Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

Discussion of "Simple Estimators for Invertible Index Models" by H. Ahn, H. Ichimura, J. Powell, and P. Ruud

S. Khan & E. Tamer

To cite this article: S. Khan & E. Tamer (2018) Discussion of "Simple Estimators for Invertible Index Models" by H. Ahn, H. Ichimura, J. Powell, and P. Ruud , Journal of Business & Economic Statistics, 36:1, 11-15, DOI: 10.1080/07350015.2017.1392312

To link to this article: https://doi.org/10.1080/07350015.2017.1392312

Discussion of "Simple Estimators for Invertible Index Models" by H. Ahn, H. Ichimura, J. Powell, and P. Ruud

S. KHAN
Department of Economics, Boston College, Chestnut Hill, MA (*shakeeb.khan@bc.edu*)

E. TAMER
Department of Economics, Harvard University, Cambridge, MA (*elietamer@fas.harvard.edu*)

This is an interesting article that considers the question of inference on unknown linear index coefficients in a general class of models where reduced form parameters are invertible function of one or more linear index. Interpretable sufficient conditions such as monotonicity and or smoothness for the invertibility condition are provided. The results generalize some work in the previous literature by allowing the number of reduced form parameters to exceed the number of indices. The identification and estimation expand on the approach taken in previous work by the authors. Examples include Ahn, Powell, and Ichimura (2004) for monotone single-index regression models to a multi-index setting and extended by Blundell and Powell (2004) and Powell and Ruud (2008) to models with endogenous regressors and multinomial response, respectively. A key property of the inference approach taken is that the estimator of the unknown index coefficients (up to scale) is computationally simple to obtain (relative to other estimators in the literature) in that it is closed form. Specifically, unifying an approach for all models considered in this article, the authors propose an estimator, which is the eigenvector of a matrix (defined in terms of a preliminary estimator of the reduced form parameters) corresponding to its smallest eigenvalue. Under suitable conditions, the proposed estimator is shown to be root-n-consistent and asymptotically normal.

KEYWORDS: Closed form; Invertibility; Linear index; Robustness.

1. ASSESSMENT

This is a very good and interesting article for a number of reasons. It adds much needed identification results and a new estimator for a class of models that have not been widely considered in the semiparametric literature. This especially pertains to multinomial choice models under general conditions. Multinomial choice models are extremely relevant in fields such as industrial organization, labor, and development and there has not been as much work on semiparametric estimation of such models. Methods that are available, and used in practice are often for parametric models, such as multinomial probit and multinomial logit along with parametric random coefficients. But parametric models are problematic for several reasons. Most rely on questionable parametric assumptions, some such as the probit model are hard to compute, and more importantly, logit-based models suffer from the IIA property that imposes unreasonably strong substitution patterns. The semiparametric estimator proposed in this article addresses all of these problems. It does not rely on distributional assumptions, nor does it require the independence within choice requirement that leads to IIA. Furthermore, it is computationally very tractable, so practitioners can implement the new methods. So overall, we think this article does make a needed contribution to the literature.

Having said that, we explain here why we are concerned with the proposed approach for two main reasons.

1.1 Robustness to Heteroscedasticity

Our first concern is the assumed statistical relationship between observed and unobserved variables. The independence/index type assumption severely restricts the type of heteroscedasticity allowed for, and this type of misspecification will result in the proposed estimator being inconsistent. In nonlinear models such as binary choice and censored regression models it is well known that ignoring heteroscedasticity can be a much more severe problem than distributional misspecification. That appears to be the case for multinomial models as well. We demonstrate this by conducting the following simulation study of a trivariate multinomial profile model, with three choices corresponding to observed outcome variables as 0,1,2. As a standard normalization we set the latent utility of the first choice

Journal of Business & Economic Statistics
January 2018, Vol. 36, No. 1
DOI: 10.1080/07350015.2017.1392312
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jbes*.

to be identically 0. For the other two choices, we generated latent utilities for each individual from the model:

$$u_{ij} = x'_{ij}\beta_0 + \epsilon_{ij} \quad j = 1, 2,$$

where $(\epsilon_{i1}, \epsilon_{i2})$ is drawn from a bivariate normal distribution with means 0, scales 1, and correlation 0.5. For the first design, x_{i1}, x_{i2} was drawn independently of ϵ_{ij} with the same bivariate normal distribution. The unknown parameter vector β_0 is set at $(1, 1)$. Observed outcomes y_i corresponded to utility maximization. From this setup, we generated outcomes of sizes 100, 200, 400, and 800 and replicated this dataset 500 times.

For the proposed estimator, we report mean bias and RMSE. To implement the estimator there were two tuning parameters to select—the first stage nonparametric estimator and the second stage matching. For each one we used a Silverman rule of thumb.

The results clearly demonstrate consistency of the new estimator, though there are finite sample problems at 100 observations. This is to be expected with any method that relies on nonparametric methods.

The second design introduces misspecification in the form of heteroscedasticity. We model this by premultiplying the error vector $\epsilon_i \equiv \epsilon_{1i}, \epsilon_{2i}$ by the 2×2 matrix with 0.5 for off diagonal elements and diagonal elements consisting of (x_{i2}, x_{i1}). The statistics demonstrate the poor performance of the proposed estimator in this setting. The biases can be large and not decline with the sample size. We are wondering whether the simple approach of the article can be altered to allow for some heteroscedasticity, while still allowing for correlation in observables across choices.

1.2 Robustness to Discrete Support

A second important issue that we mention is the required conditions on the support of the covariates. Specifically, point identification of the proposed models requires that at least one of the regressors be continuously distributed. In one sense such a condition is to be expected in semiparametric models if point identification is the goal, and is often required for a large class of univariate or multivariate "distribution free" models. Interestingly, this does demonstrate that semiparametric models do not actually nest parametric models such as multinomial probit or multinomial logit under point identification.

An inference procedure based on sufficient conditions for point identification is said to be *point robust* (to support conditions) if said inference procedure delivers a nontrivial set (by nontrivial set we mean an informative set) (or the identified set) when these support conditions fail to hold in one's data. We view this point robust property in the context of support conditions as attractive in many settings as continuity of support is an idealization. In a binary choice model, the maximum score (Manski, 1975) approach to inference is point robust since even when all the regressors are discrete, the maximum score objective function is maximized on a nontrivial set. (The argmax of the maximum score objective function coincides with the identified set for the model in (1.1) based on the usual conditional median restriction only if $P_x\{x : x\beta_0 = 0\} = 0$. See Komarova (2013).) This maximum score inference approach is *adaptive* to the support conditions that are required for point identification. On the other hand, and on the other end of the spectrum, there are inference approaches whereby when the support conditions do not hold, no information can be gained about the parameter of interest (even if the identified set in these cases is finite).

The approach in this article, based on matching when choice probabilities are close, is *not* point robust to the support conditions. For simplicity, we illustrate this point with a binary choice model of the following form:

$$y_i = I[x_{1i}\beta_0 + x_{2i} + \epsilon_i > 0], \tag{1.1}$$

where the two observed regressors (x_{1i}, x_{2i}) are binary $1/-1$ with probability $\frac{1}{2}$ distributed independently of each other as well as independently of the disturbance term ϵ_i. The parameter of interest is the scalar β_0 and the coefficient of x_{2i} was set to 1 as a scale normalization.

A simple matching estimator for β_0 was proposed in Ahn, Powell, and Ichimura (2004). For the problem at hand it can expressed as minimizing the following least-square objective function, with respect to b:

$$\frac{1}{n(n-1)} \sum_{i \neq j} \hat{\omega}_{ij}(x_{2i} - x_{2j} - (x_{1i} - x_{1j})b)^2,$$

where $\hat{\omega}_{ij}$ is an estimated weight function, $\hat{\omega}_{ij} \approx I[P_i = P_j]$ and $P_i = P(y_i = 1|x_i)$, $P_j = P(y_j = 1|x_j)$. Since the distribution of ϵ is unknown, P_i, P_j are also unknown and need to estimate in a preliminary stage.

Clearly, this estimator is not designed for cases with discrete covariates. This matching estimator is also not point robust. For example, assume that $P(y = 1|1, 1) > P(y = 1|-1, 1) > P(y = 1|-1, 1) > P(y = 1|-1, -1)$ which leads to $[0, 1]$ being the identified set (based on a rank order property). In particular, let $P(y = 1|1, 1) = 0.5$, $P(y = 1|-1, 1) = 0.25$, $P(y = 1|1, -1) = 0.2$, and $P(y = 1|-1, -1) = 0.05$ (there are other values that will work). It is not immediately clear how one would match in this case, but say we match observations with choice probabilities that are within 0.2. This means that observations with covariate values $(-1, 1)$ are matched with ones with $(1, -1)$ and observations $(1, -1)$ are matched with $(-1, -1)$. We can see that the limit objective function (it is simple to show that this limit will be $4b^2(0.05)(0.2) +$

Table 1. Homoscedastic design

Sample size	Bias	RMSE
100	0.1250	0.6611
200	0.0288	0.3081
400	0.0264	0.2209
800	0.0177	0.1439

Table 2. Heteroscedastic design

Sample size	Bias	RMSE
100	0.6524	8.3992
200	– 0.0597	15.7435
400	– 0.5653	4.1657
800	0.1657	3.7240

Table 3. Discrete design

	0.01		0.1		1.0	
Sample size	Bias	RMSE	Bias	RMSE	Bias	RMSE
100	– 0.8600	0.8617	– 0.8619	0.8633	– 0.8712	0.8726
200	– 0.8595	0.8603	– 0.8631	0.8639	– 0.8704	0.8711
400	– 0.8642	0.8646	– 0.8678	0.8682	– 0.8735	0.8738
800	– 0.8672	0.8675	– 0.8691	0.8693	– 0.8759	0.8760

Table 4. Cross-sectional estimator design 1

N	Mean	Median	RMSE
50	0.1656	0.1900	0.5366
100	0.0173	0.0100	0.4656
200	0.0410	0.0500	0.3389
400	0.0001	– 0.0010	0.2544

$4(b-1)^2(0.2)(0.25)$, which is minimized at $b = 0.05/(0.04 - 0.05) = -1.25$. There is nothing special about this example, and various combinations of probabilities and matching rules can yield parameters outside the identified set) is minimized at a point that is outside the identified set.

The question then becomes, can one obtain a point robust estimator in this setting? Consider a maximum rank type estimator as proposed by Han (1987):

$$\hat{\beta} = \arg\max_b \frac{1}{n(n-1)} \sum_{i \neq j} I[y_i > y_j] I[x_i'b > x_j'b],$$

where here $x_i \equiv (x_{1i}, x_{2i})$.

We note this estimator is defined as an objective function and does not have a closed form. Consequently, optimization methods, such as simulated annealing are used to construct the estimator. In that sense it is not as “simple” as the proposed estimator. On the other hand, given our sample design involving discrete regressors, the MRC will limit to an objective that will maximized on the identified set. So, the MRC in this case is point robust.

1.3 Point Robustness in the Multinomial Choice Model

Now, let us consider a particular version of the multinomial choice model studied in the article where we will suggest a point robust rank-based estimation approach. In particular, consider the three choice 0, 1, 2 multinomial choice model with the following utilities:

$$y_0^* = 0$$
$$y_1^* = x_1'\beta_0 + \epsilon_1$$
$$y_2^* = x_2'\beta_0 + \epsilon_2.$$

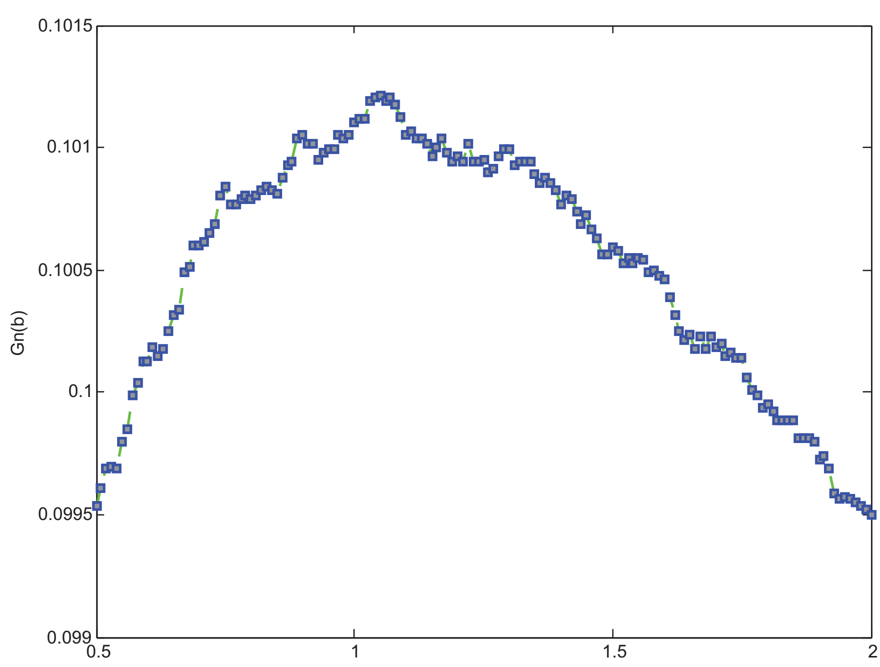

Figure 1. Objective function for cross-sectional model 1.

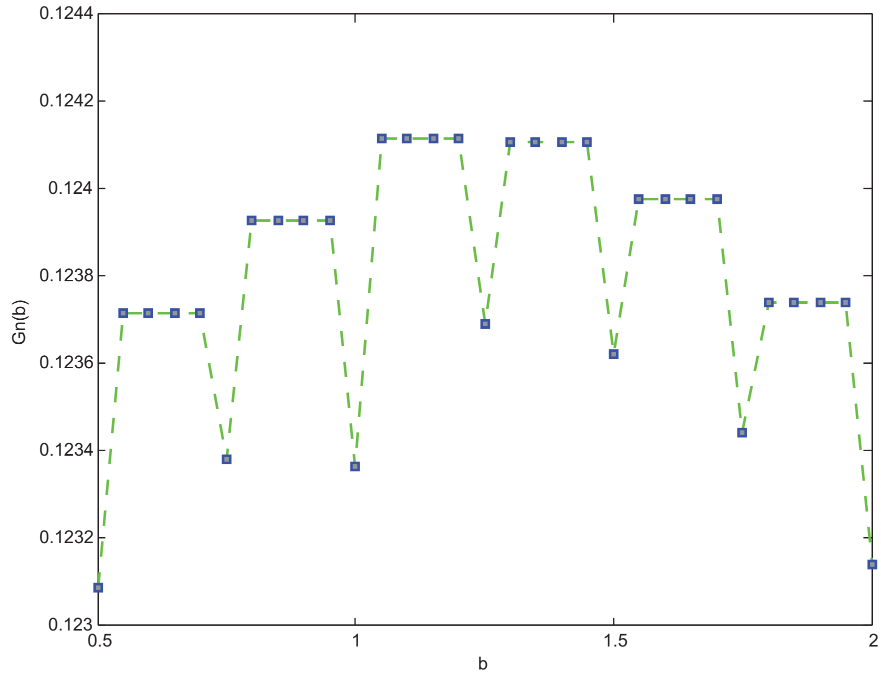

Figure 2. Objective function for cross-sectional model 2.

So by utility maximization,

$$y_0 = I[y_0^* > y_1^*, y_0^* > y_2^*]$$
$$y_1 = I[y_1^* > y_0^*, y_1^* > y_2^*]$$
$$y_2 = I[y_2^* > y_0^*, y_0^* > y_1^*].$$

Notice, for example, that, conditional on $(x_2 = \mathbf{x_2})$ (say), the choice probability for choice 1

$$P(\epsilon_1 + x_1'\beta_0 \geq 0; \epsilon_1 + x_1'\beta_0 \geq \epsilon_2 + \mathbf{x_2}'\beta_0)$$

is *increasing in* $x_1'\beta_0$ and so we can use a maximum rank correlation estimator to get β_0 local to x_2. This type of conditional rank estimator has better robustness properties to failure of the support conditions than the matching estimator. But, of course, such an estimator is tedious to compute. The table below indicates favorable finite sample properties of our proposed conditional MRC estimator, which, assuming all the regressors for the second choice utility function are discrete. The MRC procedure here only assigns positive weight to pairs of observation whose regressors match for the section choice utility match up.

Tables 1–4 and Figures 1–2 demonstrate the relative finite sample properties of the proposed rank-based procedure. We consider two designs—in the first, there are three choices; y_0, y_1, y_2. Corresponding latent utilities are

$$y_0^* = 0, \quad y_1^* = x_1'\beta_0 + \epsilon_1, \quad y_2^* = x_2'\beta_0 + \epsilon_2.$$

x_1, x_2 are each two-dimensional, as is β_0. For scale normalization, the first component of β_0 is set to 1, and the second component of β_0 was set to 1.25. For x_1 the first component was standard normal, the second component to be Bernoulli, with $p = 0.5$.

For x_2 each of the two components had a Bernoulli distribution, with $p = 0.5\, \epsilon_1, \epsilon_2$ were bivariate normal, mean 0, variance 1, correlation 0.5.

For design 2, the distributions is the same as in design 1 with one change. The first component of x_1 was discrete instead of standard normal. Specifically it took the values $-2, -1, 1, 2$ each with probability 0.25.

Note that under design 1, the regression coefficient is point identified. The table below reports mean bias, median bias, and RMSE for samples sizes of 50, 100, 200, 400 using 401 replications. We also plot the objective function for one draw of 800 observations. As these results indicate our proposed procedure is constant with RMSE declining at the parametric (root-n rate). The objective function appears to be approximately globally concave.

Under design 2, the regression coefficients are not point identified. The figure plots the objective function for one draw of 800 observations. This illustrates the advantage of our procedure, as even though point identification is lost in the model, the rank procedure produces a set estimator of (1, 1.5), which contains the true parameter value for $\beta_0 = 1.25$.

2. CONCLUSION

The article discussed makes progress on a central problem in econometrics. It provides closed-form estimators for parameters in multinomial choice models with general error structures. In our comment, we highlight an aspect of this matching estimator that may not be desirable when data are heteroscedastic or discrete. A rank-based procedure may be more desirable in discrete designs.

[Received June 2017. Revised September 2017.]

REFERENCES

Ahn, H., Powell, J., and Ichimura, H. (2004), “Simple Estimators for Monotone Index Models,” UC Berkeley Working Paper. [11,12]

Blundell, R. W., and Powell, J. L. (2004), “Endogeneity in Semiparametric Binary Response Models,” *The Review of Economic Studies*, 71, 655–679. [11]

Han, A. K. (1987), “Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator,” *Journal of Econometrics*, 35, 303–316. [13]

Komarova, T. (2013), “Binary Choice Models With Discrete Regressors: Identification and Misspecification,” *Journal of Econometrics*, 177, 14–33. [12]

Manski, C. F. (1975), “Maximum Score Estimation of the Stochastic Utility Model of Choice,” *Journal of Econometrics*, 3, 205–228. [12]

Powell, J., and Ruud, P. (2008), “Simple Estimators for Semiparametric Multinomial Choice Models,” UC Berkeley Working Paper. [11]