

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

Jeremy T. Fox
Rice University and NBER

June 2020

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 (2017). 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 semi-parametric 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.

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 .

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

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.

Let $x_i = (x_{i,1}, \dots, x_{i,J})$, $w_i = (w_{i,1}, \dots, w_{i,J})$, $\tilde{x}_i = (\tilde{x}_{i,1}, \dots, \tilde{x}_{i,J})$ and $\epsilon_i = (\epsilon_{i,1}, \dots, \epsilon_{i,J})$ all be thought of as long vectors. Also, define $\gamma_i = (\beta_i, \epsilon_i)$ and think of it as another long vector. Note that γ_i is a heterogeneous parameter vector, not a homogeneous parameter. 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 that the support of x , reordered as $x = (w, \tilde{x})$ is $\mathbb{R}^{\dim(w)} \times \tilde{X}$, where the support of w is the large support $\mathbb{R}^{\dim(w)}$ and the support of the vector \tilde{x} is \tilde{X} , a closed superset of an open subset of the real space $\mathbb{R}^{\dim(\tilde{x})}$.¹ Having support on the product space $\mathbb{R}^{\dim(w)}$ 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} .

Some version of a large support condition, while not particularly attractive, is necessary to achieve 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

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

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

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 $\Pr(y = 1 | w_1)$ 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 ϵ_1^* as follows

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

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 $\Pr(y = 0 | \bar{w}_1) + \Pr(y = 1 | \bar{w}_1) = 1$, 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 identification.

We are now ready to state the main identification theorem.

Theorem 1. *If the support of x is as stated previously, γ is independent of x , γ has finite absolute moments, and 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 (2017, 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 $\Pr(y = j | x)$ and the vector w is a subvector of x . If $\Pr(y = j | x)$ 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 $\bar{w} = (\bar{w}_1, \dots, \bar{w}_J)$.

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

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

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

Third, we use Masten (2017, Theorem 1) to identify the distribution of γ itself. From the previous step we observe $G_{\tilde{u}}(\tilde{u} | \tilde{x})$ for \tilde{x} varying in an open set, which is equivalent to the distribution that Masten (2017, Theorem

1) maintains is observed in a seemingly unrelated regression model.² Therefore, $F(\gamma)$ is identified for any J . \square

This identification argument relies heavily on large support of w , meaning w has support $\mathbb{R}^{\dim(w)}$. 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 an 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 $\mathbb{R}^{\dim(w)}$.

References

- Berry, Steven T. and Philip A. Haile**, “Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers,” 2016. Yale University working paper.
- Fox, Jeremy T. and Amit Gandhi**, “Nonparametric Identification and Estimation of Random Coefficients in Multinomial Choice Models,” *RAND Journal of Economics*, 2016, 47 (1), 118–139.
- , **Kyoo il Kim, Stephen Ryan, and Patrick Bajari**, “The Random Coefficients Logit Model is Identified,” *Journal of Econometrics*, 2012, 1662 (2), 204–212.
- Gautier, E. and Y. Kitamura**, “Nonparametric Estimation in Random Coefficients Binary Choice Models,” *Econometrica*, 2013, 81 (2), 581–607.
- Ichimura, H. and TS Thompson**, “Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution,” *Journal of Econometrics*, 1998, 86 (2), 269–295.
- Lewbel, Arthur**, “Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables,” *Journal of Econometrics*, 2000, 97 (1), 145–177.
- , “An Overview of the Special Regressor Method,” in “Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics,” Oxford University Press, 2014, pp. 38–62.
- Masten, Matthew**, “Random coefficients on endogenous variables in simultaneous equations models,” *The Review of Economic Studies*, 2017, 85 (2), 1193–1250.

²Masten (2017, Theorem 1) is stated for a seemingly unrelated regression model with two equations but this was just an editorial decision on his part and the argument applies to any number of equations. Masten (2017, 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 (2017, Theorem 1).