Journal of Econometrics 195 (2016) 236–254

Contents lists available at ScienceDirect

Journal of Econometrics

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

A simple nonparametric approach to estimating the distribution of random coefficients in structural models

Jeremy T. Fox [a,b,*], Kyoo il Kim [c,**], Chenyu Yang [d]

[a] Rice University, United States
[b] NBER, United States
[c] Michigan State University, United States
[d] Simon Business School, University of Rochester, United States

ARTICLE INFO

Article history:
Received 27 July 2011
Received in revised form
8 July 2015
Accepted 15 May 2016
Available online 19 September 2016

JEL classification:
C21
C22
C25

Keywords:
Random coefficients
Mixtures
Discrete choices
Dynamic programming
Sieve estimation

ABSTRACT

We explore least squares and likelihood nonparametric mixtures estimators of the joint distribution of random coefficients in structural models. The estimators fix a grid of heterogeneous parameters and estimate only the weights on the grid points, an approach that is computationally attractive compared to alternative nonparametric estimators. We provide conditions under which the estimated distribution function converges to the true distribution in the weak topology on the space of distributions. We verify most of the consistency conditions for three discrete choice models. We also derive the convergence rates of the least squares nonparametric mixtures estimator under additional restrictions. We perform a Monte Carlo study on a dynamic programming model.

1. Introduction

Economic researchers often work with models where the parameters are heterogeneous across the population. A classic example is that consumers may have heterogeneous preferences over a set of product characteristics in an industry with differentiated products. These heterogeneous parameters are often known as random coefficients. When working with cross sectional data, the goal is often to estimate the distribution of heterogeneous parameters. Our paper establishes the consistency and rates of convergence of "fixed grid" nonparametric estimators for a distribution of heterogeneous parameters due to Bajari et al. (2007), Train (2008, Section 6), Fox et al. (2011), and Koenker and Mizera (2014). These estimators are computationally simpler than some alternatives. We use FKRB to refer to Fox et al. (2011).

We estimate the distribution of heterogeneous parameters $F(\beta)$ in the model

$$P_j(x) = \int g_j(x, \beta)\, dF(\beta), \tag{1}$$

where j is the index of the jth out of J finite values of the outcome y, x is a vector of observed explanatory variables, β is the vector of heterogeneous parameters, and $g_j(x, \beta)$ is the probability that the jth outcome occurs for an observation with heterogeneous parameters β and explanatory variables x. Given this structure, $P_j(x)$ is the cross sectional probability of observing the jth outcome when the explanatory variables are x. The researcher picks $g_j(x, \beta)$ as the underlying model, has an i.i.d. sample of N observations (y_i, x_i), and wishes to estimate $F(\beta)$. As F is only restricted to be a valid CDF, the mixture model (1) is nonparametric.

The unknown distribution $F(\beta)$ enters (1) linearly. The estimators we analyze exploit linearity and achieve a computationally simpler estimator than some alternatives. All the fixed grid estimators divide the support of the vector β into a finite and known

* Correspondence to: Department of Economics, MS-22 Rice University, P.O. Box 1892, Houston, TX 77251, United States.
** Correspondence to: MSU Department of Economics, Marshall-Adams Hall, 486 W Circle Dr Rm 110, East Lansing MI 48824, United States.
E-mail addresses: jeremyfox@gmail.com (J.T. Fox), kyookim@msu.edu (K.i. Kim).

http://dx.doi.org/10.1016/j.jeconom.2016.05.018
0304-4076/

grid of vectors $\beta^1, \ldots, \beta^R$. Computationally, the unknown parameters are the weights $\theta^1, \ldots, \theta^R$ on the R grid points. These can be estimated using a least squares or likelihood criterion with the constraints that each $\theta^r \geq 0$ and that $\sum_{r=1}^{R} \theta^r = 1$. The estimator of the distribution $F(\beta)$ with N observations and R grid points becomes

$$\hat{F}_N(\beta) = \sum_{r=1}^{R} \hat{\theta}^r 1\left[\beta^r \leq \beta\right],$$

where $\hat{\theta}^r$'s denote estimated weights and $1[\beta^r \leq \beta]$ is equal to 1 when $\beta^r \leq \beta$. Computationally, the least squares and likelihood constrained optimization problems are globally convex and concave, respectively. Particular numerical algorithms are guaranteed to converge to a global optimum.

FKRB discuss the advantages of this estimator for complex structural models, like dynamic programming models with heterogeneous parameters. In this respect, fixed grid estimators share some computational advantages with the parametric approach in Ackerberg (2009). Our Monte Carlo study in an online appendix is to a discrete choice, dynamic programming model.

FKRB and other previous analyses assume that the R grid points used in a finite sample are indeed the true grid points that contain the finite support of the true $F_0(\beta)$. Thus, the true distribution $F_0(\beta)$ is assumed to be known up to a finite number of weights $\theta^1, \ldots, \theta^R$. As economists often lack convincing economic rationales to pick one set of grid points over another, assuming that the researcher knows the true distribution up to finite weights is unrealistic.

Instead of assuming that the distribution is known up to weights $\theta^1, \ldots, \theta^R$, this paper requires the true distribution $F_0(\beta)$ to satisfy much weaker restrictions. In particular, the true $F_0(\beta)$ can have any of continuous, discrete and mixed continuous and discrete supports. The prior approaches are parametric as the true weights $\theta^1, \ldots, \theta^R$ lie in a finite-dimensional subset of a real space. Here, the approach is nonparametric as the true $F_0(\beta)$ is known to lie only in the infinite-dimensional space of multivariate distributions on the space of heterogeneous parameters β.

In a finite sample of N observations, our estimators are still implemented by choosing a fixed grid of points $\theta^1, \ldots, \theta^R$, ideally to trade off bias and variance in the estimate $\hat{F}_N(\beta)$. We, however, recognize that as the sample increases, R and thus the fineness of the grid of points should also increase in order to reduce the bias in the approximation of $F(\beta)$. We write $R(N)$ to emphasize that the number of grid points (and implicitly the grid of points itself) is now a function of the sample size. The main theorem in our paper is that, under restrictions on the economic model and an appropriate choice of $R(N)$, our least squares and likelihood estimators $\hat{F}_N(\beta)$ converge to the true $F_0(\beta)$ as $N \to \infty$, in a function space. We use the Lévy–Prokhorov metric, a common metrization of the weak topology on the space of multivariate distributions.

We recognize that the nonparametric versions of our estimators are special cases of sieve estimators (Chen, 2007). Sieve estimators estimate functions by increasing the flexibility of the approximating class used for estimation as the sample size increases. A sieve estimator for a smooth function might use an approximating class defined by a Fourier series, for example. As we are motivated by practical considerations in empirical work, our estimators' choice of basis, a finite grid of points, is justified by the estimators' computational simplicity. Further and unlike a typical sieve estimator, we need to constrain our estimated functions to be valid distribution functions. Our constrained least squares and likelihood approaches are both computationally simple and ensure that the estimated CDFs satisfy the theoretical properties of a valid CDF.

Because our estimators are sieve estimators, we prove their consistency by satisfying high-level conditions for the consistency of sieve extremum estimators, as given in an appendix lemma in Chen and Pouzo (2012). We repeat this lemma and its proof in our paper so our consistency proof is self-contained. Our fixed grid estimators are not a special case of the two-step sieve estimators explored using lower-level conditions in the main text of Chen and Pouzo.[1]

We prove the consistency of our estimators for the distribution of heterogeneous parameters, in function space under the weak topology. We present separate theorems for mixtures of discrete grid points and mixtures of continuous densities with a grid of points over the parameters of each density. The theorem for the mixture of grid points requires the heterogeneous parameters to lie in a, not necessarily known, compact set. The theorem for a mixture of continuous densities allows for unbounded support of the heterogeneous parameters. Our consistency theorems are not specific to the economic model being estimated.

We provide the rate of convergences for a subset of the models handled by our consistency theorem, namely those that are differentiable in the heterogeneous parameters, which include the random coefficients logit model. The convergence rates, the asymptotic estimation error bounds, consist of two terms: the bias and the variance. While obtaining the variance term is rather standard in the sieve estimation literature, deriving the bias term depends on the specific approximation methods (e.g., power series or splines). Because our use of approximating functions is new in the sieve estimation literature, deriving the bias term is not trivial. We provide the bias term, which is the smallest possible approximation error of the true function using sieves for the class of models we consider.

Our rate of convergence results highlight an important practical issue with any nonparametric estimator: there is a curse of dimensionality in the dimension of the heterogeneous parameters. Larger sample sizes will be needed if the vector of heterogeneous parameters has more elements. Further, the rate results indicate that our baseline estimator is not practical when there are a large number of heterogeneous parameters. In high dimensional settings, we suggest allowing heterogeneous parameters on only a subset of explanatory variables and estimating homogeneous parameters on the remaining explanatory variables. We extend our consistency result to models where some parameters are homogeneous. However, including homogeneous parameters requires nonlinear optimization, which loses some of the computational advantages of our estimators.

We provide a Monte Carlo study in an online appendix. We estimate a dynamic programming, discrete choice model, adding heterogeneous parameters to the framework of Rust (1987). The dynamic programming problem must be solved once for each realization of the heterogeneous parameters. We present results for both the fixed grid likelihood and least squares estimators as well as, for comparison, a likelihood estimator where we estimate both the grid of points and the weights on those points. We show that our fixed grid estimators have superior speed but inferior statistical accuracy compared to the more usual approach of estimating a flexible grid.

The outline of our paper is as follows. Section 2 presents three examples of discrete choice mixture models. Section 3 introduces

[1] Note that under the Lévy–Prokhorov metric on the space of multivariate distributions, the problem of optimizing the population objective function over the space of distributions turns out to be well posed under the definition of Chen (2007). Thus, our method does not rely on a sieve space to regularize the estimation problem to address the ill-posed inverse problem, as much of the sieve literature focuses on.