Journal of Econometrics 206 (2018) 613–644

Contents lists available at ScienceDirect

Journal of Econometrics

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

Estimation of random coefficients logit demand models with interactive fixed effects[☆]

Hyungsik Roger Moon[a,b,*], Matthew Shum[c], Martin Weidner[d,e]

[a] Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089-0253, United States
[b] Department of Economics, Yonsei University, Seoul, Republic of Korea
[c] Division of Humanities and Social Sciences, California Institute of Technology, MC 228-77, Pasadena, CA 91125, United States
[d] Department of Economics, University College London, Gower Street, London WC1E 6BT, UK
[e] CeMMAP, UK

ARTICLE INFO

Article history:
Available online 11 July 2018

JEL classification:
C23
C25

Keywords:
Discrete-choice demand model
Interactive fixed effects
Factor analysis
Panel data
Random utility model

ABSTRACT

We extend the Berry, Levinsohn and Pakes (BLP, 1995) random coefficients discrete-choice demand model, which underlies much recent empirical work in IO. We add interactive fixed effects in the form of a factor structure on the unobserved product characteristics. The interactive fixed effects can be arbitrarily correlated with the observed product characteristics (including price), which accommodate endogeneity and, at the same time, capture strong persistence in market shares across products and markets. We propose a two-step least squares-minimum distance (LS-MD) procedure to calculate the estimator. Our estimator is easy to compute, and Monte Carlo simulations show that it performs well. We consider an empirical illustration to US automobile demand.

1. Introduction

The Berry et al. (1995) (hereafter BLP) demand model, based on the random coefficients logit multinomial choice model, has become the workhorse of demand modeling in empirical industrial organization and antitrust analysis. An important virtue of this model is that it parsimoniously and flexibly captures substitution possibilities between the products in a market. At the same time, the nested simulated GMM procedure proposed by BLP accommodates possible endogeneity of the observed product-specific regressors, notably price. This model and estimation approach has proven very popular (e.g. Nevo (2001), Petrin (2002); surveyed in Ackerberg et al. (2007)).

Taking a cue from recent developments in panel data econometrics (e.g. Bai and Ng (2006), Bai (2009), and Moon and Weidner (2015, 2017)), we extend the standard BLP demand model by adding interactive fixed effects to the unobserved product characteristic, which is the main "structural error" in the BLP model. This interactive fixed effect specification combines market (or time) specific fixed effects with product specific fixed effects in a multiplicative form, which is often referred to as a factor structure.

☆ We thank participants in presentations at Georgetown, Johns Hopkins, Ohio State, Penn State, Rice, Texas A&M, UC Davis, UC Irvine, UCLA, Chicago Booth, Michigan, UPenn, Wisconsin, Southampton, the 2009 California Econometrics Conference and the 2010 Econometric Society World Congress for helpful comments. Chris Hansen, Han Hong, Sung Jae Jun, Jinyong Hahn, and Rosa Matzkin provided very helpful discussions. Moon acknowledges the NSF for financial support via SES 0920903. Weidner acknowledges support from the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001, and from the European Research Council grant ERC-2014-CoG-646917-ROMIA.
* Corresponding author at: Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089-0253, United States.
E-mail addresses: moonr@usc.edu (H.R. Moon), mshum@caltech.edu (M. Shum), m.weidner@ucl.ac.uk (M. Weidner).

https://doi.org/10.1016/j.jeconom.2018.06.016
0304-4076/

Our factor-based approach extends the baseline BLP model in two ways. First, we offer an alternative to the usual moment-based GMM approach. The interactive fixed effects "soak up" some important channels of endogeneity, which may obviate the need for instrumental variables of endogenous regressors such as price. This is important as such instruments may not be easy to identify in practice. Moreover, our analysis of the BLP model with interactive fixed effects illustrates that the problem of finding instruments for price (which arises in any typical demand model) is distinct from the problem of underidentification of some model parameters (such as the variance parameters for the random components), which arises from the specific nonlinearities in the BLP random coefficients demand model. In our setting, the fixed effects may obviate the need for instruments to control for price endogeneity but, as we will point out, we still need to impose additional moment conditions in order to identify these nonlinear parameters. Second, even if endogeneity persists in the presence of the interactive fixed effects, the instruments only need to be exogenous with respect to the residual part of the unobserved product characteristics, which is not explained by the interactive fixed effect. This may expand the set of variables which may be used as instruments.

To our knowledge, the current paper presents the first application of some recent developments in the econometrics of long panels (with product and market fixed effects) to the workhorse demand model in empirical IO. Relative to the existing panel factor literature (for instance, Bai (2009), and Moon and Weidner (2015, 2017)) that assume a linear regression with exogenous regressors, the nonlinear model that we consider here poses both identification and estimation challenges. Namely, the usual principal components approach for linear factor models with exogenous regressors is inadequate due to the nonlinearity of the model and the potentially endogenous regressors. At the same time, the conventional GMM approach of BLP cannot be used for identification and estimation due to the presence of the interactive fixed effects.

We propose an alternative identification and estimation scheme which we call the *Least Squares-Minimum Distance (LS-MD)* method. It consists of two steps. The first step is a least squares regression of the mean utility on the included product-market specific regressors, factors, and the instrumental variables. The second step minimizes the norm of the least squares coefficient of the instrumental variables in the first step. This estimation approach is similar to the two stage estimation method for a class of instrumental quantile regressions in Chernozhukov and Hansen (2006). We show that under regularity conditions that are comparable to the standard GMM problem, the parameter of interest is point identified and its estimator is consistent. We also derive the limit distribution under asymptotic sequences where both the number of products and the number of markets converge to infinity. In practice, the estimator is simple and straightforward to compute. Monte Carlo simulations demonstrate its good small-sample properties.

Our work complements some recent papers in which alternative estimation approaches and extensions of the standard random coefficients logit model have been proposed, including Villas-Boas and Winer (1999), Knittel and Metaxoglou (2014), Dube et al. (2012), Harding and Hausman (2007), Bajari et al. (2011), and Gandhi et al. (2010).

We illustrate our estimator on a dataset of market shares for automobiles, inspired by the exercise in BLP. This application illustrates that our estimator is easy to compute in practice. Significantly, we find that, once factors are included in the specification, the estimation results under the assumption of exogenous and endogenous price are quite similar, suggesting that the factors are indeed capturing much of the unobservable product and time effects leading to price endogeneity.

The paper is organized as follows. Section 2 introduces the model. In Section 3 we discuss how to identify the model when valid instruments are available. In Section 4 we introduce the LS-MD estimation method. Consistency and asymptotic normality are discussed in Section 5. Section 6 contains Monte Carlo simulation results, and Section 7 discusses the empirical example. Section 8 concludes. In the appendix we list the assumptions for the asymptotic analysis and provide technical derivations and proofs of results in the main text.

Notation

We write A' for the transpose of a matrix or vector A. For column vectors v the Euclidean norm is defined by $\|v\| = \sqrt{v'v}$. For the nth largest eigenvalues (counting multiple eigenvalues multiple times) of a symmetric matrix B we write $\mu_n(B)$. For an $m \times n$ matrix A the Frobenius norm is $\|A\|_F = \sqrt{\mathrm{Tr}(AA')}$, and the spectral norm is $\|A\| = \max_{0 \neq v \in \mathbb{R}^n} \frac{\|Av\|}{\|v\|}$, or equivalently $\|A\| = \sqrt{\mu_1(A'A)}$. Furthermore, we use $P_A = A(A'A)^{\dagger}A'$ and $M_A = \mathbb{1}_m - A(A'A)^{\dagger}A'$, where $\mathbb{1}_m$ is the $m \times m$ identity matrix, and $(A'A)^{\dagger}$ denotes a generalized inverse, since A may not have full column rank. The vectorization of an $m \times n$ matrix A is denoted vec(A), which is the $mn \times 1$ vector obtained by stacking the columns of A. For square matrices B, C, we use $B > C$ (or $B \geq C$) to indicate that $B - C$ is positive (semi) definite. We use ∇ for the gradient of a function, *i.e.* $\nabla f(x)$ is the vector of partial derivatives of f with respect to each component of x. We use "wpa1" for "with probability approaching one".

2. Model

The random coefficients logit demand model is an aggregate market-level model, formulated at the individual consumer-level. Consumer i's utility of product j in market[1] t is given by

$$u_{ijt} = \delta^0_{jt} + \epsilon_{ijt} + X'_{jt} v_i, \tag{2.1}$$

[1] The t subscript can also denote different time periods.