since the equilibrium price depends, in part, on the error term in the demand equation. Hence the standard assumption that error terms and regressors are independent is violated in this (and every) system of simultaneous equations. Thus, if we attempt to take the plim of the right-hand side of (7.14), we will find that the second term is not zero. It follows that $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ will be inconsistent.

The results of this simple example are true in general. Since they are determined simultaneously, all the endogenous variables in a simultaneous equation system generally depend on the error terms in all the equations. Thus, except perhaps in a few very special cases, the right-hand side endogenous variables in a structural equation from such a system will always be correlated with the error terms. As a consequence, application of OLS to such an equation will always yield biased and inconsistent estimates.

We have now seen two important situations in which explanatory variables will be correlated with the error terms of regression equations, and are ready to take up the main topic of this chapter, namely, the method of instrumental variables. This method can be used whenever the error terms are correlated with one or more explanatory variables, regardless of how that correlation may have arisen. It is remarkably simple, general, and powerful.

7.4 Instrumental Variables: The Linear Case

The fundamental ingredient of any IV procedure is a matrix of **instrumental variables** (or simply **instruments**, for short). We will call this matrix $\boldsymbol{W}$ and specify that it is $n \times l$. The columns of $\boldsymbol{W}$ are simply exogenous and/or predetermined variables that are known (or at least assumed) to be independent of the error terms $\boldsymbol{u}$. In the context of the simultaneous equations model, a natural choice for $\boldsymbol{W}$ is the matrix of all the exogenous and predetermined variables in the model. There must be at least as many instruments as there are explanatory variables in the equation to be estimated. Thus, if the equation to be estimated is the linear regression model (7.01), with $\boldsymbol{X}$ having k columns, we require that $l \geq k$. This is an identification condition; see Section 7.8 for further discussion of conditions for identification in models estimated by IV. Some of the explanatory variables may appear among the instruments. Indeed, as we will see below, any column of $\boldsymbol{X}$ that is known to be exogenous or predetermined should be included in $\boldsymbol{W}$ if we want to obtain asymptotically efficient estimates.

The intuition behind IV procedures is the following. Least squares minimizes the distance between $\boldsymbol{y}$ and $\mathcal{S}(\boldsymbol{X})$, which leads to inconsistent estimates because $\boldsymbol{u}$ is correlated with $\boldsymbol{X}$. The n–dimensional space in which $\boldsymbol{y}$ is a point can be divided into two orthogonal subspaces, $\mathcal{S}(\boldsymbol{W})$ and $\mathcal{S}^{\perp}(\boldsymbol{W})$. Instrumental variables minimizes only the portion of the distance between $\boldsymbol{y}$ and $\mathcal{S}(\boldsymbol{X})$ that lies in $\mathcal{S}(\boldsymbol{W})$. Provided that $\boldsymbol{u}$ is independent of $\boldsymbol{W}$, as assumed, any