Journal of Econometrics 193 (2016) 113–122

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A discontinuity test for identification in triangular nonseparable models

Carolina Caetano [a], Christoph Rothe [b,*], Neşe Yıldız [a]

[a] Department of Economics, University of Rochester, 238 Harkness Hall, P.O. Box: 270156, Rochester, NY 14627, United States
[b] Department of Economics, Columbia University, 420 W 118th St., New York, NY 10027, United States

ARTICLE INFO

Article history:
Received 23 November 2014
Received in revised form
16 September 2015
Accepted 19 January 2016
Available online 20 February 2016

JEL classification:
C12
C14
C31
C36
C52

Keywords:
Nonseparable model
Triangular systems
Control variable
Instrument validity
Nonparametric identification

ABSTRACT

This paper presents a test for the validity of control variable approaches to identification in triangular nonseparable models. Assumptions commonly imposed to justify such methods include full independence of instruments and disturbances and existence of a reduced form that is strictly monotonic in a scalar disturbance. We show that if the data has a particular structure, namely that the distribution of the endogenous variable has a mass point at the lower (or upper) boundary of its support, validity of the control variable approach implies a continuity condition on an identified function, which can be tested empirically.

1. Introduction

Empirical specifications with nonseparable unobservables have become increasingly popular in econometrics in recent years (e.g. Matzkin, 2003; Chesher, 2003; Imbens and Newey, 2009; Blundell and Matzkin, 2010). In their most basic form, these models assume that an outcome variable Y is linked to a covariate X and an unobserved quantity U through the relationship

$$Y = m_1(X, U).$$

Compared to classical specifications with additively separable disturbances, these types of models can accommodate very general forms of unobserved heterogeneity. For example, they allow for heterogeneous responses to policy interventions among observationally identical individuals. Both economic theory and empirical evidence strongly suggest that such general forms of unobserved heterogeneity are a common feature of economic data (e.g. Heckman, 2001).

The additional flexibility of these models comes of course at a cost. When the covariate X is endogenous, the lack of an additively separable disturbance complicates the identification of many interesting structural parameters. Availability of an instrument, say Z, that is uncorrelated with U but correlated with X does generally not suffice for identification of meaningful quantities. Instead, one has to impose additional conditions. One popular approach is based on control variables (e.g. Blundell and Powell, 2003; Imbens and Newey, 2009). This method entails finding a random variable R that can be written as an identified function of the data, and is such that X and U are stochastically independent conditional on R; that is $X \perp U|R$. This property ensures that changes in X can be interpreted as causal after conditioning on R, and thus many structural parameters can be identified from the conditional distribution of Y given X and R.

A control variable arises for example if the model has a triangular structure. This means that the endogenous variable is assumed to be generated in a first stage as

$$X = m_2(Z, V),$$

* Corresponding author.
E-mail addresses: carol.caetano@rochester.edu (C. Caetano), cr2690@columbia.edu (C. Rothe), nese.yildiz@rochester.edu (N. Yıldız).

http://dx.doi.org/10.1016/j.jeconom.2016.01.007
0304-4076/

with Z an instrument and V an unobserved quantity. In a seminal paper, Imbens and Newey (2009) show that if Z is independent of (U, V) and m_2 depends monotonically on the continuously distributed scalar V, then $R \equiv F_{X|Z}(X, Z) = F_V(V)$ is a valid control variable, where $F_{X|Z}$ denotes the conditional CDF of X given Z and F_V denotes the unconditional CDF of V. This is because in such a model R is a one-to-one transformation of V, and conditional on V all variation in the endogenous variable X comes from variation in Z.

While this approach to identification is powerful, the postulated triangular specification and the restrictions on first stage unobserved heterogeneity impose substantial limitations for the underlying economic model. For example, the condition that X depends monotonically on V implies that individuals with common values of X and Z will react in exactly the same way to exogenous variation in Z. In many empirical contexts such a behavioral restriction can be difficult to justify through theoretical considerations alone, and thus its validity might be doubtful. Other conditions, like the full independence of Z and (U, V) might be equally questionable in practice. Unfortunately, one cannot simply adapt established methods for specification testing in models with additively separable disturbances to settings with nonseparable unobserved heterogeneity. For example, tests for instrument validity based on overidentifying restrictions, such as those of Sargan (1958) and Hansen (1982), have no direct analogue in triangular nonseparable models: if the identifying assumptions mentioned above hold with some vector of instruments Z, there is no guarantee that they will also hold (*mutatis mutandis*) with some subvector of Z.

In this paper, we propose a test of the conditions necessary for justifying a control variable approach in nonseparable models with a particular additional structure. Specifically, we study the case where the data generating process is such that the distribution of the endogenous covariate X has a mass point at some known value, is otherwise continuously distributed, and exerts a continuous effect on the outcome variable of interest. In most applications, the location of the mass point coincides with the lower or upper boundary of the support of X. We show that in this case validity of the control variable approach implies a continuity condition on a certain function at one particular point, and that this continuity condition can be tested. Validity of the control function approach can thus be potentially refuted using data and a weak set of maintained assumptions alone. To the best of our knowledge, our paper is the first to consider this type of testable implications of identifying assumptions for this type of nonseparable models.

The idea behind the derivation of our testable implication is related to that of Caetano (2015), who shows that endogeneity of a covariate X with the above-mentioned properties leads to a discontinuity in the conditional expectation function of Y given X at the mass point. In this paper, the starting point for our analysis is the insight that conditioning on a valid control variable should remove this discontinuity. If not the control variable approach must be invalid; and at least one of the assumptions that was made to justify it has to be violated. This basic idea can unfortunately not be implemented directly, because in our setting a valid control variable is only available for those individuals whose realization of the endogenous variable is different from the mass point. We address this issue by integrating out the control variable in such a way that it is no longer necessary to identify it at the mass point, which yields a continuity condition on an identified function.

We also propose a test statistic that is based on a direct sample analogue of the function whose continuity we wish to verify. Its computation only involves standard nonparametric regression techniques and a simple numerical integration step. Deriving its asymptotic properties is a non-standard problem, as it involves nonparametric regression with estimated data points. We use recent results in Mammen et al. (2012, 2015) on generated covariates in non- and semiparametric models to account for this two-stage structure. Through a Monte Carlo study, we show that this leads to a test with good size and power properties in finite samples.

1.1. Potential applications

Focusing on settings where the endogenous variable has a mass point at the lower or upper boundary of its support is clearly restrictive, but such scenarios can still be found in a wide array of applications. If the endogenous variable is a choice which must be non-negative, such as the quantity consumed of a particular good, our methods can usually be applied. One such case is the problem of estimating the effects of smoking during pregnancy on the baby's birth weight. Smoking amounts cannot be negative, and a sizable proportion of the pregnant women do not smoke. The literature on this topic is extensive (see e.g. Kramer, 1987, 1998; Vogler and Kozlowski, 2002; Almond et al., 2005; Tominey, 2007; Fertig, 2010). Papers using instrumental variable approaches to address the problem of the endogeneity of smoking include Evans and Ringel (1999) and Lien and Evans (2005), who use tax variations across locations and time, and Wehby et al. (2011), who use genetic markers that predict smoking behavior as instruments for smoking. This literature generally focuses on linear models, but our approach could be used to investigate the validity of a triangular nonseparable specification in such a context.

Another interesting class of examples are studies of the effects of labor supply, expressed as "hours of work", on different outcomes. See Blundell and MaCurdy (1999) for a survey on the extensive literature on this topic. Approaches based on instrumental variables include, for example, Heckman and MaCurdy (1980)'s study of the effect of female labor supply on female wages, which uses the wages of the husband or other non-female generated sources of income as instrument for female hours worked. Another example is Connelly and Kimmel (2009), who study of the effect of labor supply on time spent in childcare, using age and education squared, as well as spouse's age and education as instruments for hours of work. A third example can be found in Blau and Grossberg (1992), who study the effect of maternal labor supply on the child's cognitive development, using supply side determinants of maternal labor supply, such as predicted wage, as instruments for the maternal hours worked. Hours of work must of course be non-negative, and a large part of the population supply exactly zero hours, and thus our methods can be used to test the validity of a control function approach in these setups.

In other settings the presence of mass points is due to certain legal restrictions. One example is the variable "schooling". Due to minimum attendance laws (and additionally minimum working age laws), students are forced to remain in school until a certain age threshold is reached. The well known literature on the returns to schooling (see Card (1999)) is concerned with endogeneity of "schooling". Several instrumental variables have been used, such as for example Card (1995)'s proximity to college. We can therefore test the validity of a control function based approach in this setting.

Another variable which is also constrained by law is wage, which must be equal to or larger than a pre-established quantity for all legal employment. Examples of problems and instruments include Stewart and Swaffield (1997), which studies the effect of wages on the desired hours of work, and uses as instrument for wages a combination of variables including years of education and the male age-specific regional unemployment rate. Iwata and Tamada (2014) examine the effect of wages on the commuting time of married women, also using a combination of variables as instrument for wages, including the average market wage for women and the age of the respondent. Lydon and Chevalier (2002)