CHAPTER 7 INSTRUMENTAL VARIABLES REGRESSION

Key Words: Endogeneity, Instrumental variables, Hausman's test, 2SLS.

Abstract: In this chapter we first discuss possibilities that the condition $E(\varepsilon_t|X_t) = 0$ *a.s.* may fail, which will generally render inconsistent the OLS estimator for the true model parameters. We then introduce a consistent two-stage least squares (2SLS) estimator, investigating its statistical properties and providing intuitions for the nature of the 2SLS estimator. Hypothesis tests are constructed. We consider various test procedures corresponding to the cases for which the disturbance is an MDS with conditional homoskedasticity, an MDS with conditional heteroskedasticity, and a non-MDS process, respectively. The latter case will require consistent estimation of a long-run variance-covariance matrix. It is important to emphaisze that the t-test and F-test obtained from the second stage regression estimation cannot be used even for large samples. Finally, we consider some empirical applications and conclude this chapter by presenting a brief summary of the comprehensive econometric theory for linear regression models developed in Chapters 2–7.

7.1 Motivation

In all previous chapters, we always assumed that $E(\varepsilon_t|X_t) = 0$ holds even when there exist conditional heteroskedasticity and autocorrelation.

Questions: When may the condition $E(\varepsilon_t|X_t) = 0$ fail? And, what will happen to the OLS estimator $\hat{\beta}$ if $E(\varepsilon_t|X_t) = 0$ fails?

There are at least three possibilities where $E(\varepsilon_t|X_t) = 0$ may fail. The first is model misspecification (e.g., functional form misspecification or omitted variables). The second is the existence of measurement errors in regressors (also called errors in variables). The third is the estimation of a subset of a simultaneous equation system. We will consider the last two possibilities in this chapter. For the first case (i.e., model misspecification), it may not be meaningful to discuss consistent estimation of the parameters in a misspecified regression model.

Some Motivating Examples

We first provide some examples in which $E(\varepsilon_t|X_t) \neq 0$.

Example 1 [Errors of Measurements or Errors in Variables]:

Often, economic data measure concepts that differ somewhat from those of economic theory. It is therefore important to take into account errors of measurements. This is usually called errors in variables in econometrics. Consider a data generating process (DGP)

$$Y_t^* = \beta_0^o + \beta_1^o X_t^* + u_t, \tag{7.1}$$

where X_t^* is the income, Y_t^* is the consumption, and $\{u_t\}$ is i.i.d. $(0, \sigma_u^2)$ and is independent of $\{X_t^*\}$.

Suppose both X_t^* and Y_t^* are not observable. The observed variables X_t and Y_t contain measurement errors in the sense that

$$X_t = X_t^* + v_t, \tag{7.2}$$

$$Y_t = Y_t^* + w_t, \tag{7.3}$$

where $\{v_t\}$ and $\{w_t\}$ are measurement errors independent of $\{X_t^*\}$ and $\{Y_t^*\}$, such that $\{v_t\} \sim i.i.d.\ (0, \sigma_v^2)$ and $\{w_t\} \sim i.i.d.\ (0, \sigma_w^2)$. We assume that the series $\{v_t\}, \{w_t\}$ and $\{u_t\}$ are all mutually independent of each other.

Because we only observe (X_t, Y_t), we are forced to estimate the following regression model

$$Y_t = \beta_0^o + \beta_1^o X_t + \varepsilon_t, \tag{7.4}$$

where ε_t is some unobservable disturbance.

Clearly, the disturbance ε_t is different from the original (true) disturbance u_t. Although the linear regression model is correctly specified, we no longer have $E(\varepsilon_t|X_t) = 0$ due to the existence of the measurement errors. This is explained below.

Question: If we use the OLS estimator $\hat{\beta}$ to estimate this model, is $\hat{\beta}$ consistent for β^o?

From the general regression analysis in Chapter 2, we have known that the key for the consistency of the OLS estimator $\hat{\beta}$ for β^o is to check if $E(X_t\varepsilon_t) = 0$. From Eqs. (7.1) – (7.3), we have

$$\begin{aligned} Y_t &= Y_t^* + w_t \\ &= (\beta_0^o + \beta_1^o X_t^* + u_t) + w_t \\ X_t &= X_t^* + v_t. \end{aligned}$$

Therefore, from Eq. (7.4), we obtain

$$\begin{aligned} \varepsilon_t &= Y_t - \beta_0^o - \beta_1^o X_t \\ &= [\beta_0^o + \beta_1^o X_t^* + u_t + w_t] - \beta_0^o - \beta_1^o (X_t^* + v_t) \\ &= u_t + w_t - \beta_1^o v_t. \end{aligned}$$

The regression error ε_t contains the true disturbance u_t and a linear combination of measurement errors.

Now, the expectation

$$\begin{aligned} E(X_t\varepsilon_t) &= E[(X_t^* + v_t)\varepsilon_t] \\ &= E(X_t^*\varepsilon_t) + E(v_t\varepsilon_t) \\ &= 0 - \beta_1^o E(v_t^2) \\ &= -\beta_1^o \sigma_v^2 \\ &\neq 0. \end{aligned}$$

Consequently, by the WLLN, the OLS estimator

$$\begin{aligned} \hat{\beta} - \beta^o &= \hat{Q}_{xx}^{-1} n^{-1} \sum_{t=1}^{n} X_t\varepsilon_t \\ &\xrightarrow{p} Q_{xx}^{-1} E(X_t\varepsilon_t) \\ &= -\beta_1^o \sigma_v^2 Q_{xx}^{-1} \neq 0. \end{aligned}$$

In other words, $\hat{\beta}$ is not consistent for β^o due to the existence of the measurement errors in regressors $\{X_t\}$.

Question: What is the effect of the measurement errors $\{w_t\}$ in the dependent variable Y_t?

Example 2 [Errors of Measurements in Dependent Variable]: Now we consider a data generating process (DGP) given by

$$Y_t^* = \beta_0^o + \beta_1^o X_t^* + u_t,$$

where X_t^* is the income, Y_t^* is the consumption, and $\{u_t\}$ is i.i.d. $(0, \sigma_u^2)$ and is independent of $\{X_t^*\}$.

Suppose X_t^* is now observed, and Y_t^* is still not observable, such that

$$\begin{aligned} X_t &= X_t^*, \\ Y_t &= Y_t^* + w_t, \end{aligned}$$

where $\{w_t\}$ is i.i.d.$(0, \sigma_w^2)$ measurement errors independent of $\{X_t^*\}$ and $\{Y_t^*\}$. We assume that the two series $\{w_t\}$ and $\{u_t\}$ are mutually independent.

Because we only observe (X_t, Y_t), we are forced to estimate the following model

$$Y_t = \beta_0^o + \beta_1^o X_t + \varepsilon_t.$$

Question: If we use the OLS estimator $\hat{\beta}$ to estimate this model, is $\hat{\beta}$ consistent for β^o?

Answer: Yes! The measurement errors in Y_t do not cause any trouble for consistent estimation of β^o.

The measurement error in Y_t can be regarded as part of the true regression disturbance. It increases the asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta^o)$, that is, the existence of measurement errors in Y_t renders the estimation of β^o less precise.

Example 3 [Errors in Expectations] Consider a linear regression model

$$Y_t = \beta_0 + \beta_1 X_t^* + \varepsilon_t,$$

where X_t^* is the economic agent's conditional expectation of X_t at time $t-1$, and $\{\varepsilon_t\}$ is an i.i.d.$(0, \sigma^2)$ sequence with $E(\varepsilon_t|X_t^*) = 0$. The conditional expectation X_t^* is a latent variable. When the economic agent follows rational expectations, then $X_t^* = E(X_t|I_{t-1})$ and we have

$$X_t = X_t^* + v_t,$$

where

$$E(v_t|I_{t-1}) = 0,$$

where I_{t-1} is the information available to the economic agent at time $t-1$. Assume that two error series $\{\varepsilon_t\}$ and $\{v_t\}$ are independent of each other.

We can consider the following regression model

$$Y_t = \beta_0^o + \beta_1^o X_t + u_t,$$

where the error term

$$u_t = \varepsilon_t - \beta_1^o v_t.$$

Since

$$\begin{aligned} E(X_t u_t) &= E[(X_t^* + v_t)(\varepsilon_t - \beta_1^o v_t)] \\ &= -\beta_1^o \sigma_v^2 \\ &\neq 0 \end{aligned}$$

provided $\beta_1^o \neq 0$, the OLS estimator is not consistent for β_1^o.

Example 4 [Endogeneity due to Omitted variables] Consider an earning data generating process

$$Y_t = X_t' \beta^o + \gamma A_t + u_t,$$

where Y_t is the earning, X_t is a vector consisting of working experience and schooling, and A_t is ability which is unobservable, and the disturbance u_t satisfies the condition that $E(u_t|X_t, A_t) = 0$. Because one does not observe A_t, one is forced to consider the regression model

$$Y_t = X_t'\beta^o + \varepsilon_t$$

and is interested in knowing β^o, the marginal effect of schooling and working experience. However, we have $E(X_t\varepsilon_t) \neq 0$ because A_t is usually correlated with X_t.

Example 5 [Production-Bonus Causality; Groves, Hong, McMillan and Naughton 1994]: Consider a production function data generating process

$$\ln(Y_t) = \beta_0^o + \beta_1^o \ln(L_t) + \beta_2^o \ln(K_t) + \beta_3^o B_t + \varepsilon_t,$$

where Y_t, L_t, K_t are the output, labor and capital stock, B_t is the proportion of bonus out of total pay, and t is a time index. Without loss of generality, we assume that

$$\begin{aligned} E(\varepsilon_t) &= 0, \\ E[\ln(L_t)\varepsilon_t] &= 0, \\ E[\ln(K_t)\varepsilon_t] &= 0. \end{aligned}$$

Economic theory suggests that the use of bonus in addition to basic wage will provide a stronger incentive for workers to work harder in a transitional economy. This theory can be tested by checking if $\beta_3^o = 0$. However, the test procedure is complicated because there exists a possibility that when a firm is more productive, it will pay more bonus to workers regardless of the effort of its workers. In this case, the OLS estimator $\hat{\beta}_3$ cannot consistently estimate β_3^o and cannot be used to test the null hypothesis.

Why?

To reflect the fact that a more productive firm pays more bonus to its workers, we can assume a structural equation for bonus:

$$B_t = \alpha_0^0 + \alpha_1^0 \ln(Y_t) + w_t \tag{7.5}$$

where $\alpha_1^0 > 0$, and $\{w_t\}$ is an i.i.d. $(0, \sigma_w^2)$ sequence that is independent of $\{Y_t\}$. For simplicity, we assume that $\{w_t\}$ is independent of $\{\varepsilon_t\}$.

Put $X_t = [1, \ln(L_t), \ln(K_t), B_t]'$. Now, from Eq. (7.5) and then Eq. (7.4), we have

$$\begin{aligned} E(B_t\varepsilon_t) &= E[(\alpha_0^o + \alpha_1^o \ln(Y_t) + w_t)\varepsilon_t] \\ &= \alpha_1^o E[\ln(Y_t)\varepsilon_t] \\ &= \alpha_1^o \beta_3^o E(B_t\varepsilon_t) + \alpha_1^o E(\varepsilon_t^2). \end{aligned}$$

It follows that

$$E(B_t\varepsilon_t) = \frac{\alpha_1^o}{1-\alpha_1^o\beta_3^o}\sigma^2 \neq 0,$$

where $\sigma^2 = \text{var}(\varepsilon_t)$. Consequently, the OLS estimator $\hat{\beta}_3$ is inconsistent for β_3^o due to the existence of the causality from productivity $\ln(Y_t)$ to bonus B_t.

The bias of the OLS estimator for β_3^o in the above model is usually called the simultaneous equation bias because it arises from the fact that productivity function is but one of two relationships that hold simultaneously. It is a common phenomena in economics. It is the rule rather than the exception for economic relationships to be embedded in a simultaneous system of equations. We now consider two more examples with simultaneous equation bias.

Example 6 [Simultaneous Equation Bias] We consider the following simple model of national income determination:

$$\begin{aligned} C_t &= \beta_0^o + \beta_1^o I_t + \varepsilon_t, & (7.6) \\ I_t &= C_t + D_t, & (7.7) \end{aligned}$$

where I_t is the income, C_t is the consumption expenditure, and D_t is the non-consumption expenditure. The variables I_t and C_t are called endogenous variables, as they can be determined by the two-equation model. The variable D_t is called an exogenous variable, because it is determined outside the model (or the system considered). We assume that $\{D_t\}$ and $\{\varepsilon_t\}$ are mutually independent, and $\{\varepsilon_t\}$ is i.i.d. $(0, \sigma^2)$.

Question: If the OLS estimator $\hat{\beta}$ is applied to the first equation, is it consistent for β^o?

To answer this question, we have from Eq. (7.7)

$$\begin{aligned} E(I_t\varepsilon_t) &= E[(C_t + D_t)\varepsilon_t] \\ &= E(C_t\varepsilon_t) + E(D_t\varepsilon_t) \\ &= \beta_1^o E(I_t\varepsilon_t) + E(\varepsilon_t^2) + 0. \end{aligned}$$

It follows that

$$E(I_t\varepsilon_t) = \frac{1}{1-\beta_1^o}\sigma^2 \neq 0.$$

Thus, $\hat{\beta}$ is not consistent for β^o.

In fact, this bias problem can also be seen from the so-called reduced form model.

Question: What is the reduced form?

Solving for Eqs. (7.6) and (7.7) simultaneously, we can obtain the "reduced forms" that express endogenous variables in terms of exogenous variables and disturbances:

$$\begin{aligned} C_t &= \frac{\beta_0^o}{1-\beta_1^o} + \frac{\beta_1^o}{1-\beta_1^o} D_t + \frac{1}{1-\beta_1^o}\varepsilon_t, \\ I_t &= \frac{\beta_0^o}{1-\beta_1^o} + \frac{1}{1-\beta_1^o} D_t + \frac{1}{1-\beta_1^o}\varepsilon_t. \end{aligned}$$

Obviously, I_t is positively correlated with ε_t (i.e., $E(I_t\varepsilon_t) \neq 0$). Thus, the OLS estimator for the regression of C_t on I_t in Eq. (7.6) will not be consistent for β_1^o, the parameter for marginal propensity to consume. Generally speaking, the OLS estimator for the reduced form is consistent for new parameters, which are functions of original parameters.

Example 7 [Wage-Price Spiral Model] Consider the system of equations

$$W_t = \beta_0^o + \beta_1^o P_t + \beta_2^o D_t + \varepsilon_t, \tag{7.8}$$
$$P_t = \alpha_0^o + \alpha_1^o W_t + v_t, \tag{7.9}$$

where W_t, P_t, D_t are the wage, price, and excess demand in the labor market respectively. Eq. (7.8) describes the mechanism of how wage is determined. In particular, wage depends on price and excess demand for labor. Eq. (7.9) describes how price depends on wage (or income).

Suppose D_t is an exogenous variable, with $E(\varepsilon_t|D_t) = 0$. There are two endogenous variables, W_t and P_t, in the system of equations (7.8) and (7.9).

Question: Will W_t be correlated with v_t? And, will P_t be correlated with ε_t?

To answer these questions, we first obtain the reduced form equations:

$$\begin{aligned} W_t &= \frac{\beta_0^o + \beta_1^o\alpha_0}{1-\beta_1^o\alpha_1^o} + \frac{\beta_1^o}{1-\beta_1^o\alpha_1^o} D_t + \frac{\varepsilon_t + \beta_1^o v_t}{1-\beta_1^o\alpha_1^o}, \\ P_t &= \frac{\alpha_0^o}{1-\beta_1^o\alpha_1^o} + \frac{\alpha_1^o\beta_2^o}{1-\beta_1^o\alpha_1^o} D_t + \frac{\alpha_1^o\varepsilon_t + v_t}{1-\beta_1^o\alpha_1^o}. \end{aligned}$$

Conditional on the exogenous variable D_t, both W_t and P_t are correlated with ε_t and v_t. As a consequence, both the OLS estimator for β_1^o in Eq. (7.8) and the OLS estimator for α_1^o in Eq. (7.9) will be inconsistent.

In this chapter, we will consider a method called two stage least squares estimation to obtain consistent estimators for the unknown parameters in all above examples except for the parameter β_2^o in Eq. (7.8) of Example 7. No methods can deliver a consistent estimator for β_2^o in Eq. (7.8) because it is not identifiable. This is the so-called identification problem of the simultaneous equations.

A Digression: Identification Problem in Simultaneous Equation Models

To see why there is no way to obtain a consistent estimator for β_2^o in Eq. (7.8), from Eq. (7.9), we can write

$$W_t = -\frac{\alpha_1^o}{\alpha_2^o} + \frac{1}{\alpha_2^o} P_t - \frac{v_t}{\alpha_2^o}. \tag{7.10}$$

Let a and b be two arbitrary constants. We multiply Eq. (7.8) with a, and multiply Eq. (7.10) with b, and add them together:

$$(a+b)W_t = a\beta_1^o - \frac{b\alpha_1}{\alpha_2} + (a\beta_2^o + \frac{b}{\alpha_2})P_t + a\beta_3^o D_t + (a\varepsilon_t - \frac{b}{\alpha_2} v_t),$$

or

$$W_t = \left[\frac{a\beta_1^o}{a+b} - \frac{b\alpha_1^o}{(a+b)\alpha_2^o}\right] + \frac{1}{a+b}(a\beta_2^o + \frac{b}{\alpha_2^o})P_t + \frac{a\beta_3^o}{a+b} D_t + \frac{1}{a+b}(a\varepsilon_t - \frac{b}{\alpha_2^o} v_t). \tag{7.11}$$

This new equation, (7.11), is a combination of the original wage equation (7.8) and the price equation (7.9). It is of the same statistical form as Eq. (7.8). Since a and b are arbitrary, there is an infinite number of parameters that can satisfy Eq. (11) and they are all indistinguishable from Eq. (7.8). Consequently, if we use OLS to run regression of W_t on P_t and D_t, or more generally, use any other method to estimate the equation (7.8) or (7.11), there is no way to know which model, either Eq. (7.8) or Eq. (7.11), is being estimated. Therefore, there is no way to estimate β_2^o. This is the so-called identification problem with simultaneous equation models. To avoid such identification problems in simultaneous equations, certain conditions are required to make the system of simultaneous equations identifiable. For example, if an extra variable, say money supply growth rate, is added in the price equation in (7.9), we then obtain

$$P_t = \alpha_0^o + \alpha_1^o W_t + \alpha_2^o M_t + v_t, \tag{7.12}$$

then the system of equations (7.8) and (7.12) is identifiable provided $\alpha_2^o \neq 0$, and so the

parameters in Eqs. (7.8) and (7.12) can be consistently estimated. [Question: Check why the system of equations (7.8) and (7.12) is identifiable.]

We note that for the system of equations (7.8) and (7.9), although Eq. (7.8) cannot be consistently estimated by any method, Eq. (7.9) can still be consistently estimated using the method proposed below. For an identifiable system of simultaneous equations with simultaneous equation bias, we can use various methods to estimate them consistently, including 2SLS, the generalized method of moments and the maximum likelihood or quasi-maximum likelihood estimation methods. These methods will be introduced below and in subsequent chapters.

7.2 Assumptions

We now provide a set of regularity conditions for our formal analysis in this chapter.

Assumption 7.1 $\{Y_t, X_t', Z_t'\}'$ is an ergodic stationary stochastic process, where X_t is a $K \times 1$ vector, Z_t is a $l \times 1$ vector, and $l \geq K$.

Assumption 7.2 [Linearity]:

$$Y_t = X_t'\beta^o + \varepsilon_t, \qquad t = 1, ..., n,$$

for some unknown parameter β^o and some unobservable disturbance ε_t;

Assumption 7.3 [Nonsingularity]: The $K \times K$ matrix

$$Q_{xx} = E(X_t X_t')$$

is nonsingular and finite;

Assumption 7.4 [IV Conditions]:
(i) $E(X_t\varepsilon_t) \neq 0$;
(ii) $E(Z_t\varepsilon_t) = 0$;
(iii) The $l \times l$ matrix

$$Q_{zz} = E(Z_t Z_t')$$

is finite and nonsingular, and the $l \times K$ matrix

$$Q_{zx} = E(Z_t X_t')$$

is finite and of full rank.

Assumption 7.5 [CLT]: $n^{-1/2} \sum_{t=1}^n Z_t\varepsilon_t \to^d N(0, V)$ for some $K \times K$ symmetric matrix $V \equiv \text{avar}(n^{-1/2} \sum_{t=1}^n Z_t\varepsilon_t)$ finite and nonsingular.

Remarks:
Assumption 7.1 allows for i.i.d. and stationary time series observations.

Assumption 7.5 directly assumes that the CLT holds. This is often called a "high level assumption." It covers three cases: IID, MDS and non-MDS for $\{X_t\varepsilon_t\}$, respectively. For an IID or MDS sequence $\{Z_t\varepsilon_t\}$, we have $V = \text{var}(Z_t\varepsilon_t) = E(Z_tZ_t'\varepsilon_t^2)$. For a non-MDS process $\{Z_t\varepsilon_t\}$, $V = \Sigma_{j=-\infty}^{\infty}\text{cov}(Z_t\varepsilon_t, Z_{t-j}\varepsilon_{t-j})$ is a long-run variance-covariance matrix.

The random vector Z_t that satisfies Assumption 7.4 is called instruments. The condition that $l \geq K$ in Assumption 7.1 implies that the number of instruments Z_t is larger than or at least equal to the number of regressors X_t.

Question: Why is the condition of $l \geq K$ required?

Question: How to choose instruments Z_t in practice?

First of all, one should analyze which explanatory variables in X_t are endogenous or exogenous. If an explanatory variable is exogenous, then this variable should be included in Z_t, the set of instruments. For example, the constant term should always be included, because a constant is uncorrelated with any random variables. All other exogenous variables in X_t should also be included in the set of Z_t. If k_0 of K regressors are endogenous, one should find at least k_0 additional instruments.

Most importantly, we should choose an instrument vector Z_t which is closely related to X_t as much as possible. As we will see below, the strength of the correlation between Z_t and X_t affects the magnitude of the asymptotic variance of the 2SLS estimator for β_0 which we will propose, although it does not affect the consistency provided the correlation between Z_t and X_t is not zero.

In time series regression models, it is often reasonable to assume that lagged variables of X_t are not correlated with ε_t. Therefore, we can use lagged values of X_t, for example, X_{t-1}, as an instrument. This instrument is expected to be highly correlated with X_t if $\{X_t\}$ is a time series process. In light of this, we can choose the set of instruments $Z_t = (1, \ln L_t, \ln K_t, B_{t-1})'$ in estimating Eq.(7.4) in Example 5, choose $Z_t = (1, D_t, I_{t-1})'$ in estimating Eq.(7.6) in Example 6, choose $Z_t = (1, D_t, P_{t-1})'$ in estimating Eq.(7.9) in Example 7. For examples with measurement errors or expectational errors, where $E(X_t\varepsilon_t) \neq 0$ due to the presence of measurement errors or expectational errors, we can choose $Z_t = X_{t-1}$ if the measurement errors or expectational errors in X_t are serially uncorrelated (check this!). The expectational errors in X_t are MDS and so are uncorrelated in Example 3 when the economic agent has rational expectations.

7.3 Two-Stage Least Square (2SLS) Estimation

Question: Because $E(\varepsilon_t|X_t) \neq 0$, the OLS estimator $\hat{\beta}$ is not consistent for β^o. How to obtain consistent estimators for β^o in situations similar to the examples described in Section 7.1?

We now introduce a two-stage least squares (2SLS) procedure, which can consistently estimate the true parameter β^o. The 2SLS procedure can be described as follows:

Stage 1: Regress X_t on Z_t via OLS and save the predicted value $\hat{X}_t$.

Here, the artificial linear regression model is

$$X_t = \gamma' Z_t + v_t, \qquad t = 1, ..., n,$$

where γ is a $l \times K$ parameter matrix, and v_t is a $K \times 1$ regression error. From the result in Chapter 2, we have $E(Z_t v_t) = 0$ if and only if γ is the best LS approximation coefficient, i.e., if and only if

$$\gamma = [E(Z_t Z_t')]^{-1} E(Z_t X_t).$$

In matrix form, we can write

$$\mathbf{X} = \mathbf{Z}\gamma + v,$$

where $\mathbf{X}$ is a $n \times K$ matrix, $\mathbf{Z}$ is a $n \times l$ matrix, γ is a $l \times K$ matrix, and v is a $n \times K$ matrix.

The OLS estimator for γ is

$$\begin{aligned} \hat{\gamma} &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} \\ &= \left(n^{-1}\sum_{t=1}^{n} Z_t Z_t'\right)^{-1} n^{-1}\sum_{t=1}^{n} Z_t X_t'. \end{aligned}$$

The predicted value or the sample projection of X_t on Z_t is

$$\hat{X}_t = \hat{\gamma}' Z_t$$

or in matrix form

$$\hat{\mathbf{X}} = \mathbf{Z}\hat{\gamma} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}.$$

Stage 2: Use the predicted value $\hat{X}_t$ as regressors for Y_t. Regress Y_t on $\hat{X}_t$, and the resulting OLS estimator is called the 2SLS estimator, denoted as $\hat{\beta}_{2sls}$.

Question: Why use the fitted value $\hat{X}_t = \hat{\gamma}' Z_t$ as regressors?

We first consider

$$X_t = \gamma' Z_t + v_t,$$

where γ is the best linear LS approximation coefficient, and so v_t is orthogonal to Z_t in the sense $E(Z_t v_t) = 0$. Because $E(Z_t \varepsilon_t) = 0$, the population projection $\gamma' Z_t$ is orthogonal to ε. In general, $v_t = X_t - \gamma' Z_t$, which is orthogonal to Z_t, is correlated with ε_t. In other words, the auxiliary regression in stage 1 decomposes X_t into two components: $\gamma' Z_t$ and v_t, where $\gamma' Z_t$ is orthogonal to ε_t, and v_t is correlated with ε_t.

Since the best linear LS approximation coefficient γ is unknown, we have to replace it with $\hat{\gamma}$. The fitted value $\hat{X}_t = \hat{\gamma}' Z_t$ is the (sample) projection X_t onto Z_t. The regression of X_t on Z_t purges the component of X_t that is correlated with ε_t so that the projection $\hat{X}_t$ is approximately orthogonal to ε_t given that Z_t is orthogonal to ε_t. (The word "approximately" is used here because $\hat{\gamma}$ is an estimator of γ and thus contains some estimation error.)

The regression model in the second stage can be written as

$$Y_t = \hat{X}_t' \beta^o + \tilde{\varepsilon}_t$$

or in matrix form

$$Y = \hat{\mathbf{X}} \beta^o + \tilde{\varepsilon}.$$

Note that the disturbance $\tilde{\varepsilon}_t$ is not ε_t because $\hat{X}_t$ is not X_t.

Using $\hat{\mathbf{X}} = \mathbf{Z}\hat{\gamma} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$, we can write the second stage OLS estimator, namely the 2SLS estimator as follows:

$$\begin{aligned}
\hat{\beta}_{2sls} &= (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'Y \\
&= [(\mathbf{Z}\hat{\gamma})'(\mathbf{Z}\hat{\gamma})]^{-1}(\mathbf{Z}\hat{\gamma})'Y \\
&= \left\{[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]'[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]\right\}^{-1}[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]'Y \\
&= [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'Y \\
&= [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'Y \\
&= \left[\frac{\mathbf{X}'\mathbf{Z}}{n}\left(\frac{\mathbf{Z}'\mathbf{Z}}{n}\right)^{-1}\frac{\mathbf{Z}'\mathbf{X}}{n}\right]^{-1}\frac{\mathbf{X}'\mathbf{Z}}{n}\left(\frac{\mathbf{Z}'\mathbf{Z}}{n}\right)^{-1}\frac{\mathbf{Z}'Y}{n}.
\end{aligned}$$

Using the expression $Y = \mathbf{X}\beta^o + \varepsilon$ from Assumption 7.2, we have

$$\begin{aligned}
\hat{\beta}_{2sls} - \beta^o &= \left[\frac{\mathbf{X}'\mathbf{Z}}{n}\left(\frac{\mathbf{Z}'\mathbf{Z}}{n}\right)^{-1}\frac{\mathbf{Z}'\mathbf{X}}{n}\right]^{-1}\frac{\mathbf{X}'\mathbf{Z}}{n}\left(\frac{\mathbf{Z}'\mathbf{Z}}{n}\right)^{-1}\frac{\mathbf{Z}'\varepsilon}{n} \\
&= \left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right]^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{Z'\varepsilon}{n},
\end{aligned}$$

where

$$\begin{aligned}
\hat{Q}_{zz} &= \frac{\mathbf{Z}'\mathbf{Z}}{n} = n^{-1}\sum_{t=1}^{n} Z_t Z_t', \\
\hat{Q}_{xz} &= \frac{\mathbf{X}'\mathbf{Z}}{n} = n^{-1}\sum_{t=1}^{n} X_t Z_t', \\
\hat{Q}_{zx} &= \frac{\mathbf{Z}'\mathbf{X}}{n} = n^{-1}\sum_{t=1}^{n} Z_t X_t' = \hat{Q}_{xz}'.
\end{aligned}$$

Question: What are the statistical properties of $\hat{\beta}_{2sls}$?

7.4 Consistency of 2SLS

By the WLLN for a stationary ergodic process, we have

$$\begin{aligned}
\hat{Q}_{zz} &\xrightarrow{p} Q_{zz}, \quad l \times l \\
\hat{Q}_{xz} &\xrightarrow{p} Q_{xz}, \quad K \times l, \\
\frac{Z'\varepsilon}{n} &\xrightarrow{p} E(Z_t\varepsilon_t) = 0, \quad l \times 1.
\end{aligned}$$

Also, $Q_{xz}Q_{zz}^{-1}Q_{zx}$ is a $K \times K$ symmetric and nonsingular matrix because Q_{xz} is of full rank, Q_{zz} is nonsingular, and $l \geq K$. It follows by continuity that

$$\left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right]^{-1} \xrightarrow{p} [Q_{xz}Q_{zz}^{-1}Q_{zx}]^{-1}.$$

Consequently, we have

$$\hat{\beta}_{2sls} - \beta^o \xrightarrow{p} [Q_{xz}Q_{zz}^{-1}Q_{zx}]^{-1}Q_{xz}Q_{zz}^{-1} \cdot 0 = 0.$$

We now state this consistency result in the following theorem.

Theorem [Consistency of 2SLS]: *Under Assumptions 7.1-7.4, as $n \to \infty$,*

$$\hat{\beta}_{2sls} \xrightarrow{p} \beta^o.$$

To provide intuition why the 2SLS estimator $\hat{\beta}_{2sls}$ is consistent for β^o, we consider

$$Y_t = X_t'\beta^o + \varepsilon_t.$$

The OLS estimator $\hat{\beta}$ is not consistent for β^o because $E(X_t\varepsilon_t) \neq 0$. Suppose we decompose the regressor X_t into two terms:

$$X_t = \tilde{X}_t + v_t,$$

where one $\tilde{X}_t = \gamma' Z_t$ is a projection of X_t on Z_t and so it is orthogonal to ε_t. The other component, $v_t = X_t - \tilde{X}_t$, is generally correlated with ε_t. Then consistent estimation for β^o is possible if we observe v_t and run the following augmented regression

$$\begin{aligned} Y_t &= X_t'\beta^o + \varepsilon_t \\ &= \tilde{X}_t'\beta^o + (v_t'\beta^o + \varepsilon_t) \\ &= \tilde{X}_t'\beta^o + u_t, \end{aligned}$$

where $u_t = v_t'\beta^o + \varepsilon_t$ is the disturbance when regressing Y_t on $\tilde{X}_t$. Because

$$\begin{aligned} E(\tilde{X}_t u_t) &= \gamma' E(Z_t u_t) \\ &= \gamma' E(Z_t v_t')\beta^o + \gamma' E(Z_t \varepsilon_t) \\ &= 0, \end{aligned}$$

the OLS estimator of regressing Y_t on $\tilde{X}_t$ would be consistent for β^o.

However, $\tilde{X}_t = \gamma' Z_t$ is not observable, so we need to use a proxy, i.e., $\hat{X}_t = \hat{\gamma}' Z_t$, where $\hat{\gamma}$ is the OLS estimator of regressing X_t on Z_t. This results in the 2SLS estimator $\hat{\beta}_{2sls}$. The estimation error of $\hat{\gamma}$ does not affect the consistency of the 2SLS estimator $\hat{\beta}$.

7.5 Asymptotic Normality of 2SLS

We now derive the asymptotic distribution of $\hat{\beta}_{2sls}$. Write

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{2sls} - \beta^o) &= \left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right]^{-1} \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{\mathbf{Z}'\varepsilon}{\sqrt{n}}. \\ &= \hat{A} \cdot \frac{\mathbf{Z}'\varepsilon}{\sqrt{n}}, \end{aligned}$$

where the $K \times l$ matrix

$$\hat{A} = \left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right]^{-1} \hat{Q}_{xz}\hat{Q}_{zz}^{-1}.$$

By the CLT assumption (Assumption 7.5), we have

$$\frac{\mathbf{Z}'\varepsilon}{\sqrt{n}} = n^{-\frac{1}{2}} \sum_{t=1}^{n} Z_t \varepsilon_t \xrightarrow{d} N(0, V) \sim G,$$

where V is a finite and nonsingular $l \times l$ matrix, and we denote the random vector $G \sim N(0, V)$. Then by the Slutsky theorem, we have

$$\begin{aligned} &\sqrt{n}(\hat{\beta}_{2sls} - \beta^o) \xrightarrow{d} \left(Q_{xz}Q_{zz}^{-1}Q_{zx}\right)^{-1} Q_{xz}Q_{zz}^{-1} \cdot N(0, V) \\ \sim\ & N(0, AVA') \\ \sim\ & N(0, \Omega), \end{aligned}$$

where $A = (Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}Q_{xz}Q_{zz}^{-1}$. The asymptotic variance of $\sqrt{n}(\hat{\beta}_{2sls} - \beta^o)$

$$\begin{aligned}
\text{avar}(\sqrt{n}\hat{\beta}_{2sls}) &= \Omega \\
&= AVA' \\
&= \left\{[Q_{xz}Q_{zz}^{-1}Q_{zx}]^{-1}Q_{xz}Q_{zz}^{-1}\right\} V \left\{[Q_{xz}Q_{zz}^{-1}Q_{zx}]^{-1}Q_{xz}Q_{zz}^{-1}\right\}' \\
&= [Q_{xz}Q_{zz}^{-1}Q_{zx}]^{-1}Q_{xz}Q_{zz}^{-1}VQ_{zz}^{-1}Q_{zx}[Q_{xz}Q_{zz}^{-1}Q_{zx}]^{-1}.
\end{aligned}$$

Theorem [Asymptotic Normality of 2SLS]: *Under Assumptions 7.1-7.5, as $n \to \infty$,*

$$\sqrt{n}(\hat{\beta}_{2sls} - \beta^o) \xrightarrow{d} N(0, \Omega).$$

The estimation of V depends on whether $\{Z_t\varepsilon_t\}$ is an MDS. We first consider the case where $\{Z_t\varepsilon_t\}$ is an MDS process. In this case, $V = E(Z_tZ_t'\varepsilon_t^2)$ and so we need not estimate the long-run variance-covariance matrix.

Case I: $\{Z_t\varepsilon_t\}$ is a Stationary Ergodic MDS

Assumption 7.6 [MDS]: *(i) $\{Z_t\varepsilon_t\}$ is an MDS; (ii)* $\text{var}(Z_t\varepsilon_t) = E(Z_tZ_t'\varepsilon_t^2)$ *is finite and nonsingular.*

Corollary: *Under Assumptions 7.1–7.4 and 7.6, we have as $n \to \infty$,*

$$\sqrt{n}(\hat{\beta}_{2sls} - \beta^o) \xrightarrow{d} N(0, \Omega),$$

where Ω is defined as above with $V = E(Z_tZ_t'\varepsilon_t^2)$.

There is no need to estimate a long-run variance-covariance matrix but Ω involves consistent estimation of the heteroskedasticity-consistent variance-covariance matrix V.

When $\{Z_t\varepsilon_t\}$ is an MDS with conditional homoskedasticity, the asymptotic variance Ω can be greatly simplified.

Special Case: Conditional Homoskedasticity

Assumption 7.7 [Conditional Homoskedasticity]: $E(\varepsilon_t^2|Z_t) = \sigma^2$ a.s.

Note that the conditional expectation in Assumption 7.7 is conditional on Z_t, not on X_t.

Under this assumption, by the law of iterated expectations, we obtain

$$\begin{aligned} V &= E(Z_t Z_t' \varepsilon_t^2) \\ &= E[Z_t Z_t' E(\varepsilon_t^2 | Z_t)] \\ &= \sigma^2 E(Z_t Z_t') \\ &= \sigma^2 Q_{zz}. \end{aligned}$$

It follows that

$$\begin{aligned} \Omega &= (Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}Q_{xz}Q_{zz}^{-1}\sigma^2 Q_{zz}Q_{zz}^{-1}Q_{zx}(Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1} \\ &= \sigma^2(Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}. \end{aligned}$$

Corollary [Asymptotic Normality of 2SLS under MDS with Conditional Homoskedasticity] *Under Assumptions 7.1–7.4, 7.6 and 7.7, we have as* $n \to \infty$,

$$\sqrt{n}(\hat{\beta}_{2sls} - \beta^o) \xrightarrow{d} N(0, \Omega),$$

where

$$\Omega = \sigma^2 [Q_{xz}Q_{zz}^{-1}Q_{zx}]^{-1}.$$

Case II: $\{Z_t\varepsilon_t\}$ is a Stationary Ergodic non-MDS

In this general case, we have

$$V \equiv \text{avar}\left(n^{-1/2}\sum_{t=1}^{n} Z_t\varepsilon_t\right) = \sum_{j=-\infty}^{\infty} \Gamma(j)$$

where $\Gamma(j) = \text{cov}(Z_t\varepsilon_t, Z_{t-j}\varepsilon_{t-j})$. We need to use a long-run variance-covariance estimator for V. When $\{Z_t\varepsilon_t\}$ is not an MDS, there is no need (and in fact there is no way) to consider conditional homoskedasticity and conditional heteroskedasticity separately.

7.6 Interpretation and Estimation of Ω

The asymptotic variance Ω of $\hat{\beta}_{2sls}$ is so complicated that it will be highly desirable if we can find an interpretation to help understand its structure. What is the nature of $\hat{\beta}_{2sls}$? What is Ω?

Let us revisit the second stage regression model

$$Y_t = \hat{X}_t'\beta^o + \tilde{\varepsilon}_t, \qquad (7.12)$$

where the regressor

$$\hat{X}_t = \hat{\gamma}' Z_t$$

is the sample projection of X_t on Z_t, and the disturbance $\tilde{\varepsilon}_t = Y_t - \hat{X}_t'\beta^o$. Note that $\tilde{\varepsilon}_t \neq \varepsilon_t$ because $\hat{X}_t \neq X_t$. Given $Y_t = X_t'\beta^o + \varepsilon_t$ from Assumption 7.2, we have

$$\begin{aligned}
\tilde{\varepsilon}_t &= Y_t - \hat{X}_t'\beta^o \\
&= \varepsilon_t + (X_t - \hat{X}_t)'\beta^o \\
&= \varepsilon_t + \hat{v}_t'\beta^o,
\end{aligned}$$

where ε_t is the true disturbance and $\hat{v}_t \equiv X_t - \hat{X}_t = X_t - \hat{\gamma}' Z_t$. Since $\hat{v}_t$ is the estimated residual from the first OLS regression

$$\mathbf{X} = \mathbf{Z}\gamma + v,$$

we have the following FOC holds:

$$\mathbf{Z}'(\mathbf{X} - \hat{\mathbf{X}}) = \mathbf{Z}'\hat{v} = 0.$$

It follows that the 2SLS estimator

$$\begin{aligned}
\hat{\beta}_{2sls} &= (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'Y \\
&= (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'(\hat{\mathbf{X}}\beta^o + \tilde{\varepsilon}) \\
&= \beta^o + (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'[\varepsilon + \hat{v}\beta^o] \\
&= \beta^o + (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\varepsilon
\end{aligned}$$

because $\hat{\mathbf{X}}'\hat{v} = 0$ (why?). Therefore, the asymptotic properties of $\hat{\beta}_{2sls}$ are determined by

$$\begin{aligned}
\hat{\beta}_{2sls} - \beta^o &= (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\varepsilon \\
&= \left(\frac{\hat{\mathbf{X}}'\hat{\mathbf{X}}}{n}\right)^{-1}\frac{\hat{\mathbf{X}}'\varepsilon}{n}.
\end{aligned}$$

In other words, the estimated residual $\hat{v} = \mathbf{X} - \hat{\mathbf{X}}$ from the first stage regression has no impact on the statistical properties of $\hat{\beta}_{2sls}$, although it is a component of $\tilde{\varepsilon}_t$. Thus, when analyzing the asymptotic properties of $\hat{\beta}_{2sls}$, we can proceed as if we were estimating $Y = \hat{\mathbf{X}}\beta^o + \varepsilon$ by OLS.

Next, recall that we have

$$\begin{aligned}
\hat{\mathbf{X}} &= \mathbf{Z}\hat{\gamma}, \\
\hat{\gamma} &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} \\
&\to {}^p Q_{zz}^{-1} Q_{zx} = \gamma
\end{aligned}$$

By the WLLN, the sample projection $\hat{X}_t$ "converges" to the population projection $\tilde{X}_t \equiv \gamma' Z_t$ as $n \to \infty$. That is, $\hat{X}_t$ will become arbitrarily close to $\tilde{X}_t$ as $n \to \infty$. In fact, the estimation error of $\hat{\gamma}$ in the first stage has no impact on the asymptotic properties of $\hat{\beta}_{2sls}$.

Thus, we can consider the following artificial regression model

$$Y_t = \tilde{X}_t' \beta^o + \varepsilon_t, \tag{7.13}$$

whose infeasible OLS estimator

$$\tilde{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y.$$

As we will show below, the asymptotic properties of $\hat{\beta}_{2sls}$ are the same as those of the infeasible OLS estimator $\tilde{\beta}$. This helps a lot in understanding the variance-covariance structure of $\hat{\beta}_{2sls}$. It is important to emphasize that the equation in (7.13) is not derived from other equations. It is just a convenient way to understand the nature of $\hat{\beta}_{2sls}$.

We now show that the asymptotic properties of $\hat{\beta}_{2sls}$ are the same as the asymptotic properties of $\tilde{\beta}$. For the asymptotic normality, observe that

$$\begin{aligned} \sqrt{n}(\tilde{\beta} - \beta^o) &= \hat{Q}_{\tilde{x}\tilde{x}}^{-1} \frac{\tilde{X}'\varepsilon}{\sqrt{n}} \\ &\to {}^d Q_{\tilde{x}\tilde{x}}^{-1} \cdot N(0, \tilde{V}) \\ &\sim N(0, Q_{\tilde{x}\tilde{x}}^{-1} \tilde{V} Q_{\tilde{x}\tilde{x}}^{-1}) \end{aligned}$$

using the asymptotic theory in Chapters 5 and 6, where

$$\begin{aligned} Q_{\tilde{x}\tilde{x}} &\equiv E(\tilde{X}_t \tilde{X}_t'), \\ \tilde{V} &\equiv \text{avar}\left(n^{-1/2} \sum_{t=1}^{n} \tilde{X}_t \varepsilon_t \right). \end{aligned}$$

We first consider the case where $\{Z_t \varepsilon_t\}$ is MDS with conditional homoskedasticity.

Case 1: MDS with Conditional Homoskedasticity

Suppose $\{\tilde{X}_t \varepsilon_t\}$ is MDS, and $E(\varepsilon_t^2 | \tilde{X}_t) = \sigma^2$ a.s. Then we have

$$\begin{aligned} \tilde{V} &= E(\tilde{X}_t \tilde{X}_t' \varepsilon_t^2) \\ &= \sigma^2 Q_{\tilde{x}\tilde{x}} \end{aligned}$$

by the law of iterated expectations (LIE). It follows that

$$\sqrt{n}(\tilde{\beta} - \beta^o) \xrightarrow{d} N(0, \sigma^2 Q_{\tilde{x}\tilde{x}}^{-1}).$$

Because $\tilde{X}_t = \gamma' Z_t$, $\gamma = Q_{zz}^{-1} Q_{zx}$, we have

$$\begin{aligned} Q_{\tilde{x}\tilde{x}} &= E(\tilde{X}_t \tilde{X}_t') \\ &= \gamma' E(Z_t Z_t') \gamma \\ &= \gamma' Q_{zz} \gamma \\ &= Q_{xz} Q_{zz}^{-1} Q_{zz} Q_{zz}^{-1} Q_{zx} \\ &= Q_{xz} Q_{zz}^{-1} Q_{zx}. \end{aligned}$$

Therefore,

$$\begin{aligned} \sigma^2 Q_{\tilde{x}\tilde{x}}^{-1} &= \sigma^2 \left(Q_{xz} Q_{zz}^{-1} Q_{zx} \right)^{-1} \\ &= \Omega \equiv \text{avar}(\sqrt{n}\hat{\beta}_{2sls}). \end{aligned}$$

This implies that the asymptotic distribution of $\tilde{\beta}$ is indeed the same as the asymptotic distribution of $\hat{\beta}_{2sls}$ under the MDS with conditional homoskedasticity.

The asymptotic variance formula

$$\text{avar}(\sqrt{n}\hat{\beta}_{2sls}) = \sigma^2 Q_{\tilde{x}\tilde{x}}^{-1} = \sigma^2 (\gamma' Q_{zz} \gamma)^{-1}$$

indicates that the asymptotic variance of $\sqrt{n}\hat{\beta}_{2sls}$ will be large if the correlation between Z_t and X_t, as measured by γ, is weak. Thus, more precise estimation of β^o will be obtained if one chooses the instrument vector Z_t such that Z_t is highly correlated with X_t.

Question: How to estimate Ω under the MDS disturbances with conditional homoskedasticity?

Consider the asymptotic variance estimator

$$\begin{aligned} \hat{\Omega} &= \hat{s}^2 \hat{Q}_{\hat{x}\hat{x}}^{-1} \\ &= \hat{s}^2 \left(\hat{Q}_{xz} \hat{Q}_{zz}^{-1} \hat{Q}_{zx} \right)^{-1} \end{aligned}$$

where $\hat{s}^2 = \hat{e}'\hat{e}/(n-K)$, $\hat{e} = Y - \mathbf{X}\hat{\beta}_{2sls}$,

$$\hat{Q}_{\hat{x}\hat{x}} = n^{-1} \sum_{t=1}^{n} \hat{X}_t \hat{X}_t'$$

and $\hat{X}_t = \hat{\gamma}' Z_t$ is the sample projection of X_t on Z_t. Note that we have to use $\hat{X}_t$ rather than $\tilde{X}_t$ because $\tilde{X}_t = \gamma' Z_t$ is unknown.

It should be emphasized that $\hat{e}$ is not the estimated residual from the second stage regression (i.e., not from the regression of Y on $\hat{X}$). This implies that even under conditional homoskedasticity, the conventional t-statistic in the second stage regression does not converge to $N(0,1)$ in distribution, and $J \cdot \hat{F}$ does not converge to χ^2_J where $\hat{F}$ is the F-statistic in the second stage regression.

To show $\hat{\Omega} \to^p \Omega$, we shall show (i) $\hat{Q}^{-1}_{\hat{x}\hat{x}} \to^p Q^{-1}_{\tilde{x}\tilde{x}}$ and (ii) $\hat{s}^2 \to^p \sigma^2$.

We first show (i). There are two methods for proving this.

Method 1: We shall show $\hat{Q}^{-1}_{\hat{x}\hat{x}} \to^p Q^{-1}_{\tilde{x}\tilde{x}}$. Because $\hat{X}_t = \hat{\gamma}' Z_t$ and $\hat{\gamma} \to^p \gamma$, we have

$$\begin{aligned}
\hat{Q}_{\hat{x}\hat{x}} &= n^{-1} \sum_{t=1}^{n} \hat{X}_t \hat{X}_t' \\
&= \hat{\gamma}' \left(n^{-1} \sum_{t=1}^{n} Z_t Z_t' \right) \hat{\gamma} \\
&= \hat{\gamma}' \hat{Q}_{zz} \hat{\gamma} \\
&\xrightarrow{p} \gamma' Q_{zz} \gamma \\
&= E[(\gamma' Z_t)(Z_t' \gamma)] \\
&= E(\tilde{X}_t \tilde{X}_t') \\
&= Q_{\tilde{x}\tilde{x}}.
\end{aligned}$$

Method 2: We shall show $(\hat{Q}_{xz} \hat{Q}^{-1}_{zz} \hat{Q}_{zx})^{-1} \to^p (Q_{xz} Q^{-1}_{zz} Q_{zx})^{-1}$, which follows immediately from $\hat{Q}_{xz} \to^p Q_{xz}$ and $\hat{Q}_{zz} \to^p Q_{zz}$ by the WLLN. This method is more straightforward but is less intuitive than the first method.

Next, we shall show (ii) $\hat{s}^2 \to^p \sigma^2$. We decompose

$$\begin{aligned}
\hat{s}^2 &= \frac{\hat{e}'\hat{e}}{n-K} \\
&= \frac{1}{n-K}\sum_{t=1}^{n}(Y_t - X_t'\hat{\beta}_{2sls})^2 \\
&= \frac{1}{n-K}\sum_{t=1}^{n}[\varepsilon_t - X_t'(\hat{\beta}_{2sls}-\beta^o)]^2 \\
&= \frac{1}{n-K}\sum_{t=1}^{n}\varepsilon_t^2 \\
&\quad +(\hat{\beta}_{2sls}-\beta^o)'\frac{1}{n-K}\sum_{t=1}^{n}X_tX_t'(\hat{\beta}_{2sls}-\beta^o) \\
&\quad -2(\hat{\beta}_{2sls}-\beta^o)'\frac{1}{n-K}\sum_{t=1}^{n}X_t\varepsilon_t \\
&\to^p \sigma^2 + 0\cdot Q_{xx}\cdot 0 - 2\cdot 0\cdot E(X_t\varepsilon_t) \\
&= \sigma^2.
\end{aligned}$$

Note that although $E(X_t\varepsilon_t) \neq 0$, the last term still vanishes to zero in probability, because $\hat{\beta}_{2sls} - \beta^o \to^p 0$.

Question: What happens if we use $s^2 = e'e/(n-K)$, where $e = Y - \hat{\mathbf{X}}\hat{\beta}_{2sls}$ is the estimated residual from the second stage regression? Do we still have $s^2 \to^p \sigma^2$?

We have proved the following theorem.

Theorem [Consistency of $\hat{\Omega}$ under MDS with Conditional Homoskedasticity]: *Under Assumptions 7.1 – 7.4, 7.6 and 7.7, we have as $n \to \infty$,*

$$\hat{\Omega} = \hat{s}^2\hat{Q}_{\hat{x}\hat{x}}^{-1} \xrightarrow{p} \Omega = \sigma^2 Q_{\tilde{x}\tilde{x}}^{-1} = \sigma^2\left(Q_{xz}Q_{zz}^{-1}Q_{zx}\right)^{-1}.$$

Case 2: $\{Z_t\varepsilon_t\}$ is **an MDS with Conditional Heteroskedasticity**

When there exists conditional heteroskedasticity but $\{Z_t\varepsilon_t\}$ is still an MDS, the infeasible OLS estimator $\tilde{\beta}$ in the artificial regression

$$Y = \tilde{X}\beta^o + \varepsilon$$

has the following asymptotic distribution:

$$\sqrt{n}(\tilde{\beta}-\beta^o) \xrightarrow{d} N(0, Q_{\tilde{x}\tilde{x}}^{-1}\tilde{V}Q_{\tilde{x}\tilde{x}}^{-1}),$$

where

$$\tilde{V} = E(\tilde{X}_t \tilde{X}_t' \varepsilon_t^2).$$

Given $\tilde{X}_t = \gamma' Z_t$, $\gamma = Q_{zz}^{-1} Q_{zx}$, $Q_{\tilde{x}\tilde{x}} = \gamma' Q_{zz} \gamma$, and $\tilde{V} = \gamma' E(Z_t Z_t' \varepsilon_t^2)\gamma = \gamma' \ V\gamma$, where $V = E(Z_t Z_t' \varepsilon_t^2)$ under the MDS assumption with conditional heteroskedasticity, we have

$$\begin{aligned}
\text{avar}(\sqrt{n}\tilde{\beta}) &= Q_{\tilde{x}\tilde{x}}^{-1} \tilde{V} Q_{\tilde{x}\tilde{x}}^{-1} \\
&= [E(\tilde{X}_t \tilde{X}_t')]^{-1} E[\tilde{X}_t \tilde{X}_t' \varepsilon_t^2][E(\tilde{X}_t \tilde{X}_t')]^{-1} \\
&= [\gamma' E(Z_t Z_t')\gamma]^{-1} \gamma' E(Z_t Z_t' \varepsilon_t^2)\gamma[\gamma' E(Z_t Z_t')\gamma]^{-1} \\
&= \left(Q_{xz} Q_{zz}^{-1} Q_{zx}\right)^{-1} Q_{xz} Q_{zz}^{-1} V Q_{zz}^{-1} Q_{zx} (Q_{xz} Q_{zz}^{-1} Q_{zx})^{-1} \\
&= \Omega \equiv \ \text{avar}(\sqrt{n}\hat{\beta}_{2sls}).
\end{aligned}$$

This implies that the asymptotic distribution of the infeasible OLS estimator $\tilde{\beta}$ is the same as the asymptotic distribution of $\hat{\beta}_{2sls}$ under MDS with conditional heteroskedasticity. Therefore, the estimator for Ω is

$$\hat{\Omega} = \hat{Q}_{\hat{x}\hat{x}}^{-1} \hat{V}_{\hat{x}\hat{x}} \hat{Q}_{\hat{x}\hat{x}}^{-1},$$

where

$$\begin{aligned}
\hat{V}_{\hat{x}\hat{x}} &= n^{-1} \sum_{t=1}^{n} \hat{X}_t \hat{X}_t' \hat{e}_t^2 \\
&= \hat{\gamma}' \left(n^{-1} \sum_{t=1}^{n} Z_t Z_t' \hat{e}_t^2 \right) \hat{\gamma},
\end{aligned}$$

where $\hat{\gamma} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} = \hat{Q}_{zz}^{-1} \hat{Q}_{zx}$ and $\hat{e}_t = Y_t - X_t' \hat{\beta}_{2sls}$. This is a White's (1980) heteroskedasticity-consistent variance estimator for $\hat{\beta}_{2sls}$.

Now, put

$$\hat{V} \equiv n^{-1} \sum_{t=1}^{n} Z_t Z_t' \hat{e}_t^2,$$

Then

$$\hat{\Omega} = [\hat{Q}_{xz} \hat{Q}_{zz}^{-1} \hat{Q}_{zx}]^{-1} \hat{Q}_{xz} \hat{Q}_{zz}^{-1} \hat{V} \hat{Q}_{zz}^{-1} \hat{Q}_{zx} [\hat{Q}_{xz} \hat{Q}_{zz}^{-1} \hat{Q}_{zx}]^{-1},$$

where (please check it!)

$$\begin{aligned}
\hat{V} &= n^{-1} \sum_{t=1}^{n} Z_t Z_t' \hat{e}_t^2 \\
&\to {}^{p} V = E(Z_t Z_t' \varepsilon_t^2)
\end{aligned}$$

under suitable regularity conditions.

Question: How to show $\hat{\Omega} \to^p \Omega$ under MDS with conditional heteroskedasticity?

Again, there are two methods to show $\hat{\Omega} \to^p \Omega$ here.

Method 1: We shall show $\hat{Q}_{\hat{x}\hat{x}} \to^p Q_{\tilde{x}\tilde{x}}$ and $\hat{V}_{\hat{x}\hat{x}} \to^p \tilde{V}$. The fact that $\hat{Q}_{\hat{x}\hat{x}} \to^p Q_{\tilde{x}\tilde{x}}$ has been shown earlier in the case of conditional homoskedasticity. To show $\hat{V}_{\hat{x}\hat{x}} \to^p \tilde{V}$, we write

$$\begin{aligned}
\hat{V}_{\hat{x}\hat{x}} &= n^{-1}\sum_{t=1}^{n} \hat{X}_t \hat{X}_t \hat{e}_t^2 \\
&= \hat{\gamma}' \left(n^{-1} \sum_{t=1}^{n} Z_t Z_t' \hat{e}_t^2 \right) \hat{\gamma} \\
&= \hat{\gamma}' \hat{V} \hat{\gamma}.
\end{aligned}$$

Because $\hat{\gamma} \to^p \gamma$, and following the consistency proof for $n^{-1}\sum_{t=1}^{n} X_t X_t' e_t^2$ in Chapter 4, we can show (please verify!) that

$$\begin{aligned}
\hat{V} &= n^{-1}\sum_{t=1}^{n} Z_t Z_t' \hat{e}_t^2 \\
&\to {}^p E(Z_t Z_t' \varepsilon_t^2) = V,
\end{aligned}$$

under the following additional moment condition:

Assumption 7.8: *(i)* $E(Z_{jt}^4) < \infty$ *for all* $0 \le j \le l$; *and (ii)* $E(\varepsilon_t^4) < \infty$.

It follows that

$$\begin{aligned}
\hat{V}_{\hat{x}\hat{x}} &\to {}^p \gamma' E(Z_t Z_t' \varepsilon_t^2) \gamma \\
&= E(\tilde{X}_t \tilde{X}_t' \varepsilon_t^2) \\
&= \tilde{V}.
\end{aligned}$$

This and $\hat{Q}_{\hat{x}\hat{x}} \to^p Q_{\tilde{x}\tilde{x}}$ imply $\hat{\Omega} \to^p \Omega$.

Method 2: Given that

$$\hat{\Omega} = \left(\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right)^{-1} \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{V}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\left(\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right)^{-1},$$

it suffices to show $\hat{Q}_{xz} \to^p Q_{zx}, \hat{Q}_{zz} \to^p Q_{zz}$ and $\hat{V} \to^p V$. The first two results immediately follow by the WLLN. The last result follows by using a similar reasoning of the consistency proof for $n^{-1}\sum_{t=1}^{n} X_t X_t' e_t^2$ in Chapter 4 or 5.

We now summarize the result derived above.

Theorem [Consistency of $\hat{\Omega}$ under MDS with Conditional Heteroskedasticity]: *Under Assumptions 7.1-7.4, 7.6 and 7.8, we have as $n \to \infty$,*

$$\begin{aligned}
\hat{\Omega} &= \hat{Q}_{\hat{x}\hat{x}}^{-1}\hat{V}_{\hat{x}\hat{x}}\hat{Q}_{\hat{x}\hat{x}}^{-1} \\
&\to^p \Omega = Q_{\tilde{x}\tilde{x}}^{-1}\tilde{V}Q_{\tilde{x}\tilde{x}}^{-1} \\
&= (Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}Q_{xz}Q_{zz}^{-1}VQ_{zz}^{-1}Q_{zx}(Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}.
\end{aligned}$$

where $\tilde{V} = E(\tilde{X}_t\tilde{X}_t'\varepsilon_t^2)$ and $V = E(Z_tZ_t'\varepsilon_t^2)$.

CASE 3: $\{Z_t\varepsilon_t\}$ is a Stationary Ergodic non-MDS

Finally, we consider a general case where $\{Z_t\varepsilon_t\}$ is not an MDS, which may arise as in the examples discussed in Chapter 6.

In this case, we have $\sqrt{n}(\hat{\beta}_{2sls} - \beta^o) \to^d N(0, \Omega)$ as $n \to \infty$, where

$$\begin{aligned}
\Omega &= Q_{\tilde{x}\tilde{x}}^{-1}\tilde{V}Q_{\tilde{x}\tilde{x}}^{-1} \\
&= (Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}Q_{xz}Q_{zz}^{-1}VQ_{zz}^{-1}Q_{zx}(Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1},
\end{aligned}$$

with

$$\begin{aligned}
\tilde{V} &= \sum_{j=-\infty}^{\infty}\tilde{\Gamma}(j), \qquad \tilde{\Gamma}(j) = \operatorname{cov}(\tilde{X}_t\varepsilon_t, \tilde{X}_{t-j}\varepsilon_{t-j}), \\
V &= \sum_{j=-\infty}^{\infty}\Gamma(j), \qquad \Gamma(j) = \operatorname{cov}(Z_t\varepsilon_t, Z_{t-j}\varepsilon_{t-j}).
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
\operatorname{avar}(\sqrt{n}\tilde{\beta}) &= Q_{\tilde{x}\tilde{x}}^{-1}V_{\tilde{x}\tilde{x}}Q_{\tilde{x}\tilde{x}}^{-1} \\
&= (\gamma'Q_{xx}\gamma)^{-1}\gamma'V\gamma(\gamma'Q_{xx}\gamma)^{-1} \\
&= \Omega \equiv \operatorname{avar}(\sqrt{n}\hat{\beta}_{2sls}).
\end{aligned}$$

Thus, the asymptotic variance of $\sqrt{n}\hat{\beta}_{2sls}$ is the same as the asymptotic variance of $\tilde{\beta}$ under this general case.

Question: How to estimate Ω?
Answer: Use a long-run variance-covariance matrix estimator for V or $\tilde{V}$.

We directly assume that we have a consistent estimator $\hat{V}$ for V.

Assumption 7.9: $\hat{V} \to^p V \equiv \Sigma_{j=-\infty}^{\infty}\Gamma(j)$, where $\Gamma(j) = \text{cov}(Z_t\varepsilon_t, Z_{t-j}\varepsilon_{t-j})$.

Question: How to estimate $\tilde{V} = \Sigma_{j=-\infty}^{\infty}\tilde{\Gamma}(j)$?

Recall that $\tilde{\Gamma}(j) = \gamma'\Gamma(j)\gamma$. A consistent estimator for $\tilde{V}$ can be given by

$$\hat{\gamma}'\hat{V}\hat{\gamma} \to^p \tilde{V}.$$

Theorem [Consistency of $\hat{\Omega}$ under Non-MDS]: *Under Assumptions 7.1-7.4, and 7.9, we have as $n \to \infty$,*

$$\begin{aligned}
\hat{\Omega} &= \hat{Q}_{\hat{x}\hat{x}}^{-1}\hat{V}_{\hat{x}\hat{x}}\hat{Q}_{\hat{x}\hat{x}}^{-1} \\
&= (\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx})^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{V}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx})^{-1} \\
&\to {}^p\Omega = Q_{\tilde{x}\tilde{x}}^{-1}\tilde{V}Q_{\tilde{x}\tilde{x}}^{-1},
\end{aligned}$$

where $\hat{V}_{\hat{x}\hat{x}} = \hat{\gamma}\hat{V}\hat{\gamma}'$ and

$$\Omega = (Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}Q_{xz}Q_{zz}^{-1}VQ_{zz}^{-1}Q_{zx}(Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}.$$

With a consistent estimator of Ω, we can develop various confidence interval estimators and various tests for the null hypothesis $\mathbf{H}_0 : R\beta^o = r$. We will consider the latter now.

7.7 Hypothesis Testing

Now, consider the null hypothesis of interest

$$\mathbf{H}_0 : R\beta^o = r,$$

where R is a $J \times K$ nonstochastic matrix, and r is a $J \times 1$ nonstochastic vector. The test statistics will differ depending on whether $\{Z_t\varepsilon_t\}$ is an MDS, and whether $\{\varepsilon_t\}$ is conditionally homoskedastic when $\{Z_t\varepsilon_t\}$ is an MDS. For space, here we do not present the results on t-type test statistics when $J = 1$.

CASE I: $\{Z_t\varepsilon_t\}$ is an MDS with Conditional Homoskedasticity

Theorem [Hypothesis Testing]: *Put $\hat{e} \equiv Y - \mathbf{X}\hat{\beta}_{2sls}$. Then under Assumptions 7.1-7.4, 7.6 and 7.7, the Wald test statistic*

$$\begin{aligned} \hat{W} &= \frac{n(R\hat{\beta}_{2sls} - r)'[R(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}R']^{-1}(R\hat{\beta}_{2sls} - r)}{\hat{e}'\hat{e}/(n-K)} \\ &\to {}^{d}\chi_J^2 \text{ as } n \to \infty, \end{aligned}$$

under $\mathbf{H}_0$.

Proof: The result follows immediately from the asymptotic normality theorem for $\sqrt{n}(\hat{\beta}_{2sls} - \beta^o)$, $\mathbf{H}_0$ (which implies $\sqrt{n}(R\hat{\beta}_{2sls} - r) = R\sqrt{n}(\hat{\beta}_{2sls} - \beta^o)$), the consistent asymptotic variance estimation theorem, and the Slutsky theorem.

Remarks:

Question: Is $\hat{W}/J$ the F-statistic from the second stage regression?

Answer: No, because $\hat{e}$ is not the estimated residual from the second stage regression.

Question: Do we still have

$$\hat{F} = \frac{(e_r'e_r - e_u'e_u)/J}{e_u'e_u/(n-K)},$$

where e_r and e_u are estimated residuals from the restricted and unrestricted regression models in the second stage regression respectively?

Answer: No. (Why?)

Case II: $\{Z_t\varepsilon_t\}$ is a Stationary Ergodic MDS with Conditional Heteroskedasticity

Theorem [Hypothesis Testing]: *Under Assumptions 7.1-7.4, 7.6 and 7.8, the Wald test statistic*

$$\begin{aligned} \hat{W} &\equiv n(R\hat{\beta}_{2sls} - r)'[R\hat{Q}_{\hat{x}\hat{x}}^{-1}\hat{V}_{\hat{x}\hat{x}}\hat{Q}_{\hat{x}\hat{x}}^{-1}R']^{-1}(R\hat{\beta}_{2sls} - r) \\ &\to {}^{d}\chi_J^2 \end{aligned}$$

under $\mathbf{H}_0$, *where* $\hat{V}_{\hat{x}\hat{x}} = n^{-1}\Sigma_{t=1}^n \hat{X}_t\hat{X}_t'\hat{e}_t^2$ *and* $\hat{e}_t = Y_t - X_t'\hat{\beta}_{2sls}$.

Question: Suppose there exists conditional homoskedasticity but we use $\hat{W}$ above. Is $\hat{W}$ an asympotically valid procedure in this case?

Answer: Yes, $\hat{W}$ is asymptotically valid. However, the finite sample performance of $\hat{W}$ will be generally less satisfactory than the test statistic in Case I.

CASE III: $\{Z_t\varepsilon_t\}$ **is a Stationary ergodic non-MDS**

When $\{Z_t\varepsilon_t\}$ is non-MDS, we can still construct a Wald test which is robust to conditional heteroskedasticity and autocorrelation, as is stated below.

Theorem [Hypothesis Testing]: *Under Assumptions 7.1-7.5 and 7.9, the Wald test statistic*

$$\begin{aligned}\hat{W} &= n(R\hat{\beta}_{2sls}-r)'[R\hat{Q}_{\hat{x}\hat{x}}^{-1}\hat{V}_{\hat{x}\hat{x}}\hat{Q}_{\hat{x}\hat{x}}^{-1}R']^{-1}(R\hat{\beta}_{2sls}-r) \\ &\to {}^d\chi^2_J\end{aligned}$$

under $\mathbf{H}_0$, *where* $\hat{V}_{\hat{x}\hat{x}} = \hat{\gamma}'\hat{V}\hat{\gamma}, \hat{\gamma} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$ and $\hat{V}$ *is a long-run variance-covariance estimator for* $V = \Sigma_{j=-\infty}^{\infty}\Gamma(j)$ *with* $\Gamma(j) = \text{cov}(Z_t\varepsilon_t, Z_{t-j}\varepsilon_{t-j})$.

7.8 Hausman's Test

When there exists endogeneity so that $E(X_t\varepsilon_t) \neq 0$, the OLS estimator $\hat{\beta}$ is inconsistent for β^o. Instead, the 2SLS estimator $\hat{\beta}_{2sls}$ should be used, which involves the choice of the instrumental vector Z which in turn affects the efficiency of $\hat{\beta}_{2sls}$. In practice, it is not uncommon that practitioners are not sure whether there exists endogeneity. In this section, we introduce Hausman's (1978) test for endogeneity. The null hypothesis of interest is:

$$\mathbf{H}_0 : E(\varepsilon_t|X_t) = 0.$$

If this null hypothesis is rejected, one has to use the 2SLS estimator $\hat{\beta}_{2sls}$ provided that one can find a set of instruments Z_t that satisfies Assumption 7.4.

For simplicity, we impose the following conditions.

Assumption 7.10: (i) $\{(X_t', Z_t')'\varepsilon_t\}$ is an MDS process; and (ii) $E(\varepsilon_t^2|X_t, Z_t) = \sigma^2$ a.s.

Assumptions 7.10 is made for simplicity. They could be relaxed to be a non-MDS process with conditional heteroskedasticity but Hausman's (1978) test statistic to be introduced below should be generalized.

Question: How to test the conditional homoskedasticity assumption that $E(\varepsilon_t^2|X_t, Z_t) = \sigma^2$?

Answer: Put $\hat{e}_t = Y_t - \hat{X}_t'\hat{\beta}_{2sls}$. (Question: Can we use $e_t = Y_t - X_t'\hat{\beta}_{2sls}$?) Then run an auxiliary regression of $\hat{e}_t^2$ on vech(U_t), where $U_t = (X_t', Z_t')'$, a $(K+l) \times 1$ vector. Then under the condition that $E(\varepsilon_t^4|X_t, Z_t) = \mu_4$ is a constant, we have $nR^2 \to^d \chi^2_J$ under the null hypothesis of conditional homoskedasticity, where $J = (K + l)(K + l + 1)/2 - 1$.

The basic idea of Hausman's test is under $\mathbf{H}_0 : E(\varepsilon_t|X_t) = 0$, both the OLS estimator $\hat{\beta} = (X'X)^{-1}X'Y$ and the 2SLS estimator $\hat{\beta}_{2sls}$ are consistent for β^o. They converge to the same limit β^o but it can shown that $\hat{\beta}$ is an asymptotically efficient estimator while $\hat{\beta}_{2sls}$ is not. Under the alternatives to H_0, $\hat{\beta}_{2sls}$ remains to be consistent for β^o but $\hat{\beta}$ is not. Hausman (1978) considers a test for H_0 based on the difference between the two estimators:

$$\hat{\beta}_{2sls} - \hat{\beta},$$

which converges to zero under H_0 but generally to a nonzero constant under the alternatives to H_0, giving the test its power when the sample size n is sufficiently large.

To construct Hausman's (1978) test statistic, we need to derive the asymptotic distribution of $\hat{\beta}_{2sls} - \hat{\beta}$ For this purpose, we first state a lemma.

Lemma: Suppose $\hat{A} \to^p A$ and $\hat{B} = O_P(1)$. Then $(\hat{A} - A)\hat{B} \to^p 0$.

We first consider the OLS $\hat{\beta}$. Note that

$$\sqrt{n}(\hat{\beta} - \beta^o) = \hat{Q}_{xx}^{-1} n^{-1/2} \sum_{t=1}^{n} X_t \varepsilon_t$$

where $\hat{Q}_{xx}^{-1} \to^p Q_{xx}^{-1}$ and

$$n^{-1/2} \sum_{t=1}^{n} X_t \varepsilon_t \to^d N(0, \sigma^2 Q_{xx})$$

as $n \to \infty$ (see Chapter 5). It follows that $n^{-1/2} \sum_{t=1}^{n} X_t \varepsilon_t = O_P(1)$, and by Lemma, we have

$$\sqrt{n}(\hat{\beta} - \beta^o) = Q_{xx}^{-1} n^{-1/2} \sum_{t=1}^{n} X_t \varepsilon_t + o_P(1).$$

Similarly, we can obtain

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{2sls} - \beta^o) &= \hat{A} n^{-1/2} \sum_{t=1}^{n} Z_t \varepsilon_t \\ &= A n^{-1/2} \sum_{t=1}^{n} Z_t \varepsilon_t + o_P(1), \end{aligned}$$

where $\hat{A} = (\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx})^{-1}\hat{Q}_{xz}\hat{Q}_{zz} \to^p A = (Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}Q_{xz}Q_{zz}^{-1}$ and $n^{-1/2}\sum_{t=1}^n Z_t\varepsilon_t \to^d N(0, \sigma^2 Q_{zz})$ (see Theorem). It follows that

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_{2sls} - \hat{\beta}) &= n^{-1/2}\sum_{t=1}^n \left[(Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}Q_{xz}Q_{zz}^{-1}Z_t - Q_{xx}^{-1}X_t\right]\varepsilon_t + o_P(1) \\
&\to {}^d N(0, \sigma^2(Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1} - \sigma^2 Q_{xx}^{-1})
\end{aligned}$$

by the CLT for the stationary ergodic MDS process and Assumption 7.10. Therefore, under the null hypothesis $\mathbf{H}_0$, the quadratic form

$$\begin{aligned}
H &= \frac{n(\hat{\beta}_{2sls} - \hat{\beta})'\left[(\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx})^{-1} - \hat{Q}_{xx}^{-1}\right]^{-1}(\hat{\beta}_{2sls} - \hat{\beta})}{s^2} \\
&\to {}^d\chi_K^2
\end{aligned}$$

as $n \to \infty$ by the Slutsky theorem, where $s^2 = e'e/n$ is the residual variance estimator based on the OLS residual $e = Y - X\hat{\beta}$. This is called Hausman's test statistic.

Question: Can we replace the residual variance estimator s^2 by $\hat{s}^2 = \hat{e}'\hat{e}/n$, where $\hat{e} = Y - X\hat{\beta}_{2sls}$?

Theorem [Hausman's Test for Endogeneity] *Suppose Assumptions 7.1–7.4, 7.10 and $\mathbf{H}_0$ hold, and $Q_{xx} - Q_{xz}Q_{zz}^{-1}Q_{zx}$ is strictly positive definite. Then as $n \to \infty$,*

$$H \to^d \chi_K^2.$$

Remarks:

We note that in the above Theorem,

$$\begin{aligned}
\text{avar}[\sqrt{n}(\hat{\beta}_{2sls} - \hat{\beta})] &= \sigma^2(Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1} - \sigma^2 Q_{xx}^{-1} \\
&= \text{avar}(\sqrt{n}\hat{\beta}_{2sls}) - \text{avar}(\sqrt{n}\hat{\beta}).
\end{aligned}$$

This simple asymptotic variance-covariance structure is made possible under Assumption 7.10. Suppose there exists conditional heteroskedasticity (i.e., $E(\varepsilon_t^2|X_t, Z_t) \neq \sigma^2$). Then we no longer have the above simple variance-covariance structure for $\text{avar}[\sqrt{n}(\hat{\beta} - \hat{\beta}_{2sls})]$.

The variance-covariance $(Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1} - Q_{xx}^{-1}$ may become singular when its rank $J < K$. In this case, we have to modify the Hausman's test statistic by using the generalized inverse of the variance estimator:

$$H = \frac{n(\hat{\beta}_{2sls} - \hat{\beta})'\left[(\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx})^{-1} - \hat{Q}_{xx}^{-1}\right]^{-}(\hat{\beta}_{2sls} - \hat{\beta})}{s^2}$$

Note that now $H \to^d \chi^2_J$ under $\mathbf{H}_0$ where $J < K$.

Question: What is the generalized inverse A^- of matrix A?

Question: How to modify the Hausman's test statistic so that it remains asymptotically χ^2_K when there exists conditional heteroskedasticity (i.e., $E(\varepsilon_t^2|X_t, Z_t) \neq \sigma^2$) but $\{(X_t', Z_t')'\varepsilon_t\}$ is still an MDS process?

In fact, Hausman's (1978) test is a general approach to testing model specification, not merely whether endogeneity exists. For example, it can be used to test whether a fixed effect panel regression model or a random effect panel regression model should be used. In Hausman (1978), two estimators are compared, one of which is asymptotically efficient under the null hypothesis but inconsistent under the alternative and another of which is asymptotically inefficient but consistent under the alternative hypothesis. This approach was extended by White (1981) to compare any two different estimators either of which need not be asymptotically most efficient. The methods of Hausman and White were further extended by Newey (1985), Tauchen (1985) and White (1990) to construct moment-based tests for model specification.

Hausman's test is used to check whether $E(\varepsilon_t|X_t) = 0$. Suppose this orthoganality condition fails, one has to choose an instrumental vector Z_t that satisfies Assumption 7.4. When we choose a set of variables Z_t, how can we check the validity of Z_t as instruments? In particular, how to check whether $E(\varepsilon_t|Z_t) = 0$? For this purpose, we will consider a so-called overidentification test, which will be introduced in Chapter 8.

7.9 Empirical Applications

Application I: Incentives in Chinese State-owned Enterprises

Groves, Hong, McMillan and Naughton (1994, Quaterly Journal of Economics)

Application II: The Consumption Function

Campbell and Mankiw (1989, 1991)
The consumption function

$$\begin{aligned} \Delta C_t &= \mu + \lambda \Delta Y_t + \varepsilon_t, \\ \Delta Y_t &= Z_t'\delta + v_t, \end{aligned}$$

where ΔY_t is income growth, ΔC_t is consumption growth.

7.10 Summary and Concluding Remarks

In this chapter, we discuss the possibilities that the condition of $E(\varepsilon_t|X_t) = 0$ may fail in practice, which will render inconsistent the OLS estimator for the true model parameters. With the use of instrumental variables, we introduce a consistent two-stage least squares (2SLS) estimator. We investigate the statistical properties of the 2SLS estimator and provide some interpretations that can enhance deeper understanding of the nature of the 2SLS estimator. We discuss how to construct consistent estimators for the asymptotic variance of the 2SLS estimator under various scenarios, including MDS with conditional homoskedasticity, MDS with conditional heteroskedasticity, and non-MDS possibly with conditional heteroskedasticity. For the latter, consistent estimation for the long-run variance covariance matrix is needed. With these consistent asymptotic variance estimators, various hypothesis test procedures are proposed. It is important to emphasize that the conventional t-test and F-test cannot be used even for large samples. Finally, some empirical applications that employ 2SLS are considered.

In fact, the 2SLS procedure is one of several approaches to consistent estimation of model parameters when the condition of $E(\varepsilon_t|X_t) = 0$ fails. There are alternative estimation procedures that also yield consistent estimators. For example, suppose the correlation between X_t and ε_t is caused by the omitted variables problem, namely

$$\varepsilon_t = g(W_t) + u_t,$$

when $E(u_t|X_t, W_t) = 0$ and W_t is an omitted variable which is correlated with X_t. This delivers a partially linear regression model

$$Y_t = X_t'\beta^o + g(W_t) + u_t.$$

Because $E(Y_t|W_t) = E(X_t|W_t)'\beta^o + g(W_t)$, we obtain

$$Y_t - E(Y_t|W_t) = [X_t - E(X_t|W_t)]'\beta^o + u_t$$

or

$$Y_t^* = X_t^{*\prime}\beta^o + u_t,$$

where $Y_t^* = Y_t - E(Y_t|W_t)$ and $X_t^* = X_t - E(X_t|W_t)$. Because $E(X_t^* u_t) = 0$, the OLS estimator $\tilde{\beta}^*$ of regressing Y_t^* on X_t^* would be consistent for β^o. However, (Y_t^*, X_t^*) are not observable, so $\tilde{\beta}^*$ is infeasible. Nevertheless, one can first estimate $E(Y_t|W_t)$ and $E(X_t|W_t)$ nonparametrically, and then obtain a feasible OLS estimator which will be consistent for the true model parameter (e.g., Robinson 1988). Specifically, let $\hat{m}_Y(W_t)$

and $\hat{m}_X(W_t)$ be consistent nonparametric estimators for $E(Y_t|W_t)$ and $E(X_t|W_t)$ respectively. Then we can obtain a feasible OLS estimator

$$\tilde{\beta}_a^* = \left[\sum_{t=1}^{n} \hat{X}_t^* \hat{X}_t^{*\prime}\right]^{-1} \sum_{t=1}^{n} \hat{X}_t^* \hat{Y}_t^*,$$

where $\hat{X}_t^* = X_t - \hat{m}_X(W_t)$ and $\hat{Y}_t^* = Y_t - \hat{m}_Y(W_t)$. It can be shown that $\tilde{\beta}_a^* \to^p \beta^o$ and

$$\sqrt{n}(\tilde{\beta}_a^* - \beta^o) \to^d N(0, Q^{*-1} V^* Q^{*-1}),$$

where $Q^* = E(X_t^* X_t^{*\prime})$ and $V^* = \mathrm{var}(n^{-1/2}\Sigma_{t=1}^n X_t^* u_t)$. The first stage nonparametric estimation has no impact on the asymptotic properties of the feasible OLS estimator $\tilde{\beta}_a$.

Another method to consistently estimate the true model parameters is to make use of panel data. A panel data is a collection of observations for a total of n cross-sectional units and each of these units has T time series observations over the same time period. This is called a balanced panel data. In contrast, an unbalanced panel data is a collection of observations for a total of n cross-sectional units and each unit may have different lengths of time series observations with some common overlapping time periods.

With a balanced panel data, we have

$$\begin{aligned} Y_{it} &= X_{it}'\beta^o + \varepsilon_{it} \\ &= X_{it}'\beta^o + \alpha_i + u_{it}, \end{aligned}$$

where α_i is called individual-specific effect and u_{it} is called idiosyncratic disturbance such that $E(u_{it}|X_{it}, \alpha_i) = 0$. When α_i is correlated with X_{it}, which may be caused by omitted variables which do not change over time, the panel data model is called a fixed effect panel data model. When α_i is uncorrelated with X_{it}, the panel data model is called a random effect panel data model. Here, we consider a fixed effect panel data model with strictly exogenous variables X_{it}. Because ε_{it} is correlated with X_{it}, the OLS estimator of regressing Y_{it} on X_{it} is not consistent for β^o. However, one can consider the demeaned model

$$Y_{it} - \dot{Y}_{i.} = (X_{it} - \dot{X}_{i.})'\beta^o + (\varepsilon_{it} - \dot{\varepsilon}_{i.}),$$

where $\dot{Y}_{i.} = T^{-1}\Sigma_{t=1}^T Y_{it}$ and similarly for $\dot{X}_{i.}$ and $\dot{\varepsilon}_{i.}$. The demeaning procedure removes the unobservable individual-specific effect and as a result, the OLS estimator for the demeaned model, which is called the within estimator in the panel data literature, will be consistent for the true model parameter β^o. (It should be noted that for a dynamic panel data model where X_{it} is not strictly exogenous, the within estimator is not consistent for β^o when the number of the time periods T is fixed. Different estimation methods have to be used.) See Hsiao (2002) for detailed discussion of panel data econometric models.

Chapters 2 to 7 present a relatively comprehensive econometric theory for linear regression models often encountered in economics and finance. We start with a general regression analysis, discussing the interpretation of a linear regression model, which depends on whether the linear regression model is correctly specified. After discussing the classical linear regression model in Chapter 3, Chapters 4 to 7 discuss various extensions and generalizations when some assumptions in the classical linear regression model are violated. In particular, we consider the scenarios under which the results for classical linear regression models are approximately applicable for large samples. The key condition here are conditional homokedasticity and serial uncorrelatedness in the regression disturbance. When there exists conditional heteroskedasticity or serial correlation in the regression disturbance, the results for classical linear regression models are no longer applicable; we provide robust asymptotically valid procedures under these scenarios.

The asymptotic theory developed for linear regression models in Chapters 4–7 can be easily extended to more complicated, nonlinear models. For example, consider a nonlinear regression model

$$Y_t = g(X_t, \beta^o) + \varepsilon_t,$$

where $E(\varepsilon_t|X_t) = 0$ a.s. The nonlinear least squares estimator solves the minimization of the sum of squared residual problem

$$\hat{\beta} = \arg\min_{\beta} \sum_{t=1}^{n} \left[Y_t - g(X_t, \beta)\right]^2.$$

The first order condition is

$$D(\hat{\beta})'e = 0,$$

where $D(\beta)$ is a $n \times K$ matrix, with the t-th row being $\partial g(X_t, \beta)/\partial\beta$. Although one generally does not have a closed form expression for $\hat{\beta}$, all asymptotic theory and procedures in Chapters 4–7 are applicable to the nonlinear least squares estimator if one replaces X_t by $(\partial/\partial\beta)g(X_t, \beta)$. See also the discussion in Chapters 8 and 9.

The asymptotic theory in Chapters 4–7 however, cannot be directly applied to some popular nonlinear models. Examples of such nonlinear models are

- Rational Expectations Model:

$$E\left[m(Z_t, \beta^o)\right] = 0;$$

- Conditional Variance model:

$$Y_t = g(X_t, \beta^o) + \sigma(X_t, \beta^o)u_t,$$

where $g(X_t, \beta)$ is a parametric model for $E(Y_t|X_t)$, $\sigma^2(X_t, \beta)$ is a parametric model for $\text{var}(Y_t|X_t)$, and $\{u_t\}$ is i.i.d.$(0, 1)$;

- Conditional probability model of Y_t given X_t :

$$f(y|X_t, \beta).$$

These nonlinear models are not models for conditional mean or regression; they also model other characteristics of the conditional distribution of Y_t given X_t. For these models, we need to develop new estimation methods and new asymptotic theory, which we will turn to in subsequent chapters.

One important part that we do not discuss in Chapters 2–7 is model specification testing. Chapter 2 emphasizes the importance of correct model specification for the validity of economic interpretation of model parameters. How to check whether a linear regression model is correctly specified for conditional mean $E(Y_t|X_t)$? This is called model specification testing. Some popular specification tests in econometrics are Hausman's (1978) test and White's (1981) test which compares two parameter estimators for the same model parameter. Also, see Hong and White's (1995) specification test using a nonparametric series regression approach.

References

Hausman, J. (1978), Econometrica.

Hong, Y. and H. White (1995), Econometrica.

Hsiao, C. (2002), Panel Data Analysis.

Robinson, P. (1988), Econometrica.

White, H. (1981), Journal of the American Statistical Association.

EXERCISES

7.1. Consider the following simple Keynes national income model

$$\begin{aligned} C_t &= \beta_1^o + \beta_2^o(Y_t - T_t) + \varepsilon_t, & (1.1)\\ T_t &= \gamma_1^o + \gamma_2^o Y_t + v_t, & (1.2)\\ Y_t &= C_t + G_t, & (1.3) \end{aligned}$$

where C_t, Y_t, T_t, G_t are the consumption, income, tax, and government spending respectively, and $\{\varepsilon_t\}$ and $\{v_t\}$ are i.i.d. $(0, \sigma_\varepsilon^2)$ and $(0, \sigma_v^2)$ respectively. Model (1.1) is a consumption function which we are interested in, (1.2) is a tax function, and (1.3) is an income identity.

(a) Can the OLS estimator $\hat{\beta}$ of model (1.1) give consistent estimation for the marginal propensity to consume? Explain.

(b) Suppose G_t is an exogenous variable (i.e., G_t does not depend on both C_t and Y_t). Can G_t be used as a valid instrumental variable? If yes, describe a 2SLS procedure. If not, explain.

(c) Suppose the government has to maintain a budget balance such that

$$G_t = T_t + w_t, \tag{1.4}$$

where $\{w_t\}$ is i.i.d. $(0, \sigma_w^2)$. Could G_t be used as a valid instrumental variable? If yes, describe a 2SLS procedure. If not, explain.

7.2. Consider the data generating process

$$Y_t = X_t'\beta^o + \varepsilon_t, \tag{2.1}$$

where $X_t = (1, X_{1t})'$,

$$\begin{aligned} X_{1t} &= v_t + u_t, & (2.2)\\ \varepsilon_t &= w_t + u_t. & (2.3) \end{aligned}$$

where $\{v_t\}, \{u_t\}$ and $\{w_t\}$ are all i.i.d. $N(0, 1)$, and they are mutually independent.

(a) Is the OLS estimator $\hat{\beta}$ consistent for β^o? Explain.

(b) Suppose that $Z_{1t} = w_t - \varepsilon_t$. Is $Z_t = (1, Z_{1t})'$ a valid instrumental vector? Explain.

(c) Find an instrumental vector and the asymptotic distribution of $\hat{\beta}_{2sls}$ using this instrumental vector. [Note you need to find $\sqrt{n}(\hat{\beta}_{2sls} - \beta^o) \to^d N(0, V)$ for some V, where the expression of V should be given.]

(d) To test the hypothesis

$$\mathbf{H}_0 : R\beta^o = r,$$

where R is a $J \times 2$ matrix, and r is a $J \times 1$ vector. Suppose that $\tilde{F}$ is the F-statistic in the second stage regression of 2SLS. Could we use $J \cdot \tilde{F}$ as an asymptotic χ^2_J test? Explain.

7.3. Consider the following demand-supply system:

$$\begin{aligned} Y_t &= \alpha_0^o + \alpha_1^o P_t + \alpha_2^o S_t + \varepsilon_t, \\ Y_t &= \beta_0^o + \beta_1^o P_t + \beta_2^o C_t + v_t, \end{aligned}$$

where the first equation is a model for the demand of certain good, where Y_t is the quantity demanded for the good, P_t is the price of the good, S_t is the price of a substitute, and ε_t is a shock to the demand. The second equation is a model for the supply of the good, where Y_t is the quantity supplied, C_t is the cost of production, and v_t is a shock to the supply. Suppose S_t and C_t are exogenous variables, $\{\varepsilon_t\}$ is i.i.d.$(0, \sigma_\varepsilon^2)$ and $\{v_t\}$ is i.i.d.$(0, \sigma_v^2)$, and two series $\{\varepsilon_t\}$ and $\{v_t\}$ are independent of each other. We have also assumed that the market is always clear so the quantity demanded is equal to the quantity supplied.

(a) Suppose we use a 2SLS estimator to estimate the demand model with the instruments $Z_t = (S_t, C_t)'$. Describe the 2SLS procedure. Is the resulting 2SLS $\hat{\alpha}_{2sls}$ consistent for $\alpha^o = (\alpha_0^o, \alpha_1^o, \alpha_2^o)'$? Explain.

(b) Suppose we use a 2SLS estimator to estimate the supply equation with instruments $Z_t = (S_t, C_t)'$. Describe the 2SLS procedure. Is the resulting 2SLS $\hat{\beta}_{2sls}$ consistent for $\beta^o = (\beta_0^o, \beta_1^o, \beta_2^o)'$? Explain.

(c) Suppose $\{\varepsilon_t\}$ and $\{v_t\}$ are contemporaneously correlated, namely, $E(\varepsilon_t v_t) \neq 0$. This can occur when there is a common shock to both the demand and supply of the good. Does this affect the conclusions in part (a) and part (b). Explain.

7.4. Show that under Assumptions 7.1-7.4, $\hat{\beta}_{2sls} \xrightarrow{p} \beta^o$ as $n \to \infty$.

7.5. Suppose Assumptions 7.1-7.5 hold.

(a) Show that $\sqrt{n}(\hat{\beta}_{2sls} - \beta^o) \xrightarrow{d} N(0, \Omega)$ as $n \to \infty$, where

$$\Omega = [Q_{xz}Q_{zz}^{-1}Q_{zx}]^{-1}Q_{xz}Q_{zz}^{-1}VQ_{zz}^{-1}Q_{zx}[Q_{xz}Q_{zz}^{-1}Q_{zx}]^{-1},$$

and V is given in Assumption 7.5;

(b) If in addition that $\{Z_t\varepsilon_t\}$ is an ergodic stationary MDS process with $E(\varepsilon_t^2|Z_t) = \sigma^2$. Show that

$$\Omega = \sigma^2[Q_{xz}Q_{zz}^{-1}Q_{zx}]^{-1}.$$

7.6. Suppose Assumptions 7.1 – 7.4, 7.6 and 7.7 hold.

(a) Define

$$\hat{s}^2 = \frac{\hat{e}'\hat{e}}{n}$$

where $\hat{e} = Y - X\hat{\beta}_{2sls}$. Show $\hat{s}^2 \to^p \sigma^2 = \text{var}(\varepsilon_t)$ as $n \to \infty$.

(b) Define

$$s^2 = \frac{e'e}{n},$$

where $e = Y - \hat{X}\hat{\beta}_{2sls}$ is the estimated residual from the second stage regression of Y_t on $\hat{X}_t = \hat{\gamma}' Z_t$. Show that s^2 is not a consistent estimator for σ^2.

7.7. [2SLS Hypothesis Testing] Suppose Assumptions 7.1-7.5 hold. Define a F-statistic

$$F = \frac{n(R\hat{\beta}_{2sls} - r)'[R\hat{Q}_{\hat{x}\hat{x}}^{-1}R']^{-1}(R\hat{\beta}_{2sls} - r)/J}{e'e/(n-K)},$$

where $e_t = Y_t - \hat{X}_t'\hat{\beta}_{2sls}$ is the estimated residual from the second stage regression of Y_t on $\hat{X}_t$. Does $J \cdot F \to^d \chi^2_J$ under the null hypothesis $\mathbf{H}_0 : R\beta^o = \gamma$? If yes, give your reasoning. If not, provide a modification so that the modified test statistic converges to χ^2_J under $\mathbf{H}_0$.

7.8. Let

$$\hat{V} = \frac{1}{n}\sum_{t=1}^{n} Z_t Z_t' \hat{e}_t^2,$$

where $\hat{e}_t = Y_t - X_t'\hat{\beta}_{2sls}$. Show $\hat{V} \to^p V$ under Assumptions 7.1–7.8.

7.9. Suppose the following assumptions hold:

Assumption 3.1 [Linearity]: $\{Y_t, X_t'\}'$ is a stationary ergodic process with

$$Y_t = X_t'\beta^o + \varepsilon_t, \qquad t = 1, ..., n,$$

for some unknown parameter β^o and some unobservable disturbance ε_t;

Assumption 3.2 [Nonsingularity] The $K \times K$ matrix

$$Q_{xx} = E(X_t X_t')$$

is nonsingular and finite;

Assumption 3.3 [Orthogonality]

(i) $E(X_t\varepsilon_t)=0$;

(ii) $E(Z_t\varepsilon_t)=0$, where Z_t is a $l\times 1$ random vector, with $l\geq K$;

(iii) The $l\times l$ matrix

$$Q_{zz}=E(Z_tZ_t')$$

is finite and nonsingular, and the $l\times K$ matrix

$$Q_{xz}=E(Z_tX_t')$$

is finite and of full rank;

Assumption 3.4: $\{(X_t',Z_t')'\varepsilon_t\}$ is an martingale difference sequence.
Assumption 3.5: $E(\varepsilon_t^2|X_t,Z_t)=\sigma^2$ a.s.

Under these assumptions, both OLS

$$\hat{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y$$

and 2SLS

$$\hat{\beta}_{2sls}=[(\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'Y$$

are consistent for β^o.

(a) Show that $\hat{\beta}$ is a special 2SLS estimator $\hat{\beta}_{2sls}$ with some proper choice of instrumental vector Z_t.

(b) Which estimator, $\hat{\beta}$ or $\hat{\beta}_{2sls}$, is more asymptotically efficient? [Hint: if $\sqrt{n}(\hat{\beta}_1-\beta^o)\to^d N(0,\Omega_1)$ and $\sqrt{n}(\hat{\beta}_2-\beta^o)\to^d N(0,\Omega_2)$, then $\hat{\beta}_1$ is asymptotically more efficient than $\hat{\beta}_2$ if and only if $\Omega_2-\Omega_1$ or $\Omega_1^{-1}-\Omega_2^{-1}$ is positive semi-definite.]

7.10. Consider the linear regression model

$$Y_t=X_t'\beta^o+\varepsilon_t,$$

where $E(X_t\varepsilon_t)\neq 0$. Our purpose is to find a consistent estimation procedure for β^o.

First, consider the artificial regression

$$X_t=\gamma' Z_t+v_t,$$

where X_t is the regressor vector, Z_t is the instrumental vector, $\gamma=[E(Z_tZ_t')]^{-1}E(Z_tX_t')$ is the best linear LS approximation coefficient, and v_t is the $K\times 1$ regression error.

Now, suppose instead of decomposing X_t, we decompose the regression error ε_t as follows:

$$\varepsilon_t=v_t'\rho^0+u_t,$$

where $\rho^0 = [E(v_t v_t')]^{-1} E(v_t \varepsilon_t)$ is the best linear LS approximation coefficient.

Now, assuming that v_t is observable, we consider the augmented linear regression model

$$Y_t = X_t'\beta^o + v_t'\rho^0 + u_t.$$

Show $E[(X_t', v_t)'u_t] = 0$. One important implication of this orthogonality condition is that if v_t is observable then the OLS estimator of regressing Y_t on X_t and v_t will be consistent for $(\beta^o, \rho^o)'$.

7.11. In practice, v_t is unobservable. However, it can be estimated by the estimated residual

$$\hat{v}_t = X_t - \hat{\gamma}' Z_t = X_t - \hat{X}_t.$$

We now consider the following feasible augmented linear regression model

$$Y_t = X_t'\beta^o + \hat{v}_t'\rho + \tilde{u}_t,$$

and we denote the resulting OLS estimator as $\hat{\alpha} = (\hat{\beta}', \hat{\rho}')'$, where $\hat{\beta}$ is the OLS estimator for β^o and $\hat{\rho}$ is the OLS estimator for ρ.

Show $\hat{\beta} = \hat{\beta}_{2sls}$.[Hint: The following decomposition may be useful: Suppose

$$A = \begin{bmatrix} B & C' \\ C & D \end{bmatrix}$$

is a nonsingular square matrix, where B is $k_1 \times k_1, C$ is $k_2 \times k_1$ and D is $k_2 \times k_2$. Then

$$A^{-1} = \begin{bmatrix} B^{-1}(I + C'E^{-1}CB^{-1}) & -B'^{-1}C'E^{-1} \\ -E^{-1}CB^{-1} & E^{-1} \end{bmatrix},$$

where $E = D - CB^{-1}C'$.]

7.12. Suppose $\hat{Y}$ is a $n \times 1$ vector of the fitted values of regressing Y_t on Z_t, and $\hat{X}$ is a $n \times K$ matrix of fitted values of regressing X_t on Z_t. Show that $\hat{\beta}_{2sls}$ is equal to the OLS estimator of regressing $\hat{Y}$ on $\hat{X}$.

7.13 [Hausman's Test] Suppose Assumptions 3.1, 3.2, 3.3(ii, iii), 3.4 and 3.5 in Problem 7.8 hold. A test for the null hypothesis $\mathbf{H}_0 : E(X_t\varepsilon_t) = 0$ can be constructed by comparing $\hat{\beta}$ and $\hat{\beta}_{2sls}$, because they will converge in probability to the same limit β^o under H_0 and to different limits under the alternatives to $\mathbf{H}_0$. Assume $\mathbf{H}_0$ holds.

(a) Show that

$$\sqrt{n}(\hat{\beta} - \beta^o) - Q_{xx}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \varepsilon_t \to^p 0$$

or equivalently

$$\sqrt{n}(\hat{\beta} - \beta^o) = Q_{xx}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \varepsilon_t + o_P(1),$$

where $Q_{xx} = E(X_t X_t')$. [Hint: If $\hat{A} \to^p A$ and $\hat{B} = O_P(1)$, then $\hat{A}\hat{B} - A\hat{B} \to^p 0$ or $\hat{A}\hat{B} = A\hat{B} + o_P(1)$.]

(b) Show that

$$\sqrt{n}(\hat{\beta}_{2sls} - \beta^o) = Q_{\tilde{x}\tilde{x}}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{X}_t \varepsilon_t + o_P(1),$$

where $Q_{\tilde{x}\tilde{x}} = E(X_t X_t'), \tilde{X}_t = \gamma' Z_t$ and $\gamma = [E(Z_t Z_t')]^{-1} E(Z_t X_t)$.

(e) Show that

$$\sqrt{n}(\hat{\beta}_{2sls} - \hat{\beta}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ Q_{xx}^{-1} X_t - Q_{\tilde{x}\tilde{x}}^{-1} \tilde{X}_t \right\} \varepsilon_t + o_P(1).$$

(d) The asymptotic distribution of $\sqrt{n}(\hat{\beta}_{2sls} - \hat{\beta})$ is determined by the leading term only in part (c). Find its asymptotic distribution.

(e) Construct an asymptotically χ^2 test statistic. What is the degree of freedom of the asymptotic χ^2 distribution? Assume that $Q_{xx} - Q_{\tilde{x}\tilde{x}}$ is strictly positive definite.

7.14. Suppose Assumptions 3.1, 3.2, 3.3(ii, iii) and 3.4 in Problem 7.8 hold, $E(X_{jt}^4) < \infty$ for $1 \leq j \leq K$, $E(Z_{jt}^4) < \infty$ for $1 \leq j \leq l$, and $E(\varepsilon_t^4) < \infty$. Construct a Hausman's test statistic for $\mathbf{H}_0 : E(\varepsilon_t | X_t) = 0$ and derive its asymptotic distribution under $\mathbf{H}_0$.