CHAPTER 1

Finite-Sample Properties of OLS

ABSTRACT

The **Ordinary Least Squares** (OLS) estimator is the most basic estimation procedure in econometrics. This chapter covers the **finite-** or **small-sample properties** of the OLS estimator, that is, the statistical properties of the OLS estimator that are valid for any given sample size. The materials covered in this chapter are entirely standard. The exposition here differs from that of most other textbooks in its emphasis on the role played by the assumption that the regressors are "strictly exogenous."

In the final section, we apply the finite-sample theory to the estimation of the cost function using cross-section data on individual firms. The question posed in Nerlove's (1963) study is of great practical importance: are there increasing returns to scale in electricity supply? If yes, microeconomics tells us that the industry should be regulated. Besides providing you with a hands-on experience of using the techniques to test interesting hypotheses, Nerlove's paper has a careful discussion of why the OLS is an appropriate estimation procedure in this particular application.

1.1 The Classical Linear Regression Model

In this section we present the assumptions that comprise the classical linear regression model. In the model, the variable in question (called the **dependent variable**, the **regressand**, or more generically the **left-hand [-side] variable**) is related to several other variables (called the **regressors**, the **explanatory variables**, or the **right-hand [-side] variables**). Suppose we observe n values for those variables. Let y_i be the i-th observation of the dependent variable in question and let $(x_{i1}, x_{i2}, \dots, x_{iK})$ be the i-th observation of the K regressors. The **sample** or **data** is a collection of those n observations.

The data in economics cannot be generated by experiments (except in experimental economics), so both the dependent and independent variables have to be treated as random variables, variables whose values are subject to chance. A **model** is a set of restrictions on the joint distribution of the dependent and independent variables. That is, a model is a set of joint distributions satisfying a set of assumptions. The classical regression model is a set of joint distributions satisfying Assumptions 1.1–1.4 stated below.

The Linearity Assumption

The first assumption is that the relationship between the dependent variable and the regressors is linear.

Assumption 1.1 (linearity):

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i \quad (i = 1, 2, \dots, n), \tag{1.1.1}$$

where β's are unknown parameters to be estimated, and ε_i is the unobserved error term with certain properties to be specified below.

The part of the right-hand side involving the regressors, $\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_K x_{iK}$, is called the **regression** or the **regression function**, and the coefficients (β's) are called the **regression coefficients**. They represent the marginal and separate effects of the regressors. For example, β_2 represents the change in the dependent variable when the second regressor increases by one unit while other regressors are held constant. In the language of calculus, this can be expressed as $\partial y_i/\partial x_{i2} = \beta_2$. The linearity implies that the marginal effect does not depend on the level of regressors. The error term represents the part of the dependent variable left unexplained by the regressors.

Example 1.1 (consumption function): The simple consumption function familiar from introductory economics is

$$CON_i = \beta_1 + \beta_2 YD_i + \varepsilon_i, \tag{1.1.2}$$

where CON is consumption and YD is disposable income. If the data are annual aggregate time-series, CON_i and YD_i are aggregate consumption and disposable income for year i. If the data come from a survey of individual households, CON_i is consumption by the i-th household in the cross-section sample of n households. The consumption function can be written as (1.1.1) by setting $y_i = CON_i$, $x_{i1} = 1$ (a constant), and $x_{i2} = YD_i$. The error term ε_i represents other variables besides disposable income that influence consumption. They include those variables — such as financial assets — that might be observable but the researcher decided not to include as regressors, as well as those variables — such as the "mood" of the consumer — that are hard to measure. When the equation has only one nonconstant regressor, as here, it is called the **simple regression model**.

The linearity assumption is not as restrictive as it might first seem, because the dependent variable and the regressors can be transformations of the variables in question. Consider

Example 1.2 (wage equation): A simplified version of the wage equation routinely estimated in labor economics is

$$\log(WAGE_i) = \beta_1 + \beta_2 S_i + \beta_3 TENURE_i + \beta_4 EXPR_i + \varepsilon_i, \qquad (1.1.3)$$

where $WAGE$ = the wage rate for the individual, S = education in years, $TENURE$ = years on the current job, and $EXPR$ = experience in the labor force (i.e., total number of years to date on all the jobs held currently or previously by the individual). The wage equation fits the generic format (1.1.1) with $y_i = \log(WAGE_i)$. The equation is said to be in the **semi-log** form because only the dependent variable is in logs. The equation is derived from the following nonlinear relationship between the level of the wage rate and the regressors:

$$WAGE_i = \exp(\beta_1)\exp(\beta_2 S_i)\exp(\beta_3 TENURE_i)\exp(\beta_4 EXPR_i)\exp(\varepsilon_i). \qquad (1.1.4)$$

By taking logs of both sides of (1.1.4) and noting that $\log[\exp(x)] = x$, one obtains (1.1.3). The coefficients in the semi-log form have the interpretation of *percentage changes*, not changes in levels. For example, a value of 0.05 for β_2 implies that an additional year of education has the effect of raising the wage rate by 5 percent. The difference in the interpretation comes about because the dependent variable is the log wage rate, not the wage rate itself, and the change in logs equals the percentage change in levels.

Certain other forms of nonlinearities can also be accommodated. Suppose, for example, the marginal effect of education tapers off as the level of education gets higher. This can be captured by including in the wage equation the squared term S^2 as an additional regressor in the wage equation. If the coefficient of the squared term is β_5, the marginal effect of education is

$$\beta_2 + 2\beta_5 S \quad (= \partial \log(WAGE)/\partial S).$$

If β_5 is negative, the marginal effect of education declines with the level of education.

There are, of course, cases of genuine nonlinearity. For example, the relationship (1.1.4) could not have been made linear if the error term entered additively rather than multiplicatively:

$$WAGE_i = \exp(\beta_1)\exp(\beta_2 S_i)\exp(\beta_3 TENURE_i)\exp(\beta_4 EXPR_i) + \varepsilon_i.$$

Estimation of nonlinear regression equations such as this will be discussed in Chapter 7.

Matrix Notation

Before stating other assumptions of the classical model, we introduce the vector and matrix notation. The notation will prove useful for stating other assumptions precisely and also for deriving the OLS estimator of $\boldsymbol{\beta}$. Define K-dimensional (column) vectors $\mathbf{x}_i$ and $\boldsymbol{\beta}$ as

$$\underset{(K\times 1)}{\mathbf{x}_i} = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{bmatrix}, \quad \underset{(K\times 1)}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}. \tag{1.1.5}$$

By the definition of vector inner products, $\mathbf{x}_i'\boldsymbol{\beta} = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK}$. So the equations in Assumption 1.1 can be written as

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad (i = 1, 2, \ldots, n). \tag{1.1.1$'$}$$

Also define

$$\underset{(n\times 1)}{\mathbf{y}} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \underset{(n\times 1)}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \underset{(n\times K)}{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix} = \begin{bmatrix} x_{11} & \ldots & x_{1K} \\ \vdots & \ldots & \vdots \\ x_{n1} & \ldots & x_{nK} \end{bmatrix}. \tag{1.1.6}$$

In the vectors and matrices in (1.1.6), there are as many rows as there are observations, with the rows corresponding to the observations. For this reason $\mathbf{y}$ and $\mathbf{X}$ are sometimes called the **data vector** and the **data matrix**. Since the number of columns of $\mathbf{X}$ equals the number of rows of $\boldsymbol{\beta}$, $\mathbf{X}$ and $\boldsymbol{\beta}$ are conformable and $\mathbf{X}\boldsymbol{\beta}$ is an $n\times 1$ vector. Its i-th element is $\mathbf{x}_i'\boldsymbol{\beta}$. Therefore, Assumption 1.1 can be written compactly as

$$\underset{(n\times 1)}{\mathbf{y}} = \underbrace{\underset{(n\times K)}{\mathbf{X}}\underset{(K\times 1)}{\boldsymbol{\beta}}}_{(n\times 1)} + \underset{(n\times 1)}{\boldsymbol{\varepsilon}}.$$

The Strict Exogeneity Assumption

The next assumption of the classical regression model is

Assumption 1.2 (strict exogeneity):

$$\mathrm{E}(\varepsilon_i \mid \mathbf{X}) = 0 \quad (i = 1, 2, \ldots, n). \tag{1.1.7}$$

Here, the expectation (mean) is conditional on the regressors for *all* observations. This point may be made more apparent by writing the assumption without using the data matrix as

$$\mathrm{E}(\varepsilon_i \mid \mathbf{x}_1, \ldots, \mathbf{x}_n) = 0 \quad (i = 1, 2, \ldots, n).$$

To state the assumption differently, take, for any given observation i, the joint distribution of the $nK + 1$ random variables, $f(\varepsilon_i, \mathbf{x}_1, \ldots, \mathbf{x}_n)$, and consider the conditional distribution, $f(\varepsilon_i \mid \mathbf{x}_1, \ldots, \mathbf{x}_n)$. The conditional mean $\mathrm{E}(\varepsilon_i \mid \mathbf{x}_1, \ldots, \mathbf{x}_n)$ is in general a nonlinear function of $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. The strict exogeneity assumption says that this function is a constant of value zero.[1]

Assuming this constant to be zero is not restrictive if the regressors include a constant, because the equation can be rewritten so that the conditional mean of the error term is zero. To see this, suppose that $\mathrm{E}(\varepsilon_i \mid \mathbf{X})$ is μ and $x_{i1} = 1$. The equation can be written as

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i \\ &= (\beta_1 + \mu) + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + (\varepsilon_i - \mu). \end{aligned}$$

If we redefine β_1 to be $\beta_1 + \mu$ and ε_i to be $\varepsilon_i - \mu$, the conditional mean of the new error term is zero. In virtually all applications, the regressors include a constant term.

[1] Some authors define the term "strict exogeneity" somewhat differently. For example, in Koopmans and Hood (1953) and Engle, Hendry, and Richards (1983), the regressors are strictly exogenous if $\mathbf{x}_i$ is independent of ε_j for all i, j. This definition is stronger than, but not inconsistent with, our definition of strict exogeneity.

Example 1.3 (continuation of Example 1.1): For the simple regression model of Example 1.1, the strict exogeneity assumption can be written as

$$\mathrm{E}(\varepsilon_i \mid YD_1, YD_2, \ldots, YD_n) = 0.$$

Since $\mathbf{x}_i = (1, YD_i)'$, you might wish to write the strict exogeneity assumption as

$$\mathrm{E}(\varepsilon_i \mid 1, YD_1, 1, YD_2, \ldots, 1, YD_n) = 0.$$

But since a constant provides no information, the expectation conditional on

$$(1, YD_1, 1, YD_2, \ldots, 1, YD_n)$$

is the same as the expectation conditional on

$$(YD_1, YD_2, \ldots, YD_n).$$

Implications of Strict Exogeneity

The strict exogeneity assumption has several implications.

- The *un*conditional mean of the error term is zero, i.e.,

  $$\mathrm{E}(\varepsilon_i) = 0 \quad (i = 1, 2, \ldots, n). \tag{1.1.8}$$

  This is because, by the Law of Total Expectations from basic probability theory,[2] $\mathrm{E}[\mathrm{E}(\varepsilon_i \mid \mathbf{X})] = \mathrm{E}(\varepsilon_i)$.

- If the cross moment $\mathrm{E}(xy)$ of two random variables x and y is zero, then we say that x is **orthogonal** to y (or y is orthogonal to x). Under strict exogeneity, the regressors are orthogonal to the error term for *all* observations, i.e.,

  $$\mathrm{E}(x_{jk}\varepsilon_i) = 0 \quad (i, j = 1, \ldots, n;\, k = 1, \ldots, K)$$

  or

  $$\mathrm{E}(\mathbf{x}_j \cdot \varepsilon_i) = \begin{bmatrix} \mathrm{E}(x_{j1}\,\varepsilon_i) \\ \mathrm{E}(x_{j2}\,\varepsilon_i) \\ \vdots \\ \mathrm{E}(x_{jK}\,\varepsilon_i) \end{bmatrix} = \underset{(K\times 1)}{\mathbf{0}} \quad \text{(for all } i, j\text{)}. \tag{1.1.9}$$

[2] The Law of Total Expectations states that $\mathrm{E}[\mathrm{E}(y \mid \mathbf{x})] = \mathrm{E}(y)$.

The proof is a good illustration of the use of properties of conditional expectations and goes as follows.

PROOF. Since x_{jk} is an element of $\mathbf{X}$, strict exogeneity implies

$$\mathrm{E}(\varepsilon_i \mid x_{jk}) = \mathrm{E}[\mathrm{E}(\varepsilon_i \mid \mathbf{X}) \mid x_{jk}] = 0 \tag{1.1.10}$$

by the Law of Iterated Expectations from probability theory.[3] It follows from this that

$$\begin{aligned}\mathrm{E}(x_{jk}\varepsilon_i) &= \mathrm{E}[\mathrm{E}(x_{jk}\varepsilon_i \mid x_{jk})] \quad \text{(by the Law of Total Expectations)}\\ &= \mathrm{E}[x_{jk}\,\mathrm{E}(\varepsilon_i \mid x_{jk})] \quad \text{(by the linearity of conditional expectations}^4\text{)}\\ &= 0. \qquad \blacksquare\end{aligned}$$

The point here is that strict exogeneity requires the regressors be orthogonal not only to the error term from the same observation (i.e., $\mathrm{E}(x_{ik}\varepsilon_i) = 0$ for all k), but also to the error term from the other observations (i.e., $\mathrm{E}(x_{jk}\varepsilon_i) = 0$ for all k and for $j \neq i$).

- Because the mean of the error term is zero, the orthogonality conditions (1.1.9) are equivalent to zero-correlation conditions. This is because

$$\begin{aligned}\mathrm{Cov}(\varepsilon_i, x_{jk}) &= \mathrm{E}(x_{jk}\varepsilon_i) - \mathrm{E}(x_{jk})\,\mathrm{E}(\varepsilon_i) \quad \text{(by definition of covariance)}\\ &= \mathrm{E}(x_{jk}\varepsilon_i) \quad \text{(since } \mathrm{E}(\varepsilon_i) = 0\text{, see (1.1.8))}\\ &= 0 \quad \text{(by the orthogonality conditions (1.1.9)).}\end{aligned}$$

In particular, for $i = j$, $\mathrm{Cov}(x_{ik}, \varepsilon_i) = 0$. Therefore, strict exogeneity implies the requirement (familiar to those who have studied econometrics before) that the regressors be contemporaneously uncorrelated with the error term.

Strict Exogeneity in Time-Series Models

For time-series models where i is time, the implication (1.1.9) of strict exogeneity can be rephrased as: the regressors are orthogonal to the past, current, and future error terms (or equivalently, the error term is orthogonal to the past, current, and future regressors). But for most time-series models, this condition (and *a fortiori* strict exogeneity) is not satisfied, so the finite-sample theory based on strict exogeneity to be developed in this section is rarely applicable in time-series con-

[3] The Law of Iterated Expectations states that $\mathrm{E}[\mathrm{E}(y \mid \mathbf{x}, \mathbf{z}) \mid \mathbf{x}] = \mathrm{E}(y \mid \mathbf{x})$.

[4] The linearity of conditional expectations states that $\mathrm{E}[f(\mathbf{x})y \mid \mathbf{x}] = f(\mathbf{x})\,\mathrm{E}(y \mid \mathbf{x})$.

texts. However, as will be shown in the next chapter, the estimator possesses good large-sample properties without strict exogeneity.

The clearest example of a failure of strict exogeneity is a model where the regressor includes the **lagged dependent variable**. Consider the simplest such model:

$$y_i = \beta y_{i-1} + \varepsilon_i \quad (i = 1, 2, \dots, n). \tag{1.1.11}$$

This is called the **first-order autoregressive model** (AR(1)). (We will study this model more fully in Chapter 6.) Suppose, consistent with the spirit of the strict exogeneity assumption, that the regressor for observation i, y_{i-1}, is orthogonal to the error term for i so $\mathrm{E}(y_{i-1}\varepsilon_i) = 0$. Then

$$\begin{aligned}\mathrm{E}(y_i\varepsilon_i) &= \mathrm{E}[(\beta y_{i-1} + \varepsilon_i)\varepsilon_i] \quad \text{(by (1.1.11))}\\ &= \beta\,\mathrm{E}(y_{i-1}\varepsilon_i) + \mathrm{E}(\varepsilon_i^2)\\ &= \mathrm{E}(\varepsilon_i^2) \quad (\text{since } \mathrm{E}(y_{i-1}\varepsilon_i) = 0 \text{ by hypothesis}).\end{aligned}$$

Therefore, unless the error term is always zero, $\mathrm{E}(y_i\varepsilon_i)$ is not zero. But y_i is the regressor for observation $i+1$. Thus, the regressor is not orthogonal to the past error term, which is a violation of strict exogeneity.

Other Assumptions of the Model

The remaining assumptions comprising the classical regression model are the following.

Assumption 1.3 (no multicollinearity): *The rank of the $n \times K$ data matrix, $\mathbf{X}$, is K with probability 1.*

Assumption 1.4 (spherical error variance):

$$\textit{(homoskedasticity)}\;\; \mathrm{E}(\varepsilon_i^2 \mid \mathbf{X}) = \sigma^2 > 0 \quad (i = 1, 2, \dots, n),^{5} \tag{1.1.12}$$

(no correlation between observations)

$$\mathrm{E}(\varepsilon_i\varepsilon_j \mid \mathbf{X}) = 0 \quad (i, j = 1, 2, \dots, n;\, i \neq j). \tag{1.1.13}$$

[5] When a symbol (which here is σ^2) is given to a moment (which here is the second moment $\mathrm{E}(\varepsilon_i^2 \mid \mathbf{X})$), by implication the moment is assumed to exist and is finite. We will follow this convention for the rest of this book.

To understand Assumption 1.3, recall from matrix algebra that the rank of a matrix equals the number of linearly independent columns of the matrix. The assumption says that none of the K columns of the data matrix $\mathbf{X}$ can be expressed as a linear combination of the other columns of $\mathbf{X}$. That is, $\mathbf{X}$ is of **full column rank**. Since the K columns cannot be linearly independent if their dimension is less than K, the assumption implies that $n \geq K$, i.e., there must be at least as many observations as there are regressors. The regressors are said to be **(perfectly) multicollinear** if the assumption is not satisfied. It is easy to see in specific applications when the regressors are multicollinear and what problems arise.

Example 1.4 (continuation of Example 1.2): If no individuals in the sample ever changed jobs, then $TENURE_i = EXPR_i$ for all i, in violation of the no multicollinearity assumption. There is evidently no way to distinguish the tenure effect on the wage rate from the experience effect. If we substitute this equality into the wage equation to eliminate $TENURE_i$, the wage equation becomes

$$\log(WAGE_i) = \beta_1 + \beta_2 S_i + (\beta_3 + \beta_4) EXPR_i + \varepsilon_i,$$

which shows that only the sum $\beta_3 + \beta_4$, but not β_3 and β_4 separately, can be estimated.

The homoskedasticity assumption (1.1.12) says that the conditional second moment, which in general is a nonlinear function of $\mathbf{X}$, is a constant. Thanks to strict exogeneity, this condition can be stated equivalently in more familiar terms. Consider the conditional variance $\operatorname{Var}(\varepsilon_i \mid \mathbf{X})$. It equals the same constant because

$$\begin{aligned} \operatorname{Var}(\varepsilon_i \mid \mathbf{X}) &\equiv \operatorname{E}(\varepsilon_i^2 \mid \mathbf{X}) - \operatorname{E}(\varepsilon_i \mid \mathbf{X})^2 \quad \text{(by definition of conditional variance)} \\ &= \operatorname{E}(\varepsilon_i^2 \mid \mathbf{X}) \quad \text{(since } \operatorname{E}(\varepsilon_i \mid \mathbf{X}) = 0 \text{ by strict exogeneity).} \end{aligned}$$

Similarly, (1.1.13) is equivalent to the requirement that

$$\operatorname{Cov}(\varepsilon_i, \varepsilon_j \mid \mathbf{X}) = 0 \quad (i, j = 1, 2, \dots, n;\ i \neq j).$$

That is, in the joint distribution of $(\varepsilon_i, \varepsilon_j)$ conditional on $\mathbf{X}$, the covariance is zero. In the context of time-series models, (1.1.13) states that there is no **serial correlation** in the error term.

Since the (i, j) element of the $n \times n$ matrix $\boldsymbol{\varepsilon\varepsilon}'$ is $\varepsilon_i \varepsilon_j$, Assumption 1.4 can be written compactly as

$$\mathrm{E}(\boldsymbol{\varepsilon\varepsilon}' \mid \mathbf{X}) = \sigma^2 \mathbf{I}_n. \tag{1.1.14}$$

The discussion of the previous paragraph shows that the assumption can also be written as

$$\mathrm{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \sigma^2 \mathbf{I}_n.$$

However, (1.1.14) is the preferred expression, because the more convenient measure of variability is second moments (such as $\mathrm{E}(\varepsilon_i^2 \mid \mathbf{X})$) rather than variances. This point will become clearer when we deal with the large sample theory in the next chapter. Assumption 1.4 is sometimes called the **spherical** error variance assumption because the $n \times n$ matrix of second moments (which are also variances and covariances) is proportional to the identity matrix $\mathbf{I}_n$. This assumption will be relaxed later in this chapter.

The Classical Regression Model for Random Samples

The sample $(\mathbf{y}, \mathbf{X})$ is a **random sample** if $\{y_i, \mathbf{x}_i\}$ is i.i.d. (independently and identically distributed) across observations. Since by Assumption 1.1 ε_i is a function of $(y_i, \mathbf{x}_i)$ and since $(y_i, \mathbf{x}_i)$ is independent of $(y_j, \mathbf{x}_j)$ for $j \neq i$, $(\varepsilon_i, \mathbf{x}_i)$ is independent of $\mathbf{x}_j$ for $j \neq i$. So

$$\begin{aligned}
\mathrm{E}(\varepsilon_i \mid \mathbf{X}) &= \mathrm{E}(\varepsilon_i \mid \mathbf{x}_i), \\
\mathrm{E}(\varepsilon_i^2 \mid \mathbf{X}) &= \mathrm{E}(\varepsilon_i^2 \mid \mathbf{x}_i), \\
\text{and } \mathrm{E}(\varepsilon_i \varepsilon_j \mid \mathbf{X}) &= \mathrm{E}(\varepsilon_i \mid \mathbf{x}_i)\, \mathrm{E}(\varepsilon_j \mid \mathbf{x}_j) \quad (\text{for } i \neq j).
\end{aligned} \tag{1.1.15}$$

(Proving the last equality in (1.1.15) is a review question.) Therefore, Assumptions 1.2 and 1.4 reduce to

$$\text{Assumption 1.2:}\ \mathrm{E}(\varepsilon_i \mid \mathbf{x}_i) = 0 \quad (i = 1, 2, \ldots, n), \tag{1.1.16}$$

$$\text{Assumption 1.4:}\ \mathrm{E}(\varepsilon_i^2 \mid \mathbf{x}_i) = \sigma^2 > 0 \quad (i = 1, 2, \ldots, n). \tag{1.1.17}$$

The implication of the identical distribution aspect of a random sample is that the joint distribution of $(\varepsilon_i, \mathbf{x}_i)$ does not depend on i. So the *un*conditional second moment $\mathrm{E}(\varepsilon_i^2)$ is constant across i (this is referred to as **unconditional homoskedasticity**) and the functional form of the conditional second moment $\mathrm{E}(\varepsilon_i^2 \mid \mathbf{x}_i)$ is the same across i. However, Assumption 1.4 — that the *value* of the conditional