Define $P_A = A(A'A)^{-1}A'$ and $M_A = I - P_A$ for any matrix A with full column rank. Let $X = (x_1, \ldots, x_N)'$, and similarly for X_1, X_2, y, Z_1, Z_2 and Z. Define $\widehat{X} = P_{[Z_1]}X$ and $\widehat{\beta} = (\widehat{X}'\widehat{X})^{-1}\widehat{X}'y$, so that $\widehat{\beta}$ is the instrumental variables (IV) estimator of (1) using Z_1 as instruments. Similarly, define $\widetilde{X} = P_Z X$ and $\widetilde{\beta} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'y$, so that $\widetilde{\beta}$ is the IV estimator of (1) using Z as instruments.

Show that $\widehat{\beta}_1 = \widetilde{\beta}_1$ if $Z_2' M_{[Z_1]}[X_1 - X_2(X_2'P_1X_2)^{-1}X_2'P_1X_1] = 0$.

28. For the crime in North Carolina example given in section 11.6, replicate the results in Tables 11.2–11.6 using the data for 1987. Do the same using the data for 1981. Are there any notable differences in the results as we compare 1981 to 1987?

29. *Spatial Lag Test with Equal Weights.* This is based on Baltagi and Liu (2009). Consider the spatial lag dependence model described in Section 11.2.1, with the $N \times N$ spatial weight matrix W having zero elements across the diagonal and *equal elements* $1/(N-1)$ off the diagonal. The LM test for zero spatial lag, i.e., $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$, is given in Anselin (1988). This takes the form

$$LM = \frac{\left[\hat{u}'Wy/\left(\hat{u}'\hat{u}/N\right)\right]^2}{\left(WX\hat{\beta}\right)'\overline{P}_X WX\hat{\beta}/\hat{\sigma}^2 + tr\left(W^2 + W'W\right)}.$$

where $\overline{P}_X = I - X\left(X'X\right)^{-1}X'$, $\hat{\beta}$ is the restricted mle, which in this case is the least squares estimator of β. Similarly, $\hat{\sigma}^2$ is the corresponding restricted mle of σ^2, which in this case is the least squares residual sums of squares divided by N, i.e., $(\hat{u}'\hat{u}/N)$, where $\hat{u}$ denotes the least squares residuals. Show that for the equal weight spatial matrix, this LM test statistic will always equal $N/2\left(N-1\right)$ no matter what ρ is.

30. *Growth and Inequality Reconsidered.* For the Lundberg and Squire (2003) *Growth and Inequality* example considered in section 11.6.

    (a) Estimate these equations using 3SLS and verify the results reported in Table 1 of Lundberg and Squire (2003, p. 334). These results show that openness enhances growth and education reduces inequality, see Table 11.9. How do these 3SLS results compare with 2SLS? Are the over-identification restrictions rejected by the data?

(b) Lundberg and Squire (2003) allow growth to enter the inequality equation, and inequality to enter the growth equation. Estimate these respecified equations using 3SLS and verify the results reported in Table 2 of Lundberg and Squire (2003, p. 336). How do these 3SLS results compare with 2SLS? Are the over-identification restrictions rejected by the data?

31. *Married Women Labor Supply.* Mroz (1987) questions the exogeneity assumption of the wife's wage rate in a simple specification of married women labor supply. Using the PSID for 1975, his sample consists of 753 married white women between the ages of 30 and 60 in 1975, with 428 working at some time during the year. The wife's annual hours of work (hours) is regressed on the logarithm of her wage rate (lwage); the nonwife income (nwifeinc); the wife's age (age), her years of schooling (educ), the number of children less than six years old in the household (kidslt6), and the number of children between the ages of five and nineteen (kidsge6). The data set was obtained from the data web site of Wooldridge (2009).

(a) Replicate Table III of Mroz (1987, p. 769) which gives the descriptive statistics of the data.

(b) Replicate Table IV of Mroz (1987, p. 770) which runs OLS and 2SLS using a variety of instrumental variables for lwage. These instrumental variables are described in Table V of Mroz (1987, p. 771).

(c) Run the over-identification test for each 2sls regression in (b), as well as the diagnostics for the first stage regression.

References

This chapter is influenced by Johnston (1984), Kelejian and Oates (1989), Maddala (1992), Davidson and MacKinnon (1993), Mariano (2001) and Stock and Watson (2003). Additional references include the econometric texts cited in Chapter 3 and the following:

Anderson, T.W. and H. Rubin (1950), "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 21: 570–582.

Baltagi, B.H. (1989), "A Hausman Specification Test in a Simultaneous Equations Model," *Econometric Theory*, Solution 88.3.5, 5: 453–467.

Baltagi, B.H. and L. Liu (2009), "Spatial Lag Test with Equal Weights," *Economics Letters*, 104: 81–82.

Basmann, R.L. (1957), "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," *Econometrica*, 25: 77–83.

Basmann, R.L. (1960), "On Finite Sample Distributions of Generalized Classical Linear Identifiability Tests Statistics," *Journal of the American Statistical Association*, 55: 650–659.

Bekker, P.A. (1994), "Alternative Approximations to the Distribution of Instrumental Variable Estimators," *Econometrica* 62: 657–681.

Bekker, P.A. and T.J. Wansbeek (2001), "Identification in Parametric Models," Chapter 7 in Baltagi, B.H. (ed.) *A Companion to Theoretical Econometrics* (Blackwell: Massachusetts).

Bound, J., D.A. Jaeger and R.M. Baker (1995), "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Exogenous Explanatory Variable is Weak," *Journal of American Statistical Association* 90: 443–450.

Cornwell, C. and W.N. Trumbull (1994), "Estimating the Economic Model of Crime Using Panel Data," *Review of Economics and Statistics*, 76: 360–366.

Cragg, J.G. and S.G. Donald (1996), "Inferring the Rank of Matrix," *Journal of Econometrics*, 76: 223–250.

Deaton, A. (1997), *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy* (Johns Hopkins University Press: Baltimore).

Durbin, J. (1954), "Errors in Variables," *Review of the International Statistical Institute*, 22: 23–32.

Farebrother, R.W. (1985), "The Exact Bias of Wald's Estimator," *Econometric Theory*, Problem 85.3.1, 1: 419.

Farebrother, R.W. (1991), "Comparison of t-Ratios," *Econometric Theory*, Solution 90.1.4, 7: 145–146.

Fisher, F.M. (1966), *The Identification Problem in Econometrics* (McGraw-Hill: New York).

Gao, C. and K. Lahiri (2000), "Degeneration of Feasible GLS to 2SLS in a Limited Information Simultaneous Equations Model," *Econometric Theory*, Problem 00.2.1, 16: 287.

Gronau, R. (1973), "The Effects of Children on the Housewife's Value of Time," *Journal of Political Economy*, 81: S168–199.

Haavelmo, T. (1944), "The Probability Approach in Econometrics," *Supplement to Econometrica*, 12.

Hall, A. (1993), "Some Aspects of Generalized Method of Moments Estimation," Chapter 15 in *Handbook of Statistics*, Vol. 11 (North Holland: Amsterdam).

Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50: 646–660.

Hausman, J.A. (1978), "Specification Tests in Econometrics," *Econometrica*, 46: 1251–1272.

Hausman, J.A. (1983), "Specification and Estimation of Simultaneous Equation Models," Chapter 7 in Griliches, Z. and Intriligator, M.D. (eds.) *Handbook of Econometrics*, Vol. I (North Holland: Amsterdam).

Holly, A. (1987), "Identification and Estimation of a Simple Two-Equation Model," *Econometric Theory*, Problem 87.3.3, 3: 463–466.

Holly, A. (1988), "A Hausman Specification Test in a Simultaneous Equations Model," *Econometric Theory*, Problem 88.3.5, 4: 537–538.

Holly, A. (1990), "Comparison of t-ratios," *Econometric Theory*, Problem 90.1.4, 6: 114.

Kapteyn, A. and D.G. Fiebig (1981), "When are Two-Stage and Three-Stage Least Squares Estimators Identical?," *Economics Letters*, 8: 53–57.

Koopmans, T.C. and J. Marschak (1950), *Statistical Inference in Dynamic Economic Models* (John Wiley and Sons: New York).

Koopmans, T.C. and W.C. Hood (1953), *Studies in Econometric Method* (John Wiley and Sons: New York).

Laffer, A.B., (1970), "Trade Credit and the Money Market," *Journal of Political Economy*, 78: 239–267.

Lott, W.F. and S.C. Ray (1992), *Applied Econometrics: Problems with Data Sets* (The Dryden Press: New York).

Lundberg, M. and L. Squire (2003), "The Simultaneous Evolution of Growth and Inequality," *The Economic Journal*, 113: 326–344.

Manski, C.F. (1995), *Identification Problems in the Social Sciences* (Harvard University Press: Cambridge)

Mariano, R.S. (2001), "Simultaneous Equation Model Estimators: Statistical Properties and Practical Implications," Chapter 6 in Baltagi, B.H. (ed.) *A Companion to Theoretical Econometrics* (Blackwell: Massachusetts).

Mroz, T.A. (1987), "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica,* 55: 765–799.

Nelson, C.R. and R. Startz (1990), "The Distribution of the Instrumental Variables Estimator and its t-Ratio when the Instrument is a Poor One," *Journal of Business*, 63: S125–140.

Sapra, S.K. (1997), "Equivariance of an Instrumental Variable (IV) Estimator in the Linear Regression Model," *Econometric Theory*, Problem 97.2.5, 13: 464.

Singh, N. and A N. Bhat (1988), "Identification and Estimation of a Simple Two-Equation Model," *Econometric Theory*, Solution 87.3.3, 4: 542–545.

Staiger, D. and J. Stock (1997), "Instrumental Variables Regression With Weak Instruments," *Econometrica,* 65: 557–586.

Stock, J.H. and M.W. Watson (2003), *Introduction to Econometrics* (Addison Wesley: Boston).

Theil, H. (1953), "Repeated Least Squares Applied to Complete Equation Systems," *The Hague, Central Planning Bureau* (Mimeo).

Theil, H. (1971), *Principles of Econometrics* (Wiley: New York).

Wald, A. (1940), "Fitting of Straight Lines if Both Variables are Subject to Error," *Annals of Mathematical Statistics*, 11: 284–300.

Wooldridge, J.M. (1990), "A Note on the Lagrange Multiplier and F Statistics for Two Stage Least Squares Regression," *Economics Letters*, 34: 151–155.

Wooldridge, J.M. (2009), *Introductory Econometrics: A Modern Approach* (South-Western: Ohio).

Wu, D.M. (1973), "Alternative Tests of Independence Between Stochastic Regressors and Disturbances," *Econometrica*, 41: 733–740.

Zellner, A. and Theil, H. (1962), "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations," *Econometrica*, 30: 54–78.

Appendix: The Identification Problem Revisited: The Rank Condition of Identification

In section 11.1.2, we developed a *necessary* but not sufficient condition for identification. In this section we emphasize that model identification is crucial because only then can we get meaningful estimates of the parameters. For an *under-identified* model, different sets of parameter values agree well with the statistical evidence. As Bekker and Wansbeek (2001, p. 144) put it, preference for one set of parameter values over other ones becomes arbitrary. Therefore, "Scientific conclusions drawn on the basis of such arbitrariness are in the best case void and in the worst case dangerous." Manski (1995, p. 6) also warns that "negative identification findings imply that statistical inference is fruitless. It makes no sense to try to use a sample of finite size to infer something that could not be learned even if a sample of infinite size were available."

Consider the simultaneous equation model

$$By_t + \Gamma x_t = u_t \qquad t = 1, 2, \ldots, T. \tag{A.1}$$

which displays the whole set of equations at time t. B is $G\times G$, Γ is $G\times K$ and u_t is $G\times 1$. B is square and nonsingular indicating that the system is complete, i.e., there are as many equations as there are endogenous variables. Premultiplying (A.1) by B^{-1} and solving for y_t in terms of the exogenous variables and the vector of errors, we get

$$y_t = \Pi x_t + v_t \quad t = 1,2,\ldots,T. \tag{A.2}$$

where $\Pi = -B^{-1}\Gamma$, is $G \times K$, and $v_t = B^{-1}u_t$. Note that if we premultiply the structural model in (A.1) by an arbitrary nonsingular $G \times G$ matrix F, then the new structural model has the same reduced form given in (A.2). In this case, each new structural equation is a linear combination of the original structural equations, but the reduced form equations are the same. One idea for identification, explored by Fisher (1966), is to note that (A.1) is completely defined by B, Γ and the probability density function of the disturbances $p(u_t)$. The specification of the structural model which comes from economic theory, imposes a lot of zero restrictions on the B and Γ coefficients. In addition, there may be cross-equations or within equation restrictions. For example, constant returns to scale of the production function, or homogeneity of a demand equation, or symmetry conditions. In addition, the probability density function of the disturbances may itself contain some zero covariance restrictions. The structural model given in (A.1) is identified if these restrictions are enough to distinguish it from any other structural model. This is operationalized by proving that the only nonsingular matrix F which results in a new structural model that satisfies the same restrictions on the original model is the identity matrix. If after imposing the restrictions, only certain rows of F resemble the corresponding rows of an identity matrix, up to a scalar of normalization, then the corresponding equations of the system are identified. The remaining equations are not identified. This is the same concept of taking a linear combination of the demand and supply equations and seeing if the linear combination is different from demand or supply. If it is, then both equations are identified. If it looks like demand but not supply, then supply is identified and demand is not identified. Let us look at an example.

Example (A.1): Consider a demand and supply equations with

$$Q_t = a - bP_t + cY_t + u_{1t} \tag{A.3}$$
$$Q_t = d + eP_t + fW_t + u_{2t} \tag{A.4}$$

where Y is income and W is weather. Writing (A.3) and (A.4) in matrix form (A.1), we get

$$B = \begin{bmatrix} 1 & b \\ 1 & -e \end{bmatrix} \qquad \Gamma = \begin{bmatrix} -a & -c & 0 \\ -d & 0 & -f \end{bmatrix} \qquad y_t = \begin{pmatrix} Q_t \\ P_t \end{pmatrix} \tag{A.5}$$
$$x_t' = [1, Y_t, W_t] \qquad u_t' = (u_{1t}, u_{2t})$$

There are two zero restrictions on Γ. The first is that income does not appear in the supply equation and the second is that weather does not appear in the demand equation. Therefore, the order condition of identification is satisfied for both equations. In fact, for each equation, there is one excluded exogenous variable and only one right hand side included endogenous variable. Therefore, both equations are just-identified. Let $F = [f_{ij}]$ for $i,j = 1,2$, be a nonsingular matrix. Premultiply this system by F. The new matrix B is now FB and the new matrix Γ is now $F\Gamma$. In order for the transformed system to satisfy the same restrictions as the original model, FB must satisfy the following normalization restrictions:

$$f_{11} + f_{12} = 1 \qquad f_{21} + f_{22} = 1 \tag{A.6}$$

Also, $F\Gamma$ should satisfy the following zero restrictions

$$-f_{21}c + f_{22}0 = 0 \qquad f_{11}0 - f_{12}f = 0 \tag{A.7}$$

Since $c \neq 0$, and $f \neq 0$, then (A.7) implies that $f_{21} = f_{12} = 0$. Using (A.6), we get $f_{11} = f_{22} = 1$. Hence, the only nonsingular F that satisfies the same restrictions on the original model is the identity matrix, provided $c \neq 0$ and $f \neq 0$. Therefore, both equations are identified.

Example (A.2): If income does not appear in the demand equation (A.3), i.e., $c = 0$, the model looks like

$$Q_t = a - bP_t + u_{1t} \tag{A.8}$$
$$Q_t = d + eP_t + fW_t + u_{2t} \tag{A.9}$$

In this case, only the second restriction given in (A.7) holds. Therefore, $f \neq 0$ implies $f_{12} = 0$, however f_{21} is not necessarily zero without additional restrictions. Using (A.6), we get $f_{11} = 1$ and $f_{21} + f_{22} = 1$. This means that only the first row of F looks like the first row of an identity matrix, and only the demand equation is identified. In fact, the order condition for identification is not met for the supply equation since there are no excluded exogenous variables from that equation but there is one right hand side included endogenous variable. See problems 6 and 7 for more examples of this method of identification.

Example (A.3): Suppose that $u \sim (0, \Omega)$ where $\Omega = \Sigma \otimes I_T$, and $\Sigma = [\sigma_{ij}]$ for $i, j = 1, 2$. This example shows how a variance-covariance restriction can help identify an equation. Let us take the model defined in (A.8), (A.9) and add the restriction that $\sigma_{12} = \sigma_{21} = 0$. In this case, the transformed model disturbances will be $Fu \sim (0, \Omega^*)$, where $\Omega^* = \Sigma^* \otimes I_T$, and $\Sigma^* = F\Sigma F'$. In fact, since Σ is diagonal, $F\Sigma F'$ should also be diagonal. This imposes the following restriction on the elements of F:

$$f_{11}\sigma_{11}f_{21} + f_{12}\sigma_{22}f_{22} = 0 \tag{A.10}$$

But, $f_{11} = 1$ and $f_{12} = 0$ from the zero restrictions imposed on the demand equation, see example 4. Hence, (A.10) reduces to $\sigma_{11}f_{21} = 0$. Since $\sigma_{11} \neq 0$, this implies that $f_{21} = 0$, and the normalization restriction given in (A.6), implies that $f_{22} = 1$. Therefore, the second equation is also identified.

Example (A.4): In this example, we demonstrate how cross-equation restrictions can help identify equations. Consider the following simultaneous model

$$y_1 = a + by_2 + cx_1 + u_1 \tag{A.11}$$
$$y_2 = d + ey_1 + fx_1 + gx_2 + u_2 \tag{A.12}$$

and add the restriction $c = f$. It can be easily shown that the first equation is identified with $f_{11} = 1$, and $f_{12} = 0$. The second equation has no zero restrictions, but the cross-equation restriction $c = f$ implies:

$$-cf_{11} - ff_{12} = -cf_{21} - ff_{22}$$

Using $c = f$, we get

$$f_{11} + f_{12} = f_{21} + f_{22} \tag{A.13}$$

But, the first equation is identified with $f_{11} = 1$ and $f_{12} = 0$. Hence, (A.13) reduces to $f_{21}+f_{22} = 1$, which together with the normalization condition $-f_{21}b + f_{22} = 1$, gives $f_{21}(b + 1) = 0$. If $b \neq -1$, then $f_{21} = 0$ and $f_{22} = 1$. The second equation is also identified provided $b \neq -1$.

Alternatively, one can look at the problem of identification by asking whether the structural parameters in B and Γ can be obtained from the reduced form parameters. It will be clear from the following discussion that this task is impossible if there are no restrictions on this simultaneous model. In this case, the model is hopelessly unidentified. However, in the usual case where there are a lot of zeros in B and Γ, we may be able to retrieve the remaining non-zero coefficients from Π. More rigorously, $\Pi = -B^{-1}\Gamma$, which implies that

$$B\Pi + \Gamma = 0 \tag{A.14}$$

or

$$AW = 0 \quad \text{where} \quad A = [B, \Gamma] \quad \text{and} \quad W' = [\Pi', I_K] \tag{A.15}$$

For the first equation, this implies that

$$\alpha_1' W = 0 \quad \text{where} \quad \alpha_1' \text{ is the first row of } A. \tag{A.16}$$

W is known (or can be estimated) and is of rank K. If the first equation has no restrictions on its structural parameters, then α_1' contains $(G+K)$ unknown coefficients. These coefficients satisfy K homogeneous equations given in (A.16). Without further restrictions, we cannot solve for $(G+K)$ coefficients (α_1') with only K equations. Let ϕ denote the matrix of R zero restrictions on the first equation, i.e., $\alpha_1'\phi = 0$. This together with (A.16) implies that

$$\alpha_1'[W, \phi] = 0 \tag{A.17}$$

and we can solve uniquely for α_1' (up to a scalar of normalization) provided the

$$\text{rank } [W, \phi] = G + K - 1 \tag{A.18}$$

Economists specify each structural equation with the left hand side endogenous variable having the coefficient one. This normalization identifies one coefficient of α_1', therefore, we require only $(G + K - 1)$ more restrictions to uniquely identify the remaining coefficients of α_1. $[W, \phi]$ is a $(G + K) \times (K + R)$ matrix. Its rank is less than any of its two dimensions, i.e., $(K + R) \geq (G + K - 1)$, which results in $R \geq G - 1$, or the *order condition* of identification. Note that this is a necessary but not sufficient condition for (A.18) to hold. It states that the number of restrictions on the first equation must be greater than the number of endogenous variables minus one. If all R restrictions are zero restrictions, then it means that the number of excluded exogenous plus the number of excluded endogenous variables should be greater than $(G - 1)$. But the G endogenous variables are made up of the left hand side endogenous variable y_1, the g_1 right hand side included endogenous variables Y_1, and $(G - g_1 - 1)$ excluded endogenous variables. Therefore, $R \geq (G - 1)$ can be written as $k_2 + (G - g_1 - 1) \geq (G - 1)$ which reduces to $k_2 \geq g_1$, which was discussed earlier in this chapter.

The *necessary and sufficient condition* for identification can now be obtained as follows: Using (A.1) one can write

$$Az_t = u_t \quad \text{where} \quad z_t' = (y_t', x_t') \tag{A.19}$$

and from the first definition of identification we make the transformation $FAz_t = Fu_t$, where F is a $G \times G$ nonsingular matrix. The first equation satisfies the restrictions $\alpha_1'\phi = 0$, which can be rewritten as $\iota' A\phi = 0$, where ι' is the first row of an identity matrix I_G. F must satisfy the restriction that (first row of $FA)\phi = 0$. But the first row of FA is the first row of F, say f_1', times A. This means that $f_1'(A\phi) = 0$. For the first equation to be identified, this condition on the transformed first equation must be equivalent to $\iota' A\phi = 0$, up to a scalar constant. This holds if and only if f_1' is a scalar multiple of ι', and the latter condition holds if and only if the rank $(A\phi) = G - 1$. The latter is known as the *rank condition* for identification.

Example (A.5): Consider the simple Keynesian model given in example 1. The second equation is an identity and the first equation satisfies the order condition of identification, since I_t is the excluded exogenous variable from that equation, and there is only one right hand side included endogenous variable Y_t. In fact, the first equation is just-identified. Note that

$$A = [B, \Gamma] = \begin{bmatrix} \beta_{11} & \beta_{12} & \gamma_{11} & \gamma_{12} \\ \beta_{21} & \beta_{22} & \gamma_{21} & \gamma_{22} \end{bmatrix} = \begin{bmatrix} 1 & -\beta & -\alpha & 0 \\ -1 & 1 & 0 & -1 \end{bmatrix} \tag{A.20}$$

and ϕ for the first equation consists of only one restriction, namely that I_t is not in that equation, or $\gamma_{12} = 0$. This makes $\phi' = (0, 0, 0, 1)$, since $\alpha_1'\phi = 0$ gives $\gamma_{12} = 0$. From (A.20), $A\phi = (\gamma_{12}, \gamma_{22})' = (0, -1)'$ and the rank $(A\phi) = 1 = G - 1$. Hence, the rank condition holds for the first equation and it is identified. Problem 8 reconsiders example (A.1), where both equations are just-identified by the order condition of identification and asks the reader to show that both satisfy the rank condition of identification as long as $c \neq 0$ and $f \neq 0$.

The reduced form of the simple Keynesian model is given in equations (11.4) and (11.5). In fact,

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \alpha & 1 \end{bmatrix} /(1 - \beta \tag{A.21}$$

Note that

$$\begin{aligned} \pi_{11}/\pi_{22} &= \alpha \quad \text{and} \quad \pi_{21}/\pi_{22} = \alpha \\ \pi_{12}/\pi_{22} &= \beta \quad \text{and} \quad (\pi_{22} - 1)/\pi_{22} = \beta \end{aligned} \tag{A.22}$$

Therefore, the structural parameters of the consumption equation can be retrieved from the reduced form coefficients. However, what happens if we replace Π by its OLS estimate $\widehat{\Pi}_{OLS}$? Would the solution in (A.22) lead to two estimates of (α, β) or would this solution lead to a unique estimate? In this case, the consumption equation is just-identified and the solution in (A.22) is unique. To show this, recall that

$$\widehat{\pi}_{12} = m_{ci}/m_{ii}; \quad \text{and} \quad \widehat{\pi}_{22} = m_{yi}/m_{ii} \tag{A.23}$$

Solving for $\widehat{\beta}$, using (A.22), one gets

$$\widehat{\beta} = \widehat{\pi}_{12}/\widehat{\pi}_{22} = m_{ci}/m_{yi} \tag{A.24}$$

and

$$\widehat{\beta} = (\widehat{\pi}_{22} - 1)/\widehat{\pi}_{22} = (m_{ci} - m_{yi})/m_{yi} \tag{A.25}$$

(A.24) and (A.25) lead to a unique solution because equation (11.2) gives

$$m_{yi} = m_{ci} + m_{ii} \tag{A.26}$$

In general, we would not be able to solve for the structural parameters of an unidentified equation in terms of the reduced form parameters. However, when this equation is identified, replacing the reduced form parameters by their OLS estimates would lead to a unique estimate of the structural parameters, only if this equation is just-identified, and to more than one estimate depending upon the degree of over-identification. Problem 8 gives another example of the just-identified case, while problem 9 considers a model with one unidentified and another over-identified equation.

Example (A.6): Equations (11.13) and (11.14) give an unidentified demand and supply model with

$$B = \begin{bmatrix} 1 & -\beta \\ 1 & -\delta \end{bmatrix} \quad \text{and} \quad \Gamma = \begin{bmatrix} -\alpha \\ -\gamma \end{bmatrix} \tag{A.27}$$

The reduced form equations given by (11.16) and (11.17) yield

$$\Pi = -B^{-1}\Gamma = \begin{bmatrix} \pi_{11} \\ \pi_{21} \end{bmatrix} = \begin{bmatrix} \alpha\delta - \gamma\beta \\ \alpha - \gamma \end{bmatrix} / (\delta - \beta) \tag{A.28}$$

Note that one cannot solve for (α, β) nor (γ, δ) in terms of (π_{11}, π_{21}) without further restrictions.