Non-parametric Identification and Testable Implications of the Roy Model

Francisco J. Buera*
Northwestern University

January 2006

Abstract

This paper studies non-parametric identification and the testable implications of the Roy Model. Results in the literature are generalized in three directions. First, my model allows fully general functions that characterize the income of different occupations. Second, identification is obtained using strictly weaker exclusion restrictions. For example, the model with additive heterogeneity is non-parametrically identified without any exclusion restriction. Third, testable implications of a general occupational choice model are derived. The analysis draws on results on the existence and uniqueness of solutions to systems of partial differential equations.

*I thank Fernando Alvarez, Pierre-Andre Chiappori, Ivar Ekeland, Xavier Gine, James Heckman, Chuck Manski, Rosa Matzkin, Phil Reny, Thomas Sargent, Hugo Sonnenschein, Chris Taber, Elie Tamer, Robert Townsend and Ivan Werning for valuable comments and suggestions. I have also benefited from the comments of participants at seminars at the University of Chicago, Universidad Torcuato Di Tella and Northwestern University.

1 Introduction

Several economic models share this basic structure: (i) individuals choose among a finite set of alternatives to obtain the highest income; (ii) potential incomes depend on observable and unobservable characteristics of the individuals; and (iii) unobservable characteristics are distributed in the population according to an unknown distribution. This setup is shared by many discrete choice models. Examples include the studies of women's choices between market and non-market work (Gronau (1974), Heckman (1974)), the analysis of choice of schooling levels (Willis and Rosen (1979)) and models of entrepreneurship with borrowing constraints (Evans and Jovanovic (1989)). The general framework can be applied to a broader set of problems, such as empirical studies of auction models and competing risk models, among others (Athey and Haile (2002), Tsiastis (1975)).

Heckman and Honore (1990) study a version of this basic setup for the case the outcome of the individual choice are observed, the Roy model. In their paper, they derive the testable implications and the identification restrictions of the lognormal model and its log-concave extensions. They also study the identification of the non-normal model with additive unobservable characteristics using the variation in agents' observable characteristics.[1] They show that the structure of a two sector version of the model is identified if the econometrician observe two excluded variables. It is necessary to observe a variable that affects the income in Sector 1 but not the income of Sector 2, and another variable that affects the income of Sector 2 but not the income of Sector 1.

Unfortunately, the results in the literature cannot be applied to the study of important economic models, because often the theory provides fewer exclusion restrictions. For example, in the model of entrepreneurship with borrowing constraints considered in Evans and Jovanovic (1989), the theory provides only one exclusion restriction, namely, that individual's wealth affects their profits as entrepreneurs but does not affect their wages.

In addition, the literature have restricted the analysis to models with additively separable unobservable characteristics. However, often times, the unobservable char-

[1] Other authors have studied the problem of identification when only the choice is observed, as opposed to the choice and the outcome associated with it, e.g. Cameron and Heckman (1998), Manski (1985), Matzkin (1992, 1994).

acteristics interact with the observable variables in a non-linear fashion[2].

In this paper, I overcome these limitations by generalizing the identification analysis in two respects. First, I allow the functions that characterize the income of different occupations to be fully general, following Matzkin (2001). Second, I prove identification using strictly weaker exclusion restrictions.[3] A striking example of the latter is that the model with additive unobservable heterogeneity (e.g., Heckman and Honore (1990)) is non-parametrically identified without the need for any exclusion restrictions.

In addition, I study the testable implications of the non-additive Roy model. For the case more than two individual characteristics (or two individual characteristics that only affect the outcome of one of the choices) are observed, then the model has testable implications. In particular, I provide an equation that must be satisfied by data.

An important methodological contribution of this paper is the application of results from the theory of Partial Differential Equations (PDE) to the study of identification and testable implications of discrete choice models with observable outcomes. I show that the Roy Model implies a system of PDE in a subset of the structure of the model and the joint distribution of data, but not involving the joint distribution of unobserved heterogeneity (Lemma 4). The analysis of the testable implications and identification of the model then simplifies to the study of existence and uniqueness of solutions of a system of PDE.

In this respect, this paper extends the use of techniques that have proven to be extremely useful in extending our understanding of the identification of preferences in consumer theory (e.g., Chiappori (1992, 1997), Chiappori and Browning (1998)). An advantage of the analysis in this paper relative to the standard identification analysis in consumer theory is that it explicitly considers an stochastic structure of heterogeneity.

The rest of the paper is organized as follows. In section 2, I introduce the structure of the Roy model. In section 3, I study a model with one exclusive attribute to highlight the basic intuition of the more general results. In section 4, I present

[2] Again, the model of Evans and Jovanovic (1989) provides an example of a model with this feature.

[3] There have been others papers generalizing the analysis in other important dimensions. For example Heckman (1990) and Heckman and Smith (1998) study the identification of models where selection is based on the maximization of utility as opposed to the maximization of income.

the general results. In particular, I give a local characterization of the relationship between the observable characteristics and the structure of discrete choice model in terms of a partial differential equation (PDE). I then apply recent results on the existence and uniqueness of solutions to PDE to study the identification and to derive testable implications of the general model. Section 5 applies these results to analyze the identification of the entrepreneurial choice model. Section 6 concludes and discusses directions for future research.

2 The Roy Model

Consider an economy populated by a continuum of agents. In this economy, each agent is described by a vector of observable characteristics, $x \in X \subseteq R^k$, $k \geq 1$, and a vector of unobserved characteristics, $\varepsilon = (\varepsilon_1, \varepsilon_2) \in [\underline{\varepsilon}_1, \bar{\varepsilon}_1] \times [\underline{\varepsilon}_2, \bar{\varepsilon}_2] \subseteq R^2$, the vector of individual abilities, that is unobserved to the econometrician but observe by agent when making their decisions. The ability vector is distributed across the population according to the joint distribution $G(\varepsilon_1, \varepsilon_2)$. Throughout the paper, I assume that ε is independent of x.[4]

The income that an agent with sector i specific ability ε_i and covariates x can earn in sector i is given by the following function:[5]

$$y_i = v^i(\varepsilon_i, x) : [\underline{\varepsilon}_i, \bar{\varepsilon}_i] \times X \to Y_i \subseteq R,\ i = 1, 2.$$

Without loss of generality the analysis can be restricted to strictly increasing functions of the unobserved heterogeneity, ε_i. In particular, define the function

$$h^i(y_i, x) :\ Y_1 \times X \to \epsilon_i,\ i = 1, 2$$

i.e. the inverse of v^i with respect to ε_i.

Agents in this economy face a simple decision problem. Agents choose among two mutually exclusive alternatives to maximize their income, y_i. In particular, they

[4]See Buera (2005) for dynamic model of occupational choice where this assumption is relaxed.

[5]This function may arise from an underlying maximization problem, e.g., the profit maximization problem of an entrepreneur as describe latter in this section.

choose Sector 1 iff:

$$y_1 \geq y_2$$
$$v^1(\varepsilon_1, x) \geq v^2(\varepsilon_2, x)$$

This inequality implies the following condition over the unobservable characteristics:

$$\varepsilon_1 \geq h^1\left(v^2(\varepsilon_2, x), x\right),$$

i.e., individuals with relatively high Sector 1 ability choose sector 1.

The set of agents with ability pairs $(\varepsilon_1, \varepsilon_2)$ that are indifferent between Sector 1 and Sector 2 is given by the curve:

$$\varepsilon_2 = h^2\left(v^1(\varepsilon_1, x), x\right)$$

that divides the $(\varepsilon_1, \varepsilon_2)$ space in two sections. Figure 1 illustrates this function and the corresponding preference areas.

The following example is used as a leading application to illustrate the results in the paper.

An Example

Consider an economy where individuals have access to a technology that can be operated at a variable scale, k. In particular, the revenues that an agent obtains from operating this technology are indexed by her entrepreneurial specific ability ε_1, i.e., $f(\varepsilon_1, k)$, a strictly increasing function of entrepreneurial ability. In addition, agents face a borrowing constraint that restricts the amount of capital that can be invested in the project, thereby limiting the scale of the project. In its more general form, this restriction is summarized by the condition:

$$k \leq c(\varepsilon_1, b, r)$$

where $c(\varepsilon_1, b, r)$ gives the maximum feasible scale as a function of their entrepreneurial ability ε_1, their wealth b, and the interest rate r. A common specification for the financial constraint is $c(\varepsilon_1, b, r) = \lambda b$, where b corresponds to the wealth of an agent (e.g., Evans and Jovanovic (1989)).

Conditional on the agent choosing to start a business, his or her income is given by the profit function:

$$\begin{aligned} y_1 &= v^1(\varepsilon_1, b, r) \\ &= \max_k \{f(\varepsilon_1, k) - rk\}. \\ &\qquad s.t. \\ &\qquad k \leq c(\varepsilon_1, b, r) \end{aligned}$$

Alternatively, an agent may choose to be a wage earner. I assume that in this case his or her income is not a function of his of her observable characteristics:

$$y_2 = v^2(\varepsilon_2, b, r) = \varepsilon_2,$$

where I have chosen a particular normalization for Sector 2's specific ability.

The general framework encompasses a broad set of possible specifications: models where entrepreneurial ability enters additively ($f(\varepsilon_1, k) = \tilde{f}(k) + \varepsilon_1$); models with fixed cost ($f(\varepsilon_1, k) = \tilde{f}(k) - \varepsilon_1$, $c(\varepsilon_1, b, r) = \tilde{c}(b - \varepsilon_1, r)$); the Cobb-Douglas model with multiplicative ability and linear borrowing constraints as in Evans and Jovanovic (1989) ($f(\varepsilon_1, k) = \varepsilon_1 k^\alpha$, $c(\varepsilon_1, b, r) = \lambda b$).

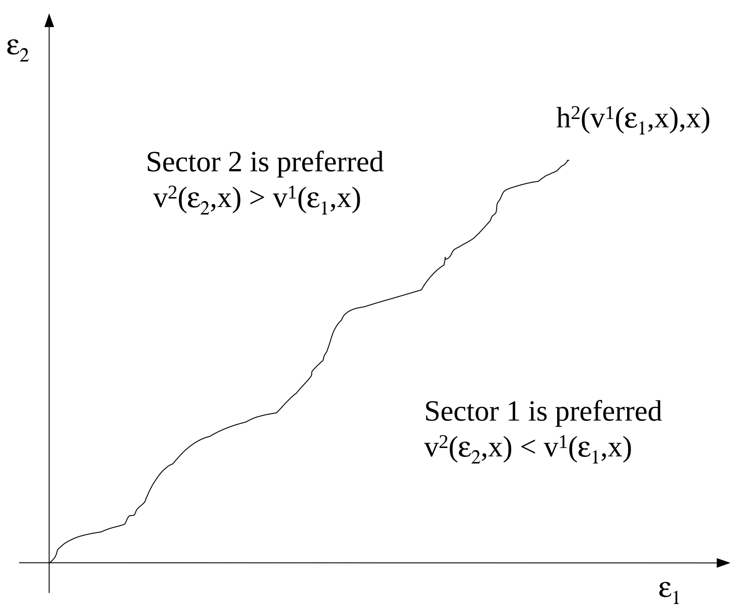

Figure 1: Preference Map

2.1 Data

It is assumed that the vector of observable characteristics, x, the income that individuals earn and the sector where they earn their income are observed.[6] This information can be summarized by two functions, $F^1(y_1, x) = \Pr(i = 1, y \leq y_1|x)$ and $F^2(y_2, x) = \Pr(i = 2, y \leq y_2|x)$, giving the fraction of people in the population earning an income smaller that y_1 in Sector 1 and the fraction with income smaller than y_2 in Sector 2 conditional on a vector of observable characteristics x.

2.2 Identification and Testable Implications

Identification and the existence of testable implications are statements about the properties of the mapping between the set of structures of the model and the set of joint distributions of data. Identification corresponds to the question of when this mapping is injective. The existence of testable implications is concerned with the question: can all of the possible joint distributions of data can be generated by varying the structure of the model? i.e., whether this mapping is *onto* the set of data (surjective).

Let's define the set $\mathcal{H}$ to be a set of functions (including joint distributions of latent variables (e.g. sector-specific abilities)) defining the structure of the model (e.g., $v^i(\varepsilon_i, x)$ and $G(\varepsilon_1, \varepsilon_2)$). Similarly, let $\mathcal{F}$ be the set of joint distributions over the vector of observed random variables (e.g. functions $F^i(y, x)$). Then, define a model as a mapping $\gamma : \mathcal{H} \to \mathcal{F}$. The following are standard definitions of identification and testable implications.

Definition 1: A model γ is identified iff for every $(H, H') \in \mathcal{H}$, $\gamma(H) = \gamma(H')$ implies $H = H'$.
Definition 2: A model γ has testable implications iff $\cup_{H \in \mathcal{H}} \gamma(H) \neq \mathcal{F}$.

[6] Alternatively, we may assume that we observe only the choice without observing the outcome associated with that choice, or we may observe the income an agent earns without knowing the sector where the agent earned his or her income. Manski (1985) and Matzkin (1992) considered the former case, while Heckman and Honore (1990) deal partially with the latter.

2.3 The Relation Between the Structure and Data

In the choice model introduced in section 2, the mapping defining the model is given by the following integral equations:[7]

$$\begin{aligned} F^1(y_1,x) &= P(i=1, y \leq y_1 | x) \\ &= \int_{\underline{\varepsilon}_1}^{h^1(y_1,x)} \int_{\underline{\varepsilon}_2}^{h^2\left(v^1(\varepsilon_1,x),x\right)} g(\varepsilon_1,\varepsilon_2)\, d\varepsilon_2 d\varepsilon_1 \end{aligned} \quad (1)$$

and

$$\begin{aligned} F^2(y_2,x) &= P(i=2, y \leq y_2 | x) \\ &= \int_{\underline{\varepsilon}_2}^{h^2(y_2,x)} \int_{\underline{\varepsilon}_1}^{h^1\left(v^2(\varepsilon_2,x),x\right)} g(\varepsilon_1,\varepsilon_2)\, d\varepsilon_1 d\varepsilon_2 \end{aligned} \quad (2)$$

The first expression corresponds to the integral over those agents that have chosen Sector 1 (i.e., those with relatively low ε_2 and high ε_1), and that earn an income lower or equal to y_1. As can be seen graphically in Figure 1, the integration is over the lower "triangle". Similarly, the second equation corresponds to the integral of the upper "triangle" in figure 1.

The rest of the paper studies the properties of this mapping.

2.4 Normalization

A normalization is needed to study the identification of the general choice model. This amounts to normalizing the units of the ability vector, an intrinsically unobserved object. In other words, ability can be redefined by transforming it with an arbitrary monotone function, and still obtaining the same observables.

Formally, by choosing an arbitrary strictly monotone vector function, $m : R^2 \rightarrow R^2$, a new ability measure can be defined, $(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2) = (m_1(\varepsilon_1), m_2(\varepsilon_2))$, and also new elements of the structure of the model can be obtained $\tilde{G}(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2) = G\left(m_1^{-1}(\tilde{\varepsilon}_1), m_2^{-1}(\tilde{\varepsilon}_2)\right)$, $\tilde{v}^1(\tilde{\varepsilon}_1, x) = v^1\left(m_1^{-1}(\tilde{\varepsilon}_1), x\right)$ and $\tilde{v}^2(\tilde{\varepsilon}_2, x) = v^2\left(m_2^{-1}(\tilde{\varepsilon}_2), x\right)$, and still obtaining the same joint distribution of income and observable characteristics. Therefore, normal-

[7] To simplify the expressions, the arguments of these functions are sometimes omitted, written as F^1 and F^2.

ization of the ability vector is needed. A particularly simple one is to define the units of sector-specific abilities to correspond to the income an agent with observable characteristics, $\bar{x}$, obtains in the corresponding sectors, i.e.

$$h^1(y_1, \bar{x}) = y_1$$
$$h^2(y_2, \bar{x}) = y_2.$$

Proposition 1 summarizes this discussion.[8]

Proposition 1 (Normalization): *Take any functions $G(\varepsilon_1, \varepsilon_2)$, $v^1(\varepsilon_1, x)$ and $v^2(\varepsilon_2, x)$, that imply observables $F^1(y, x)$ and $F^2(y, x)$. Then any strictly increasing transformation of the ability index, $(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2) = (m_1(\varepsilon_1), m_2(\varepsilon_2))$, and corresponding functions $\tilde{G}(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2) = G\left(m_1^{-1}(\tilde{\varepsilon}_1), m_2^{-1}(\tilde{\varepsilon}_2)\right)$, $\tilde{v}^1(\tilde{\varepsilon}_1, x) = v^1\left(m_1^{-1}(\tilde{\varepsilon}_1), x\right)$ and $\tilde{v}^2(\tilde{\varepsilon}_2, x) = v^2\left(m_2^{-1}(\tilde{\varepsilon}_2), x\right)$ imply the same observables, $F^1(y, x)$ and $F^2(y, x)$.*

Proposition 1 provides a fundamental non-identification result. It provide the boundaries of point identification. The rest of the paper derives and studies necessary conditions for a model to rationalize the data. As long as the necessary conditions imply that the model is identified up to the normalization described in Proposition 1, the conditions will also be sufficient.

3 Testable Implications and Identification of a Model with One Exclusive Attribute

I study a special case of the choice model introduced in the previous section. The analysis of this model is intended to help build intuition for the more general result derived later in the paper. After introducing the simple model, I derive its testable implications, and then I study its identification.

Consider a model with a single observable characteristic that only affects Sector 1, i.e. $x \in R$ and

$$y_1 = v^1(\varepsilon_1, x),$$
$$y_2 = \varepsilon_2.$$

[8] This result is an extension of Lemma 1 in Matzkin (2001) pertaining to the Roy Model.

Furthermore, assume that this observable characteristic has a positive effect on Sector 1 income, i.e., $\partial v^1(\varepsilon_1, x)/\partial x > 0$. Intuitively, this characteristic corresponds to an attribute only valuable for Sector 1 (e.g. wealth in the case of entrepreneurs, an observable sector specific skill, etc.).

As in the more general model, the observables are summarized by two functions $F^1(y,x)$ and $F^2(y,x)$ giving the fraction of people that are in Sector 1 (Sector 2) and have income less than y conditional on x (see equations (1) and (2)).

The model makes no predictions about the effect of the attribute on the function $F^1(y,x)$. $F^1(y,x)$ can increase or decrease as I look at people with better attributes. This is the case since there are two offsetting effects. On one hand, an increase in the attribute causes those that were already working in Sector 1 to earn a higher income. I refer to this effect as the *positive direct effect*. On the other hand, with an increase in the attribute, more people are going to choose Sector 1 at all income levels. These are people that were indifferent between Sector 1 and Sector 2. Moreover, since individuals are heterogeneous in their Sector 1 and Sector 2 abilities, there will be individuals that are indifferent at various Sector 1 income levels. I refer to the second effect as the *negative selection effect*.

Formally:

$$-\frac{\partial F^1(y,x)}{\partial x} = \underbrace{-\frac{\partial h^1(y,x)}{\partial x}\int_{\underline{\varepsilon}_2}^{y} g\left(h^1(y,x), \varepsilon_2\right) d\varepsilon_2}_{\substack{>0 \\ \text{(positive direct effect)}}} \underbrace{-\int_{\underline{\varepsilon}_1}^{h^1(y_1,x)} \frac{\partial v^1(\varepsilon_1,x)}{\partial x} g\left(\varepsilon_1, v^1(\varepsilon_1,x)\right) d\varepsilon_1}_{\substack{<0 \\ \text{(negative selection effect)}}} \gtreqless 0 \quad \forall y,x \tag{3}$$

The predictions of the model are associated with the function $F^2(y,x)$. In particular, the main prediction of the model is that the function $F^2(y,x)$ increases as the relevant attribute for Sector 1 is raised. Intuitively, as the attribute for Sector 1 increases more people find it profitable to work in Sector 1. Formally:

$$-\frac{\partial F^2(y,x)}{\partial x} = \underbrace{-\int_{\underline{\varepsilon}_2}^{y} \frac{\partial h^1(\varepsilon_2,x)}{\partial x} g\left(h^1(\varepsilon_2,x), \varepsilon_2\right) d\varepsilon_2}_{\substack{>0 \\ \text{positive selection effect}}} > 0 \quad \forall y,x \tag{4}$$

This implication is related to the natural prediction that the probability of choosing Sector 1 is increasing in Sector 1's attribute.[9] It says that there will be fewer people working in Sector 2 at all income strata.

The positive selection effect in (4) is the mirror image of the negative selection effect in (3). This follows from the fact that the individuals who change sectors in response to a small change in the attribute are marginal agents, i.e. they have the same income in Sector 1 as in Sector 2. Therefore, the number of people entering Sector 1 in a particular income bracket equal the number of people leaving the same income bracket for Sector 2. The following Lemma states this result.

Lemma 1:

$$\int_{\underline{\varepsilon}_1}^{h^1(y_1,x)} \frac{\partial v^1(\varepsilon_1,x)}{\partial x} g\left(\varepsilon_1, v^1(\varepsilon_1,x)\right) d\varepsilon_1 = -\int_{\underline{\varepsilon}_2}^{y} \frac{\partial h^1(\varepsilon_2,x)}{\partial x} g\left(h^1(\varepsilon_2,x), \varepsilon_2\right) d\varepsilon_2$$

This implies the second testable implication of the model. The distribution of the maximum income is decreasing in x (i.e. the distribution improves when considering people with better attributes).

Lemma 2:

$$\begin{aligned}\frac{\partial \Pr\left(\max\{y_1,y_2\} \le y|x\right)}{\partial x} &= \frac{\partial h^1(y,x)}{\partial x} \int_{\underline{\varepsilon}_2}^{y} g\left(h^1(y,x), \varepsilon_2\right) d\varepsilon_2 \\ &< 0.\end{aligned}$$

Proof. See appendix. ■

Lemma 2 suggests a method of learning the value function $v^1(\varepsilon_1,x)$ independently of the joint distribution of the abilities. This is make clear in the following lemma:

Lemma 3:

$$\begin{aligned}\frac{\partial v^1\left(h^1(y,x),x\right)}{\partial x} &= -\frac{\frac{\partial h^1(y,x)}{\partial x}}{\frac{\partial h^1(y,x)}{\partial y}} \\ &= -\frac{\frac{\partial F^1(y,x)}{\partial x} + \frac{\partial F^2(y,x)}{\partial x}}{\frac{\partial F^1(y,x)}{\partial y}}\end{aligned}$$

Proof. The first equality follows from the Implicit Function Theorem. The second

[9] In particular, $P(\text{sector } 1|x) = 1 - \lim_{y\to\infty} F^2(y|x)$

follows from Lemma 1 and that $\frac{\partial F^1(y,x)}{\partial y} = \frac{\partial h^1(y,x)}{\partial y} \int_{\underline{\varepsilon}_2}^{y} g\left(h^1\left(y,x\right),\varepsilon_2\right) d\varepsilon_2$. ■

Section 5 presents a generalizatio of Lemma 3. Results on the existence and uniqueness of partial differential equations are used to study the identification and testable implications of the general model.

4 General Results

Five results are presented in this section. First, I prove Lemma 4, a generalization of Lemma 3, which gives a partial characterization of the mapping describing the relation between the structure of a general occupational choice model and the observables (i.e., equations (1) and (2)). In particular, I derive a system of PDE that must be satisfied by any pair of functions v^1 and v^2 rationalizing the observables F^1 and F^2. I then study this system of PDE. Proposition 2 shows that the model with additive heterogeneity is non-parametrically identified without any exclusion restrictions. Propositions (3)-(5) present identification results for the case of non-additive heterogeneity and establish the testable implications of the general Roy model. Finally, Proposition 6 shows that the joint distribution of abilities is identified as long as there is "enough" variation within the observable characteristics of the agents.

4.1 A Useful Characterization of the Relation Between the Structure and the Obsevables

The goal is to obtain a relation between a subset of the structure and the observables that will prove useful for model identification. The main idea is very simple and can be described using figure 1. In terms of figure 1, the integral in (1) corresponds to the integral of $g\left(\varepsilon_1,\varepsilon_2\right)$ over the lower "triangle" (i.e., the integral between the curve $h^2\left(v^1\left(\varepsilon_1,x\right),x\right)$, $h^1\left(y,x\right)$ and the ε_1-axis). Similarly, the integral in (2) corresponds to the area of the upper "triangle". The key step is to realize that the derivatives of $F^1\left(y,x\right)$ and $F^2\left(y,x\right)$ contain common elements. In particular, the derivative of $F^1\left(y,x\right)$ and $F^2\left(y,x\right)$ involve the area over the two dotted segments in figure 1 and over the indifference segment (i.e. $h^2\left(v^1\left(\varepsilon_1,x\right),x\right)$, see Lemma 1). When viewed as a function of these areas, the derivatives of $F^1\left(y,x\right)$ and $F^2\left(y,x\right)$ define a system of 4 equations in 3 unknowns. For this system to be consistent a partial differential equation must be satisfied between the functions $h^1\left(y,x\right)$ and $h^2\left(y,x\right)$

and the observables, i.e. equation (5) must be satisfied. This is formally stated in the following Lemma:

Lemma 4: *The following are necessary conditions for a pair of functions* $v^1(\varepsilon_1, x)$ *and* $v^2(\varepsilon_2, x)$ *to rationalize the observables given by the functions* $F^1(y_1, x)$ *and* $F^2(y_2, x)$*:*

$$\frac{h^1_{x_l}(y,x)}{h^1_{y_1}(y,x)}F^1_{y_1}(y,x)+\frac{h^2_{x_l}(y,x)}{h^2_{y_2}(y,x)}F^2_{y_2}(y,x)=F^1_{x_l}(y,x)+F^2_{x_l}(y,x) \quad \text{all } l=1,...,k. \tag{5}$$

where $h^1_{x_l}(y,x)=\frac{\partial h^1(y,x)}{\partial x_l}$, ..., $F^1_{y_1}(y,x)=\frac{\partial F^1(y,x)}{\partial y_1}$,....

Proof. See appendix. ■

To highlight the usefulness of the above result, I next consider how the identification and testable implication of versions of the model of entrepreneurship with borrowing constraints introduced in section 2 can be studied using equation (5).

Example 1: Consider the partial generalization of the model in Evans and Jovanovic (1989): $f(\varepsilon_1, k) = \varepsilon_1 k^\alpha$ and $c(\varepsilon_1, k, b, r) = k - \lambda b.$ Assume that the econometrician only observes variation in agents' wealth levels (e.g. a single cross-section of agents facing a common interest rate is observed).

In this example the system of PDE in (5) simplifies to:

$$r\lambda(1-\alpha) - \frac{y}{b}\alpha = -\frac{F^1_b(y,b)+F^2_b(y,b)}{F^1_{y_1}(y,b)}$$

For fixed y and b, this is a linear equation in $\lambda(1-\alpha)$ and α. Therefore, as long as two points (y,b) and (y',b') with $\frac{y}{b} \neq \frac{y'}{b'}$ are observed, we have two equations in two unknowns that can be uniquely solved for λ and α. This model imposes many testable restrictions on the data represented by F^1 and F^2.

The identification relies on very strong assumptions about the structure of technologies and borrowing constraints. I illustrate this point with the following example.

Example 2: Assume the true model is given by the following structure: $f(\varepsilon_1, k) = \varepsilon_1 k^\alpha$ and $c(b) = \lambda b,$ i.e. the observables are generated using this structure. Also, assume the econometrician only knows that the borrowing constraint is of the linear type, i.e. $c(b) = \lambda' b$, but has no information about the production function, i.e. it can be of the general type $f(\varepsilon, k)$. Then, production functions of the type $f(\varepsilon_1, k) =$

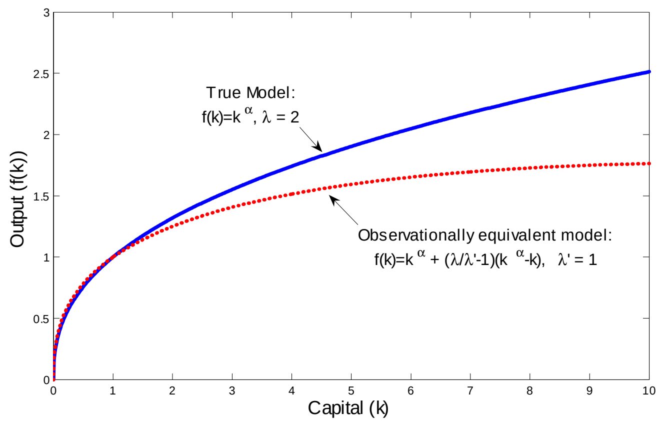

Figure 2: Production Function, True Model (solid) and Observationally Equivalent Model (dashed)

$\varepsilon_1 k^\alpha + r\left(\frac{\lambda-\lambda'}{\lambda'}\right)(k^\alpha - k)$ and an arbitrary borrowing constraint parameter λ rationalize the data.

Figure (2) illustrates this example. A model with a tighter borrowing constraint, $\lambda' = 1$, and technology with stronger decreasing returns to scale is observationally equivalent to the true model, $\lambda = 2$. See the Appendix for a detail analysis of this example.[10]

4.2 Identification and Testable Implications

In this section, I study the identification and testable implications of the Roy Model by applying results on existence and uniqueness results of solutions to systems of PDE. I first show that the model with additive heterogeneity is non-parametrically identified without any exclusion restrictions. I then study the non-additive model. Identification on the non-additive model is obtained using exclusion restriction. It also shown that for the model to have testable implications more than two individual characteristics (or two variables that only affect the income of one of the sectors) need to be observed.

[10] If minor assumptions about the monotonicity or the concavity of technologies were to be imposed, it would be possible to identify the model.

4.2.1 Additive Heterogeneity

I consider first the Roy model with additive unobservable characteristics (e.g., Heckman and Honore (1990)). In particular, let the function describing the income that an agent with sector i specific ability ε_i and covariates x can earn in sector i take the following separable in unobservable specification:

$$v^i(\varepsilon_i, x) = g^i(x) + \varepsilon_i, i = 1, 2. \tag{6}$$

The following proposition proves the non-parametric identification of this model.
Proposition 2: *If i) functions $F^i(y, x)$, $i = 1, 2$, are observed, ii) there exist two values of income y and y' such that the following matrix has full rank:*

$$\begin{pmatrix} F_y^1(y, x) & -F_y^2(y, x) \\ F_y^1(y', x) & -F_y^2(y', x) \end{pmatrix}, \tag{7}$$

and iii) the value of the functions $g^i(\bar{x})$ is known at a point $\bar{x}$, then the functions $g^i(x)$, $i = 1, 2$, are non-parametrically identified.

This is an striking example of the generality of results that are obtained with this method. The nonparametric identification of the model with additive heterogeneity is obtained without any exclusion restriction! Furthermore, notice the rank condition on (7) is generically satisfied.[11]

This model have many testable implications, i.e., restrictions on the admissible data F^1 and F^2.

4.2.2 Non-additive Heterogeneity

Next, we study the identification and testable implications of the non-additive model. Exclusion restrictions are required for the non-parametric identification of the non-additive model. The different propositions illustrate how different results on existence and uniqueness of solutions to PDE can be used to answer identification questions.

[11] The elements of this matrix are $F_y^1(y, x) = \int_{\underline{\varepsilon}_2}^{y - g^2(x)} g\left(y - g^1(x), \varepsilon_2\right) d\varepsilon_2$ and $F_y^2(y, x) = \int_{\underline{\varepsilon}_1}^{y - g^1(x)} g\left(\varepsilon_1, y - g^2(x)\right) d\varepsilon_1$. For most joint distribution, $g(\varepsilon_1, \varepsilon_2)$, the rank condition will be satisfied. For a pathological example, consider suppose ε_1, ε_2 are distributed independent and uniformly distributed in the unit square, $[0, 1] \times [0, 1]$.

1-dimensional x, $x \in \mathbb{R}$ The following proposition show that a model with a single observable characteristic is identified but has no testable implications. Intuitively, the existence of solution to a first order PDE guarantees that we can always solve for a function $h^1(y,x)$ ($v^1(\varepsilon_1,x)$) that satisfy equation (5), and therefore, implies that the model has not testable implications. Uniqueness of solutions to a single PDE given a boundary condition ($h^1(y,\bar{x}) = y$) guarantees identification.

Proposition 2: *Assume that i) $x \in R$,* ii) $h^1(y,\bar{x}) = y\ \forall y$ *(normalization), $h^2(y,x) = \varphi^2(y,x)\ \forall y,x$ where φ^2 is a known function (e.g., $\varphi^2(y,x) = y$, i.e., normalization plus an exclusion restriction), and* iv) *F^1, F^2 and $\varphi^2 \in C^1$;*[12] *then a) the model has no testable implications, and b) the function $v^1(\varepsilon_1,x)$ is locally identified.*

Proof. See appendix. ■

2-dimensinal x, x $\in \mathbb{R}^2$ When the dimension of the vector of observable characteristics x is larger than one then (5) correspond to a system of PDE. Existence of solutions of systems of PDE is not guarantee unless strong restriction on the coefficient of the equations are imposed. This conditions are the testable implications of the theory. I present two results for the case x is a 2-dimensional vector, $x \in R^2$. First, I study the case in which one of the variables affects exclusively Sector 1 and the other affects exclusively Sector 2.[13]

Proposition 3: *Assume that: i) $x \in R^2$,* ii) *$h^1(y,\bar{x}_1,x_2) = \varphi^1(y,x_2)\ \forall y,x_2$ (φ^1 known) (e.g., $\varphi^1(y,x_2) = y$, i.e. normalization and exclusion restrictions),iii) $h^2(y,x_1,\bar{x}_2) = \varphi^2(y,x_1)\ \forall y,x_1$ (φ^2 known) (e.g., $\varphi^2(y,x_1) = y$, i.e. normalization and exclusion restrictions), iv) F^1, F^2, φ^1, and φ^2 are analytic functions; then, a) the model has no testable implications, b) the functions $v^1(\varepsilon_1,x)$ and $v^2(\varepsilon_2,x)$ are locally identified.*

Proof. In this case the system of equations in (5) simplifies to a system of two PDEs in two unknown, $h^1(y,x)$ and $h^2(y,x)$. Existence and uniqueness of solutions given the boundary conditions, $h^2(y,x_1,\bar{x}_2) = \varphi^2(y,x_1)$ and $h^1(y,\bar{x}_1,x_2) = \varphi^1(y,x_2)$, follows from the Cauchy-Kovalevska Theorem (see for example Krzyzanski (1971)). ■

Next, I consider the case in which x only affects Sector 1, i.e. x_1 and x_2 only affect the income in the first sector. In this case, the model has testable implications.

[12] C^n corresponds to the space of n times continuously differentiable functions.

[13] More generally, we have information about h^1 and h^2 in a one-dimensional subspace of its domain.

Proposition 4: *Assume that: i) $x \in R^2$, ii) $h^1(y, \bar{x}) = y$ (normalization), iii) $h^2(y, x) = \varphi^2(y, x)$ (φ^2 known) (e.g. $\varphi^2(y, x) = y$, i.e. normalization and exclusion restrictions), iv) F^1, F^2, and $\varphi^2 \in C^\infty$; then, a) the model implies the following testable restriction on the observables*

$$\left(F^1_{x_1} - F^2_{x_1}\right) F^2_{x_2 y} = \left(F^1_{x_2} - F^2_{x_2}\right) F^2_{x_1 y}, \tag{8}$$

and the function $v^1(\varepsilon_1, x)$ is locally identified.

Proof:. In this case the system of equations in (5) simplifies to a system of two PDEs in one unknown, $h^1(y, x)$,

$$\begin{aligned} h^1_{x_1}(y, x) &= \left[\frac{F^1_{x_1} - F^2_{x_1}}{F^1_{y_1}} + \frac{h^2_{x_1}(y, x)}{h^2_{y_2}(y, x)} \frac{F^2_{y_2}}{F^1_{y_1}}\right] h^1_{y_1}(y, x) \\ h^1_{x_2}(y, x) &= \left[\frac{F^1_{x_2} - F^2_{x_2}}{F^1_{y_1}} + \frac{h^2_{x_2}(y, x)}{h^2_{y_2}(y, x)} \frac{F^2_{y_2}}{F^1_{y_1}}\right] h^1_{y_1}(y, x) \end{aligned}$$

Frobenius' Theorem can be used to prove that if the condition on the observables is satisfied (i.e. $\left(F^1_{x_1} - F^2_{x_1}\right) F^2_{x_2 y} = \left(F^1_{x_2} - F^2_{x_2}\right) F^2_{x_1 y}$), then there exists a unique function satisfying the system of PDEs and the boundary conditions. The testable implication is a condition for the two equations to commute, i.e., for $h^1_{x_1 x_2} = h^1_{x_2 x_1}$. ■

What is particular about this case is that the system of PDE is overdetermined, i.e., there are two equations and one unknown function $h^1(y, x)$. A solution to this system of PDE will exist only if the coefficient of this system of equations satisfy condition (8). In general, the additional conditions required for the system of equations to be consistent provide additional information that can be used to obtain stronger identification results (e.g., Chiappori and Ekeland (2002)). For the case of the Roy model, the extra condition does not involve the unknown function $h^1(y, x)$, and therefore, do not provide additional information that can be used for identifying the structure of the model.

For the case $x \in \mathbb{R}^k$, $k > 2$, the system of equations (5) will be overdetermined, implying a system of equations that must be satisfied by data to be rationalized by a Roy model. Identification is granted provided the function $h^2(y, x)$ is known.

4.2.3 Identifying the Joint Distribution of Abilities

Once functions $h^i(y,x)$ have being identified using equation (5) and the boundary information (e.g., $h^1(y,\bar{x}) = y$), the observed function $F^i(y,x)$ can be used to non-parametrically identified the joint distribution of the unobserved heterogeneity, $g(\varepsilon_1, \varepsilon_2)$.

Proposition 5 (Heckman & Honore (1990)): *Assume that in a neighborhood of the point $(\bar{y}, \bar{x})$ the following matrix has full rank:*

$$\begin{pmatrix} h^1_{y_1}(y,x) & h^2_{y_2}(y,x) \\ h^1_{x_l}(y,x) & h^2_{x_l}(y,x) \end{pmatrix}. \tag{9}$$

Then, the joint distribution of abilities, $G(\varepsilon_1, \varepsilon_2)$, is identified in the neighborhood of the point $(h^1(\bar{y},\bar{x}), h^2(\bar{y},\bar{x}))$.

Proof. Functions $F^1(y,x)$ and $F^2(y,x)$ are observed. In particular, the following function is observed

$$\begin{aligned} F^1(y,x) + F^2(y,x) &= \int_{\underline{\varepsilon}_1}^{h^1(y,x)} \int_{\underline{\varepsilon}_2}^{h^2\left(v^1(\varepsilon_1,x),x\right)} g(\varepsilon_1,\varepsilon_2)\, d\varepsilon_2 d\varepsilon_1 \\ &+ \int_{\underline{\varepsilon}_2}^{h^2(y,x)} \int_{\underline{\varepsilon}_1}^{h^1\left(v^2(\varepsilon_2,x),x\right)} g(\varepsilon_1,\varepsilon_2)\, d\varepsilon_1 d\varepsilon_2 \\ &= G\left(h^1(y,x), h^2(y,x)\right) \end{aligned}$$

The rank condition on (9) guarantees that the image of the vector $(h^1(y,x), h^2(y,x))$ is a 2-dimensional subset of R^2. ■

5 Application: Identifying the Structure of an Entrepreneurial Choice Model

In this section, I apply the results of section 4 to study the identification of the entrepreneurial choice model with borrowing constraints introduced at the beginning of the paper.

Example 4: Consider the general version of the entrepreneurial choice model with borrowing constraints introduced earlier. The structure of the entrepreneur's problem is given by the production function $f(\varepsilon_1, k)$ and the borrowing constraint function $c(\varepsilon_1, b, r)$.

Proposition 3 guarantees that profit function of an entrepreneur, $v^1(\varepsilon_1, b, r)$, is identified up to a normalization, e.g. $v^1\left(\varepsilon_1, \bar{b}, \bar{r}\right) = \varepsilon_1$. However, I still need to establish the relation between the profit function and the underling structure given by the functions $f(\varepsilon_1, k)$ and $c(\varepsilon_1, b, r)$.

From the definition of the profit function, I know that the following equations hold:

$$v_b^1(\varepsilon_1, b, r) = [f_k(\varepsilon_1, c(\varepsilon_1, b.r)) - r]\, c_b(\varepsilon_1, b, r)$$
$$v_r^1(\varepsilon_1, b, r) = [f_k(\varepsilon_1, c(\varepsilon_1, b.r)) - r]\, c_r(\varepsilon_1, b, r) - c(\varepsilon_1, b, r)$$

Substituting for $f_k(\varepsilon_1, c(\varepsilon_1, b.r)) - r$ I obtain a single PDE in the unknown function $c(\varepsilon_1, b, r)$:

$$v_r^1(\varepsilon_1, b, r) - v_b^1(\varepsilon_1, b, r)\frac{k_r(\varepsilon_1, b, r)}{k_b(\varepsilon_1, b, r)} + k(\varepsilon_1, b, r) = 0. \tag{10}$$

Then, the borrowing constraint function is identified up to a function of two variables, e.g. $c(\varepsilon_1, b, \bar{r}) = \phi(\varepsilon_1, b)$.

Next, I discuss two particular cases: $c(\varepsilon_1, b, r) = \tilde{c}(b, r)$ and $c(\varepsilon_1, b, r) = \tilde{c}(b)$.

Example 4.a: Assume that the borrowing constraint is not a function of the agents' abilities, i.e. $c(\varepsilon_1, b, r) = \tilde{c}(b, r)$.

In this example, (10) simplifies to:

$$v_r^1(\varepsilon_1, b, r) - v_b^1(\varepsilon_1, b, r)\frac{\tilde{c}_r(b, r)}{\tilde{c}_b(b, r)} + \tilde{c}(b, r) = 0. \tag{11}$$

Proposition 6 characterize this PDE.

Proposition 6: *Assume that:*

(a) $c(\varepsilon_1, b, r) = \tilde{c}(b, r)$,

(b) $v_{b\varepsilon_1}^1(\varepsilon_1, b, r) \neq 0$.

Then, $f(\varepsilon_1, k)$ and $\tilde{c}(b, r)$ are identified. Furthermore, the maximum scale is given by the following expression:

$$\tilde{c}(b, r) = -v_r^1(\varepsilon_1, b, r) + \frac{v_{r\varepsilon_1}^1(\varepsilon_1, b, r)}{v_{b\varepsilon_1}^1(\varepsilon_1, b, r)} v_b^1(\varepsilon_1, b, r)$$

Next, I consider a more restrictive case.

Example 4.b: Assume that the borrowing constraint is not a function of the agents' abilities, i.e. $c(\varepsilon_1, b, r) = \tilde{c}(b)$.

In this example, (10) simplifies to:

$$v_r^1(\varepsilon_1, b, r) - k(b) = 0.$$

In this example, the maximum scale is clearly identified.

6 Conclusion

I study the identification and testable implications of the Roy model. I generalize the results in the literature in three respects. First, I allow the functions characterizing the income associate with different choices to be fully general. Second, I prove identification using strictly weaker exclusion restrictions. Proposition 2 is an striking example of these generalizations. There, I prove the non-parametric identification of the model with additive heterogeneity does not require any exclusion restrictions! Finally, testable implications of a general non-additive model are derived.

The analysis exploits the information contained in equation (5), a necessary condition implied by the Roy model. This equation is derived by studying the local implications of the model. Nevertheless, the question of whether equation (5) provides a sufficient condition for the model to rationalize the data remains open. Global implications of the Roy model have proven useful in identifying the structure of the model in the past, e.g., identification at infinity arguments used in Heckman (1990). Further research is still required to obtain a complete characterization of the implications of the Roy model and fully understand the limits of point identification.

I apply the general results to analyze the identification of a model of entrepreneurial choice with liquidity constraints. In doing so, I deepen the understanding of the validity of structural estimation of this model, which has been done by various authors in the literature of entrepreneurship with borrowing constraints (e.g. Evans and Jovanovic (1989), Paulson and Townsend (2001)). In particular, the analysis in this paper implies that assumptions on the distribution of unobservables were not key to identifying the model. On the other hand, assumptions on the technologies or the form of the borrowing constraints were important, depending on the information observed.

An obvious next step is to investigate if the conclusions of the empirical works previously cited are robust to the functional forms that were used: was the normality assumption important in the results of Evans and Jovanovic (1989)? Is a linear borrowing constraint specification consistent with the data? Can more be learned about the shape of borrowing constraints? I leave these questions for future research.

A Proof of the Results in the Paper

Proof of Lemma 2.

$$\Pr\left(\max\left\{y_1, y_2\right\} \leq y|x\right) = F^1\left(y, x\right) + F^2\left(y, x\right).$$

Therefore,

$$\begin{aligned}
\frac{\partial \Pr\left(\max\left\{y_1, y_2\right\} \leq y|x\right)}{\partial x} &= \frac{\partial F^1\left(y, x\right)}{\partial x} + \frac{\partial F^2\left(y, x\right)}{\partial x} \\
&= \frac{\partial h^1\left(y, x\right)}{\partial x} \int_{\underline{\varepsilon}_2}^{y} g\left(h^1\left(y, x\right), \varepsilon_2\right) d\varepsilon_2 \\
&+ \int_{\underline{\varepsilon}_1}^{h^1(y_1, x)} \frac{\partial v^1\left(\varepsilon_1, x\right)}{\partial x} g\left(\varepsilon_1, v^1\left(\varepsilon_1, x\right)\right) d\varepsilon_1 \\
&+ \int_{\underline{\varepsilon}_2}^{y} \frac{\partial h^1\left(\varepsilon_2, x\right)}{\partial x} g\left(h^1\left(\varepsilon_2, x\right), \varepsilon_2\right) d\varepsilon_2
\end{aligned}$$

using Lemma 1

$$\begin{aligned}
\frac{\partial \Pr\left(\max\left\{y_1, y_2\right\} \leq y|x\right)}{\partial x} &= \frac{\partial F^1\left(y, x\right)}{\partial x} + \frac{\partial F^2\left(y, x\right)}{\partial x} \\
&= \frac{\partial h^1\left(y, x\right)}{\partial x} \int_{\underline{\varepsilon}_2}^{y} g\left(h^1\left(y, x\right), \varepsilon_2\right) d\varepsilon_2 < 0
\end{aligned}$$

■

Proof of Lemma 4:. The mapping defining the model is given by:

$$F^1\left(y_1, x\right) = \int_{\underline{\varepsilon}_1}^{h^1(y_1, x)} \int_{\underline{\varepsilon}_2}^{h^2\left(v^1(\varepsilon_1, x), x\right)} g\left(\varepsilon_1, \varepsilon_2\right) d\varepsilon_2 d\varepsilon_1 \tag{12}$$

$$F^2(y_2, x) = \int_{\underline{\varepsilon}_2}^{h^2(y_2,x)} \int_{\underline{\varepsilon}_1}^{h^1\left(v^2(\varepsilon_2,x),x\right)} g(\varepsilon_1, \varepsilon_2)\, d\varepsilon_1 d\varepsilon_2 \tag{13}$$

The goal is to obtain a relation between a subset of the structure and the observables that will prove useful for identification of the model. The main idea is very simple and can be described using figure 1. In terms of figure 1, the integral in (12) corresponds to the integral of $g(\varepsilon_1, \varepsilon_2)$ over the lower triangle (i.e., the integral between the curve $h^2(v^1(\varepsilon_1, x), x)$, $h^1(y, x)$ and the ε_1-axis). Similarly, the integral in (13) corresponds to the integral over the upper triangle. The key step is to realize that the derivatives of $F^1(y, x)$ and $F^2(y, x)$ contain common elements. In particular, the derivative of $F^1(y, x)$ and $F^2(y, x)$ involve the area over the two dotted segments in figure 1 and over the indifference segment (i.e. $h^2(v^1(\varepsilon_1, x), x)$). When viewed as a function of these areas, the derivatives of $F^1(y|x)$ and $F^2(y|x)$ define a system of 4 equations in 3 unknowns. Therefore, for this system to be consistent a partial differential equation must be satisfied between the functions $h^1(y, x)$ and $h^2(y, x)$ and the observables, i.e. equation (5) must be satisfied.

Locally, the system can be described by the following equations:

$$F^1_{y_1}(y_1,x) = h^1_{y_1}(y_1,x)\int_{\underline{\varepsilon}_2}^{h^2(y_1,x)} g\left(h^1(y_1,x),\varepsilon_2\right)d\varepsilon_2 \tag{14}$$

$$\begin{aligned} F^1_{x_l}(y_1,x) &= h^1_{x_l}(y_1,x)\int_{\underline{\varepsilon}_2}^{h^2(y_1,x)} g\left(h^1(y_1,x),\varepsilon_2\right)d\varepsilon_2 \\ &+ \int_{\underline{\varepsilon}_1}^{h^1(y_1,x)} h^2_{y_2}\left(v^1(\varepsilon_1,x),x\right)v^1_{x_l}(\varepsilon_1,x)\,g\left(\varepsilon_1,h^2\left(v^1(\varepsilon_1,x),x\right)\right)d\varepsilon_1 \\ &+ \int_{\underline{\varepsilon}_1}^{h^1(y_1,x)} h^2_{x_l}\left(v^1(\varepsilon_1,x),x\right)g\left(\varepsilon_1,h^2\left(v^1(\varepsilon_1,x),x\right)\right)d\varepsilon_1 \end{aligned} \tag{15}$$

$$F^2_{y_2}(y_2,x) = h^2_{y_2}(y_2,x)\int_{\underline{\varepsilon}_1}^{h^1(y_2,x)} g\left(\varepsilon_1,h^2(y_2,x)\right)d\varepsilon_1 \tag{16}$$

$$\begin{aligned} F^2_{x_l}(y_2,x) &= h^2_{x_l}(y_2,x)\int_{\underline{\varepsilon}_1}^{h^1(y_2,x)} g\left(\varepsilon_1,h^2(y_2,x)\right)d\varepsilon_1 \\ &+ \int_{\underline{\varepsilon}_2}^{h^2(y_2,x)} h^1_{y_1}\left(v^2(\varepsilon_2,x),x\right)v^2_{x_i}(\varepsilon_2,x)\,g\left(h^1\left(v^2(\varepsilon_2,x),x\right),\varepsilon_2\right)d\varepsilon_2 \\ &+ \int_{\underline{\varepsilon}_2}^{h^2(y_2,x)} h^1_{x_l}\left(v^2(\varepsilon_2,x),x\right)g\left(h^1\left(v^2(\varepsilon_2,x),x\right),\varepsilon_2\right)d\varepsilon_2. \end{aligned} \tag{17}$$

Using the last two terms of equation (17), the integral is taken along the curve $(h^1(v^2(\varepsilon_2,x),x),\varepsilon_2)$ (i.e., over the $(\varepsilon_1,\varepsilon_2)$ pairs, such that in $v^1(\varepsilon_1,x) = v^2(\varepsilon_2,x)$) we can use a change in the variables of integration, $d\varepsilon_2 = h^2_{y_2}(v^1(\varepsilon_1,x),x)\,v^1_{\varepsilon_1}(\varepsilon_1,x)\,d\varepsilon_1$, and rewrite equation (17) as:

$$\begin{aligned}
&\int_{\underline{\varepsilon}_2}^{h^2(y_2,x)} h^1_{y_1}\left(v^2\left(\varepsilon_2,x\right),x\right) v^2_{x_l}\left(\varepsilon_2,x\right) g\left(h^1\left(v^2\left(\varepsilon_2,x\right),x\right),\varepsilon_2\right) d\varepsilon_2 \\
&= \int_{\underline{\varepsilon}_1}^{h^1(y,x)} h^1_{y_1}\left(v^1\left(\varepsilon_1,x\right),x\right) v^2_{x_l}\left(h^2\left(v^1\left(\varepsilon_1,x\right),x\right),x\right) g\left(\varepsilon_1,h^2\left(v^1\left(\varepsilon_1,x\right),x\right)\right) \\
&h^2_{y_2}\left(v^1\left(\varepsilon_1,x\right),x\right) v^1_{\varepsilon_1}\left(\varepsilon_1,x\right) d\varepsilon_1 \\
&= \int_{\underline{\varepsilon}_1}^{h^1(y,x)} \frac{1}{v^1_{\varepsilon_1}\left(\varepsilon_1,x\right)} v^2_{x_l}\left(h^2\left(v^1\left(\varepsilon_1,x\right),x\right),x\right) g\left(\varepsilon_1,h^2\left(v^1\left(\varepsilon_1,x\right),x\right)\right) \\
&\frac{1}{v^2_{\varepsilon_2}\left(h^2\left(v^1\left(\varepsilon_1,x\right),x\right),x\right)} v^1_{\varepsilon_1}\left(\varepsilon_1,x\right) d\varepsilon_1 \\
&= \int_{\underline{\varepsilon}_1}^{h^1(y,x)} \frac{v^2_{x_l}\left(h^2\left(v^1\left(\varepsilon_1,x\right),x\right),x\right)}{v^2_{\varepsilon_2}\left(h^2\left(v^1\left(\varepsilon_1,x\right),x\right),x\right)} g\left(\varepsilon_1,h^2\left(v^1\left(\varepsilon_1,x\right),x\right)\right) \\
&= -\int_{\underline{\varepsilon}_1}^{h^1(y,x)} h^2_{x_l}\left(v^1\left(\varepsilon_1,x\right),x\right) g\left(\varepsilon_1,h^2\left(v^1\left(\varepsilon_1,x\right),x\right)\right).
\end{aligned}$$

The first equality follows from the change in variables along the $h^2\left(v^1\left(\varepsilon_1,x\right),x\right)$ curve (replacing $d\varepsilon_2$ by $h^2_{y_2}\left(v^1\left(\varepsilon_1,x\right),x\right) v^1_{\varepsilon_1}\left(\varepsilon_1,x\right) d\varepsilon_1$, using that along that curve $v^1\left(\varepsilon_1,x\right) = v^2\left(\varepsilon_2,x\right)$ and similarly that $\varepsilon_2 = h^2\left(v^1\left(\varepsilon_1,x\right),x\right)$). The second and fourth equalities use that $h^i_y\left(y,x\right) = \frac{1}{v^i_{\varepsilon_i}\left(h^i(y,x),x\right)}$ and $h^i_{x_l}\left(y,x\right) = -\frac{v^i_{x_l}\left(h^i(y,x),x\right)}{v^i_{\varepsilon_i}\left(h^i(y,x),x\right)}$. The third equality follows from rearranging the existing terms.

Similarly for the third integral in (17):

$$\begin{aligned}
&\int_{\underline{\varepsilon}_2}^{h^2(y_2,x)} h^1_{x_l}\left(v^2\left(\varepsilon_2,x\right),x\right) g\left(h^1\left(v^2\left(\varepsilon_2,x\right),x\right),\varepsilon_2\right) d\varepsilon_2 \\
&= -\int_{\underline{\varepsilon}_1}^{h^1(y,x)} h^2_{y}\left(v^1\left(\varepsilon_1,x\right),x\right) v^1_{x_l}\left(\varepsilon_1,x\right) g\left(\varepsilon_1,h^2\left(v^1\left(\varepsilon_1,x\right),x\right)\right).
\end{aligned}$$

Therefore:

$$\begin{aligned}
F^2_{x_l}(y_2,x) = {}& h^2_{x_l}(y_2,x)\int_{\underline{\varepsilon}_1}^{h^1(y_2,x)} g\left(\varepsilon_1, h^2(y_2,x)\right) d\varepsilon_1 \\
& - \int_{\underline{\varepsilon}_1}^{h^1(y_1,x)} h^2_{x_l}\left(v^1(\varepsilon_1,x),x\right) g\left(\varepsilon_1, h^2\left(v^1(\varepsilon_1,x),x\right)\right) d\varepsilon_1 \\
& - \int_{\underline{\varepsilon}_1}^{h^1(y_1,x)} h^2_{y_2}\left(v^1(\varepsilon_1,x),x\right) v^1_{x_l}(\varepsilon_1,x)\, g\left(\varepsilon_1, h^2\left(v^1(\varepsilon_1,x),x\right)\right) d\varepsilon_1 .
\end{aligned}$$

Defining:

$$\begin{aligned}
a^1(y,x) &= \int_{\underline{\varepsilon}_2}^{h^2(y,x)} g\left(h^1(y_1,x), \varepsilon_2\right) d\varepsilon_2 \\
a^2(y,x) &= \int_{\underline{\varepsilon}_1}^{h^1(y,x)} g\left(\varepsilon_1, h^2(y_2,x)\right) d\varepsilon_1 \\
b^l(y,x) &= \int_{\underline{\varepsilon}_1}^{h^1(y_1,x)} h^2_{y_2}\left(v^1(\varepsilon_1,x),x\right) v^1_{x_l}(\varepsilon_1,x)\, g\left(\varepsilon_1, h^2\left(v^1(\varepsilon_1,x),x\right)\right) d\varepsilon_1 \\
&\quad + \int_{\underline{\varepsilon}_1}^{h^1(y_1,x)} h^2_{x_l}\left(v^1(\varepsilon_1,x),x\right) g\left(\varepsilon_1, h^2\left(v^1(\varepsilon_1,x),x\right)\right) d\varepsilon_1 ,
\end{aligned}$$

we can rewrite the previous system as:

$$\begin{aligned}
F^1_{y_1}(y,x) &= h^1_{y_1}(y,x)\, a^1(y,x) \\
F^2_{y_2}(y,x) &= h^2_{y_2}(y,x)\, a^2(y,x) \\
F^1_{x_l}(y,x) &= h^1_{x_l}(y,x)\, a^1(y,x) + b^l(y,x) \\
F^2_{x_l}(y,x) &= h^2_{x_l}(y,x)\, a^2(y,x) - b^l(y,x) .
\end{aligned}$$

Viewed as a system of equations in $a^1(y,x)$, $a^2(y,x)$, and $b^l(y,x)$, the previous system has $2+k$ linearly independent equations. Then, for any set of functions $a^1(y,x)$, $a^2(y,x)$, and $b^i(y,x)$ (in particular, for any joint distribution of the heterogeneity across the population, $G(\varepsilon_1,\varepsilon_2)$), it imposes k restriction on the functions $h^1(y,x)$ and $h^2(y,x)$ of the form:

$$\frac{h^1_{x_l}(y,x)}{h^1_{y_1}(y,x)} F^1_{y_1}(y,x) + \frac{h^2_{x_l}(y,x)}{h^2_{y_2}(y,x)} F^2_{y_2}(y,x) = F^1_{x_l}(y,x) + F^2_{x_l}(y,x) \quad \text{for all } l = 1,\ldots,k.$$

■

Proof of Proposition 2:. For this case, the system of PDE in (5) simplifies to

$$-g^1_{x_l}(x)\,F^1_{y_1}(y,x) - g^2_{x_l}(x)\,F^2_{y_2}(y,x) = F^1_{x_l}(y,x) + F^2_{x_l}(y,x) \quad l = 1, ..., k.$$

For any pair of values y and y' we obtain two equations that can be solved for $g^1_{x_l}(x)$ and $g^2_{x_l}(x)$, provided that the matrix

$$\begin{pmatrix} F^1_y(y,x) & -F^2_y(y,x) \\ F^1_y(y',x) & -F^2_y(y',x) \end{pmatrix}$$

has full rank. Given information about the value of $g^i(x)$ for some $\bar{x}$, we can extend the function over a neighborhood with Taylor expansion around the point $\bar{x}$. This guarantees uniqueness, and therefore, identification of the functions $g^i(x)$ up to a constants $g^i(\bar{x})$. Existence of solutions is obviously not guarantee unless strong restrictions are imposed on the observed functions $F^i(y,x)$. These conditions correspond to the testable implications of the model. ■

Proof of Proposition 3:. As we mentioned in Section 2.2, a model has no testable implications if we can always find a structure that rationalizes the data. In our case, the model is described by a single PDE, so to prove that the model has no testable implications amounts to a proof that there exists a solution to the PDE:

$$\frac{h^1_x(y,x)}{h^1_{y_1}(y,x)}F^1_{y_1}(y,x) + \frac{h^2_x(y,x)}{h^2_{y_2}(y,x)}F^2_{y_2}(y,x) = F^1_x(y,x) + F^2_x(y,x) \tag{18}$$

that also satisfy the boundary condition:

$$h^1(y,\bar{x}) = \varphi^1(y). \tag{19a}$$

This question is tackled in the first and second steps of the proof. The problem of identification corresponds to the question of uniqueness of a solution to the above problem. In the third step, I prove identification.

The basic idea is to use the method of characteristics to construct a candidate solution by solving the ordinary differential equations given by the characteristics equations of the PDE (5). These are standard arguments in the PDE literature. (e.g., see Theorem 1.1 in Tran (2000)). ■

Proof of Proposition 6:. Consider two ability levels ε_1 and ε_1' such that $v_b^1(\varepsilon_1, b, r) \neq v_b^1(\varepsilon_1', b, r)$. From (11) we obtain two independent equations:

$$\begin{aligned} v_r^1(\varepsilon_1, b, r) - v_b^1(\varepsilon_1, b, r)\frac{c_r(b,r)}{c_b(b,r)} + c(b,r) = 0 \\ v_r^1(\varepsilon_1', b, r) - v_b^1(\varepsilon_1', b, r)\frac{c_r(b,r)}{c_b(b,r)} + c(b,r) = 0, \end{aligned}$$

that can be solved uniquely for $\frac{c_r(b,r)}{c_b(b,r)}$ and $c(b,r)$. In particular, we obtain

$$c(b,r) = \frac{-v_b^1(\varepsilon_1', b, r)\, v_r^1(\varepsilon_1, b, r) + v_b^1(\varepsilon_1, b, r)\, v_r^1(\varepsilon_1', b, r)}{-v_b^1(\varepsilon_1, b, r) + v_b^1(\varepsilon_1', b, r)}.$$

Summing and substracting $v_b^1(\varepsilon_1, b, r)\, v_r^1(\varepsilon_1, b, r)$ from the numerator of the right hand side we obtain:

$$\begin{aligned} c(b,r) &= \frac{-v_b^1(\varepsilon_1', b, r)\, v_r^1(\varepsilon_1, b, r) + v_b^1(\varepsilon_1, b, r)\, v_r^1(\varepsilon_1', b, r)}{-v_b^1(\varepsilon_1, b, r) + v_b^1(\varepsilon_1', b, r)} \\ &= \frac{\left[v_b^1(\varepsilon_1, b, r) - v_b^1(\varepsilon_1', b, r)\right] v_r^1(\varepsilon_1, b, r) - v_b^1(\varepsilon_1, b, r)\left[v_r^1(\varepsilon_1, b, r) - v_r^1(\varepsilon_1', b, r)\right]}{-v_b^1(\varepsilon_1, b, r) + v_b^1(\varepsilon_1', b, r)} \\ &= \frac{\left[v_b^1(\varepsilon_1, b, r) - v_b^1(\varepsilon_1', b, r)\right] v_r^1(\varepsilon_1, b, r) - v_b^1(\varepsilon_1, b, r)\left[v_r^1(\varepsilon_1, b, r) - v_r^1(\varepsilon_1', b, r)\right]}{-v_b^1(\varepsilon_1, b, r) + v_b^1(\varepsilon_1', b, r)}, \end{aligned}$$

then,

$$\begin{aligned} c(b,r) &= \frac{-v_{b\varepsilon_1}^1(\varepsilon_1, b, r)\, v_r^1(\varepsilon_1, b, r) + v_{r\varepsilon_1}^1(\varepsilon_1, b, r)\, v_b^1(\varepsilon_1, b, r)}{v_{b\varepsilon_1}^1(\varepsilon_1, b, r)} \\ &= -v_r^1(\varepsilon_1, b, r) + \frac{v_{r\varepsilon_1}^1(\varepsilon_1, b, r)}{v_{b\varepsilon_1}^1(\varepsilon_1, b, r)} v_b^1(\varepsilon_1, b, r). \end{aligned}$$

■

B Derivations associated with Example 2

Assume the true model is given by the following structure: $f(k) = k^\alpha$ and $c(b) = \lambda b$, i.e. the observables are generated using this structure. Also, assume the econometrician only knows that the borrowing constraint is of the linear type, i.e. $c(b) = \lambda' b$, but has no information about the production function, i.e. it can be of the general type $f(\varepsilon, k)$. In this example, (5) takes the following form:

$$h_b^1 + a(y,b)\, h_y^1 = 0$$

where $h^1(y,b)$ solve

$$v^1\left(h^1(y,b), b\right) = f\left(h^1(y,b), \lambda b\right) - r\lambda b$$

and

$$\begin{aligned} a(y,b) &= -\frac{F_b^1(y,b) + F_b^2(y,b)}{F_y^1(y,b)} \\ &= \lambda\left[(y + r\lambda b)\,\alpha\,(\lambda b)^{-1} - r\right]. \end{aligned}$$

This give a simple linear PDE that can be solve analytically by the method of characteristics.

The characteristic curve is given by the following ODE:

$$\frac{\partial y}{\partial b} = \alpha\frac{y}{b} - r\lambda(1-\alpha)$$

whose exact solution is:

$$y(b) = C_1 b^{\alpha} - r\lambda b$$

where C_1 is an arbitrary constant that is pinned down by specifying the income level associated with an initial value of wealth, e.g., $y(1/\lambda)$ implying

$$\begin{aligned} y(1/\lambda') &= C_1 (1/\lambda')^{\tilde{\alpha}} - r\lambda(1/\lambda') \\ &\Leftrightarrow \\ C_1 &= y\left(\frac{1}{\lambda'}\right)\left(\frac{1}{\lambda'}\right)^{-\alpha} + r\frac{\lambda}{\lambda'}\left(\frac{1}{\lambda'}\right)^{-\alpha}. \end{aligned}$$

The boundary condition is given by the following normalization of ability:

$$v(\varepsilon, 1/\lambda) = \varepsilon - r$$

The value function is then given by:

$$\begin{aligned} v(\varepsilon, b) &= -r\lambda b + \left((\varepsilon - r) \left(\frac{1}{\lambda'} \right)^{-\alpha} + r\frac{\tilde{\lambda}}{\lambda} \left(\frac{1}{\lambda'} \right)^{-\alpha} \right) b^{\alpha} \\ &= \varepsilon \left(\frac{1}{\lambda'} \right)^{-\alpha} b^{\alpha} + \left(-r \left(\frac{1}{\lambda'} \right)^{-\alpha} + r\frac{\lambda}{\lambda'} \left(\frac{1}{\lambda'} \right)^{-\alpha} \right) b^{\alpha} - r\lambda b \end{aligned}$$

implying

$$\begin{aligned} f(\varepsilon, \lambda' b) &= \varepsilon \left(\frac{1}{\lambda'} \right)^{-\alpha} b^{\alpha} + \left(-r \left(\frac{1}{\lambda'} \right)^{-\alpha} + r\frac{\lambda}{\lambda'} \left(\frac{1}{\lambda'} \right)^{-\alpha} \right) b^{\alpha} - r\lambda b + r\lambda' b \\ &= \varepsilon \left(\frac{1}{\lambda'} \right)^{-\alpha} b^{\alpha} + b^{\alpha} \left(\frac{1}{\lambda'} \right)^{-\alpha} r \left(\frac{\lambda - \lambda'}{\lambda'} \right) - (\lambda - \lambda') rb \end{aligned}$$

or

$$\begin{aligned} f(\varepsilon, k) &= \varepsilon k^{\alpha} + k^{\alpha} r \left(\frac{\lambda - \lambda'}{\lambda'} \right) - \left(\frac{\lambda - \lambda'}{\lambda'} \right) rk \\ &= \varepsilon k^{\alpha} + r \left(\frac{\lambda - \lambda'}{\lambda'} \right) (k^{\alpha} - k). \end{aligned}$$

C Example With Linear Value Functions and Uniform Heterogeneity

Assume that the value function given the income an agent obtains from choosing sector i takes the following form:

$$y_i = v^i(\varepsilon_i, x) = \beta^i x + \varepsilon_i$$

Also assume that the heterogeneity is distributed uniformly in the unit square, $[0,1] \times [0,1]$, i.e.

$$G(\varepsilon_1, \varepsilon_2) = \varepsilon_1 \varepsilon_2, \quad \varepsilon_i \in [0,1].$$

In this case we can solve explicitly for the mapping defining the model:

$$
\begin{aligned}
F^1(y_1|x) &= \int_0^{y_1-\beta_1 x}\int_0^{\beta_1 x+\varepsilon_1-\beta_2 x} d\varepsilon_2 d\varepsilon_1 \\
&= \int_0^{y_1-\beta_1 x} (\beta_1 x+\varepsilon_1-\beta_2 x)\, d\varepsilon_1 \\
&= \frac{1}{2}(y_1-\beta_1 x)^2 + (y_1-\beta_1 x)(\beta_1-\beta_2)\, x \\
F^2(y_2|x) &= \int_{(\beta_1-\beta_2)x}^{y_2-\beta_2 x}\int_0^{\beta_2 x+\varepsilon_2-\beta_1 x} d\varepsilon_1 d\varepsilon_2 \\
&= \int_{(\beta_1-\beta_2)x}^{y_2-\beta_2 x} (\beta_2 x+\varepsilon_2-\beta_1 x)\, d\varepsilon_2 \\
&= \frac{1}{2}(y_2-\beta_2 x)^2 - \frac{1}{2}(\beta_1-\beta_2)^2 x^2 + (y_2-\beta_2 x)(\beta_2-\beta_1)\, x + (\beta_1-\beta_2)^2 x^2 \\
&= \frac{1}{2}(y_2-\beta_2 x)^2 + (y_2-\beta_2 x)(\beta_2-\beta_1)\, x + \frac{1}{2}(\beta_1-\beta_2)^2 x^2.
\end{aligned}
$$

Similarly, we can solve for the relationship between the derivatives of the observables and the structure of the model:

$$
\begin{aligned}
F^1_{y_1}(y_1|x) &= y_1 - \beta_1 x + (\beta_1-\beta_2)\, x \\
&= y_1 - \beta_2 x \\
F^1_x(y_1|x) &= -\beta_1(y_1-\beta_1 x) - \beta_1(\beta_1-\beta_2)\, x \\
&\quad + (y_1-\beta_1 x)(\beta_1-\beta_2) \\
&= -\beta_1(y_1-\beta_2 x) \\
&\quad + y_1(\beta_1-\beta_2) - \beta_1^2 x + \beta_1\beta_2 x \\
F^2_{y_2}(y_2|x) &= y_2 - \beta_2 x + (\beta_2-\beta_1)\, x \\
&= y_2 - \beta_1 x \\
F^2_x(y_2|x) &= -\beta_2(y_2-\beta_2 x) - \beta_2(\beta_2-\beta_1)\, x \\
&\quad + (y_2-\beta_2 x)(\beta_2-\beta_1) + (\beta_1-\beta_2)^2 x \\
&= -\beta_2(y_2-\beta_1 x) \\
&\quad - y_2(\beta_1-\beta_2) + \beta_1^2 x - \beta_1\beta_2 x.
\end{aligned}
$$

In this case, the rank condition in Example 1 will not be satisfied but provided $\beta_1 \neq \beta_2$

for any two point such that $yx' \neq y'x$ a related rank condition is satisfied:

$$\begin{pmatrix} F^1_{y_1}(y|x) & -F^2_{y_2}(y|x) \\ F^1_{y_1}(y'|x') & -F^2_{y_2}(y'|x') \end{pmatrix}$$
$$= \begin{pmatrix} y - \beta_2 x & -(y - \beta_1 x) \\ y' - \beta_2 x' & -(y' - \beta_1 x') \end{pmatrix}.$$

This matrix is full rank iff

$$\det \begin{pmatrix} y - \beta_2 x & -(y - \beta_1 x) \\ y' - \beta_2 x' & -(y' - \beta_1 x') \end{pmatrix} \neq 0$$

i.e., iff

$$\begin{aligned} -(y - \beta_2 x)(y' - \beta_1 x') + (y' - \beta_2 x')(y - \beta_1 x) &\neq 0 \\ y\beta_1 x' + \beta_2 x y' - y'\beta_1 x - \beta_2 x' y &\neq 0 \\ yx'(\beta_1 - \beta_2) - y'x(\beta_1 - \beta_2) &\neq 0 \\ (\beta_1 - \beta_2)(yx' - y'x) &\neq 0. \end{aligned}$$

References

[1] Aghion, P. and P. Bolton (1997) “A Trickle-down Theory of Grwoth and Development.” The Review of Economic Studies, 64, 151-172.

[2] Athey, S. and P. Haile (2002) “Identification of Standard Auction Models”, Econometrica, 70, 2107-2140.

[3] Banerjee, A. and A. Newman (1993) “Occupational Choice and the Process of Development.” Journal of Political Economy, 101, 274-298.

[4] Browning, M. and P. A. Chiappori. "Efficient Intra-Household Allocation : A General Characterization and Empirical Tests" Econometrica, 66 6, 1998, 1241-78.

[5] Buera, Francisco (2005). “A Dynamic Model of Entrepreneurship with Borrowing Constraints: Theory and Evidence.” mimeo, Northwestern University.

[6] Chiappori, P. A. "Collective labour supply and welfare", Journal of Political Economy, 100, 1992, 437-67.

[7] Chiappori, P. A. "Introducing Household Production in Collective Models of Labor Supply", Journal of Political Economy, 105, 1997, 191-209.

[8] Chiappori, P. A. and I. Ekeland (2002) "On Simultaneous Equations." mimeo, University of Chicago.

[9] Evans, D. and B. Jovanovic (1989) "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints", Journal of Political Economy, 97, 808-827.

[10] Gine, X. and R. Townsend " Evaluation of Financial Liberalization: A General Equilibrium Model with Constrained Occupation Choice." forthcoming Journal of Development Economics.

[11] Heckman, J. (1990) "Varieties of Selection Bias", American Economics Review, 80, 313-318.

[12] Heckman, J. and B. Honore (1990) "The Empirical Content of the Roy Model", Econometrica, 58, 1121-49.

[13] Heckman, J. and J. Smith (1998) "Evaluating the Welfare State", in Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium, S. Strom Ed.

[14] Jeong, H. and R. Townsend. "Growth and Inequality: Model Evaluation Based on an Estimation-Calibration Strategy," mimeo, University of Chicago (2003).

[15] Jeong, H. and R. Townsend. "Sources of TFP Growth: Occupational Choice and Financial Deepening," mimeo, University of Chicago (2005).

[16] Krzyzanski, M. (1971) *Partial Differential Equations of Second Order*, Polish Scientific Publishers, Warszawa.

[17] Lloyd-Ellis, H. and D. Bernhardt (2000) "Enterprise, Inequality and Economic Development." The Review of Economic Studies, 67, 147-168.

[18] Manski, C. (1985) "Identification of Binary Response Models", Journal of the American Statistical Association, 83, 729-738.

[19] Matsuyama, K. (2000). "Endogenous Inequality." The Review of Economic Studies, 67, 743-759.

[20] Matzkin, R.L. (1992) "Nonparametric and Distribution-Free Estimation of the Binary Choice and the Threshold-Crossing Models", Econometrica, 58, 137-168.

[21] Matzkin, R. L. (1994) "Restrictions of Economic Theory in Nonparametric Methods", Handbook of Econometrics, Vol. IV.

[22] Matzkin, R. L. (2001) "Nonparametric Estimation of Nonadditive Random Functions." mimeo, Northwestern University.

[23] Paulson A. and R. Townsend (2001) "The Nature of Financial Constraints: Distinguishing the Micro underpinning of Macro Models", mimeo University of Chicago.

[24] Tran, D. (2000) *The Characteristic Method and Its Generalizations for First-Order Nonlinear Partial Differential Equations.*

[25] Tsiastis, A. (1975) "A Nonidentifiability Aspect of the Problem of Competing Risks," *Proceedings of the National Academy of Sciences*, 72, 20-22.