**Chapter 19 Summary**

One of the critical assumptions of the classical linear regression model is that the error term and regressor(s) are uncorrelated. But if they are correlated, then we call such regressor(s) stochastic or endogenous regressors. In this situation the OLS estimators are biased and the bias does not disappear even if the sample size increases indefinitely. In other words, the OLS estimators are not even consistent. As a result, tests of significance and hypothesis testing become suspect.

If we can find proxy variables such that they are uncorrelated with the error term, but are correlated with the stochastic regressors and are not candidates in their own right in the regression model, we can obtain consistent estimates of the coefficients of the suspected stochastic regressors. Such variables, if available, are called instrumental variables, or instruments for short.

In large samples IV estimators are normally distributed with mean equal to the true population value of the regressor under stress and the variance that involves the population correlation coefficient of the instrument with the suspect stochastic regressor. But in small, or finite, samples, IV estimators are biased and their variances are less efficient than the OLS estimators.

The success of IV depends on how strong the instruments are – that is, how strongly they are correlated with the stochastic regressor. If this correlation is high, we say such IVs are strong, but if it is low, we call them weak instruments. If the instruments are weak, IV estimators may not be normally distributed even in large samples.

Finding “good” instruments is not easy. It requires intuition, introspection, familiarity with prior empirical work, or sometimes just luck. That is why it is important to test explicitly whether the chosen instrument is weak or strong, using tests like the Hausman test.

We need one instrument per stochastic regressor. But if we have more than one instrument for a stochastic regressor, we have a surfeit of instruments and we need to test their validity. Validity here means whether the surfeit instruments have high correlation with the regressor but are uncorrelated with the error term. Fortunately, several tests are available to test for this.

If there is more than one stochastic regressor in a model, we will have to find an instrument(s) for each stochastic regressor. Again, we need to test the instruments for their validity.

One practical reason why IVs have become popular is that we have excellent statistical packages, such as *Stata* and *Eviews*, which make the task of estimating IV regression models very easy.

The topic of IV is still evolving and considerable research is being done on it by various academics. It pays to visit their websites to learn more about the recent developments in the field. Of course, the Internet is a source of information on IV and other statistical techniques.