ELSEVIER

Journal of Econometrics 96 (2000) 201–229

JOURNAL OF Econometrics

www.elsevier.nl/locate/econbase

Semiparametric identification and heterogeneity in discrete choice dynamic programming models

Christopher R. Taber

Department of Economics and Institute for Policy Research Northwestern University, 2003 Sheridan Road, Evanston, IL 60208, USA

Received 1 January 1997; received in revised form 1 June 1999

Abstract

Empirical discrete choice dynamic programming models have become important empirical tools. A question that arises in estimation and interpretation of the results from these specifications is which combination of data and assumptions are needed to overcome problems of heterogeneity, selection, and omited variables bias. This paper addresses this question by considering nonparametric identification of a version of the model that allows for quite general forms of unobservable and information structures. I show that the model can be identified under conditions similar to a static polychotomous choice model. Using a stochastic version of an 'identification of infinity' argument, utility can be identified up to a monotonic transformation of the observables under strong support conditions and two types of exclusion restriction. The first type is similar to a standard static exclusion restriction: a variable that influences the first period decision, but does not enter the second period decision directly. The second type requires a variable that does not affect the utility of the first option directly, but is known during the first period, and has predictive power on the choice during the second. I also provide two specifications under which the full error structure can be identified. This requires the additional assumption of stochastic innovations in the observables. I then use the model to estimate schooling decisions in which students deciding whether to drop out of high school account for the option value of attending college.

JEL classification: C14; C35; C51

Keywords: Identification; Discrete choice; Dynamic programming

0304-4076/00/$ - see front matter
PII: S0304-4076(99)00057-3

1. Introduction

Empirical discrete choice dynamic programming models have become important empirical tools. In some applications of these models, problems of substantial heterogeneity/selection/omitted variable bias arise (see, e.g. Keane and Wolpin (1997) or Eckstein and Wolpin (1997)). The source of these biases is potentially more complex in dynamic models than static ones in that agents may have heterogeneity not only in outcomes, but also in expectations about future outcomes. A question that arises in estimation and interpretation of the results in these cases is which combination of data and assumptions are needed to overcome these problems. This paper addresses this question by considering nonparametric identification of a version of the model that allows for quite general forms of unobservable and information structures. Despite the added complexity of the model, I show that it can be identified under conditions similar to a static polychotomous choice model. Using a stochastic version of an 'identification of infinity' argument, utility can be nonparametrically identified up to a monotonic transformation of the observables under strong support conditions and two types of exclusion restriction. The first type is similar to a standard static exclusion restriction: a variable that influences the first period decision, but does not enter the second period decision directly. The second type requires a variable that does not affect the utility of the first option directly, but is known during the first period and has predictive power on the choice during the second. I also provide two specifications under which the full error structure can be identified. This requires the additional assumption of stochastic innovations in the X's: a variable known at time one that helps predict the second period decision, but conditional on second period observables, has no influence on the decision.

The specification I develop is a generalization of a dynamic 'Roy' type model, and I focus on schooling decisions. In deciding whether to drop out of high school a student takes into account both the direct value of graduating from high school as well as the value of the option to attend college. While making this decision, a student does not know whether he will attend college. Heterogeneity bias is likely to be important in that students with high returns or tastes for high school are also likely to have high returns or tastes for college. While there is a substantial literature addressing the selection/heterogeneity issue in schooling models, the previous work has typically ignored the complexity of the heterogeneity. The problem is not just that the returns to college are likely to be correlated with returns to high school, but also that agents may have additional information about their own private returns to college which is unobservable to the econometrician. For example, a high school student may know that he has excellent teaching skills. While this information may be correlated with the returns to high school, since teachers must have a college degree it is much more important for the decision about whether to attend