12.3 Checking Instrument Validity

Instrument Relevance

Instruments that explain little variation in the endogenous regressor X are called *weak instruments*. Weak instruments provide little information about the variation in X that is exploited by IV regression to estimate effect of interest: the estimate of the coefficient on the endogenous regressor is estimated inaccurately. Moreover, weak instruments cause the distribution of the estimator to deviate considerably from a normal distribution even in large samples such that the usual methods for obtaining inference about the true coefficient on X may produce wrong results. See Chapter 12.3 and Appendix 12.4 of the book for a more detailed argument on the undesirable consequences of using weak instruments in IV regression.

A Rule of Thumb for Checking for Weak Instruments

Key Concept 12.5

Consider the case of a single endogenous regressor X and m instruments $Z_1, \ldots, Z_m$. If the coefficients on all instruments in the population first-stage regression of a TSLS estimation are zero, the instruments do not explain any of the variation in the X which clearly violates assumption 1 of Key Concept 12.2. Although the latter case is unlikely to be encountered in practice, we should ask ourselves "to what extent" the assumption of instrument relevance should be fulfilled.

While this is hard to answer for general IV regression, in the case of a *single* endogenous regressor X of may use the following rule of thumb:

Compute the F-statistic which corresponds to the hypothesis that the coefficients on $Z_1, \ldots, Z_m$ are zero in the first-stage regression. If the F-statistic is less than 10, the instruments are weak such that the TSLS estimate of the coefficient on X is biased and no valid statistical inference about its true value can be made. See also Appendix 12.5 of the book.

The rule of thumb of Key Concept 12.5 is easily implemented in R. Run the first-stage regression using `lm(` subsequently compute the heteroskedasticity-robust F-statistic by means of `linearHypothesis()`. T part of the application to the demand for cigarettes discussed in Chapter 12.4.

If Instruments are Weak

There are two ways to proceed if instruments are weak:

When the Assumption of Instrument Exogeneity is Violated

If there is correlation between an instrument and the error term, IV regression is not consistent (this is sho
Appendix 12.4 of the book). The overidentifying restrictions test (also called the J-test) is an approach to t
hypothesis that *additional* instruments are exogenous. For the J-test to be applicable there need to be *m*
instruments than endogenous regressors. The J-test is summarized in Key Concept 12.5.

J-Statistic / Overidentifying Restrictions Test

Key Concept 12.6

Take $\widehat{u}_i^{TSLS}$, $i = 1, \ldots, n$, the residuals of the TSLS estimation of the general IV regression model
(12.5). Run the OLS regression

$$\widehat{u}_i^{TSLS} = \delta_0 + \delta_1 Z_{1i} + \cdots + \delta_m Z_{mi} + \delta_{m+1} W_{1i} + \cdots + \delta_{m+r} W_{ri} + e_i \qquad (1$$

and test the joint hypothesis

$$H_0 : \delta_1 = 0, \ldots, \delta_m = 0$$

which states that all instruments are exogenous. This can be done using the corresponding F-statistic
computing

$$J = mF.$$

This test is the overidentifying restrictions test and the statistic is called the J-statistic with

$$J \sim \chi^2_{m-k}$$

in large samples under the null and the assumption of homoskedasticity. The degrees of freedom m –
state the degree of overidentification since this is the number of instruments m minus the number of
endogenous regressors k.

It is important to note that the J-statistic discussed in Key Concept 12.6 is only χ^2_{m-k} distributed when the
term e_i in the regression (12.9) is homoskedastic. A discussion of the heteroskedasticity-robust J-statistic
beyond the scope of this chapter. We refer to Section 18.7 of the book for a theoretical argument.

As for the procedure shown in Key Concept 12.6, the application in the next section shows how to apply th
test using `linearHypothesis()`.