Identification of Joint Distributions with Applications to the Roy Model and Competing Auctions

Christopher P. Adams*
Federal Trade Commission
Email: cadams@ftc.gov

June 3, 2013

Abstract

The paper presents conditions under which Sklar's copula formula can be used to approximate the full joint distribution from observed marginal distributions and a partially observed joint distribution. The first part presents conditions under which the parameters of a finite polynomial copula function are uniquely determined. The second part presents conditions under which the joint probability of interest can be arbitrarily approximated by a known polynomial of the observed marginal probabilities. The result is applied to the extended Roy model providing an alternative approach to non-parametric identification. The result is also applied to identifying value distribution for differentiated goods using auction platform data.

*Thanks to Matthew Chesnes, Eric French, Ian Gale, Bruce Hansen, James Heckman, Hiro Kasahara, Kyoo il Kim, Greg Lewis, Rob McMillan, Chris Metcalf, Byoung Park, Katja Seim, Art Shneyerov, Nathan Wilson, Jeff Wooldridge, Robert Zeithammer, David Zimmer, participants at the 2009 IIOC, Wisconsin, Minnesota, DOJ, Cornell Johnson School, Duke, as well as the many others that I have talked to about this question. I am also grateful to an anonymous reviewer who pointed to the connection between the auction problem and the Roy model. The views expressed in this article are those of the author and do not necessarily reflect those of the Federal Trade Commission. All remaining errors are my own.

1 Introduction

Patients that receive Treatment A have an observed outcome for that treatment as well as an unobserved or "latent" outcome for Treatment B. Given this pair of outcomes we would like a policy that assigns treatments based on which treatment will induce the better outcome. If a patient has a poor outcome for Treatment A, what would their outcome be on the alternative treatment? If a patient does well on Treatment A, would they do just as well on Treatment B? To answer these questions we would like to know the joint distribution across the observed outcome and the latent outcome. Of course we cannot observe the outcome in the treatment that the patient did not receive. We can, however, infer information about what is not observed from what is observed. Information about the option not chosen may be inferred through "revealed preference". If we observe a consumer choose an apple over an orange we may assume that the consumer prefers the apple to the orange. Similarly if the patient choose Treatment A, their outcome on Treatment B may have been worse than their outcome on Treatment A. Exactly what information may be inferred about the treatment not chosen depends upon the data available and the behavioral assumptions that the econometrician is willing to make.

In the extended Roy model the decision maker chooses between some S treatment options. The decision maker chooses the treatment based on the outcome they will receive in each particular treatment and other costs and benefits associated with each treatment choice. The observed outcome, the observed choice and observed information about other costs and benefits of the choices allow the joint distribution of outcomes to be partially identified. Importantly, the model does not allow unobservable characteristics of either the treatment or the decision maker to determine treatment assignment. This restriction may limit the applicability of the results presented.

The paper also considers the situation where bidders choose between competing auctions for similar products. The bidder is assumed to know her own valuations for competing products and have information about the distribution of prices for the auctions. She chooses the auction with the highest expected return. The paper adapts orders statistic reasoning to this choice setting and shows that the joint distribution of valuations over the differentiated products can be partially identified from observed prices.

Sklar's theorem states that there exists a mapping between marginal distributions and joint distributions. Under continuity assumptions, this copula function is unique. Therefore, if we know the marginal distributions and the we know the copula function then we know the joint distribution. The question of interest is under what circumstances the copula function is identified from observing the marginal distributions and partially observing the joint distribution.

Some what confusingly the paper presents two notions of identification corresponding to the two parts of the main result. The first notion asks, given assumptions, when is a finite set of parameters uniquely determined? The first result presents conditions under which the vector of coefficients of a finite polynomial that maps between the marginal distributions and the observed joint probabilities is uniquely determined. The result is analogous to the identification condition for OLS. The second notion asks under what assumptions is the probability of interest equal to a known function of observed probabilities. The second part of the result presents conditions under which any joint probability can be approximated by a known finite polynomial of observed marginal probabilities.

In the Monte Carlo results a quadratic polynomial is estimated with little variation in observed prices. However the quadratic polynomial induces probability estimates which differ from the actual probabilities by an average of 0.0026.[1] If the amount of price variation is increased such that a 4th order polynomial can be estimated, then the approximation error is reduced by half.

This paper contributes to the literature on non-parametric identification of the joint distribution of treatment outcomes in the extended Roy model. Heckman and Honore (1990) present a number of results for identification in the two sector Roy model. The authors show the joint distribution is

[1] To be clear a probability is a number between 0 and 1.

partially identified for a given set of prices. The authors further show that if we observe a large enough set of prices we can non-parameterically identify the full joint distribution. This result has been generalized to other utility representations and other notions of prices (see for example Heckman and Vytlacil (2007), French and Taber (2011), and Park (2012)). All these results have two steps. In the first step, it is shown that the "indifference curve" is identified for a given set of observable characteristics of the market. This is the locus of points in the outcome space such that the decision maker is indifferent between receiving the outcome in each of the two sectors. This step identifies a mapping between the observed outcome in the chosen sector and the unobserved outcome in the alternative sector. Given this mapping it is possible to calculate joint probabilities from the observed distribution of outcomes in each of the sectors. The second step shows that if there is enough variation in some observable characteristic of the market, usually called an "instrumental variable," the complete joint distribution can be mapped out by the observed joint probabilities.

The results presented in this paper provide an alternative to the second step. Instead of requiring enough variation in the instrumental variable to move the indifference curve through the whole outcome space, the result requires only enough variation to identify coefficients of the finite polynomial copula function. Monte Carlo analysis suggest that the required amount of variation is not large relative to the accuracy of the approximation. The price of requiring less variation in observed market characteristics is requiring identification of the marginal distributions. These may be available from other data sets such as randomized control trials or they may be available through an "identification at infinity" argument. That is, there may be some group of patients who are forced to choose Treatment A and some group who are forced to choose Treatment B, given observed characteristics.

A limitation of the approach presented here and used in the literature mentioned above is that it requires that the econometrician observe all information used by the decision maker to select the treatment. The generalized Roy model allows the econometrician to observe only a subset of the signals used by the decision maker (Heckman and Vytlacil (2007)). It is unclear whether complete non-parametric identification is possible in the generalized Roy model. Given this, Manski and co-authors advocate partial identification (see Manski (2013) for an overview). It is an open question whether such assumptions provide empirical content above and beyond the Fréchet-Hoeffding bounds on the copula function. Henry and Mourifie (2012) points out that if we are able to identify the marginal distributions then in general the Fréchet-Hoeffding bounds can be used to bound the copula function and thus the joint distribution (by Sklar's theorem). Moreover, relatively weak Roy model style assumptions can further sharpen the bounds on the joint distribution.[2]

The copula function has a sordid history. David Li proposed a parametric copula function to estimate the joint distribution over default rates that used information on the marginal distributions and a correlation estimate (Li (2000)). Li's formula became a mainstay in pricing CDOs and is now blamed for the financial crisis of 2008 and 2009 (Salmon (2009)).[3] While there is a literature on using various approximations of the copula function to estimate the joint distribution (Trivedi and Zimmer (2007)), there does not seem to be a lot of work where the copula is used as part of the identification procedure itself. A notable exception is Brendstrup and Paarsch (2007) who consider a case where bidders bid in two auctions for two differentiated goods and their utility is additively separable over the two items (fish in their case). Brendstrup and Paarsch (2007) assume that the copula has certain properties, in particular it is assumed to be from the family of Archimedean copulas.

For the most part the empirical auction literature has assumed that auctions occur in isolation and that a series of observed prices or bid vectors can be used to estimate the underlying one-dimensional value distribution. This may be a reasonable assumption in some situations possibly including timber auctions. However, it does not seem particularly reasonable when

[2] In Henry and Mourifie (2012), the Roy model assumption is equivalent to selecting into the treatment that first-order stochastically dominates.

[3] Zimmer (2012) shows that the Gaussian copula proposed by Li predicts greater independence between housing prices across states than the data warrants.

modeling demand on an auction platform like eBay. A notable exception is the literature analyzing bidding dynamics across auctions (Backus and Lewis (2012), Zeithammer (2006), Jofre-Bonet and Pesendorfer (2003)). There has been little empirical work explicitly account for and using auction choice in the estimation procedure.[4]

The paper proceeds as follows. Section 2 presents the general identification result that applies Sklar's theorem to the problem of identifying the joint distribution when the marginal distributions are known and the joint distribution is partially known. Section 3 presents the identification results for the extended Roy model. Section 4 presents the identification results for a model of private value sealed bid competing auctions. Section 5 concludes.

2 Identification of the Joint Distribution

The section considers a situation where the econometrician is interested in determining the joint distribution $F(Y)$ for all $Y \in \mathcal{Y}$ but only knows the marginal distributions $F_s(Y_s)$ for all $Y_s \in \mathcal{Y}_s$ and each $s \in \{1, ..., S\} = \mathcal{S}$ and joint probabilities for a subset of the domain ($F(Y^*)$ for all $Y^* \in \mathcal{Y}^* \subset \mathcal{Y}$). Under what circumstances is this enough information to identify the joint distribution?

Assumption 1 $\{Y_1, ..., Y_S\} \sim F$ *where* $Y_s \in \mathcal{Y}_s \subset \Re$. $F : \mathcal{Y} \to [0, 1]$ *where* $\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_S$ *and* F *is continuous.*

Assumption (1) states that the joint distribution F is a continuous mapping from the S-dimensional reals to $[0, 1]$. Continuity is important for the results presented below. Hopefully, it not so restrictive as to limit the applicability of those results.

Definition (1) defines the set $\mathcal{Y}^*$ as the set on which the joint distribution is observed.

Definition 1 $\mathcal{Y}^* \subset \mathcal{Y} \subset \Re^S$ *is such that*

[4] Backus and Lewis (2012) develops on ideas presented in earlier drafts of this paper and explicitly accounts for auction choice.

1. *$\mathcal{Y}^* = \mathcal{Y}_1^* \times ... \times \mathcal{Y}_S^*$*

2. *$F(Y^*)$ is observed for all $Y^* \in \mathcal{Y}^*$*

2.1 Sklar's Theorem

The following theorem is due to Sklar. It states that there exists a unique mapping between the joint distribution and the marginal distributions.

Theorem 1 *(Sklar (1959)) Let F be a S-dimensional distribution with marginals $F_1, ..., F_S$. Then there exists a S-dimensional copula C such that for all $(Y_1, ..., Y_S) \in \mathcal{Y}$*

$$F(Y_1, ..., Y_S) = C(F_1(Y_1), ..., F_S(Y_S)) \tag{1}$$

If $F_1, ..., F_S$ are continuous then C is unique. Conversely, if C is a S-dimensional copula and $F_1, ..., F_S$ are distribution functions then F is a S-dimensional distribution function with marginals $F_1, ..., F_S$.

On first viewing Sklar's theorem is a surprising result. It is a little less surprising when one notes that joint distribution function $F(x, y)$ can be re-written as a function of the inverse of the marginal distributions, $F(G^{-1}(u), H^{-1}(v))$ where $u = G(x)$ and $v = H(y)$. That is, the joint distribution itself is a copula function. The existence proof then relies on the existence of appropriate inverse functions and their existence in turn relies on the monotonicity properties of distribution and the joint distribution functions. See Nelson (1999) for an overview.

2.2 Finite Polynomial Copula

The first part of the result is to present conditions under which the coefficients of a polynomial copula function are identified. That is, under what conditions are the coefficients of a particular K-order polynomial uniquely determined by the data? Note, the actual copula may not be a finite polynomial. In this section we are determining the K-order polyomial that "best fits" the observed probabilities.

Let Q_K be the following K-order polynomial over S dimensions that maps from S marginal probabilities to one joint probability. If $Y \in \mathcal{Y}^*$ then we can write

$$F(Y) = Q_K(F_1(Y_1), ..., F_S(Y_S)) = \sum_{k_1=0}^{K} \sum_{k_2=0}^{K} ... \sum_{k_S=0}^{K} q_{k_1 k_2 ... k_S} F_1^{k_1}(Y_1) \cdots F_S^{k_S}(Y_S) \tag{2}$$

where $Q_K : [0,1]^S \rightarrow [0,1]$ has $(K+1)^S$ parameters. This function can be re-written in matrix form

$$Q_K = x_K q_K \tag{3}$$

where x_K is a $1 \times (K+1)^S$ row vector with a typical element $F_1^{k_1} \cdots F_S^{k_S}$ and q_K is a $(K+1)^S \times 1$ vector with typical element $q_{k_1 k_2 ... k_S}$. If we have $(K+1)^S$ such elements of $\mathcal{Y}^*$ then

$$F_K = X_K q_K \tag{4}$$

where F_K is a $(K+1)^S \times 1$ vector where the typical element is $F(Y)$, X_K is a $(K+1)^S \times (K+1)^S$ matrix where each row is the x_k row vector determined by $\{F_1(Y_1), ..., F_S(Y_S)\}$.

In order to identify these $(K+1)^S$ parameters of the polynomial copula we will need sufficient information on the joint distribution. In Theorem (2) (below) we can see immediately that a necessary condition for identification is that matrix X_K is full-rank. This is analogous to what Greene (1997) calls the "identification condition" for ordinary least squares regression. The identification condition is satisfied if there are both enough observations (at least $(K+1)^S$ elements of $\mathcal{Y}^*$) and given these observations, the variables are *not* co-linear. Here the "variables" are the elements $F_1^{k_1}(Y_{1k}) \cdots F_S^{k_S}(Y_{Sk})$. Definition (2) presents sufficient conditions on the $\mathcal{Y}^*$ to find a matrix X_K that is full-rank.

Definition 2 *$\mathcal{Y}^{**} \subset \mathcal{Y}^*$ is such that for any $\kappa > 0$*

$$|F_s(Y_s') - F_s(Y_s'')| \neq \kappa |F_t(Y_t') - F_t(Y_t'')| \tag{5}$$

for all $s \neq t \in \mathcal{S}$ and all $Y' \neq Y'' \in \mathcal{Y}^{**}$.

Definition (2) states that changes in the marginal distribution in one dimension are not proportional to changes in the marginal distribution any other dimension. The following theorem states that given enough information about the joint distribution (as described in Definition (2)), the coefficients of any finite polynomial copula function are uniquely determined. The theorem also shows how to estimate such a copula.

Theorem 2 *If*

1. *Assumption (1) holds,*
2. *$F_s(Y_s)$ is observed for all $Y_s \in \mathcal{Y}_s$ for all $s \in \mathcal{S}$, and*
3. *$\mathcal{Y}_K = \{Y_1, ..., Y_{(K+1)^S}\} \subset \mathcal{Y}^{**}$.*

then there exists a K-order polynomial Q_K such that for all $Y_k \in \mathcal{Y}_K$

$$F(Y_k) = Q_K(F_1(Y_{1k}), ..., F_S(Y_{Sk})) \tag{6}$$

where

$$q_K = X_K^{-1} F_K \tag{7}$$

and X_K is an observed $(K+1)^S \times (K+1)^S$ matrix of full-rank with typical element $F_1^{k_1}(Y_{1k}) \cdots F_S^{k_S}(Y_{Sk})$ and F_K is an observed $(K+1)^S \times 1$ vector with typical element $F(Y_k)$.

Proof. The proof proceeds in four steps. Step (1) sets up the problem by writing the polynomial as a function of the observed probabilities. Step (2) finds points such that the vectors of the marginal probabilities in each dimension are different from each other. Step (3) shows the vector in dimension s constructed in Step (2) can be written as a best linear predictor of the vector of points in dimension 1. Step (4) shows that each element of a row k can be written as some multiple of the marginal distribution in dimension 1 and "error" terms to a unique set of power terms.

Step 1. By Definitions (1) and (2), F is identified on $\mathcal{Y}^{**}$ and by (2) F_s is identified on $\mathcal{Y}_s$ for all $s \in \mathcal{S}$. By (1) and Theorem (1) we can write

$$F(Y) = C(Y) = Q_K(F_1(Y_1), ..., F_S(Y_S)) \tag{8}$$

for any $Y \in \mathcal{Y}^{**}$.

By (3) there are $(K+1)^S$ elements of $\mathcal{Y}^{**}$ for which we can write Equation (8). In matrix form that system of equations is

$$F_K = X_K q_K \tag{9}$$

where F_K, X_K and q_K are defined above.

Step 2. By (3) and Definition (2) there exist $\mathcal{Y}_K = \{Y_1, ..., Y_{(K+1)^S}\} \subset \mathcal{Y}^{**}$ such that for any $Y_k \in \{Y_2, ..., Y_{(K+1)^S}\}$

$$F_s(Y_{sk}) = F_s(Y_{s(k-1)}) + a_{sk} \tag{10}$$

where $a_{sk} \neq \kappa a_{tk}$ for all $s, t \in \{1, ..., S\}$ and $\kappa > 0$.

Given this we can write

$$F_s(Y_{sk}) = F_s(Y_{s1}) + \sum_{j=2}^{k} a_{sj} \tag{11}$$

Step 3. We can write F_{sK} as a best linear approximation of F_{tK}

$$F_s(Y_{sk}) = b_0 + b_1 F_t(Y_{tk}) + \epsilon_{stk} \tag{12}$$

where ϵ_{st} be a $(K+1)^S \times 1$ vector and b_0 and b_1 are the solution that gives the "best fit".

Claim $\epsilon_{stk} \neq 0$ for at least one $k \in \{1, ..., (K+1)^S\}$.

The claim is proved by contradiction. Suppose $\epsilon_{stk} = 0$ for all $k \in \{1, ..., (K+1)^S\}$. Let

$$F_{sK} = \begin{bmatrix} F_s(Y_{s1}) \\ F_s(Y_{s2}) \\ \vdots \\ F_s(Y_{s,(K+1)^S}) \end{bmatrix} \tag{13}$$

By Step (2),

$$F_{sK} = \begin{bmatrix} F_s(Y_{s1}) \\ F_s(Y_{s1}) + a_{s2} \\ \vdots \\ F_s(Y_{s1}) + \sum_{j=2}^{(K+1)^S} a_{sj} \end{bmatrix} \tag{14}$$

By assumption

$$F_s(Y_{sk}) = b_0 + b_1 F_t(Y_{tk}) \tag{15}$$

for all $k \in \{1, ..., (K+1)^S\}$. So b_0 and b_1 are the solution to the following simultaneous equations (assuming $(K+1)^S > 1$)

$$\begin{aligned} F_s(Y_{s1}) &= b_0 + b_1 F_t(Y_{t1}) \\ F_s(Y_{s1}) + a_{s2} &= b_0 + b_1(F_t(Y_{t1}) + a_{t2}) \end{aligned} \tag{16}$$

Solving by substitution we find that $a_{s2} = b_1 a_{t2}$. A contradiction.

Note ϵ_{st} must have both postive and negative elements, otherwise b_0 and b_1 were not chosen correctly.

So we can recursively define F_{sK} as a best linear approximation of $F_{(s-1)K}$

$$F_s(Y_{sk}) = b_{s0} + b_{s1} F_{s-1}(Y_{(s-1)k}) + \epsilon_{sk} \tag{17}$$

and therefore

$$F_s(Y_{sk}) = d_{s0} + d_{s1} F_1(Y_{1k}) + \upsilon_{sk} \tag{18}$$

where $d_{s0}, d_{s1} \in \Re$ and υ_s is a vector with some positive elements and some negative elements.

Step 4. For any element of row k of X_K we can write

$$\begin{aligned} F_1^{j_1}(Y_{1k}) \cdots F_S^{j_S}(Y_{Sk}) &= \sum_{j=1}^{J} c_j F_1^j(Y_{1k}) \\ &+ \sum_{i_1=0}^{j_1} \cdots \sum_{i_S=0}^{j_S} c_{i_1 \cdots i_S} F_1^{i_1}(Y_{1k}) \epsilon_{2k}^{i_2} \upsilon_{3k}^{i_3} \cdots \upsilon_{Sk}^{i_S} \end{aligned} \tag{19}$$

where $J = \sum_{s=1}^{S} j_s$. Note that one of the elements of this equation is $F_1^{j_1}(Y_{1k})\epsilon_{2k}^{j_2}\upsilon_{3k}^{j_3} \cdots \upsilon_{Sk}^{j_S}$. Consider any other element of the kth row $F_1^{l_1}(Y_{1k}) \cdot \cdot \cdot F_S^{l_S}(Y_{Sk})$. The equation that can be constructed will only contain this element if $l_s \geq j_s$ for all $s \in \mathcal{S}$ and in this case it will contain elements not in Equation (19) unless $l_s = j_s$ for all $s \in \mathcal{S}$. That is, the two will only have the same elements in the equation when they are the same element of the row. Therefore it is not possible to write this element as a linear equation of any other set of elements of X_K in the same row. X_K is full rank. Q.E.D.

Theorem (2) states that like with OLS, identification of the parameters of the finite polynomial is possible because there is assumed to be enough variation in the data. Here it is assumed that it is possible to find points such that changes in the marginal probabilities in one dimension are not proportional to changes in marginal probabilities in any other dimension. To show that this is sufficient the proof constructs a vector in dimension 2 by taking a linear function of the marginal probabilities in dimension 1 and adding some “error” terms. The proof shows that using this construction it is possible to create “variables” that are not co-linear. Theorem (2) also shows that the copula is easy to estimate as it simply involves inverting X_K.

2.3 Approximating the Actual Copula

From the previous section we know that with enough variation in the data we can estimate any polynomial copula function. What if the actual copula is not equal the estimated polynomial or not even a polynomial, how accurate is the estimate?

Given Sklar’s theorem, the continuity assumptions and conditions on the observed joint probabilities the following theorem states that there exists a sequence of polynomials whose parameters are known that converge uniformly to the actual joint distribution.

Theorem 3 *If*

1. *Assumption (1) holds,*
2. $F_s(Y_s)$ *is observed for all* $Y_s \in \mathcal{Y}_s$ *for all* $s \in \mathcal{S}$*, and*
3. $\mathcal{Y}^{**}$ *has infinitely many elements.*

then for any $\epsilon > 0$ *there exists a* K*-order polynomial* Q_K *is such that*

$$|F(Y) - Q_K(Y)| < \epsilon \tag{20}$$

for all $Y \in \mathcal{Y}$ *and whose* $(K+1)^S$ *coefficients are uniquely determined.*

Proof. The proof proceeds in three steps. Step (1) sets up the problem, showing that by the triangle inequality the distance between F and Q_K is less than the distance between Q_K and some other polynomial P_K plus the distance between P_K and the copula function C. Step (2) shows there is a P_K that is close to C, and then determines the difference between the coefficients of Q_K and P_K on $\mathcal{Y}^*$. Step (3) uses the results of lemmas presented in the Appendix I to show that the difference between the coefficients of Q_K and P_K does not get arbitrarily large as the order of the polynomials increases. Therefore, it is possible to find a large enough polynomial so that the distance between F and Q_K is small.

Step 1. From (1), (2) and (3), by Theorem (1) and (2)

$$F(Y^*) = C(Y^*) = Q_K(Y^*) \tag{21}$$

for all $Y^* \in \mathcal{Y}^*$, where the $(K+1)^S$ parameters of Q_K are known. For any $Y \in \mathcal{Y}$, by (1) and Theorem (1) and the triangle inequality

$$\begin{aligned} |F(Y) - Q_K(Y)| &= |C(Y) - Q_K(Y)| \\ &\leq |P_K(Y) - Q_K(Y)| + |P_K(Y) - C(Y)| \end{aligned} \tag{22}$$

where P_K is some unknown K-order polynomial with $(K+1)^S$ parameters.

Step 2. Choose $\epsilon > 0$. Let $\epsilon' = \frac{\epsilon}{2M}$, where $M > 2$ is a large positive number (defined in Lemma (5), in Appendix I). By Assumption (1) and Theorem 2.10.7 (Nelson (1999)), C is uniformly continuous on $[0,1]^S$. By Stone-Weierstrass Theorem (Browder (1996)), there exists a K-order polynomial P_K such that

$$|C(Y) - P_K(Y)| < \frac{\epsilon'}{2} \tag{23}$$

for all $Y \in \mathcal{Y}$. Also from Step (1) substitute Q_K for C,

$$|Q_K(Y^*) - P_K(Y^*)| < \frac{\epsilon'}{2} \tag{24}$$

for all $Y^* \in \mathcal{Y}^*$. By (3), Theorem (2) and the definition of Q_K and P_K

$$X_K q_K - X_K p_K = e \tag{25}$$

where e is a $(K+1)^S \times 1$ vector such that $e_k \in (-\frac{\epsilon'}{2}, \frac{\epsilon'}{2})$ for all $k \in \{1, ..., (K+1)^S\}$. By Theorem (2), X_K has full rank. Rewriting

$$q_K - p_K = X_K^{-1} e \tag{26}$$

Step 3. For some $Y \in \mathcal{Y}$

$$|Q_K(Y) - P_K(Y)| = |x_k q_k - x_k p_k| \tag{27}$$

where x_k is a $1 \times (K+1)^S$ row vector where a typical element is $F_1^{k_1}(Y_1) \cdots F_S^{k_S}(Y_S)$. Substituting the result from Step (2).

$$|Q_K(Y) - P_K(Y)| = |x_k X_K^{-1} e| \tag{28}$$

Let $B(a,b) = a^T X_K^{-1} b$ define the bilinear map for X_K^{-1} given vectors a and b, where $B : [0,1]^{(K+1)^S} \times [0,1]^{(K+1)^S} \to \Re$. Also note we can re-define e such that that $e_k = (u_k - \frac{1}{2})\epsilon'$ where $u_k \in [0,1]$ for all $k \in \{1, ..., (K+1)^S\}$. So

$$\begin{aligned} |Q_K(Y) - P_K(Y)| &= |B(x_k^T, e)| \\ &= |B(x_k^T, \epsilon'(u_k - \tfrac{1}{2}))| \end{aligned} \tag{29}$$

where h is a $(K+1)^S \times 1$ vector of 1's. By Lemma (5) and Lemma (2) (in Appendix I).

$$\begin{aligned} |Q_K(Y) - P_K(Y)| &= |B(x_k^T, u_k) - \tfrac{1}{2}B(x_k^T, h)|\epsilon' \\ &\leq |B(x_k^T, u_k)|\epsilon' + \tfrac{1}{2}|B(x_k^T, h)|\epsilon' \\ &\leq M\epsilon' + \tfrac{\epsilon'}{2} \end{aligned} \tag{30}$$

From Step (1)

$$\begin{aligned} |F(Y) - Q_K(Y)| &\leq |Q_K(Y) - P_K(Y)| + |P_K(Y) - C(Y)| \\ &\leq M\epsilon' + \tfrac{\epsilon'}{2} + \tfrac{\epsilon'}{2} \\ &= \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2M} + \tfrac{\epsilon}{2M} \\ &< \epsilon \end{aligned} \tag{31}$$

Q.E.D.

Theorem 3 presents sufficient conditions for any joint probability of interest to be approximated by known function of observed marginal probabilities. Note that much of the proof is buried deep in Appendix I. The intuition is that any copula may be approximated by a finite polynomial and that arbitrary finite polynomial can, in-turn, be approximated by a finite polynomial that is estimated using the observed parts of the joint distribution. The trick is determining that the two polynomials are “close” to each other on the observed parts of the joint distribution, implies that they cannot move “too far” from each other over the rest of the domain. The proof uses this fact twice. First, it uses this fact to determine the maximum distance between the coefficients of the two polynomials. While this provides the distance for two polynomials of a given order, it doesn’t provide information about what happens to this distance as the order of the polynomial gets large. However, we know that for two polynomials of *any* order, the distance between them is small on the observed parts of the domain. The implication is that the distance between the two polynomials cannot get arbitrarily large as the order of the polynomials gets large. Given that this distance between the two polynomials is bounded, it is possible to find a large enough polynomial such that it converges the actual distribution uniformly.

2.4 Partially Identifying the Joint Probabilities

Before going to the applications it is necessary to present one more general result. The result presents conditions under which the joint distribution F is partially identified in the sense of Definition (1).

Definition 3 *Let $\mathcal{H}$ be such that for each $\mathcal{Y}^h \in \mathcal{H}$ and $h \in \mathcal{Y}^h$, we have*

1. $\mathcal{Y}^h \subset \mathcal{Y}$
2. $h_{st} : \mathcal{Y}^h_s \rightarrow \mathcal{Y}^h_t$
3. $h_{ss}(y) = y$
4. $h_{st}(h_{tu}(y)) = h_{su}(y)$, *and*
5. $h = \{h_{s1}(y), ..., h_{s(s-1)}(y), y, h_{s(s+1)}(y), ..., h_{sS}(y)\}$

for some $y \in \mathcal{Y}^h_s$ and all $s, t, u \in \mathcal{S}$

Definition (3) defines a locus of points that “bisects” the space into S blocks. In the applications, this locus of points is an indifference curve.

Definition 4 *For some $y \in \mathcal{Y}^h$ let*

$$G_s(y_s) = \int_{\underline{Y}}^{y_s} F_{-s}(h_{s1}(Y), ..., h_{s(s-1)}(Y), h_{s(s+1)}(Y), ..., h_{sS}(Y)|Y_s = Y) f_s(Y) \tag{32}$$

where F_{-s} is the joint distribution in every dimension except s for any $s \in \mathcal{S}$.

Definition (4) describes the probability of observing an outcome in dimension s less than y given that the outcomes in every other dimension must be less than some function of the outcome in dimension s. In the applications this probability is the observed probability of the outcome of interest conditional on the selected treatment.

Lemma 1 *If*

1. *$\mathcal{Y}^h$ is observed*
2. *$G_s(y_s)$ is observed for all $y_s \in \mathcal{Y}_s^h$ and all $s \in \mathcal{S}$*

then $\mathcal{Y}^h \subset \mathcal{Y}^*$.

Proof. By (1) let $Y \in \mathcal{Y}^h$, from Definition (3)

$$F(Y_1, ..., Y_S) = \sum_{s=1}^{S} \int_{\underline{Y}}^{Y_s} F_{-s}(h_{s1}(y), ..., h_{s(s-1)}(y), h_{s(s+1)}(y), ..., h_{sS}(y)|Y_s = y) f_s(y) dy \tag{33}$$

From Definition (4)

$$F(Y_1, ..., Y_S) = \sum_{s=1}^{S} G_s(Y_s) \tag{34}$$

By (2) $Y \in \mathcal{Y}^*$. Q.E.D.

Lemma (1) presents conditions under which the joint distribution is partially identified. That is, along the indifference curve the joint probabilities are known functions of the observed probabilities. Below these conditions are shown to hold in the Roy model and in the auction model allowing the results presented above to be used to identify the joint distribution up to an approximation.

3 The Roy Model

To illustrate the results presented below consider a two-sector Roy model. In this model the agent ω maximizes her return by choosing sector A over sector B if and only if

$$R_{\omega A} > R_{\omega A} \tag{35}$$

where

$$R_{\omega A} = g(Y_{\omega A}, X_\omega, Z_{\omega A}, \upsilon_{\omega A}) \tag{36}$$

and $Y_{\omega A}$ is the outcome of interest, X_ω are observable characteristics of the agent that don't vary with the sector, $Z_{\omega A}$ are a vector observable characteristics of the sector and the agent that vary with the sector, and $\upsilon_{\omega A}$ is a vector of unobservable characteristics of the sector and the agent and g is some possibly unknown function. The outcome of interest may be months lived on a particular drug treatment. The observable characteristics of the agent may include age and gender. Observable characteristics of the sector and agent may include the distance to the treatment location.

Our interest is determining the joint distribution of the outcome ($F(Y_A, Y_B)$) given that econometrician observes

- $Y_{j\omega}$ in the chosen sector $j \in \{A, B\}$
- X_ω
- $Z_{j\omega}$ for all $j \in \{A, B\}$

and agent ω behaves as stated above. This is the generalized Roy model.

In order to simplify the presentation the observable characteristics (X) are suppressed. As with the previous results the results presented below are conditional on some set of observed characteristics of the agent. Given this framework, what combination of assumptions and data is sufficient for F to be inferred from the observed probabilities?

Non-parametric identification results in the literature can be characterized by the following four steps.[5]

[5]See Heckman and Vytlacil (2007) for a description of the Roy model and a summary of results.

1. Assume there exists a set $\mathcal{Z}$ such that $F(Y|Z_i) = F(Y|Z_j) = F(Y)$ for all $Z_i \neq Z_j \in \mathcal{Z}$ and $Z_i \in \mathcal{Z}$ is observed for all $Y \in \mathcal{Y}$.
2. Show $\mathcal{Y}^h(Z) \in \mathcal{H}$ is observed for all $Z \in \mathcal{Z}$ given the necessary assumptions.
3. Show $F(Y)$ is observed for all $Y \in \mathcal{Y}^h(Z)$ for all $Z \in \mathcal{Z}$ given the observed distribution of Y_s conditional on the sector chosen and Step (2).
4. Assume $\mathcal{Z}$ is such that $\mathcal{Y} \subset \cup_{Z \in \mathcal{Z}} \mathcal{Y}^h(Z)$.

The set of observable characteristics $\mathcal{Z}$ is generally called the set of "instruments" or "instrumental variables". These variables have a couple of important features. First, they are such that conditioning on them doesn't affect the distribution of outcomes of interest. The instruments are "orthogonal" to the outcomes of interest. Second, these variables in association with other assumptions allow the locus of points $\mathcal{Y}^h$ to be identified. In the model presented above $\mathcal{Y}^h(Z)$ is the implicit function defined by

$$g(Y_{A\omega}(Y_{B\omega}), Z_{A\omega}) = g(Y_{B\omega}, Z_{B\omega}) \tag{37}$$

That is, the loci of points is mapped out by agent ω's indifference curve between the two sectors. Park (2012) shows that with sufficient variation in the Z's the implicit function can be mapped out even if g is unknown. French and Taber (2011) shows the observed sector choice can identify the "slope" of the indifference curve when Y and Z are separable. Heckman and Honore (1990) assumes the Z variable is the observed wages in each sector and assume g is a known log-linear function. Note that in all of these results it is assumed that g is not a function of the unobserved characteristics of the agent or the market (υ). That is, the econometrician observes the exact same signals that the agent uses to make the sector choice. Step (3) is an application of Lemma (1) where G's are the observed distribution of outcomes conditional upon the sector chosen. Step (4) then assumes that there is enough variation in the instrument or the sector prices for the indifference curves to vary across the whole domain.

The results presented above allow Step (4) to be replaced with an alternative step. The obvious advantage of the alternative is that it is not necessary to assume that $\mathcal{Z}$ is such that the indifference curves span the whole outcome space. The observed prices or the observed instrumental variables do not have vary so much that every joint probability can be inferred. That said, it is necessary to assume that there is enough variation in instrument or the price so that coefficients of the polynomial copula are identified. Obviously, it is also necessary to observe the marginal distributions (possibly from some other data set). This may either be done through variation in the instrument such that the marginals are "identified at infinity" or there is an instrument that forces the agent into one sector or the other such as for randomized control trials.

To simplify the exposition, the rest of the section considers a multi-sector version of the model in Heckman and Honore (1990). Consider a model with Ω agents and S choices. For the most part the paper refers to these choices as "products" and the decision maker as the "consumer". Consumer ω observes her vector of outcomes over the S products, $Y_\omega \in \mathcal{Y}$ and a set of vectors of exogenously determined prices $p \in P \subset \mathcal{Y}$.

Assumption 2 *$Y_\omega = \{Y_{\omega 1}, ..., Y_{\omega S}\} \sim F$ where Y_ω is iid for all $\omega \in \Omega$.*

Assumption 3 *Consumer ω has utility $R_{\omega s} = Y_{\omega s} - p_s$ if she chooses product s.*

Assumption 4 *Consumer ω chooses one item s if and only if*

$$R_{\omega s} = \max\{R_{\omega 1}, ...R_{\omega S}\} \tag{38}$$

Assumption 2 simplifies the problem by not allowing correlation of outcomes across consumers. The other two assumptions state that each consumer's utility is linear in "money" and each consumer chooses the product that gives her the highest net utility. So $\mathcal{Z} = P$ and $g(Y_\omega, Z_{\omega s}) = Y_\omega - p_s$. As stated above, these assumptions simplify the exposition of how the results from the previous section can be applied to the extended Roy model.

Definition 5 *Let $\tilde{\mathcal{Y}}^*$ for a given set of prices P be such that for all $\tilde{Y}_s \in \tilde{\mathcal{Y}_s}^*$ where $\tilde{\mathcal{Y}}^* = \Pi_{s=1}^S \tilde{\mathcal{Y}}_s^*$,*

1. *$Y_s' - p_s > \tilde{Y}_t - p_t$ for all $Y_s' > \tilde{Y}_s$ and all $t \in \mathcal{S}$,*
2. *$Y_s'' - p_s < \tilde{Y}_t - p_t$ for all $Y_s'' < \tilde{Y}_s$ and at least one $t \in \mathcal{S}$.*

Definition 5 is the locus of "cut off" values for a given vector of prices. It shows at what outcome consumer ω will switch from product s to product t for given prices p_s and p_t. It is the consumer's "indifference curve" described in Definition (3).

Proposition 1 *If Assumptions (1) - (4) hold and for all ω in a large Ω we observe*

1. *$Y_{\omega s}$ for ω's product choice $s \in \mathcal{S}$, and*
2. *a set of prices $P \subset \mathcal{Y}$ is observed,*

then $F(Y_1, ..., Y_S)$ is identified where $\{Y_1, ..., Y_S\} \in \tilde{\mathcal{Y}}^* \subset \mathcal{Y}$.

Proof The proof shows that the conditions of Lemma (1) are satisfied.

Step 1. By Assumptions (2), (3) and (4), and for a large Ω the probability of observing outcomes less Y_1 in product 1 is given by

$$G_1 = \int_{\underline{Y}}^{Y_1} F_{-1}(\tilde{Y}_{12}(y), ..., \tilde{Y}_{1S}(y)|Y_1 = y) f_1(y) \tag{39}$$

where $\tilde{Y}_{us}(y) = y - p_u + p_s$ (see Definition (5)).

Step 2. $\tilde{\mathcal{Y}}^* \in \mathcal{H}$. By Definition (5) and Step (1), (1), (2) and (5) of Definition (3) hold. $Y_{ss}(y) = y - p_s + p_s = y$ so (3) of Definition (3) holds. Property (4) of Definition (3) holds as $\tilde{Y}_{st}(y) = y - p_s + p_t$ and $\tilde{Y}_{tu}(y) = y - p_t + p_u$, so $\tilde{Y}_{st}(\tilde{Y}_{tu}(y)) = y - p_t + p_u - p_s + p_t = y - p_s + p_u = \tilde{Y}_{su}(y)$.

Step 3. The result from Step (1) holds for any s and from Step (2), G_s is as defined in Definition (4) and is observed from Step (1). Q.E.D.

Proposition 1 states that the joint distribution F is identified for a subset of the domain, namely, $\tilde{\mathcal{Y}}^*$. This is the first identification step described in the previous literature. This is not enough to use Theorem 3 so the statement of the following proposition states that $\mathcal{Y}^{**} \cap \tilde{\mathcal{Y}}^*$ has infinitely many elements. That is, the locus of points on which the joint probabilities are observed must be "wiggly" enough so that X_K is of full-rank. The following proposition to summarizes the identification result for this version of the Roy model.

Proposition 2 *If Assumptions (1) - (4) hold, and for all ω in a large Ω we observe*

1. *$F_s(Y_s)$ for all $Y_s \in \mathcal{Y}_s$ and $s \in \mathcal{S}$*
2. *$Y_{\omega s}$ for ω's product choice $s \in S$ given a price vector p.*
3. *$P \subset \mathcal{Y}$ is observed.*
4. *$\mathcal{Y}^{**} \cap \tilde{\mathcal{Y}}^*$ has infinitely many elements.*

then for all $\epsilon > 0$ there exits a K-order polynomial Q_K where

$$|F(Y) - Q_K(Y)| \leq \epsilon \tag{40}$$

for all $Y \in \mathcal{Y}$ and all $(K+1)^S$ parameters of Q_K are identified.

Proof. The proof follows in a straight forward fashion from the results derived above. Q.E.D.

Proposition (2) state that it is possible to get non-parametric identification in the extended Roy model by replacing the second identification steop in the literature with the result presented in the previous section. The cost is that the identification is limited to an approxmation result and requires the observation of the marginal distributions.

Appendix II presents results from a Monte Carlo analysis of a two-sector version of this model. The objective is to give the reader some idea of the trade-off between reducing requirements on the varaition in the data with the accuracy of the approximation. In the analysis it is assumed that there are two different data sets of the same underlying population. In one data set, agents choose between the two sectors after observing their outcomes in each sector and prices in each sector. From this data set the econometrician observes the outcome in the selected sector and the prices observed by the agent. This data set has 500,000 observations, 100,000 observations for each of five sets of prices. The second data set is a "randomized control trial" data set. Each arm has 100,000 observations. The outcomes are assumed to be distributed bivariate normal with two different σ's and a positive ρ. The absolute difference between the actual joint probability and the approximated joint probability are calculated across a 10×10 array of points distributed across the whole domain. The results show that when there is no variation in prices across the five sets, it is possible to identify and estimate a quadratic copula function. In this case the error rate averages at 26 in 10,000. When prices vary it is possible to identify higher order polynomials. An estimated 4th-order polynomial copula function has an error rate that averages 13 in 10,000.

4 Identification in an Auction Model

This section applies the results presented above to an auction setting where there are simultaneous sealed-bid auctions for differentiated products. Much of the intuition from the previous section carries over to this section, however things are more complicated because prices are determined endogenously and only auction prices are observed.

Consider a bidder ω who can bid in up to S simultaneous sealed bid auctions for S differentiated items.

Assumption 5 *Let $-\infty < \underline{Y} < \bar{Y} < \infty$ and $\mathcal{Y}_s = [\underline{Y}, \bar{Y}]$ for all $s \in \mathcal{S}$ and $\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_S$.*

Assumption (5) states that the "outcome" space is closed and bounded. This obviously limits the settings for which the results may be applied but it is necessary and relatively standard in the auction literature.

Assumption 6 *$M \in \mathbf{N}$ bidders face between 1 and S simultaneous sealed-bid auctions.*

Assumption 7 *Each bidder chooses 1 auction.*

Assumptions 2 and 6 are the private values assumption and the simultaneous auctions assumption, respectively. Assumption 7 states that each bidder will only enter one auction. This is equivalent to Assumption 4 presented above for the extended Roy model and simplifies the problem. It also seems relatively realistic for the case of online car auctions for example.

The game is formally described in the following existence theorem.

Theorem 4 *(Myerson (1991)) There exists a Bayes Nash equilibrium to the game $\Gamma = (M, (T_\omega)_{\omega \in M}, (Y_\omega)_{\omega \in M}, (F_\omega)_{\omega \in M}, (R_\omega)_{\omega \in M})$ where*

- $T_\omega \in \mathcal{T} = \{0,1\}^S \times [\underline{Y}, \bar{Y}]^S$, *s.t.* $\sum_{s=1}^S t_{\omega s1} = 1, \forall t_\omega \in T_\omega$, *and* $t_{\omega s2} = b_{\omega s}$.
- $Y_\omega \in \mathcal{Y} = [\underline{Y}, \bar{Y}]^S \subset \Re^S$
- $F_\omega = F^{M-1}, \forall \omega \in M$
- $R_\omega(t_\omega) = R_{\omega s} = Pr(win|b_{\omega s})(Y_{\omega s} - E(p_s|win, b_{\omega s}))$ *where* $t_{\omega s1} = 1$.

and is represented as a randomized-strategy profile $\sigma \in \times_{\omega \in M} \times_{Y_\omega \in \mathcal{Y}} \Delta(\mathcal{T})$

$$\sigma = ((\sigma_\omega(T_\omega|Y_\omega))_{T_\omega \in \mathcal{T}})_{Y_\omega \in \mathcal{Y}, \omega \in M} \tag{41}$$

$$\sigma_\omega(T_\omega|Y_\omega) \geq 0, \forall T_\omega \in \mathcal{T}, \forall Y_\omega \in \mathcal{Y}, \forall \omega \in M \tag{42}$$

$$\sum_{T_\omega \in \mathcal{T}} \sigma_\omega(T_\omega|Y_\omega) = 1, \forall Y_\omega \in \mathcal{Y}, \forall \omega \in M \tag{43}$$

Proof The proof follows in a straightforward fashion from the results presented in Myerson (1991). Step (1) shows that there is a type-agent equilibrium of a similar finite game. Step (2) argues that the equilibrium of the finite game is an equilibrium of the original game.

Step 1. Consider a modified game Γ_1 which differs from Γ in that $\mathcal{T}_1 = \{0,1\}^S$ and $b_{\omega s} = b(Y_{\omega s})$, where $b'(Y_{\omega s}) > 0$. That is, it is assumed bidders bid a monotonic function of their value once they enter auction s. Following Harsayni and Selten this modified game can be represented as a finite type-agent game and so a Nash equilibrium exists by Theorem 3.1 (Myerson (1991)).

Step 2. Γ can be represented as a two-period game in which bidder's choose the auction in the first period and the amount to bid in the second period. $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2\}$ where $\mathcal{T}_1$ is as in Step (1) and $\mathcal{T}_2 = \mathcal{Y}$. Consider the sub-game $\Gamma_{2s} = \{M, (T_\omega)_{\omega \in M}, (Y_\omega)_{\omega \in M}, (G_{\omega s})_{\omega \in M}, (R_\omega)_{\omega \in M_s}\}$ where G is the distribution of valuations such that if a bidder chooses some auction $-s$ they are assigned a valuation of $\underline{Y}$ for auction s. If the bidder chooses auction s they are assigned their valuation $Y_{\omega s}$ for auction s. G is determined by the equilibrium actions of the game and is known to the bidders. Note that the bidders do not observe actual entry choices just the equilibrium strategies. Given standard results, the Nash equilibrium of Γ_{2s} is such that $b_{\omega s} = b(Y_{\omega s})$.

Step 3. From steps (1) and (2), there is exists a sub-game perfect Nash equilibrium of Γ, therefore there exists a Nash equilibrium. Q.E.D.[6]

Theorem (4) formalizes the model and states that there is at least one Bayes Nash equilibrium. The proof uses the fact that there are well known results for the equilibrium of the second-stage auction game, which allows the game to be represented as a finite game for which existence results are well known. As in the previous section, the consumer maximizes net utility, although here is it is the expected utility given their expectation over the endogenously determined distribution of prices.

The following definition presents the analogous set of "cut off" values for the auction case. Note that here the indifference curve values are determined by expected prices which are in turn determined by the cut off values. The result above tells us that there is in fact is an equilibrium set of values.

[6]I'm grateful to Greg Lewis for suggesting this idea and for Chris Metcalf for helping with the implementation.

Definition 6 *Let* $\hat{\mathcal{Y}}^*$ *for all* $\hat{Y}_s \in \hat{\mathcal{Y}}^*_s$ *where* $\hat{\mathcal{Y}}^* = \Pi_{s=1}^S \hat{\mathcal{Y}}^*_s$*,*

1. $\Pr(win|b(Y_s'))(Y_s' - E(p_s|win, b(Y_s'))) > \Pr(win|b(\hat{Y}_u))(\hat{Y}_u - E(p_u|win, b(\hat{Y}_u)))$ *for all* $Y_s' > \hat{Y}_s$ *and all* $u \in \mathcal{S}$*,*

2. $\Pr(win|b(Y_s''))(Y_s'' - E(p_s|win, b(Y_s''))) < \Pr(win|b(\hat{Y}_u))(\hat{Y}_u - E(p_u|win, b(\hat{Y}_u)))$ *for all* $Y_s'' < \hat{Y}_s$ *and at least one* $u \in \mathcal{S}$*.*

Given this definition, the proposition shows that we can identify $F(Y_1, ..., Y_S)$ for $\{Y_1, ..., Y_S\} \in \hat{\mathcal{Y}}^*$ when the number of bidders is observed. The main result uses this proposition to present sufficient conditions for the identification of $F(.)$.

Proposition 3 *Given Assumptions (1) - (2) and (5) - (7), if we observe*

1. p_s *is a series of auctions for all* $s \in \mathcal{S}$

2. M *bidders bidding on* S *simultaneous sealed-bid auctions*

then $F(Y_1, ..., Y_S)$ *is identified where* $\{Y_1, ..., Y_S\} \in \hat{\mathcal{Y}}^* \subset \mathcal{Y}$*.*

Proof. The proof has four steps. Step (1) sets things up by partitioning the valuation vectors into a finite set K. Step (2) writes down the "almost" order statistic probabilities given the observed price. Step (3) uses the result in Step (2) and induction to identify the probability a particular bidder will bid p_k in auction s. Step (4) shows that the remaining condition of Lemma (1) holds.

Step 1. Let $[\underline{Y}, \bar{Y}]$ be segmented in to K disjoint sets of equal length such that the union is equal to the original set (partitioned). Let $Y_{ks} = \underline{Y} + \frac{k-1}{K}(\bar{Y} - \underline{Y})$ and

$$f_s(Y_k) = \int_{y=Y_k}^{Y_{k+1}} f_s(y)dy \tag{44}$$

and

$$F_s(Y_k) = \begin{cases} 0 & \text{if } k = 1 \\ \sum_{h=1}^{k-1} f_s(Y_h) & \text{if } k > 1 \end{cases} \tag{45}$$

That is, we will approximate $[\underline{Y}, \bar{Y}]$ with a discrete set of K elements such that $Y_{ks} \in \{Y_1, Y_2, ..., Y_K\}$ where $Y_1 = \underline{Y}$ and $Y_K = \underline{Y} + \frac{K-1}{K}(\bar{Y} - \underline{Y})$. Further, note the observed sequence of prices $\{p_s\}$ will also be approximated in this way, where $\{p_{sK}\}$ denotes the observed sequence from discrete set $\{Y_1, Y_2, ..., Y_K\}$.

Step 2. Given this approximation, without loss of generality consider the probability distribution of the price of item 1 (of S), p_1.

$$\Pr(p_1 = p_K | M, M \geq I) = P(M, I) H_{K1}^I \tag{46}$$

where $P(M, I)$ is the appropriate permutation formula, I denotes the order statistic of interest, i.e. $I = 2$ for a second-price sealed bid auction and

$$\Pr(p_1 = p_k | M, M \geq I) = P(M, I) H_{k1} (H_{k1} + G_{(k+1)} 1)^{I-1} \left(1 - G_{(k+1)1}\right)^{M-I} \tag{47}$$

where

$$H_{k1} = F_{-1}(\hat{Y}_{k2}, ..., \hat{Y}_{kS} | Y_{k1} = b^{-1}(p_k)) f_1(b^{-1}(p_k)) \tag{48}$$

where $\hat{Y}_{ks}$ is such that $R_{\omega s}(\hat{Y}_{ks}) = R_{\omega 1}(b^{-1}(p_k))$ for all $s \geq 2$ and $b(Y_s)$ is the bid function. Note that if $S = 1$, let $F_{-1} = 1$.

$$G_{k1} = \sum_{l=k}^{K} H_{k1} \tag{49}$$

G_{k1} is the probability that a single bidder will bid greater or equal to p_k for item 1. Note that the probability that all the other M bidders have values for item 1 that are equal to or less than p_K or do not bid on item 1 is 1. These equations have a similar flavor to the standard order statistics equations, unfortunately they are not quite the same and the identification argument is not as straightforward.[7]

Step 3. This step of the proof is by induction. The first induction step comes from Equation (46) and it is straightforward to rearrange the equation

[7] To see the difference consider G_{11}. If G_{k1} was simply the distribution of Y then $G_{11} = 1$, but here it is not. Rather it is "the probability of bidding more than p_k for item 1", which is determined by both the value for item 1 and the value for the other items. That is, observing the lowest possible price for item 1 is not particularly informative.

to identify H_{K1}. The second induction step assumes that $G_{(k+1)1}$ is known, the question is whether H_{k1} is identified. H_{k1} is the solution to Equation (47) and is polynomial equation of the form $ax(x+b)^n + c = 0$ where

$$\begin{aligned} a &= P(M, I)\left(1 - G_{(k+1)1}\right)^{M-I} \\ b &= G_{(k+1)1} \\ c &= -\Pr(p_1 = p_k | M, M \geq I) \end{aligned} \tag{50}$$

If $G_{(k+1)1} < 1$, there exists a real positive solution as $c < 0$ and the LHS is strictly increasing in x.[8] From Equation (48) and (49),

$$G_{k1} = \sum_{y=Y_k}^{Y_K} F_{-1}(\hat{Y}_{k2}, ..., \hat{Y}_{kS} | Y_{k1} = b_1^{-1}(p_k)) f_1(b^{-1}(p_k)) \tag{51}$$

Letting $K \to \infty$

$$G_1(Y) = \int_{Y}^{\bar{Y}} F_{-1}(\hat{Y}_2(y), ..., \hat{Y}_S(y) | Y_1 = y) f_1(y) \tag{52}$$

Similarly, $G_s(Y)$ is identified for all $s \in \mathcal{S}$.

Step 4.

Case 1. Let $S = 1$. We have the result from Step (3).

Case 2. Let $S > 1$. From Definition (6), (1), (2) and (5) of Definition (3) hold. Note that $\hat{Y}_s(y)$ is the implicit function defined by $R_{\omega s}(\hat{Y}_s) = R_{\omega 1}(y)$ and this is the same for all $\omega \in M$. The utility of entering auction s is

$$R_s = \Pr(win | b_s)(Y_s - E(p_s | win, b_s)) \tag{53}$$

Take the derivative of R_s with respect to Y_s. From Equation (53) and the Envelope Theorem,

$$\frac{dR_s}{dY_s} = \Pr(win | b(Y_s)) > 0 \tag{54}$$

So $\hat{Y}_s(y)$ is monotonic for all s and (3) of Definition (3) holds. To see that (4) of Definition (3) holds, note that $\hat{Y}_s(y)$ is defined by $R_s(\hat{Y}_s(y)) = R_1(y)$ and $\hat{Y}_u(y)$ is defined by $R_u(\hat{Y}_u(y)) = R_1(y)$, so $R_s(\hat{Y}_s(y)) = R_u(\hat{Y}_u(y))$ and

[8] Note for the positive solution to be less than or equal to 1, it must be that $ab^n + c \geq 0$.

by monotonicity we have the result. Q.E.D.

Proposition (3) states that the joint distribution F is identified for a subset of the domain, namely, $\hat{\mathcal{Y}}^*$. The argument is similar to the argument presented for the equivalent proposition in the Roy model section (Proposition 1). The argument is complicated by the fact that relative expected prices are determined endogenously in the auction model and outcomes (i.e. valuations) are not observed. Identifying the probability over outcomes (valuations) conditional on auction choice requires a non-trivial adaption of the order statistic approach used in Athey and Haile (2002). As was the case for the Roy model, this result is not quite enough to use Theorem 3 (see condition (4) in Proposition (4)).

The following proposition states the main result of the section. Given Sklar's Theorem and the continuity assumptions it is possible to identify the joint value distribution up to an arbitrarily close approximation.

Proposition 4 *If Assumptions (1 - 2) and (5 - 7) hold, if we observe for all $s \in \mathcal{S}$*

1. *p_s and M bidders observed in a series of auctions with no competing simultaneous auctions.*
2. *p_s and M observed in a series of auctions with $S > 1$ competing simultaneous auctions.*
3. *$\mathcal{Y}^{**} \cap \hat{\mathcal{Y}}^*$ has infinitely many elements.*

then for any $\epsilon > 0$ there exists a K-order polynomial such that

$$|F(Y) - Q_K(Y)| \leq \epsilon \tag{55}$$

for all $Y \in \mathcal{Y}$ and where Q_K has $(K+1)^S$ parameters are uniquely determined.

Proof. The proof follows from the results derived above. Q.E.D.

The proposition states that if we observe bidders bidding in simultaneous auctions for S differentiated products it is possible to estimate an approximation to the joint value distribution of the S differentiated goods. Note that the assumption that the number of bidders are observed is made for simplicity. There is a growing literature on identification when the number of bidders in the auction is not observed. I assume those results can be applied to this setting. See for example Adams (2007), Zeithammer and Adams (2010) and Adams et al. (2011) among others.

5 Conclusion

The paper shows that Sklar's theorem can be used to identify the full distribution when the marginal distributions are observed and the joint distribution is partially observed. This result is used to present additional results for the extended Roy model. In particular, the paper shows that it is possible to identify an approximation to the full distribution with data on the marginal probabilities and partial identification of the distribution from structural assumptions. The paper argues the data requirements are modest and it is not necessary to make strong parametric assumptions. That said, the extended Roy model assumes the decision maker and the econometrician have access to the same set of signals as to the likely outcomes of the treatments, which may limit the applicability of the results.

The results are also applied to the auction context. In particular, the results are applied to a setting where the bidder must choose between simultaneous sealed bid auctions for differentiated products. While the flavor of the results is similar to the Roy model, things are more complicated because expected prices are determined endogenously. In addition, in the auction context it is only possible to observe an order statistic of the distribution of outcomes.

References

Christopher P. Adams. Estimating demand from ebay prices. *International Journal of Industrial Organization*, 25(6):1213–1232, December 2007.

Christopher P. Adams, Laura Hosken, and Peter Newberry. vettes and lemons on ebay. *Quantitative Marketing and Economics*, 9(2):109–127, 2011.

Susan Athey and Philip Haile. Identification in standard auction models. *Econometrica*, 70(6):2170–2140, 2002.

Matthew Backus and Gregory Lewis. A demand system for a dynamic auction market with directed search. Harvard University, October 2012.

Bjarne Brendstrup and Harry J. Paarsch. Semiparametric identification and estimation in multi-object, english auctions. *Journal of Econometrics*, 141: 84–108, 2007.

Andrew Browder. *Mathematical Analysis: An Introduction*. Springer, 1996.

Eric French and Christopher Taber. *Handbook of Labor Economics*, volume 4A, chapter Identification of Models of the Labor Market, pages 537–617. North-Holland, 2011.

William Greene. *Econometric Analysis*. Prentice Hall, third edition, 1997.

James Heckman and B Honore. The Empirical Content of the Roy Model. *Econometrica*, 58:1128–1149, 1990.

James J. Heckman and Edward J. Vytlacil. *Handbook of Econometrics*, volume 6B, chapter Econometric Evaluation of Social Programs, Part 1: Causal models, structural models and econometric policy evaluation, pages 4780–4874. Elsevier, 2007.

Marc Henry and Ismael Mourifie. Sharp bounds in the binary roy model. University of Montreal, May 2012.

Mireia Jofre-Bonet and Martin Pesendorfer. Estimation of a dynamic auction game. *Econometrica*, 71(5):1443–1489, September 2003.

David X. Li. On default correlation: A copula function approach. *Journal of Fixed Income*, 9:43–54, 2000.

Charles Manski. *Public Policy in an Uncertain World*. Harvard University Press, 2013.

Roger Myerson. *Game Theory: Analysis of conflict*. Harvard University Press, Cambridge MA, 1991.

Roger B Nelson. *An Introduction to Copulas*. Number 139 in Lecture Notes in Statistics. Springer, 1999.

Byoung G. Park. Non-parametric identification and estimation of the extended roy model. Yale, November 2012.

Felix Salmon. Recipe for disaster ”the formula that killed wall street”. *Wired Magazine*, 17(3), March 2009. http://www.wired.com/techbiz/it/magazine/17-03/.

P K Trivedi and David Zimmer. Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics*, 1:1–111, 2007.

Robert Zeithammer. Forward-looking bidding in online auctions. *Journal of Marketing Research*, 43(3):462–476, 2006.

Robert Zeithammer and Christopher Adams. The sealed-bid abstraction in online auctions. *Marketing Science*, 29(6):964–987, 2010.

David Zimmer. The role of copulas in the housing crisis. *Review of Economics and Statistics*, 94:607–620, 2012.

6 Appendix I

The appendix presents a series of results sufficient to prove Theorem (3).

Lemma 2 *Given the assumptions of Theorem (3), for the bilinear map, B, and vector x_k (both defined above), $B(x_k^T, h) = 1$, where h is a vector of 1's.*

Proof. Note that for any $1 \times K$ vector b and $K \times K$ square matrix A with full rank such that the first column is all 1's ($a_{k1} = 1$ for all $k \in \{1, ..., K\}$) then

$$bA^{-1}h = b_1 \tag{56}$$

where b_1 is the first element of the vector b. Q.E.D.

Lemma (2) presents a linear algebra result that is useful in the proof of Theorem (3) and the proofs of the lemmas that follow.

Lemma 3 *Given the assumptions of Theorem (3), for all $Y^* \in \mathcal{Y}^*$ and $u \in [0, 1]^{(K+1)^S}$*

$$0 < B(x_k^{*T}, u) < 1 \tag{57}$$

where x_k^* is the $(K+1)^S \times 1$ vector with the typical element $F_1^{k_1}(Y_1^*) \cdots F_S^{k_S}(Y_S^*)$ and B is the bilinear map defined in the proof to Theorem (3).

Proof. By the proof to Theorem (3), for any $\epsilon > 0$ there exists two K-order polynomial Q_K and P_K such that

$$|Q_K(Y^*) - P_K(Y^*)| < \epsilon \tag{58}$$

for all $Y^* \in \mathcal{Y}^*$. Rewriting

$$|Q_K(Y^*) - P_K(Y^*)| = |x_k^{T*}(q_K - p_K)| = |B(x_k^{T*}, e)| \tag{59}$$

where $e_k \in (-\epsilon, \epsilon)$ for all $k \in \{1, ..., (K+1)^S\}$ and that $e_k = (2u_k - 1)\epsilon$ where $u_k \in [0, 1]$

$$\begin{aligned} |B(x_k^{T*}, e)| &= |B(x_k^{*T}, 2u - 1)|\epsilon \\ &= |2B(x_k^{*T}, u) - B(x_k^{*T}, h)|\epsilon \end{aligned} \tag{60}$$

By Lemma (2)

$$|2B(x_k^{*T}, u) - 1| < 1 \tag{61}$$

Q.E.D.

Lemma (3) presents the result that the two polynomials never move too far from each other on the set of outcomes where the joint probabilities are observed.

Lemma 4 *Let $B(x_k^T, u) = t_L$ where t_L is the following sum*

$$t_L = \sum_{i=1}^{L} x_{ki} \sum_{j=1}^{L} a_{ij} u_j \tag{62}$$

where x_{ki} is the ith element of the vector x_k, a_{ij} is the i, jth element of X_K^{-1} and u_j is the jth element of u. Given the assumptions of Theorem (3), there exists a $M' > 0$ such that for all L and all $i \in \{1, ..., L\}$

$$|\sum_{j=1}^{L} a_{ij} u_j| < M' \tag{63}$$

Proof. Suppose not. There exists an L and some i such that

$$|\sum_{j=1}^{L} a_{ij} u_j| = M'' > M' \tag{64}$$

where M'' can be arbitrarily large. Let there be $1 \leq L_1 \leq L$ elements such that Equation (64) holds.

Case 1. Lemma (3) does not hold. A contradiction.

Case 2. Lemma (3) holds. By Lemma (3), (3) and Definition (2) there exists a vector x_l^* such that

$$\sum_{i=1}^{L} x_{li}^* \sum_{j=1}^{L} a_{ij} u_j = \beta_1 \tag{65}$$

where $\beta_1 \in (0,1)$. So by assumption that M'' can be arbitrarily large

$$\sum_{k=1}^{L_1} x_{lk}^* b_k = 0 \tag{66}$$

where $b_k M'' = \sum_{j=1}^{L} a_{kj} u_j$. By (3) and Definition (2) there exist L_1 distinct vectors x_l^* such that Equation (66) holds. In matrix form we have

$$X^* b = o \tag{67}$$

where X^* is a $L_1 \times L_1$ matrix where each row is x_l^*, b is a $L_1 \times 1$ vector with typical element b_k, and o is a $L_1 \times 1$ vector of zeros. By the same argument used in the proof of Theorem (2), X^* is of full-rank. So $b = o$. A contradiction. Q.E.D.

Lemma (4) shows that given Lemma (3) then for any order of the two polynomials, the distance between them cannot get too far apart over the whole domain.

Lemma 5 *Given the assumptions of Theorem (3), there exists a $M > 2$ such that for all K*

$$|B(x_k^T, u)| < M \tag{68}$$

Proof. Step 1. Let $B(x_k^T, u) = t_L$ where t_L is the following sum

$$t_L = \sum_{i=1}^{L} x_{ki} \sum_{j=1}^{L} a_{ij} u_j \tag{69}$$

where x_{ki} is the ith element of the vector x_k, a_{ij} is the i,jth element of X_K^{-1} and u_j is the jth element of u.

$$\begin{aligned} |t_L| &= |\textstyle\sum_{i=1}^{L} x_{ki} \sum_{j=1}^{L} a_{ij} u_j| \\ &\leq \textstyle\sum_{i=1}^{L} |x_{ki} \sum_{j=1}^{L} a_{ij} u_j| \\ &\leq \textstyle\sum_{i=1}^{L} |x_{ki}| |\sum_{j=1}^{L} a_{ij} u_j| \\ &= \textstyle\sum_{i=1}^{L} x_{ki} |\sum_{j=1}^{L} a_{ij} u_j| \end{aligned} \tag{70}$$

Step 2. From Lemma (4) we can rewrite

$$\begin{aligned} |t_L| &\leq \sum_{i=1}^{L} x_{ki} |\sum_{j=1}^{L} a_{ij} u_j| \\ &\leq M' \sum_{i=1}^{L} x_{ki} \end{aligned} \tag{71}$$

Note that $\lim_{L\to\infty} \sum_{i=1}^{L} x_{ki} \to N < \infty$ and so there exits N' such that for all L

$$\begin{aligned} |t_L| &\leq M' \sum_{i=1}^{L} x_{ki} \\ &\leq M'N' \end{aligned} \tag{72}$$

Let $M = M'N'$. Q.E.D.

Lemma (5) shows that the distance between the polynomials is bounded as the order of the polynomial increases.

7 Appendix II

The Monte Carlo analysis considers a two-sector Roy model in which agent observes their actual outcome in both sectors and prices in both sectors. The agent chooses the sector with the highest net utility $(Y_s - p_s)$. There are five sets of observed prices. Outcomes are distributed by a bivariate normal function $N(\mu, \Sigma)$ where $\mu = \{-1, -0.5\}$ and

$$\Sigma = \left\{ \begin{array}{cc} 1.5 & 0.4 \\ 0.4 & 1 \end{array} \right\} \tag{73}$$

Table 1 presents the average over 10 trials for the mean and median absolute difference between the estimated joint distribution and the actual joint distribution over 10×10 square of points covering the outcome space.[9] The table shows the results for five different price sets. In all the price sets, the price of the alternative is zero ($p_B = 0$). In each case consumers may face one of five different prices for sector A. Over the different sets, the prices are

[9]For each trial the points are chosen by picking the first 10 outcomes randomly generated for sector A in the RCT data. The idea is to choose representative points rather than a fixed grid of the outcome space (which is infinite).

varying by a greater degree from (1) to (5). In set (1), there is no variation in the price of sector A. In set (5), prices vary between -1.6 through to 1.6.

1. $p_A = \{0, 0, 0, 0, 0\}$
2. $p_A = \{-0.2, -0.1, 0, 0.1, 0.2\}$
3. $p_A = \{-0.4, -0.2, 0, 0.2, 0.4\}$
4. $p_A = \{-0.8, -0.4, 0, 0.4, 0.8\}$
5. $p_A = \{-1.6, -0.8, 0, 0.8, 1.6\}$

The table presents the results for different estimated polynomials from quadratic (K-order 2) up to polynomials of order 25. The mean refers to the average of the absolute differences across the 100 points checked. The median is the median of the absolute differences across the 100 points.

Table 1: Difference Between Estimated and Actual Probabilities

	Price Set	1	2	3	4	5
K-order						
2	Mean	0.0025	0.0026	0.0026	0.0025	0.0026
	Median	0.0022	0.0022	0.0023	0.0022	0.0023
3	Mean	0.0148	0.0045	0.0035	0.0030	0.0022
	Median	0.0085	0.0025	0.0020	0.0020	0.0020
4	Mean	0.3443	0.0135	0.0051	0.0023	0.0013
	Median	0.3087	0.0056	0.0020	0.0012	0.0010
5	Mean	0.3838	0.1036	0.0246	0.0050	0.0020
	Median	0.4020	0.0467	0.0072	0.0019	0.0010
10	Mean	0.2972	0.1558	0.1500	0.0467	0.0137
	Median	0.2772	0.0667	0.0321	0.0061	0.0018
25	Mean	0.3538	0.2585	0.1774	0.0984	0.0360
	Median	0.3735	0.1482	0.0440	0.0148	0.0033

Table 1 shows that the quadratic copula can be estimated with no variation in prices and further the results do not get better as the variation in prices increases. A quadratic polynomial copula provides an estimate that is different from the actual probability of around 25 in 10,000. As the order of the polynomials increase they become more difficult to estimate. Note that differences in probabilities over .2 suggest that the coefficients are not identified. However these higher order polynomials can be estimated if there is enough variation in the prices. The table shows a 4th order polynomial can be estimated well with enough variation in prices and the difference between the estimated and actual probability is 13 in 10,000.