A Note on Nonparametric Identification of Distributions of Random Coefficients in Multinomial Choice Models

Author(s): Jeremy T. Fox

Source: *Annals of Economics and Statistics*, June 2021, No. 142 (June 2021), pp. 305-310

Published by: GENES on behalf of ADRES

Stable URL: https://www.jstor.org/stable/10.15609/annaeconstat2009.142.0305

Annals of Economics and Statistics, Number 142, June 2021

A NOTE ON NONPARAMETRIC IDENTIFICATION OF DISTRIBUTIONS OF RANDOM COEFFICIENTS IN MULTINOMIAL CHOICE MODELS

JEREMY T. FOX[a]

I prove the point identfication of the joint distribution of the vector of random coefficients and additive, good-specific errors in a multinomial choice model. The identification theorem extends the binary choice results of Ichimura and Thompson (1998) as well as Gautier and Kitamura (2013) from two to two or more choices.

JEL Codes: C14, C25, C35.
Keywords: Discrete Choice, Multinomial Choice, Random Coefficients, Demand.

1. INTRODUCTION

Two key papers on identifying the distribution of random coefficients and additive errors in binary (two) choice have not been extended to multinomial (two or more, importantly three or more) choice (Ichimura and Thompson, 1998; Gautier and Kitamura, 2013). Under large support on choice-specific explanatory variables, one can always turn a multinomial choice model with three or more goods into a binary choice model by setting the explanatory variables for all but two goods to be minus infinity. This identification at infinity approach does not identify the joint distribution of all unobservables in the model if there are unobservables that enter only the utilities of certain goods. Establishing the identification of the joint distribution of random coefficients and additive unobservables in the multinomial choice model is the point of this note.

The argument in this note might be considered an extension to multinomial choice of the brief argument about identifying a distribution of random coefficients in binary choice in Lewbel (2014, Section 8). Our identification proof uses the results on identifying distributions of random coefficients in seemingly unrelated regressions by Masten (2018). A key step of our identification proof is also found in Berry and Haile (2016) and Fox and Gandhi (2016).

There are some related papers on multinomial choice. Lewbel (2000) considers multinomial choice in a semiparametric setting but does not explicitly identify a distribution of random coefficients. Fox and Gandhi (2016) study the topic of this note: nonparametric identification in multinomial choice models, where the example of a linear-in-random-coefficients model is a special case of their analysis. However, Fox and Gandhi assume that the distribution of random coefficients and additive errors takes on unknown finite support in the appropriate real space. This note avoids the unknown finite support assumption. Fox, Kim, Ryan, and Bajari (2012) nonparametrically identify a distribution of random coefficients on continuous explanatory variables but rely on the additive, good-specific unobservables having a known distribution such as the type I extreme value or logit distribution. In this note, the joint distribution of the good-specific, additive unobservables is identified.

[a]Rice University and NBER. jeremyfox@gmail.com

2. MODEL

Consider a multinomial choice model with random coefficients. Let i index a consumer. There are J inside goods and one outside good, called choice 0. The outside good has a utility normalized to $u_{i,0} = 0$. The inside goods have utilities

$$u_{i,j} = \beta_i' x_{i,j} + \epsilon_{i,j},$$

where $x_{i,j}$ is a vector of observables for choice j and consumer i, β_i is consumer i's vector of random coefficients on the explanatory variables, and $\epsilon_{i,j}$ is an additive unobservable for choice j and consumer i. We do not impose that $\epsilon_{i,j}$ has mean zero and we do not allow intercept terms in $x_{i,j}$. Note that β_i is common to all choices.

We impose the scale normalization that one element of the vector β_i has the value ± 1 for each i. This rules out that this coefficient can be zero. Let $w_{i,j}$ be the corresponding scalar element of $x_{i,j}$ and let $\tilde{x}_{i,j}$ be all other elements of $x_{i,j}$, so that $x_{i,j} = (w_{i,j}, \tilde{x}_{i,j})$. Then the utility to choice j can be rewritten as

$$u_{i,j} = \pm 1 w_{i,j} + \tilde{\beta}_i' \tilde{x}_{i,j} + \epsilon_{i,j},$$

where $\tilde{\beta}_i$ corresponds to the random coefficients on only the items in $\tilde{x}_{i,j}$. We further impose that the coefficient on $w_{i,j}$ is either $+1$ for all consumers i or is -1 for all consumers i.

Let $x_i = (x_{i,1}, \ldots, x_{i,J})$, $w_i = (w_{i,1}, \ldots, w_{i,J})$, $\tilde{x}_i = (\tilde{x}_{i,1}, \ldots, \tilde{x}_{i,J})$ and $\epsilon_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,J})$ all be long vectors. Also, define $\gamma_i = (\beta_i, \epsilon_i)$ as another long vector. Note that γ_i is a heterogeneous parameter vector, not a homogeneous parameter.

We consider i.i.d. observations on (y_i, x_i), where y_i is the choice that maximizes $u_{i,j}$ over $\{0, 1, \ldots, J\}$. Given this, we can nonparametrically identify conditional choice probabilities $\Pr(y_i = j \mid x_i)$.

The sign of the coefficient on $w_{i,j}$ is learned in identification; a positive coefficient on $w_{i,j}$ corresponds to higher values of $w_{i,j}$ increasing the choice probability of good j, other explanatory variables held constant. We make this formal in the proof of identification. As the choice probability of good j is monotone in $w_{i,j}$, a failure of monotonicity in estimated choice probabilities means that the model is rejected by the data. Researchers can test for this monotonicity property before estimating the model parameters.

We assume that x_i is independent of γ_i. In principle, one can discuss endogeneity with various methods in the literature, such as Lewbel (2000), although this paper does not discuss endogeneity. In what follows, we drop the i subscript.

The only unknown object in this model is $F(\gamma)$, the joint distribution of the additive unobservables ϵ and the random coefficients β. Because we will not restrict the support of each ϵ_j and β, a sufficient condition for identification of F, as shown in this note, will be as follows.

ASSUMPTION *The support of x, reordered as $x = (w, \tilde{x})$ is $\mathbb{R}^J \times \tilde{X}$, where the support of w is the large support $\mathbb{R}^J$ and the support of the vector $\tilde{x}$ is $\tilde{X}$, a closed, weak superset of an open subset of the real space $\mathbb{R}^{\dim(\tilde{x})}$.*[1]

[1] Often the mathematical term support is defined to be a closed set.

$\tilde{X} = \mathbb{R}^{\dim(\tilde{x})}$ is allowed but certainly not required. Having support on the product space $\mathbb{R}^J$ for w and a superset of an open subset of $\mathbb{R}^{\dim(\tilde{x})}$ for $\tilde{x}$ rules out the entire vector x containing polynomial terms of other elements in x, interactions of two elements also in x, or the same element of x entering the utility of different goods. Continuous support on all elements in x rules out discrete x. Large support is needed for w but not for $\tilde{x}$. The support assumptions are discussed in more detail below.

If consumer demographics z_i are in the data, one approach is for the researcher to simply condition on z_i and identify $F(\gamma \mid z_i)$ separately for each demographic group z_i. This requires the support of x to satisfy the above restrictions conditional on z_i.

3. IDENTIFICATION

We are now ready to state the main identification theorem.

THEOREM 1 *If i) the support of x is as stated in the formal assumption, ii) γ is independent of x, iii) γ has finite absolute moments, and iv) the distribution of γ is uniquely determined by its moments, then $F(\gamma)$ is identified.*

PROOF: The argument works in three steps. First, we identify the sign of the coefficient on the explanatory variables w_j that form the scale normalization. Second, we use w to trace out the CDF of utility values (other than from w) conditional on the other explanatory variables in $\tilde{x}$. Third, we cite work by Masten (2018, Theorem 1) on seemingly unrelated regressions with random coefficients to identify the distribution of the random coefficients and additive unobservables.

First, we identify the sign of the coefficient on each w_j, which has been normalized to ± 1. We observe conditional choice probabilities $\Pr(y = j \mid x)$ and the vector w is a subvector of x. If $\Pr(y = j \mid x)$ is monotonically increasing in w_j for some j, then the coefficient on each w_j is $+1$. If it is decreasing, then the coefficient on each w_j is -1. Let $\bar{w}_j = +w_j$ or $\bar{w}_j = -w_j$, as appropriate based on the sign of the coefficient on w_j. Let $\bar{w} = (\bar{w}_1, \ldots, \bar{w}_J)$.

Second, we use arguments motivated by Lewbel (2000) to trace a CDF of utility values (other than from w) conditional on the explanatory variables in $\tilde{x}$. Define

$$\tilde{u}_j = \tilde{\beta}' \tilde{x}_j + \epsilon_j$$

and $\tilde{u} = (\tilde{u}_1, \ldots, \tilde{u}_J)$. Let the CDF of the vector $\tilde{u}$ conditional on the vector $\tilde{x}$ be $G_{\tilde{u}}(\tilde{u} \mid \tilde{x})$. Let $\tilde{u}^\star$ be a point of evaluation of the CDF. Then, for arbitrary $\tilde{x}$,

$$G_{\tilde{u}}(\tilde{u}^\star \mid \tilde{x}) = \Pr(\tilde{u} \leq \tilde{u}^\star \mid \tilde{x}) = \Pr(\tilde{u} \leq -\bar{w} \mid -\bar{w} = \tilde{u}^\star, \tilde{x}) = \Pr(y = 0 \mid -\bar{w} = \tilde{u}^\star, \tilde{x}).$$

All the lower case letters in the above display equation are vectors, except for y. This argument identifies the CDF $G_{\tilde{u}}(\tilde{u}^\star \mid \tilde{x})$ at all points of evaluation because w has full support on $\mathbb{R}^J$. This argument is not new. It appears in Berry and Haile (2016) and Fox and Gandhi (2016).

Third, we use Masten (2018, Theorem 1) to identify the distribution of γ itself. From the previous step we observe $G_{\tilde{u}}(\tilde{u} \mid \tilde{x})$ for $\tilde{x}$ varying in an open set, which is equivalent to the distribution that Masten (2018, Theorem 1) maintains is observed in a seemingly unrelated regression model.[2] Therefore, $F(\gamma)$ is identified for any J. *Q.E.D.*

The proof uses Masten (2018, Theorem 1). This theorem states conditions for the identification of a distribution of random coefficients in a system of seemingly unrelated regression equations of the form $\tilde{u}_j = \tilde{\beta}'\tilde{x}_j + \epsilon_j$ for equation j. The conditions in the theorem include that $\gamma = (\beta, \epsilon)$ is independent of x, γ has finite absolute moments, and the distribution of γ is uniquely determined by its moments.

4. DISCUSSION

This note uses large support for w and continuous but possibly bounded support for $\tilde{x}$. Both assumptions are prevalent in the previous literature. As the following discussion indicates, the large support for w originates from the discrete choice aspect of the problem, as also found in the binary choice models of Lewbel (2000), Ichimura and Thompson (1998) and Gautier and Kitamura (2013). The continuous but possibly bounded support for $\tilde{x}$ originates in the random coefficients aspect of the problem, as in Ichimura and Thompson (1998), Gautier and Kitamura (2013) and Masten (2018), as well as earlier work on regression models cited by Masten (2018).

4.1. *Discussion of Large Support for w*

This note's identification argument relies heavily on large support of w, meaning w has support $\mathbb{R}^J$. The argument does not use identification at infinity. In multinomial choice, we can precisely define identification at infinity to mean a step of an identification proof that sets $w_j = -\infty$ for $J-1$ inside goods and uses an analysis from binary choice on the resulting pair of an inside good and the outside good. One can inspect the identification proof to see that this type of argument is not used. Indeed, the second step of the proof is explicitly incompatible with identification at infinity, as the vector w is used over its full support $\mathbb{R}^J$.

Some version of a large support condition on w, while not particularly attractive, is necessary to achieve point identification. Consider the special case of our model where there is one inside good ($J = 1$), one outside good and no $\tilde{x}$. This is binary choice, which has been extensively studied in the literature. Then the utility of good 1 is

$$u_1 = \pm 1 w_1 + \epsilon_1$$

and the utility of good 0 is still $u_0 = 0$. We can identify the sign of the common coefficient on w_1 by seeing whether $\Pr(y = 1 \mid w_1)$ is monotonically increasing or decreasing in the scalar w_1. Let $\bar{w}_1 = +w_1$ or $\bar{w}_1 = -w_1$, as appropriate. In this example, $\gamma = \epsilon_1$. We can identify the distribution $F_\epsilon(\epsilon_1)$ at the point of evaluation $\epsilon_1^\star$ as follows

$$F_\epsilon(\epsilon_1^\star) = \Pr(\epsilon_1 \leq \epsilon_1^\star) = \Pr(\epsilon_1 \leq -\bar{w}_1 \mid -\bar{w}_1 = \epsilon_1^\star) = \Pr(y = 0 \mid -\bar{w}_1 = \epsilon_1^\star).$$

If ϵ takes on support on $\mathbb{R}$, then this argument shows that $\bar{w}_1$ must also take on support on all of $\mathbb{R}$ to identify $F_\epsilon(\epsilon_1)$ over its entire support. As $\Pr(y = 0 \mid \bar{w}_1) + \Pr(y = 1 \mid \bar{w}_1) =$

[2] Masten (2018, Theorem 1) directly applies to a model with a good-specific vector of random coefficients β_j for each inside good j, in addition to the good-specific additive error ϵ_j. The typical restriction is used in this paper: $\beta_{j_1} = \beta_{j_2}$ for all inside goods j_1, j_2. This is a special case of Masten (2018, Theorem 1).

1, all available data on conditional choice probabilities is used in the identification argument. This binary choice example is a special case of our model, so it shows that some large support is necessary for point identification. If the large support condition on w_1 does not hold, then $F_\epsilon(\epsilon_1)$ is point identified for a subset of its values.

Keeping the example of binary choice but adding back $\tilde{x}$ so that $u_1 = \pm 1 w_1 + \tilde{\beta}' \tilde{x}_1 + \epsilon_1$, Magnac and Maurin (2007) show that $E\left[\tilde{\beta}\right]$ is point identified with possibly bounded support for w_1 under a tail symmetry condition on F_ϵ. The argument using tail symmetry to identify $E\left[\tilde{\beta}\right]$ in Magnac and Maurin (2007) has not been extended to multinomial choice.

Even if this extension to three or more choices occurred, this would not allow the identification result for seemingly unrelated regression models with random coefficients in Masten (2018, Theorem 1) to be used to identify the higher moments of the random coefficient distribution, as in the proof of the theorem above. Masten's result for regression models requires point identification of the joint distribution of the independent and dependent variables. In the proof of the theorem above, the regression model in the proof for inside good 1 is $\tilde{u}_1 = \tilde{\beta}' \tilde{x}_1 + \epsilon_1$. If w_1 has bounded support, then the CDF of the inside good utility $\tilde{u}_1$ conditional on the vector $\tilde{x}$, $G_{\tilde{u}_1}(\tilde{u}_1 \mid \tilde{x}_1)$, will be identified over a subset of values $\tilde{u}$. The deeper mathematical tool behind the identification results in Masten (2018) as well as Ichimura and Thompson (1998) and Gautier and Kitamura (2013) is the well-known device of Cramér and Wold (1936). Using the Cramér and Wold device for identification requires the joint distribution of the independent and dependent variables and extending this device to the case where the CDF of the dependent variable conditional on the independent variables is identified at only a subset of dependent variable values is an ambitious mathematical task outside the scope of this note.

4.2. *Discussion of Continuous Support for w and $\tilde{x}$*

Consider the previous binary choice example of the outside good and one inside good with utility $u_1 = \pm 1 w_1 + \tilde{\beta}' \tilde{x}_1 + \epsilon_1$. The analysis in this note requires both w_1 and $\tilde{x}_1$ to have continuous support. If w_1 has large and continuous support and only $\tilde{x}_1$ has finite support, the argument in Lewbel (2000) still identifies the conditional distribution $G_{\tilde{u}_1}(\tilde{u}_1 \mid \tilde{x}_1)$ and $E\left[\tilde{\beta}\right]$. However, the results in Masten (2018) and earlier papers he cites do not show point identification of the higher order moments of $\gamma = \left(\tilde{\beta}, \epsilon_1\right)$, as those results rely on $\tilde{x}_1$ having support on a closed superset of an open subset of the real space $\mathbb{R}^{\dim(\tilde{x}_1)}$.

If instead it is w_1 that has finite support in the example of binary choice, then Magnac and Maurin (2008) show that $E\left[\tilde{\beta}\right]$ is only set identified. Nothing is known about identification of the higher order moments of $\gamma = \left(\tilde{\beta}, \epsilon_1\right)$. Point identification of the distribution of γ would seemingly require a version of the Cramér and Wold (1936) device to apply to discrete covariates.

Return to multinomial choice. Say two scalar elements a_1 and a_2 of the long vector $\tilde{x}$ exhibit power dependence if $a_1 = a_2^l$ for a positive integer l. Then the statement of non-identification for power dependence in Masten (2018, Theorem 1) combined with the proof steps above imply that $F(\gamma)$ is not point identified in multinomial choice.

REFERENCES

BERRY, S. T., AND P. A. HAILE (2016): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," Yale University working paper.

CRAMÉR, H., AND H. WOLD (1936): "Some Theorems on Distribution Functions," *Journal of the London Mathematical Society*, 1(4), 290.

FOX, J. T., AND A. GANDHI (2016): "Nonparametric Identification and Estimation of Random Coefficients in Multinomial Choice Models," *RAND Journal of Economics*, 47(1), 118–139.

FOX, J. T., K. KIM, S. RYAN, AND P. BAJARI (2012): "The Random Coefficients Logit Model is Identified," *Journal of Econometrics*, 1662(2), 204–212.

GAUTIER, E., AND Y. KITAMURA (2013): "Nonparametric Estimation in Random Coefficients Binary Choice Models," *Econometrica*, 81(2), 581–607.

ICHIMURA, H., AND T. THOMPSON (1998): "Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution," *Journal of Econometrics*, 86(2), 269–295.

LEWBEL, A. (2000): "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 97(1), 145–177.

——— (2014): "An Overview of the Special Regressor Method," in *Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, ed. by A. Ullah, J. Racine, and L. Su, pp. 38–62. Oxford University Press.

MAGNAC, T., AND E. MAURIN (2007): "Identification and information in monotone binary models," *Journal of Econometrics*, 139(1), 76–104.

——— (2008): "Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data," *The Review of Economic Studies*, 75(3), 835– 864.

MASTEN, M. (2018): "Random coefficients on endogenous variables in simultaneous equations models," *The Review of Economic Studies*, 85(2), 1193–1250.