Summary in English

As a rule the estimation of a single econometric relationship, which describes how a dependent variable relates to a number of explanatory variables and an error term, requires that the researcher has obtained observations on at least as many exogenous variables as the number of unknown coefficients accompanying the explanatory variables in the model. The explanatory variables can either be exogenous or endogenous. When investigating relationships for which the observations have not been obtained in carefully designed laboratory experiments, but instead are deducted - completely beyond the researcher's control - from real economic behaviour, then often some of the explanatory variables will be endogenous. This endogeneity means that the explanatory variable itself will partially depend on the dependent variable, and hence will be correlated with the error term. The presence of endogenous explanatory variables requires for the usual estimation techniques the availability of a sufficient number of variables which do not belong to the explanatory variables and which do not depend on the dependent variable either. This renders them exogenous. All variables labeled as exogenous for the relationship under examination, both the explanatory and the non-explanatory exogenous variables, form together the set of so-called instrumental variables.

This thesis studies techniques to classify variables as either exogenous or endogenous, as well as the consequences of a wrong classification. Several sequential classification procedures are designed such that they remain feasible in case of an abundance of potential instrumental variables, as often is the case in practice. The properties of such classification procedures in finite samples, also when enhanced in a certain way by bootstrap methods, turn out to be rather limited. Therefore undetected endogeneity and the subsequent use of incorrect exogeneity conditions can be common. This underlines the relevance of

studying the consequences of using instrumental variables that are in fact endogenous. This thesis assesses these consequences through simulation and in addition by determining the asymptotic distribution of estimators for linear models also when these are inconsistent due to the use of invalid instruments. The obtained asymptotic distribution appears to be very precise in approximating the actual distribution in finite samples, provided that the validly exploited exogeneity is not too weak. The weakness of an instrumental variable is determined by the extent to which this variable (in as far as not coinciding with further instrumental variables) is related with the explanatory variables in the investigated relationship.

Both the analytical results and the simulation findings lead to the remarkable conclusion that inconsistent least-squares estimators are often more attractive than consistent instrumental variable estimators. The reason is that least-squares estimators always use the strongest possible instruments. If some of those instruments are actually endogenous they result in more precise though inconsistent estimators than valid but weak instrumental variables estimators do. To what extent the designed classification procedures and the theoretical findings on inconsistent instrumental variable estimators can support and improve actual empirical inference is illustrated in an analysis of cross-section data concerning the effects of schooling on earnings.