Peter Kennedy. A Guide to Econometrics, 5th Edition, 2003. pp. 175-176.

How many instruments should be found? This turns out to be an awkward question. On the one hand, if the number of instruments (including variables that can 'serve as their own instrument) is just equal to the number of explanatory variables (i.e., one instrument for each explanatory variable). $\beta_{I\varpi}$ has neither mean nor variance so we would expect it in some cases to have poor properties in finite samples, as evidenced in Nelson and Stare (1990a, l990b). Adding an extra instrument allows it to have a mean, and one more allows it to have a variance, so it would seem desirable to have at least two more instruments than explanatory variables. On the other hand, as we add more and more instruments, in small samples it becomes closer and closer to X and so begins to introduce the bias that the IV procedure is trying to eliminate. This bias is proportional to the inverse of the F test statistic for testing the significance of the instrumental variables in explaining the explanatory variable for which it is to serve as an instrument. Lee (2001) offers some suggestions for how to alleviate the finite-sample bias of 2SLS.

The use of extra instruments beyond the bare minimum of one for each explanatory variable should be tested. Davidson and MacKinnon (1993, pp. 232-7) suggest a means of doing so by testing the joint hypothesis that the model is correctly specified and that the instruments used are valid. This test statistic is calculated as N (or N - It where K is the number of regressors) times the uncentered R2 from regressing the IV residuals on all the instruments (including a1l variables serving as their own instruments), and is distributed as a chi-square with degrees of freedom equal to the number of instruments in excess of the number of explanatory variables. (The uncentered $R^2 = 1 - \sum \hat{e}_i^2 / \sum y_i^2$ instead of $R^2 = 1 - \sum \hat{e}_i^2 / \sum (y_i - \bar{y})$. This test is sometimes called the Sargan test.

Bound et al. (1995) find that whenever there is weak correlation between the error and an explanatory variable, and also weak correlation between the instrument and this explanatory variable, IV estimation exhibits large bias, even when the sample size is very large. They recommend that the quality of the instrument be checked, for example by testing significance of the instrument in the first stage of IV estimation. A rule of thumb here has become common. Regress the variable requiring an instrument on all the exogenous variables appearing in that equation, plus all the instruments, and calculate the F-statistic for testing this latter set of variables (the instruments) against zero. The inverse of this F value indicates the bias of IV relative to OLS. Bartels (1991) reaches a related conclusion, noting that even if an instrument is not independent of the error it may be superior on the mean square error criterion to an instrument that is independent of the error. He offers some rules of thumb for selecting IV versus OLS. Good summaries of the weak IV problem, and the difference between asymptotic and small-sample properties of IV estimators, can be found in Shea (1997), Zivot, Stare, and Nelson (1998), and Woglom (2001) .