MEASURING THE IMPACT OF FINANCIAL INTERMEDIATION: LINKING CONTRACT THEORY TO ECONOMETRIC POLICY EVALUATION

ROBERT M. TOWNSEND
Massachusetts Institute of Technology

SERGIO S. URZUA
Northwestern University

We study the impact that financial intermediation can have on productivity through the alleviation of credit constraints in occupation choice and/or an improved allocation of risk, using both static and dynamic structural models as well as reduced-form OLS and IV regressions. Our goal in this paper is to bring these two strands of the literature together. Even though, under certain assumptions, IV regressions can accurately recover the true model-generated local average treatment effect, this is quantitatively different, in order of magnitude and even sign, from other policy impact parameters (e.g., ATE and TT). We also show that laying out clearly alternative models can guide the search for instruments. On the other hand, adding more margins of decision, that is, occupation choice and intermediation jointly, or adding more periods with promised utilities as key state variables, as in optimal multiperiod contracts, can cause the misinterpretation of IV as the causal effect of interest.

Keywords: Contract Theory, Financial Intermediation, Econometric Policy Evaluation

1. INTRODUCTION

This paper links contract theory models of financial intermediation with econometric policy evaluation. We study a variety of static and dynamic models in which financial intermediation has an impact on productivity through the alleviation of credit constraints in occupation choice and/or an improved allocation of risk. These models of intermediation are structural choice models that are known in the

268

literature, and, more recently, estimated with cross-sectional or panel data from developing countries (e.g., Thailand). On the other hand, there is a large empirical literature that takes advantage of natural experiments, or instruments, to assess the impact that policy variation and financial institutions are having on incomes, occupations, risk sharing, and a variety of other variables (some also in Thailand). Our goal in this paper is to bring these two strands of the literature together. Even though, under certain assumptions, an instrumental variable (IV) strategy can recover accurately the true model-generated local average treatment effects (LATE), these are quantitatively different, in order of magnitude and even sign, from other policy impact parameters (e.g., treatment on the treated, TT, the average treatment effect, ATE). We also show that clearly laying out alternative models can guide the search for instruments. Mechanism design can deliver natural lotteries or randomization that can be used as a source of identification in empirical analyses. On the other hand, adding more margins of decision, such as occupation choice and intermediation jointly, or adding more periods with promised utilities as key state variables, as in optimal multiperiod contracts, can cause the researcher to lose key identifying assumptions associated with the IV strategy (e.g., uniformity), so that IV and LATE might no longer coincide. Our objective is to help researchers and policy makers assess the impact of financial intermediation accurately.

The models we use are simple models of discrete choice when there are credit constraints. Typically some households are in financial autarky and others in a fully intermediated sector. There is a cost for entering the financial system, and this is imagined to pick up both the actual cost of traveling to a financial institution (bank) and policy distortions that limit access for some agents. We imagine that there is variation in the cost/policy in the data, so that some households are financial sector participants and others are not. Indeed, we can generate cross-sectional or panel data from a given model (sometimes using parameters that have been estimated from emerging market economies) and ask whether those data would allow an accurate quantification of the gain in the population produced by different policy variations (which emerging market countries had actually experienced). Of course, in the model itself, we implement the envisioned policy and compare various techniques that assess impact.

A key ingredient in this exercise is heterogeneity in the population, both observed and potentially unobserved. This means that there can be a nontrivial distribution of gains and/or losses in the population, depending on the policy. This is what can make the LATE (identified using the subsidy as instrument) different from TT and ATE, and at realistic parameter values, these can be quite distinct. The logic of the models also makes clear why this is likely to happen. For example, a subsidy can induce relatively inefficient households to enter business, whereas the larger population of businesses consists of talented households who were not on the margin of decision. This makes LATE negative and ATE positive. In other instances, heterogeneity in one dimension destroys monotonicity in another. A new, nearby branch of a bank can facilitate intermediation by lowering costs, for those on the margin, and though some talented households will borrow to go

into business, other richer, inefficient households will withdraw from low-return businesses and put their money into savings in the bank. Talent is not observed. This makes it difficult without the economic model to assess the impact of intermediation on the profits of entrepreneurs. This also means that widely used econometric techniques can potentially give misleading estimates, depending on what one is willing to assume and what one is trying to measure.

In Section 2, we focus first on observable and unobservable characteristics such as wealth and talent in a simple model of occupation choice, to clarify some key issues. The credit constraint is extreme: self-finance only. The utility functions are linear, but the financing constraint (sale of wealth) makes the problem nonlinear. Indeed, to gauge the impact of this, and for expositional clarity, we begin with exogenous variation in business subsidies for those in financial autarky, computing various measures of welfare gains and comparing the numbers to IV estimates. Section 3 then introduces the full model with intermediation costs and policy variation, distinguishing which instruments are valid for intermediation, and which are valid for occupation choice only.

In Section 4 we adopt a long-horizon dynamic programming formulation to study endogenous financial deepening in a model with unobserved preferences and financial participation costs. We show that unobserved preference heterogeneity can create the need for instruments, as the decision to go to a bank and the outcome of being banked or unbanked, can depend on what for an econometrician would be a common error. Importantly, participation costs can be used as instruments. Here IV and LATE coincide if policy variation in the participation costs comes as a surprise or if the participation decisions are made initially and the unobserved shocks in the model are independent and serially uncorrelated. But even in those cases, the identification of other treatment effects, such as TT or ATE, requires much more work. Of course, anticipated policy changes lowering costs cause the researcher to lose the validity of the instrument, as is well known, and we provide a clear example of this.

Section 5 introduces a model of financial intermediation with moral hazard and unobserved talent. In the model, unobserved talent is an input in the production technology, determining (counterfactual) consumption levels and individual preferences for financial intermediation. The key role played by unobserved heterogeneity, a feature shared by all the models considered in this paper, generates heterogeneity in impact parameters. We discuss under what assumptions the economic model generates instrumental variables that can be used to identify a causal effect of financial intermediation on consumption. In other words, we use the model to discuss its consequences for policy evaluations. We study its static and dynamic versions. We show how, in the static case, random assignment of wealth through a lottery can help us to recover instrumental variables at least over specified ranges of ex ante wealth. Intuitively, we show how individual specific variables affecting the probability of winning the lottery, but independent from potential outcomes associated with intermediation (e.g., costs of entering the randomization), can be used to identify a causal effect of financial intermediation on consumption.

Section 5.2 shows, however, that in dynamic mechanism-design problems the levels of promised utilities in the future matter for choices today, and so one typically loses the availability of instrumental variables even with random assignment of wealth. Essentially, promised utilities for tomorrow depend on outcomes today, to induce proper incentives, along with rewards today. But those promises for tomorrow vary with the costs of intermediation, so we lose the independence of the outcomes from the instrument.

Section 6 presents our conclusions.

2. A STANDARD MODEL OF OCCUPATION CHOICE

We start the analysis with a static model of occupational choice without intermediation. We use this simplified financial autarky framework to illustrate some of the general issues which arise later in the paper. This occupation choice model originated with Lloyd-Ellis and Bernhardt (2000) and has been used by Gine and Townsend (2004) and Jeong and Townsend (2008) to understand how occupation choice and the spread of financial infrastructure can create growth in per capita income and movements in inequality, and more generally, to quantify the welfare gains in the population from the spread of financial intermediation.

Let us assume that the individual has linear preferences over current period consumption, of the form $u(c) = c$, that is, $u'(c) > 0$ and $u''(c) = 0$. The individual faces the budget constraint $c \leq W$, where end-of-period wealth W depends on the within-period occupational choice of the agent.[1] The individual has beginning-of-period wealth b_i, assumed to be observed perfectly by the econometrician, so the initial distribution of wealth is known. This is the source of observable heterogeneity. The individual has an unobserved (from the point of the analyst) business entry cost θ_i^E. Such entry costs are standard in the industrial organization literature. See Salop (1979) for an early example. The individual also has an unobserved talent as a wage earner θ_i^W. These two unobserved talents are as if randomly assigned in the population, again a source of unobserved heterogeneity. For simplicity, we assume that θ^W and θ^E are independent. We denote by $f_{\theta^j}(\cdot)$ the density function of θ^j with $j = \{E, W\}$, and we assume that $E(\theta^W) = E(\theta^E) = 0$. We put additional structure on these densities in future sections. The literature cited earlier did not include unobserved talent in wage work.

The occupational choice of the individual is between enterprise and wage work. These two alternatives can be described by their associated potential outcomes. Specifically, for individual i we have that end-of-period wealth is the sum of initial wealth plus within-period earnings,

$$
W_i = \begin{cases} w + \theta_i^W + b_i & \text{if wage earner} \\ \pi\left(\theta_i^E, b_i, w\right) + b_i & \text{if entrepreneur.} \end{cases} \tag{1}
$$

Here w is the market wage for (unskilled) labor[2] and $\pi(\theta_i^E, b_i, w)$ represents the profit function obtained after solving the production/profit maximization problem

$$\pi\left(\theta_i^E, b_i, w\right) = \max_{\{k,l\}} f(k, l) - wl - k - \theta_i^E \tag{2}$$

$$\text{subject to } 0 \leq k \leq b_i - \theta_i^E. \tag{3}$$

The production function technology $f(k, l)$ is common to all potential firms. Here labor hired l is measured in efficiency units, not number of people per se. k is the level of capitalization measured in units of wealth. In financial autarky, the unobserved entry cost and capital k must be self-financed from wealth b_i. A household is said to be constrained when capital is equal to total wealth minus setup costs, that is, $k = b_i - \theta_i^E$, and this is binding. Indeed, in the original model, if $\theta_i^E > b_i$ it is simply not possible to establish a business. In this case we cannot ask what would be the earnings of someone who has not entered business for that reason. We modify the model below to take this into account. On the other hand, this constraint has been used in structural estimation via likelihood methods, as it provides a source of identification. We discuss this point in Section 2.2.

The decision rule associated with this occupation choice model can be presented as follows:

If $\pi(\theta_i^E, b_i, w) > w + \theta_i^W$, then the individual becomes an entrepreneur.
If $\pi(\theta_i^E, b_i, w) \leq w + \theta_i^W$, then the individual becomes a wage earner.

Therefore, if we denote by D a binary variable such that $D = 1$ if the agent becomes an entrepreneur, and 0 otherwise, we can write

$$D\left(\theta_i^E, \theta_i^W, b_i, w\right) = \begin{cases} 1 & \text{if } \pi\left(\theta_i^E, b_i, w\right) > w + \theta_i^W \\ 0 & \text{if } \pi\left(\theta_i^E, b_i, w\right) \leq w + \theta_i^W. \end{cases}$$

This model is standard in the development literature. The model can be interpreted more generally as a Roy model [Roy (1951)] in which the occupational selection is based, given the individual's talents θ_i^E and θ_i^W and wealth b_i, on a comparison of the potential gains.[3]

2.1. Standard Econometric Approaches to the Analysis of the Impact of Occupational Decisions

We focus on a simple issue: whether we can identify the effect of occupation choice on earnings using a reduced-form approach instead of the full structural model.

In this static model the econometrician observes either $\pi(\theta_i^E, b_i, w) + b_i$ or $w + \theta_i^W + b_i$, depending on whether the choice $D_i = 1$ or $D_i = 0$ is taken by the individual i. Thus, if we denote by Y_i the end-of-period observed outcome, we have

$$Y_i \equiv D_i\left(\pi\left(\theta_i^E, b_i, w\right) + b_i\right) + (1 - D_i)\left(w + \theta_i^W + b_i\right),$$

where without additional structure, profits are nonlinear in entrepreneur talent (θ_i^E), wealth (b_i), and market wage (w). However, the empirical literature primarily uses linear and separable models. That is,

$$\pi\left(\theta_i^E, b_i, w\right) \simeq \phi_w w + \phi_\theta \theta_i^E + \phi_b b_i. \tag{4}$$

This setup is particularly attractive if one notes that

$$Y_i = D_i\left[\phi_w w + \phi_\theta \theta_i^E + \phi_b b_i + b_i\right] + (1 - D_i)\left[w + \theta_i^W + b_i\right]$$
$$= w + b_i + (\phi_b b_i + (\phi_w - 1)\, w)\, D_i + \theta_i^W + \left(\phi_\theta \theta_i^E - \theta_i^W\right) D_i,$$

which can be expressed as a linear regression model,

$$Y_i = w + b_i + (\phi_b b_i + (\phi_w - 1)\, w)\, D_i + \varepsilon_i, \tag{5}$$

where $\varepsilon_i = \theta_i^W + (\phi_\theta \theta_i^E - \theta_i^W) D_i$, and the term in parentheses, $(\phi_b b_i + (\phi_w - 1) w)$, represents the gain in gross income that does not depend on unobserved talent. Notice that the random variable D_i is by construction correlated with ε_i, so the OLS regression of observed earnings onto an occupational dummy (conditioning on wealth),

$$\widehat{\phi}_1^{\text{OLS}} = \frac{\text{Cov}\,(Y, D \mid b_i = b)}{\text{Var}\,(D \mid b_i = b)},$$

would provide a biased estimator of this gain, $\phi_b b_i + (\phi_w - 1)\, w$. We illustrate the consequences of this selection problem below. Importantly, the interaction between unobserved talents, potential outcomes, and occupational choice that generates the selection problem is not a result of a linear and separable profit function but a general consequence of the theoretical framework with unobserved talent and endogenous selection.

A widely used alternative is the instrumental variable method. To consider this approach, we introduce a policy distortion (instrument) into the model. This distortion affects occupation choices in a simple way. Specifically, we assume the existence of an exogenous subsidy that increases ex post profits at the end of period by ψ. This subsidy is randomly assigned in the population, so that ψ is a random variable with $\psi > 0$ and known to the econometrician even if the choices of the household is to be a wage earner. Intuitively, it can be interpreted as an experiment or exogenous policy treatment affecting the occupation choices of the individuals but received only if the choice is to set up a firm. However, this subsidy cannot be used to finance k, and so the constraint $0 \leq k \leq b - \theta^E$ is unaltered.

The policy distortion impacts the decision rule

$$D\left(\theta_i^E, \theta_i^W, b_i, \psi_i, w\right) = \begin{cases} 1 & \text{if } \pi\left(\theta_i^E, b_i, w\right) + \psi_i > w + \theta_i^W \\ 0 & \text{if } \pi\left(\theta_i^E, b_i, w\right) + \psi_i \leq w + \theta_i^W, \end{cases}$$

where ψ_i represents the subsidy to agent i in the event of becoming an entrepreneur. More simply, and to emphasize the role of ψ_i, we use the notation $D(\psi_i)$ below,

but clearly this binary variable is a function of other observable and unobservable variables. We assume that talents (θ^E, θ^W) and subsidy ψ are independent. Indeed, the government cannot see θ^E (or θ^W) but has total control over the random subsidy. The subsidy ψ affects the decision rule, but not the potential outcomes net of the subsidy, as it enters additively. Therefore, the maximization problem of the household as a firm, if it becomes a firm, and its choice of k and l are unaltered. It gets the subsidy independent of its behavior as a firm.

The subsidy ψ_i appears to be a valid instrument. It influences choices but not the potential outcomes.[4] Additionally, in this setup, the subsidy satisfies the uniformity/monotonicity condition [Imbens and Angrist (1994); Heckman et al. (2006)]. That is, for each individual, an increase (decrease) in the subsidy unambiguously increases (reduces) the chances of becoming an entrepreneur. Indeed, suppose that the subsidy can take on two values, $\overline{\psi}$ and $\overline{\overline{\psi}}$. In this case, and without imposing a linear separable model for profits, we can use the instrument ψ to estimate

$$\Delta^{\mathrm{IV}}(\overline{\overline{\psi}}, \overline{\psi}; b) = \frac{E(Y_i \mid \psi_i = \overline{\overline{\psi}}, b_i = b) - E(Y_i \mid \psi_i = \overline{\psi}, b_i = b)}{E(D_i \mid \psi_i = \overline{\overline{\psi}}, b_i = b) - E(D_i \mid \psi_i = \overline{\psi}, b_i = b)},$$

which, under the assumption of uniformity, identifies the local average treatment effect (LATE) in income for those in the population induced to enter entrepreneurship due to the change of ψ from $\overline{\psi}$ to $\overline{\overline{\psi}}$ (the treatment here is to become an entrepreneur), or more formally

$$\Delta^{\mathrm{LATE}}(\overline{\overline{\psi}}, \overline{\psi}; b) = E\big[\pi\big(\theta_i^E, b_i, w\big) - w - \theta_i^W \mid D_i(\overline{\overline{\psi}}) = 1, D_i(\overline{\psi}) = 0, b_i = b\big].$$

This parameter does not pick up the earnings difference for those who would be entrepreneurs versus wage earners, regardless of the value of the instrument. Instead, the local average treatment effect Δ^{LATE} naturally provides the answer to a policy experiment.[5]

Given that the model features heterogeneous treatment effects, we can complete the analysis by computing two alternative treatment effects: the treatment on the treated, Δ^{TT} (average benefits of becoming an entrepreneur for individuals who actually decide to become entrepreneurs), and the average treatment effect, Δ^{ATE} (the earnings gain or loss of becoming an entrepreneur versus a wage earner in the entire population). Specifically, and presenting the treatment parameters for a particular wealth level b, we have:

$$\Delta^{\mathrm{TT}}(b) = E\left(\pi\big(\theta_i^E, b_i, w\big) - \big(w + \theta_i^W\big) \,\middle|\, D_i = 1, b_i = b\right) \qquad (6)$$

$$\Delta^{\mathrm{ATE}}(b) = E\left(\pi\big(\theta_i^E, b_i, w\big) - \big(w + \theta_i^W\big) \,\middle|\, b_i = b\right). \qquad (7)$$

If there were no heterogeneity or if all heterogeneity were observed, then all these effects (including LATE) would be equivalent [see Heckman and Vytlacil (2001)]. Otherwise, $\Delta^{\mathrm{TT}}(b)$ and $\Delta^{\mathrm{ATE}}(b)$ depend on counterfactual wages and
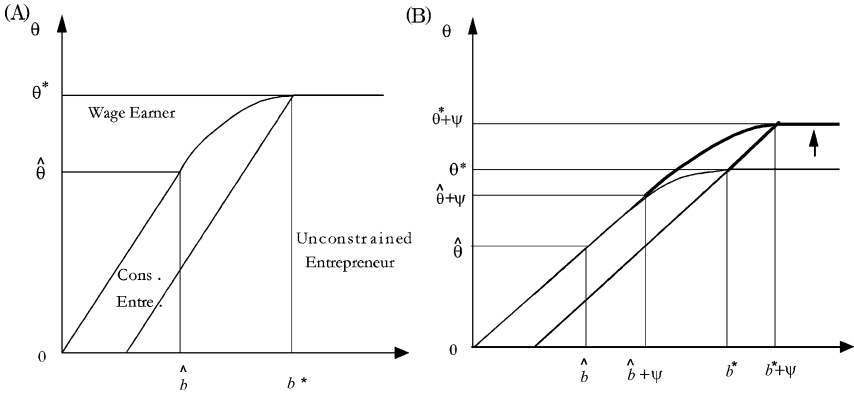
FIGURE 1. Occupational choice maps and the effect of the subsidy. (A) Without subsidy; (B) the effect of the subsidy.

profits for a given wealth level b, and the estimation of these parameters is not straightforward.

2.2. Parametric and Semiparametric Identification of Treatment Effect Parameters

Following Gine and Townsend (2004), we assume that

$$f(k, l) = \alpha k - \frac{1}{2}\beta k^2 + \sigma kl + \xi l - \frac{1}{2}\rho l^2,$$

and the profit function can be written

$$\pi\left(\theta_i^E, w, k\right) = C_0(w) + C_1(w)k + C_2 k^2 - \theta_i^E, \tag{8}$$

where $C_0(w) = \frac{(\xi - w)^2}{2\rho}$, $C_1(w) = \alpha - 1 + \sigma(\frac{\xi - w}{\rho})$, and $C_2 = \frac{1}{2}(\frac{\sigma^2}{\rho} - \beta)$. The model delivers a quadratic occupation partition as depicted in Figure 1A and a nonlinear profit function.

For expositional simplicity, we set $\theta^W = 0$ and assume that $\pi(\theta_i^E, b_i, w) = b_i - \theta^E > 0$ in Figure 1A. The points (θ^{E*}, b^*) and $(\widehat{\theta^E}, \widehat{b})$ determine the shape of the curves entirely. These points can be expressed as functions of $C_0(w)$, $C_1(w)$, and C_2.

This framework also allows us to illustrate the effect of the subsidy. Figure 1B shows the effect of moving ψ from $\overline{\psi}$ to $\overline{\overline{\psi}}$. This change essentially shifts the line of indifference vertically upward as the subsidy simply adds to the net profits of entrepreneurs. (This upward shift is not present when the household is constrained by beginning-of-period wealth.) Now for every value of wealth b there exists a group of θ^E households that weakly shift into business. The impact of the subsidy is "uniform" [or "monotone" in the language of Imbens and Angrist (1994)]; that is, the movement is (at most) in one direction only. This is the group of individuals who provide the source of variation used when estimating Δ^{LATE}.

Finally, under the assumption $\sigma^2/\rho = \beta$ and optimal capital ($k^* = b - \theta$), we can obtain linear profit functions. We want to emphasize that this approximation is not designed to exactly characterize the economic model but to show how to link the theory with common econometric practice. Therefore, from this point forward, we follow the traditional econometric approach and assume a linear and additively separable approximation for the profit function.

By itself the assumption of linear and additive separable profit functions is not sufficient for the computation of treatment effects. We need additional structure to deal with the selection problems. Consider first the case of independent and normally distributed unobserved talents, that is, $\theta^E \sim N(0, \sigma_E^2), \theta^W \sim N(0, \sigma_W^2)$. In this context, we can define the probability of being an entrepreneur in our model as

$$
\Pr\left(\pi\left(\theta_i^E, b_i, w\right) + \psi_i > \left(w + \theta_i^W\right)\right)
$$
$$
= \Pr\left(\phi_w w + \phi_\theta \theta_i^E + \phi_b b_i + \psi_i > \left(w + \theta_i^W\right)\right)
$$
$$
\equiv \Phi\left(\frac{(\phi_w - 1)w + \phi_b b_i + \psi_i}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}}\right)
$$

where $\Phi\left(\cdot\right)$ represents the cumulative distribution function of a standard normal distribution. Therefore, given the normality assumption and the structure of this last expression and with information on occupational choice (D), subsidy (ψ), wealth (b), the observed average wage in the economy (w), and profits (π) for those households with $D = 1$, we can use a probit model to identify the parameters $(\phi_w - 1)$, ϕ_b, and $\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}$.

The mean observed profit (conditional on b_i and ψ_i) can be written as follows:

$$
E\left(\pi\left(\theta_i^E, b_i, w\right) \mid D_i = 1, b_i, \psi_i\right)
$$
$$
= E\left(\phi_w w + \phi_\theta \theta_i^E + \phi_b b_i \mid \phi_w w + \phi_\theta \theta_i^E + \phi_b b_i + \psi_i > \left(w + \theta_i^W\right)\right)
$$
$$
= \phi_w w + \phi_b b_i + \phi_\theta \sigma_E E
$$
$$
\times \left(\frac{\theta_i^E}{\sigma_E} \,\middle|\, \frac{\theta_i^W - \phi_\theta \theta_i^E}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}} < \frac{(\phi_w - 1)\,w + \phi_b b_i + \psi_i}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}}\right).
$$

Given that θ_i^E/σ_E and $(\theta_i^W - \phi_\theta \theta_i^E)/\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}$ are standard jointly normally distributed random variables, we have that

$$
E\left(\pi\left(\theta_i^E, b_i, w\right) \mid D_i = 1, b_i, \psi_i\right)
$$
$$
= \phi_w w + \phi_b b_i - \frac{\phi_\theta^2 \sigma_E^2}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}} \lambda\left(\frac{(\phi_w - 1)\,w + \phi_b b_i + \psi_i}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}}\right), \tag{9}
$$

where $\lambda(\cdot)$ represents the Mills' ratio.[6] Expression (9) justifies the estimation of a linear regression model of observed profits/earnings onto the wage w (intercept), wealth b_i, and the Mills' ratio $\lambda(.)$, to obtain $\phi_\theta^2 \sigma_E^2$ (because $\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}$ is known from the probit), and also σ_W^2. The parameters ϕ_θ and σ_E^2 cannot be identified separately from this regression.

On the other hand, although unobserved, average wages among those choosing to be entrepreneurs can be written as

$$
\begin{aligned}
&E\left(w + \theta_i^W \mid D_i = 1, b_i, \psi_i\right) \\
&= E\left(w + \theta_i^W \mid \phi_w w + \phi_\theta \theta_i^E + \phi_b b_i + \psi_i > \left(w + \theta_i^W\right)\right) \\
&= w + \frac{\sigma_W^2}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}} \lambda\left(\frac{(\phi_w - 1)w + \phi_b b_i + \psi_i}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}}\right),
\end{aligned} \tag{10}
$$

which depends only on identified parameters, so it can be constructed for any value of wealth and subsidy. Thus, we can compute the treatment on the treated $(\Delta^{\mathrm{TT}}(b, \psi))$ as

$$
\Delta^{\mathrm{TT}}(b, \psi) = \underbrace{E\left(\pi\left(\theta_i^E, b_i, w\right) \mid D_i = 1, b_i = b, \psi_i = \psi\right)}_{\text{Identified from expression (9)}}
$$

$$
- \underbrace{E\left(w + \theta_i^W \mid D_i = 1, b = b_i, \psi_i = \psi\right)}_{\text{Identified from expression (10)}}
$$

and, likewise, the average treatment effect $\Delta^{\mathrm{ATE}}(b_i)$

$$
\begin{aligned}
\Delta^{\mathrm{ATE}}(b) &= E\left(\pi\left(\theta_i^E, b_i, w\right) - \left(w + \theta_i^W\right) \mid b_i = b\right) \\
&= (\phi_w - 1)w + \phi_b b.
\end{aligned}
$$

The unconditional version of $\Delta^{\mathrm{TT}}(b, \psi_i)$, that is, $\Delta^{\mathrm{TT}}(b)$, can be obtained by simply integrating out ψ over the relevant region.

The normality assumption for the identification of treatment parameters can be relaxed at the price of additional conditions. In particular, let θ^E and θ^W be two independent random variables distributed according to a (general) joint distribution function $f_{\theta^E, \theta^W}(\cdot, \cdot)$. As shown in the context of the economic model, these variables, which are unobserved by the analyst, determine profits, wages, and occupational choices.

On the other hand, and given enough data variation, we can nonparametrically estimate the probability of $D_i = 1$ using information on b_i, w, ψ_i, and actual choices D_i (Matzkin 1992). Let $p(w, b_i, \psi_i)$ denote this probability, also known in the literature as the propensity score. We can then write the conditional expectation

of observed outcome Y_i as a function of the probability of selection and wealth,

$$E(Y_i \mid p(w, b_i, \psi_i), b_i) = w + b_i + (\phi_b b_i + (\phi_w - 1) w) E\left(D_i \mid p(w, b_i, \psi_i)\right)$$

$$+ E\left(\theta_i^W + \left(\phi_\theta \theta_i^E - \theta_i^W\right) D_i \mid p(w, b_i, \psi_i)\right)$$

$$= w + b_i + (\phi_b b_i + (\phi_w - 1) w) p_i + \Lambda(p_i, b_i), \tag{11}$$

where $\Lambda(p_i, b_i) \equiv E(\theta_i^W + (\phi_\theta \theta_i^E - \theta_i^W) D_i \mid p(w, b_i, \psi_i), b_i)$, and for notational convenience, we use p_i instead of $p(w, b_i, \psi_i)$. As shown by Heckman and Vytlacil (2001), we can use this conditional expectation to form $\Delta^{TT}(b)$ and $\Delta^{ATE}(b)$, expressions (6) and (7), respectively, without imposing normality. In particular, these authors show how by computing

$$\Delta^{LIV}(p, b) = \left.\frac{\partial E(Y_i \mid p_i, b_i = b)}{\partial p_i}\right|_{p_i = p},$$

usually called the local instrumental variable estimator, the analyst can identify the treatment parameter

$$\Delta^{MTE}(p, b) \equiv E\left(\pi\left(\theta_i^E, b_i, w\right) - \left(w + \theta_i^W\right) \mid b_i = b, \theta_i^W - \phi_\theta \theta_i^E = p\right),$$

where $\Delta^{MTE}(p, b)$ represents the treatment effect for those individuals indifferent between occupations given a particular value (p) for the random variable $\theta_i^W - \phi_\theta \theta_i^E$ (conditional on wealth level b).[7] Finally, Heckman and Vytlacil (2001) show that $\Delta^{ATE}(b)$ and $\Delta^{TT}(b)$ can be obtained as weighted averages of $\Delta^{MTE}(p, b)$ according to the expressions

$$\Delta^{TT}(b) = \int \Delta^{MTE}(u, b) \omega^{TT}(u, b) \, du$$

$$\Delta^{ATE}(b) = \int \Delta^{MTE}(u, b) \omega^{ATE}(u) \, du,$$

where $\omega^{ATE}(u) = 1$ and $\omega^{TT}(u, b) = \Pr(p(w, b, \psi) > u) / \int \Pr(p(w, b, \psi) > u) du$. The argument of integration u is associated with the random variable $U = F_{\theta_i^W - \phi_\theta \theta_i^E}(\theta_i^W - \phi_\theta \theta_i^E)$, which is uniformly distributed.[8]

The question then becomes how to compute $\Delta^{LIV}(p, b)$. We can use formal semi-parametric techniques to estimate $E(Y_i \mid p_i, b_i)$ [expression (11)], and its derivative with respect to p. An alternative and simpler way to estimate this function is by approximating it using a polynomial on p_i [see Heckman et al. (2006)].[9]

2.3. Measuring the Impact of Occupations on Income

We illustrate the importance of the previous discussion by computing and comparing different estimates of the effect of occupational decisions on income. To do this, we simulate data from our model. We utilize the quadratic production function

TABLE 1. Parameter values

Parameter	Value
α	0.54561
β	−0.39064
σ	0.1021
ξ	0.2582
ρ	−0.03384
w	0.048

Note: The numbers in this table are obtained from Gine and Townsend (2004). Using data from Thailand, Gine and Townsend (2004) estimate a model of occupational choice with similar characteristics to the one studied in this paper.

described above and consider the parameterization in Table 1. These parameter values are taken directly from Gine and Townsend (2004).[10] We assume a discrete subsidy. Specifically, we assume that the subsidy ψ can take two values: 0 or 1. The value of the subsidy is randomly assigned in the population.

Wealth (b), talents (θ^W and θ^E), and the subsidy (ψ) are assumed to be distributed as follows:

$$b \sim \text{Log } N(0, 1),$$

$$\theta^W \sim N(0, 1), \theta^E \sim N(0, 1),$$

$$\psi \sim \text{Binomial}(0, 1).$$

We first reproduce the analysis that a researcher could carry out using a cross-sectional data set with information on wealth, occupation, and observed (factual) incomes.

Using cross-sectional information to estimate the effect of occupational choice. Table 2 presents the sorting into occupations obtained from model-generated cross sections of 25,000 individuals.

Consider an agnostic empirical approach in which the researcher tries to estimate the effect of the occupational choice on outcomes using a simple regression model. In particular, suppose that observed income (profits/wages) Y is written as

$$Y_i = \kappa_0 + \kappa_1 b_i + \kappa_2 b_i D_i + \kappa_3 D_i + \varepsilon_i,$$

where D_i takes a value of 1 if the individual i is an entrepreneur, and 0 otherwise.[11] Notice that in this equation we do not incorporate talent explicitly. This is because in practice the analyst does not observe this variable, so it must be excluded from the list of controls (and contained in the error term).

Table 3 presents the estimated effect of D_i on Y_i obtained using OLS and IV.[12] The results show large differences between the two approaches. The results from

TABLE 2. Sorting by occupational choices (model of occupational choice—simulated cross-sectional data)

Occupational choice	Sample size
Wage earner	6,109
Entrepreneur	18,891
Constrained	14,519
Unconstrained	4,372
Total	25,000

Note: The number of observations in each occupation is the result of the endogenous decision process faced by each of the 25,000 simulated individuals.

IV deliver negative impacts, whereas OLS would suggest a positive impact. The large discrepancies are a clear manifestation of the biases caused by the selection process. A researcher would draw dramatically different conclusions depending on how s/he interpreted the policy impact coefficients in the IV regression. In practice, likely instruments can show up as correlated with unobserved error, producing the misinterpretation of the results. The analyst must understand the economics behind the selection mechanism before drawing conclusions.

Interestingly, the negative effect estimated by IV is intuitively correct because the individuals switching occupations as a result of the variation in the instruments are those with lower profits (and higher wages). That is, because the subsidy is not included in the income gains, it induces inefficient choices. On the other hand, there are others who benefit from the subsidy but would have chosen to

TABLE 3. OLS and IV estimates (model of occupational choice—estimates from cross-sectional data)

Parameter	Estimates	
	Δ^{OLS}	Δ^{IV}
κ_0	0.606**	1.189**
κ_1	1.155**	1.142**
κ_2	−0.136**	−0.082
κ_3	0.457**	−0.356*
Average effect ($\kappa_2 \bar{b} + \kappa_3$)	0.303**	−0.450

Note: This table presents the parameters obtained from a linear regression of observed income (profits or wages depending on individual's occupation) on wealth, the occupational dummy, and the interaction between wealth and occupation (dummy). In addition, the column Δ^{IV} presents the estimates when ψ is used as instrument. Overall these results illustrate what the analyst can obtain using information produced from the model (observed outcome, wealth, and occupation) using a reduced-form strategy.
* Statistical significance at 5%.
** Statistical significance at 1%.

TABLE 4. Model-generated treatment paramaters (model of occupational choice—the causal effects of occupation on income)

| Parameter	Value
Δ^{ATE}	0.619
Δ^{TT}	1.270
$\Delta^{\text{LATE}}(1, 0)$	−0.459

Note: The numbers in this table represents the model's underlying treatment parameters associated with the effect of occupation on income (or model-generated treatment parameters). To obtain them, we use the structure of the model to simulate data on wages, profits, and choices for 25,000 individuals. The analyst would need to characterize the structure of the model (counterfactual outcomes and decision rule) before producing these treatment parameters (as opposed to a reduced form strategy).

be entrepreneurs in any event, and they had efficient rents that dominate wage earnings.

Using the structure of the model to generate counterfactual outcomes and the causal effects of occupational choices. Given our knowledge of the model, we can study the consequences of exogenous policy changes. Specifically, we provide individuals who did not receive a subsidy when it was originally assigned with the subsidy. We then use the sample of individuals switching occupation due to the change in subsidy status (from $\psi = 0$ to $\psi = 1$) to compute the model-generated local average treatment effect $[\Delta^{\text{LATE}}(1, 0)]$ (i.e., the average effect of the treatment for those individuals switching occupations as a result of a change in the instrument). However, because occupation status also depends on wealth, we first compute $\Delta^{\text{LATE}}(1, 0; b_k)$, where b_k represent the kth percentile of the wealth distribution, and then we compute $\Delta^{\text{LATE}}(1, 0)$ as the (weighted) averages of $\Delta^{\text{LATE}}(1, 0; b_k)$.[13]

As a result of our experiment, 1,861 of our original wage earners become entrepreneurs. This is precisely the group for which we can compute the model-generated local average treatment effect. Additionally, from our knowledge of the model, we can directly compute the average treatment effect (ATE) and the treatment effect of those treated (TT).

Table 4 presents the model-generated treatment effects. Notice the similarities between the model-generated LATEs (Δ^{LATE} in Table 4) and the IV effect estimated using the cross-sectional data sets (Δ^{IV} in Table 3). The discrepancies can
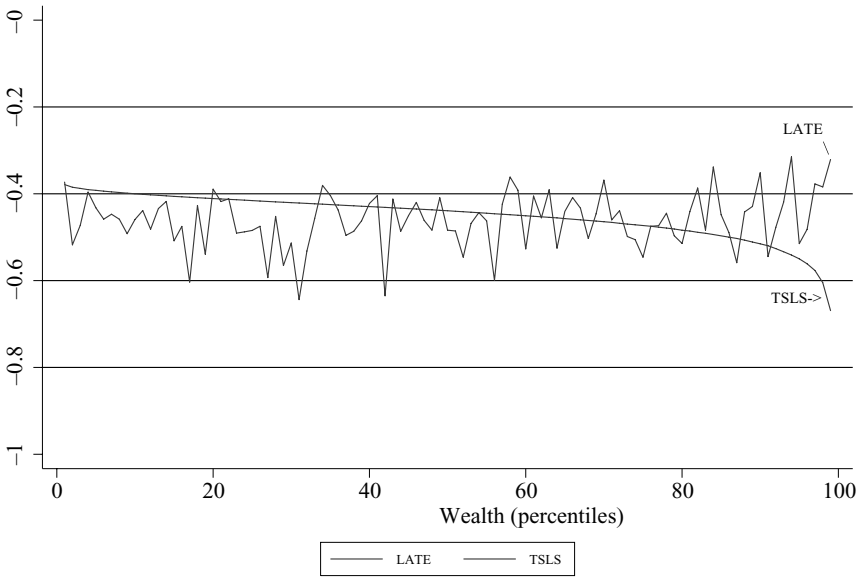
FIGURE 2. Model-generated and estimated local average treatment effect by percentile of the wealth distribution.

be attributed to the linear approximation used in the regression model. We relax this assumption in the next sections.[14] In our model, Δ^{LATE} is negative, as the subsidy induces low-productivity individuals to enter business and the subsidy is not counted as part of the gain. This is exactly the same conclusion drawn in the context of Table 3.

TT and ATE, on the other hand, are positive numbers, indicating the positive benefits associated with entrepreneurship.

Finally, Figure 2 presents the local average treatment effects by percentile of the wealth distribution. The figure presents the model-generated LATE ($\Delta^{\text{LATE}}(1, 0; b)$) and the estimated IV [$\Delta^{\text{IV}(\psi)}(1, 0; b)$] by wealth level. As expected, although the model-generated LATE fluctuates across levels of wealth (a result of our sample size), on average it is close to what the standard econometric technique delivers.

This example illustrates how the economic model delivers a valid instrument, how this instrument allows the identification of a causal effect of interest, and how this causal effect can differ from other relevant treatment parameters.

3. OCCUPATIONAL CHOICE UNDER FINANCIAL INTERMEDIATION

The simple model presented in Section 2, with the subsidy to firms, can easily be extended to incorporate an intermediated sector. The analysis in Gine and Townsend (2004) does exactly that. We follow their approach. The underlying

model in this case is similar to the model in Section 2, but now there is borrowing and lending of capital and wealth. We denote by Q_i the individual-specific cost of using the intermediated sector. Examples of Q_i include travel time to a district center or branch office, whether or not a particular intermediary has been active in a city or village according to history, particular policies of financial institutions that vary in effectiveness, and new credit in a city or village divided by the number of households. See Kaboski and Townsend (2005, 2009) and Alem and Townsend (2008) for examples.

We take the initial distribution of Q as given and, for simplicity, focus on a binary Q. The analysis can be extended directly to a continuous-valued Q. We assume Q independent from ψ, and denote by r the (equilibrium) interest rate.

An entrepreneur using the intermediated sector solves the following problem:

$$\max_{k,l} f(k, l, \theta_i^E) - wl - (1 + r)(k + \theta_i^E). \tag{12}$$

There is a neoclassical separation between production and household wealth. In effect, the agent can put all his wealth b_i in financial markets and earn interest r. Meanwhile the firm (individual) can borrow what it needs to finance k and set up cost θ_i^E. There is a lot of indeterminacy in between, in financing, that is, self-investing and borrowing/lending the difference with wealth, but real quantities and net income are all pinned down.

The wage is common to both sectors, as households are allowed to work wherever they prefer. They can join an intermediary and put their money in a saving account if they do not become firms.

As before, denote by D_i a binary variable such that $D_i = 1$ if agent i decides to become an entrepreneur, and 0 otherwise. Thus, the occupation choice when the agent is participating in the intermediated sector can be described by

$$D(\theta_i^E, \theta_i^W, w, r) = \begin{cases} 1 & \text{if } \pi(\theta_i^E, w, r) + b_i(1 + r) - Q_i + \psi_i \\ & \quad > w + \theta_i^W + b_i(1 + r) - Q_i \\ 0 & \text{otherwise,} \end{cases}$$

where $\pi(\theta_i^E, w, r)$ denotes the resulting profits after (12), is solved.

In this context, the researcher would observe $\pi(\theta_i^E, b_i, w, r) + b_i(1 + r)$ or $w + b_i(1 + r) + \theta_i^W$, depending on the value of $D(\theta_i^E, \theta_i^W, w, r)$. Thus, if we denote by $Y_I(\theta_i^E, \theta_i^W, b_i, w, r)$ the outcome observed under intermediation, we have

$$Y_I(\theta_i^E, \theta_i^W, b_i, w, r) = D(\theta_i^E, \theta_i^W, w, r)\left(\pi(\theta_i^E, w, r) + b_i(1 + r)\right)$$
$$+ \left(1 - D(\theta_i^E, \theta_i^W, w, r)\right)\left(w + b_i(1 + r) + \theta_i^W\right) \tag{13}$$

and the cost Q_i and the subsidy ψ_i are not subtracted or added, respectively, from or to Y_I; that is, we have gross gains.

On the other hand, recall that without financial intermediation the occupation choice model is

$$D\left(\theta_i^E, \theta_i^W, b_i, w\right) = \begin{cases} 1 & \text{if } \pi\left(\theta_i^E, b_i, w\right) + \psi_i > w + \theta_i^W \\ 0 & \text{otherwise.} \end{cases}$$

so that the observed outcome under financial autarky, $Y_A(\theta_i^E, \theta_i^W, b_i, w)$ (not counting the subsidy), can be presented as

$$\begin{aligned} Y_A\left(\theta_i^E, \theta_i^W, b_i, w\right) &= D\left(\theta_i^E, \theta_i^W, b_i, w\right)\left(\pi\left(\theta_i^E, b_i, w\right) + b_i\right) \\ &+ \left(1 - D\left(\theta_i^E, \theta_i^W, b_i, w\right)\right)\left(w + \theta_i^W + b_i\right). \end{aligned} \tag{14}$$

In sum, the subindex k in Y_k indicates the sector (financial autarky A or intermediation I). We use this notation in what follows.

The choice of sector, autarky versus intermediation, is made by a simple comparison of the potential associated outcomes $Y_A(\theta_i^E, \theta_i^W, b_i, w)$ and $Y_I(\theta_i^E, \theta_i^W, b_i, w, r)$, but adjusting in the choice for the subsidy ψ_i and entry cost Q_i. Note that, in general, the heterogeneity (θ_i^E, θ_i^W) does not enter additively into $Y_A(\theta_i^E, \theta_i^W, b_i, w)$ or $Y_I(\theta_i^E, \theta_i^W, b_i, w, r)$. Thus, let Υ_i be a binary variable that takes a value of 1 if the individual decides to use the financial intermediary, and 0 otherwise. Then

$$\begin{aligned} &\Upsilon_i\left(\theta_i^E, \theta_i^W, b_i, w, r, \psi_i, Q_i\right) \\ &= \begin{cases} 1 & \text{if } \begin{pmatrix} \left[Y_I\left(\theta_i^E, \theta_i^W, b_i, w, r\right) - Y_A\left(\theta_i^E, \theta_i^W, b_i, w\right)\right] \\ + \left[D\left(\theta_i^E, \theta_i^W, w, r\right) - D\left(\theta_i^E, \theta_i^W, b_i, w\right)\right]\psi_i \\ -Q_i \end{pmatrix} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

This simple framework allows us to analyze policies regarding the access to financial intermediation.

3.1. Identifying the Effects of Financial Intermediation

In the context of our model, the effect of having access to financial intermediation at the individual level (agent i) is defined as

$$\Delta_i^\Upsilon = Y_I\left(\theta_i^E, \theta_i^W, b_i, w, r\right) - Y_A\left(\theta_i^E, \theta_i^W, b_i, w\right),$$

the average treatment effect (ATE) associated with financial intermediation is

$$E\left(\Delta_i^\Upsilon\right) = E\left(Y_I\left(\theta_i^E, \theta_i^W, b_i, w, r\right) - Y_A\left(\theta_i^E, \theta_i^W, b_i, w\right)\right),$$

and the average effect of the treatment on those treated (TT) equals

$$E\left(\Delta_i^\Upsilon \mid \Upsilon_i = 1\right) = E\left(Y_I\left(\theta_i^E, \theta_i^W, b_i, w, r\right) - Y_A\left(\theta_i^E, \theta_i^W, b_i, w\right) \mid \Upsilon_i = 1\right),$$

where again for simplicity we use Υ_i instead of $\Upsilon(\theta_i^E, \theta_i^W, b_i, w, r, Q_i)$. Additionally, in what follows, we use D_i and $D_i(r)$ to denote the occupation choices $D(\theta_i^E, \theta_i^W, w, b_i)$ and $D(\theta_i^E, \theta_i^W, w, r)$ under financial autarky and the intermediated sector, respectively.

In order to analyze whether conventional econometric methods (OLS and IV) allow the identification of any of these effects, we first denote by ξ_i the observed outcome, that is,

$$\xi_i = \Upsilon_i \times Y_I\left(\theta_i^E, \theta_i^W, b_i, w, r\right) + (1 - \Upsilon_i) \times Y_A\left(\theta_i^E, \theta_i^W, b_i, w\right),$$

which after substituting expressions (13) and (13) can be written as:

$$\xi_i = \Upsilon_i \times \left[\begin{array}{c} D_i(r)\left(\pi\left(\theta_i^E, w, r\right) + b_i(1+r)\right) \\ +(1 - D_i(r))\left(w + \theta_i^W + b_i(1+r)\right) \end{array}\right]$$

$$+ (1 - \Upsilon_i) \times \left[\begin{array}{c} D_i\left(\pi\left(\theta_i^E, b_i, w\right) + b_i\right) \\ +(1 - D_i)\left(w + \theta_i^W + b_i\right) \end{array}\right]. \tag{15}$$

This expression illustrates the fact that all potential choices and outcomes play a role even when the researcher is only interested in the impact of having access to financial intermediation.

Following the conventional empirical strategy, we assume profit functions of the forms

$$\pi\left(\theta_i^E, b_i, w\right) = \gamma_w w + \gamma_b b_i + \gamma_\theta \theta_i^E \quad \text{(financial autarky)}$$

$$\pi\left(\theta_i^E, w, r\right) = \delta_w w + \delta_r r + \delta_\theta \theta_i^E \quad \text{(intermediation)}.$$

Substituting these expressions into equation (15), and after some algebra, we obtain

$$\xi_i = w + b_i + r\Upsilon_i b_i$$
$$+ \left((\gamma_w - 1)w\right) D_i (1 - \Upsilon_i) + \gamma_b b_i D_i (1 - \Upsilon_i)$$
$$+ \left((\delta_w - 1)w + \delta_r r\right) \Upsilon_i D_i(r) + \delta_b b_i \Upsilon D_i(r)$$
$$+ \eta_i\left(\theta_i^E, \theta_i^W, r, Q_i\right), \tag{16}$$

where $\eta_i(\theta_i^E, \theta_i^W, r, Q_i) = (\delta_\theta \theta_i^E - \theta_i^W)\Upsilon_i D_i(r) - (\gamma_\theta \theta_i^E - \theta_i^W)\Upsilon_i D_i + (\gamma_\theta \theta_i^E - \theta_i^W)D_i + \theta_i^W$, so $\eta_i(\theta_i^E, \theta_i^W, r, Q_i)$ contains all the terms involving unobserved talents θ_i^E and θ_i^W. Using expression (16), we can define the individual effect of having access to financial intermediation, Δ_i^Υ, as

$$\Delta_i^\Upsilon = \frac{\Delta\xi_i}{\Delta\Upsilon_i}$$
$$= rb_i + \left((\delta_w - 1)w + \delta_r r\right) D_i(r)$$
$$- \left((\gamma_w - 1)w - \gamma_b b_i\right) D_i + \frac{\Delta\eta_i\left(\theta_i^E, \theta_i^W, r, Q_i\right)}{\Delta\Upsilon_i}. \tag{17}$$

Notice that Δ_i^{Υ} (conditional on wealth b) depends on the occupation of the individual under each regime and the unobserved talents.

On empirical grounds, expression (16) suggests the estimation of the parameters defining Δ_i^{Υ} through a regression of ξ_i on b_i, $\Upsilon_i b_i$, $D_i(1 - \Upsilon_i)$, $D_i(r)\Upsilon_i$, $b_i D_i(1 - \Upsilon_i)$, and $b_i \Upsilon_i D_i(r)$. However, because unobserved talents (contained in the error term) affect both choices and potential outcomes, without further assumptions, conventional OLS estimates will not provide unbiased estimates of the parameters in the model.

An alternative approach is the instrumental variable method. The economic model provides one natural instrument for Υ_i, namely Q_i. The cost Q_i affects the choice of sector but does not affect the potential outcomes. In addition, notice that changes in Q_i produce uniform (monotonic) responses in choice Υ_i. Consequently, given two values for the instrument Q_i, $\overline{Q}$, and $\overline{\overline{Q}}$ (lowering the cost so that $\overline{\overline{Q}} < \overline{Q}$), and conditioning on wealth b, we can estimate

$$\Delta^{\text{IV}(Q)}(\overline{\overline{Q}}, \overline{Q}; b) = \frac{E(\xi_i | Q_i = \overline{\overline{Q}}, b_i = b) - E(\xi_i | Q_i = \overline{Q}, b_i = b)}{E(\Upsilon_i | Q_i = \overline{\overline{Q}}, b_i = b) - E(\Upsilon_i | Q_i = \overline{Q}, b_i = b)} \quad \textbf{(18)}$$

to identify the local treatment effect of financial intermediation on income,

$$\Delta^{\text{LATE}(Q)}(\overline{\overline{Q}}, \overline{Q}; b) = E\big(Y_I\big(\theta_i^E, \theta_i^W, b_i, w, r\big)$$
$$- Y_A\big(\theta_i^E, \theta_i^W, b_i, w\big) \,\big|\, b_i = b, \Upsilon_i(\overline{\overline{Q}}) = 1, \Upsilon_i(\overline{Q}) = 0\big), \quad \textbf{(19)}$$

where $\Upsilon(\overline{Q}) = \Upsilon(\theta_i^E, \theta_i^W, b_i, w, r, \psi_i, \bar{Q})$ and $\Upsilon(\overline{\overline{Q}}) = \Upsilon(\theta_i^E, \theta_i^W, b_i, w, r, \psi_i, \overline{\overline{Q}})$. Intuitively, in this case the local IV (expression (18)) identifies the gains in outcomes (including profits and wages but not the subsidy nor the intermediary cost) for those individuals induced to join the financial system as a consequence of the reduction in intermediation cost.

Importantly, one cannot interpret this parameter as the effect of financial intermediation on profits for entrepreneurs or on income for wage earners. This is because the change in Q also induces changes in occupational decisions in a nonuniform way. That is, changes in Q may endogenously induce individuals to switch from the wage sector to entrepreneurship and vice versa.

Additionally, although in principle the analyst could use the information on occupations to compute versions of $\Delta^{\text{IV}(Q)}(\overline{\overline{Q}}, \overline{Q}; b)$ among wage earners and/or entrepreneurs, in general these estimates would not identify the local causal effects of financial intermediation [as defined in (19)] in those populations. This is again a consequence of the nonuniform responses in occupational decisions induced by the changes in Q. Intuitively, by restricting the estimation of $\Delta^{\text{IV}(Q)}$ to entrepreneurs (wage earners) the analyst would be erroneously excluding the gains on outcomes from those initial entrepreneurs (wage earners) who would become wage earners (entrepreneurs) as a result of the change in Q. In other words, the analyst would be identifying the effect of financial intermediation for those entrepreneurs (wage

earners) who would have not switched occupations as a result of the change in the instrument. Given the economic incentives operating in the model, the $\Delta^{\mathrm{IV}(Q)}$ estimated in this way would only provide a partial response to the question of the effect of financial intermediation among entrepreneurs (wage earners).[15] We illustrate this point below.

We can use the same logic to identify the local average treatment effect of occupation (entrepreneurship) on income through the local IV estimator

$$\Delta^{\mathrm{IV}(\psi)}(\overline{\overline{\psi}}, \overline{\psi}; b) = \frac{E(\xi_i \mid \psi_i = \overline{\overline{\psi}}, b_i = b) - E(\xi_i \mid \psi_i = \overline{\psi}, b_i = b)}{E(\widetilde{D}_i \mid \psi_i = \overline{\overline{\psi}}, b_i = b) - E(\widetilde{D}_i \mid \psi_i = \overline{\psi}, b_i = b)},$$

where $\widetilde{D}_i$ is $D_i(r)\Upsilon_i + D_i(1 - \Upsilon_i)$. Under uniformity of ψ on $\widetilde{D}$, this parameter identifies the local treatment effect of occupation on income.

Analogously to the case of financial intermediation, one cannot use $\Delta^{\mathrm{IV}(\psi)}(\overline{\overline{\psi}}, \overline{\psi}; b)$ to determine the gains in income for those induced to enter to the financial system as a result of the subsidy. This is because the change in the subsidy does not produce necessarily uniform (or monotonic) movements with respect to intermediation choice.[16]

The complications of identifying $\Delta^{\mathrm{LATE}(Q)}(\overline{\overline{Q}}, \overline{Q}; b)$ by occupation or $\Delta^{\mathrm{LATE}(\psi)}(\overline{\overline{\psi}}, \overline{\psi}; b)$ by status of financial intermediation are due to the presence of two margins of choice in the model. Strictly speaking, the model includes four categories or possible treatments: wage sector and financial autarky, wage sector and financial intermediation, entrepreneurship and financial autarky, and entrepreneurship and financial intermediation. Indeed, we could phrase our discussion in the context of a model with multiple treatments and multiple instruments. In this framework, the definition of treatment effects is not as straightforward as in the binary case. Specifically, the pairwise comparison of the outcomes associated with different alternatives needs to be supplemented by considerations of the alternatives left out from the comparison. This adds a new level of complexities to the definition of treatment effects. As an example, notice that we can define the effect of financial intermediation on profits (i.e., the effect of intermediation among businesses) but for individuals effectively participating in the wage sector. This is not an intuitive treatment effect, but it is well defined in the context of a model with multiple treatments.

Heckman et al. (2006) analyze the identification power of instrumental variables in the context of models with multiple treatments and unobserved heterogeneity. They show that provided a variable (instrument) determining the preferences for a particular alternative but excluded from its potential outcome (e.g., instrument Z^j determining utility associated with alternative/option j, V_j), in models such as the one considered in this section, a local IV strategy (using Z_j as the instrument and based on a regression of observed income on dummy variables describing an individual's observed decisions) would identify the effect of option j versus the next best alternative.[17] This result complements our discussion about the

TABLE 5. Sorting by occupational choices and access to financial intermediation (model of occupational choice and financial intermediation—simulated cross-sectional data)

Occupational choice	Financial intermediation	Autarky
Wage earners	940	5,072
Entrepreneurs	2,678	16,310
Constrained	—	14,015
Unconstrained	—	2,295

Note: The number of observations in each cell is the result of the endogenous decision process faced by each of the 25,000 simulated individuals.

difficulties of further interpreting $\Delta^{IV(Q)}$ or $\Delta^{IV(\psi)}$.[18] See Heckman et al. (2006) for additional discussion.

3.2. Example

Following the same logic utilized in our previous example (Section 2.3), we investigate the consequences of using different econometric techniques when estimating the effects of financial intermediation and occupational choices. We use the parameterization presented in Table 1. Our main results are robust to different parameterizations. In addition to the structure presented in Table 1, we assume

$$Q_i \sim \text{Binomial}(0.25, 1)$$

with Q_i independent of $\psi_i, \theta_i^E, \theta_i^W$ and b_i. Table 5 presents the sorting simulated from the model for a sample size of 25,000 individuals. Given our parameterization, approximately one-fourth of the individuals become wage earners (most of them working under financial autarky), more than half of the individuals are entrepreneurs under autarky (most of whom are unconstrained), and the rest are entrepreneurs with access to financial intermediation.

Using cross-sectional information to estimate the effect of financial intermediation and occupational choices.   Suppose the econometrician focuses first on the impact of financial intermediation, proposing the linear model

$$Y_i = \kappa_0 + \kappa_1 b_i + \kappa_2 b_i \Upsilon_i + \kappa_3 \Upsilon_i + \varepsilon_i, \tag{20}$$

where Υ_i takes a value of 1 if the individual i has access to financial intermediation, and 0 otherwise.

Table 6 presents the results from OLS and IV on equation (20). The results suggest positive average effects of financial intermediation. However, because of the selection bias, the effect suggested by OLS is almost double the effect estimated by IV.

TABLE 6. OLS and IV estimates of the effect of financial intermediation on income (model of occupational choice and financial intermediation)

	κ_0	κ_1	κ_2	κ_3	Average effect ($\kappa_2\overline{b} + \kappa_3$)
Δ^{OLS}	1.015**	0.954**	0.313**	0.227**	0.585**
$\Delta^{\text{IV}(Q)}$	0.933**	1.071**	0.076*	0.236**	0.323*

Note: This table presents the parameters obtained from a linear regression of observed income (profits or wages depending on individual's occupation) on wealth, the financial intermediation dummy, and the interaction between wealth and financial intermediation (dummy). In addition, the row $\Delta^{\text{IV}(Q)}$ presents the estimates when Q is used as an instrument for endogenous financial intermediation. Overall these results illustrate what the analyst can obtain using information produced from the model (observed outcome, wealth, and financial intermediation) using a reduced-form strategy.
* Statistical significance at 5%.
** Statistical significance at 1%.

On the other hand, suppose that the analyst proposes the following linear model to investigate the effects of occupation on income:

$$Y_i = \tau_0 + \tau_1 b_i + \tau_2 b_i D_i + \tau_3 D_i + \varepsilon_i,$$

where D_i takes a value of 1 if the individual i is an entrepreneur, and 0 otherwise. These model follows closely the one presented in Section 2.3. Table 7 presents the IV and OLS of the effect of D_i on Y_i from our model-generated data. As in the previous section, the OLS estimate delivers a positive effect whereas the IV suggests a negative effect of occupation on income/profit.[19]

Using the structure of the model to generate counterfactual outcomes and the causal effect of financial intermediation and choices. Table 8 presents the model-generated local average treatment effects (LATE). These LATEs are not obtained using econometric techniques, but generated using the structure of the model. The table displays both $\Delta^{\text{LATE}(\psi)}(1, 0)$ (LATE associated with the effect of occupation) and $\Delta^{\text{LATE}(Q)}(0.25, 1)$ (LATE associated with the effect of financial intermediation).

TABLE 7. OLS and IV estimates of the effect of occupation on income (model of occupational choice and financial intermediation)

	τ_0	τ_1	τ_2	τ_3	Average effect ($\tau_2\overline{b} + \tau_3$)
Δ^{OLS}	0.578**	1.259**	−0.146**	0.433**	0.266**
$\Delta^{\text{IV}(\psi)}$	1.212**	1.177**	−0.027	−0.426**	−0.458*

Note: This table presents the parameters obtained from a linear regression of observed income (profits or wages depending on individual's occupation) on wealth, the occupational dummy, and the interaction between wealth and occupation (dummy). In addition, the row $\Delta^{\text{IV}(\psi)}$ presents the estimates when ψ is used as an instrument for the endogenous occupational decision. Overall these results illustrate what the analyst can obtain using information produced from the model (observed outcome, wealth, and occupation) using a reduced-form strategy.
* Statistical significance at 5%.
** Statistical significance at 1%.

TABLE 8. Model-generated local average treatment effects (model of occupational choice and financial intermediation)

Parameter	Value	Number of movers	Direction
$\Delta^{\text{LATE}(\psi)}(1, 0)$	**−0.466**	**2,219**	**From wage earner to entrepreneur**
	−0.444	1,548	From wage worker under autarky to entrepreneur under autarky
	−0.278	278	From wage worker under autarky to entrepreneur under financial intermediation
	−0.724	322	From wage worker under financial intermediation entrepreneur under autarky
	−0.519	71	From wage worker under financial intermediation to entrepreneur under financial intermediation
$\Delta^{\text{LATE}(Q)}(0.25, 1)$	**0.388**	**3,757**	**From autarky to financial intermediation**
	0.355	911	From wage worker under autarky to wage worker under financial intermediation
	−0.203	176	From wage worker under autarky to entrepreneur under financial intermediation
	0.752	75	From entrepreneur under autarky to wage worker under financial intermediation
	0.430	2,595	From entrepreneur under autarky to entrepreneur under financial intermediation

Note: The numbers in the table are obtained using the factual and counterfactual information on income generated by the economic model. Specifically, for each individual in the sample, we analyze the consequences of modifying the values of the instruments initially assigned. We study the individual's changes in occupational choices as well as the changes in decisions involving the financial system. Then, for each individual modifying her decisions as a result of the changes in Q or ψ, we compute the associated effects on income. This table presents the average effects on income generated using this logic. It also displays the number of individuals switching decisions as a result of the changes in the instrument (column *Number of movers*).

Importantly, our knowledge of the model allows us to generate not only an overall local average treatment effect (bold numbers in Table 8) but also the local effects of the treatment for specific groups of individuals. For example, in the case of financial intermediation, we obtain the local treatment effects for individuals switching from "wage-earner under autarky" to "wage-earner with access to financial system" (as a result of the exogenous change in the instrument) as well as the local effect for those "wage-earners under autarky" becoming "entrepreneurs under intermediation" (also as a result of a change in the instrument). This cannot be done without a structural analysis.

Notice that, as expected, the model-generated overall local treatment effects are very close to the effects estimated using the IV strategy (Tables 6 and 7).

TABLE 9. Model-generated treatment parameters associated with occupational choices and financial intermediation

Treatment parameter	Alternatives considered in the comparison	Value
Δ^{ATE}	Entrepreneurship vs. wage sector under financial autarky	0.619
Δ^{ATE}	Entrepreneurship vs. wage sector under financial intermediation	0.607
Δ^{ATE}	Financial intermediation vs. autarky for wage earners	0.227
Δ^{ATE}	Financial intermediation vs. autarky for entrepreneurs	0.215
Δ^{TT}	Entrepreneurship vs. wage sector under financial autarky	1.205
Δ^{TT}	Entrepreneurship vs. wage sector under financial intermediation	1.734
Δ^{TT}	Financial intermediation vs. autarky for wage earners	0.364
Δ^{TT}	Financial intermediation vs. autarky for entrepreneurs	0.433

Note: The table presents the treatment parameters associated with the pairwise comparison of different alternatives in the model, conditional on a specific alternative for the margin not considered in the comparison. For example, the first row presents the mean difference between profits and wages for individuals not participating in the financial system. The other rows can be interpreted using the same logic.

Table 8 also displays how the individuals in our model react to changes in the instrument. Interestingly, we observe how changes in Q induces people to move away from entrepreneurship and into the wage sector. In particular, and given our parameterization, 75 entrepreneurs would have become wage earners as a result of a change in Q. This illustrates our previous comment about the difficulty of interpreting $\Delta^{\text{IV}(Q)}$ as the effect of financial intermediation on profits for entrepreneurs and income for wage earners. The change in Q induces (nonuniform) changes in occupation. A similar logic prevents interpreting $\Delta^{\text{IV}(\psi)}$ as the effect of occupation on income for individual using the financial system or for individuals under financial autarky. As Table 8 shows, a change in ψ induces (nonuniform) changes in the financial participation decisions of the individuals in the model.

Finally, Table 9 presents the model-generated ATE and TT for the effect of financial intermediation and occupational choice. These causal parameters are presented for all the different groups of interest. It is worth noting the significant differences between these treatment effects and the local effects reported in Table 8. This illustrates the potential discrepancies between the different treatment parameters. All these parameters represent causal effects, but in our model with selection based on unobserved talents and gains, they all answer different economic questions.

4. DYNAMICS, RISK SHARING, UNOBSERVED HETEROGENEITY AND OCCUPATIONAL CHOICE

In this section we follow the analysis of Greenwood and Jovanovic (1999) (hereafter GJ), Townsend and Ueda (2006, in press), Felkner and Townsend (2007),

and Jeong Townsend (2008), with additional modifications. This literature discusses endogenous financial deepening and how well it fits both microeconomic and macroeconomic data, examining the evidence for targeting of government development banks and interest rate distortions that created a crisis and increased government involvement in the banking sector.

Consider a dynamic problem with an infinite horizon. Household i maximizes discounted expected utility,

$$E_0 \sum_{t=0}^{\infty} \beta_i^t u(c_{it}),$$

where $u(\cdot)$ is strictly concave and initial wealth is $k_{i,0} = b_{i,0}$. $E_0(\cdot)$ denotes the expectation given the information available at $t = 0$. We incorporate unobserved heterogeneity by allowing the individuals to differ in their discount factors. Specifically, we assume that $\beta_i = \overline{\beta} + \theta_i$, where θ_i is an individual-specific component known to the agent only, and $\overline{\beta}$ is common knowledge.

In autarky there is a law of motion for wealth as a function of savings, investment in specific occupations, and an exogenous random endowment. Let s_{it} denote the savings rate of household i at date t expressed as a fraction of wealth k_{it} at date t. Let Ψ_t^E be the proportion of the savings invested in a risky enterprise sector and let Ψ_t^W be the proportion invested in wage sector activities. Additionally, one unit of wealth invested in enterprise E yields $\delta_t^E + \varepsilon_{it}^E$ units of capital (wealth), whereas one unit invested in wage activity W yields an ex post rate of return of $\delta_t^W + \varepsilon_{it}^W$. The returns δ_t^E and δ_t^W are realized at the end of date t and are unknown when within-period decisions are made.

The law of motion for wealth in autarky is thus

$$k_{it+1} = s_{it} \times \left[\Psi_t^E \times \left(\delta_t^E + \varepsilon_{it}^E \right) + \Psi_t^W \times \left(\delta_t^W + \varepsilon_{it}^W \right) \right] \times k_{it}. \tag{21}$$

Consumption in autarky at t, c_{it}^A, is the residual; that is, $c_{it}^A = (1 - s_{it}) k_{it}$.

The value function W_0 associated with financial autarky, A, exists under standard regularity conditions. It satisfies the Bellman equation

$$W_0(k_{it}, \theta_i) = \max_{\Psi_i^E, \Psi_i^W, c_{it}, s_{it}} u(c_{it}) + \beta_i E \left(W_0(k_{it+1}, \theta_i) \right)$$

subject to (21). The function $W_0(k_{it}, \theta_i)$ is strictly concave in k_{it}. Under general preferences, the saving and investment policies are functions of wealth k_{it}. However, for CRRA preferences ($u(c_{it}) = c_{it}^\gamma$), they are constant. More precisely, under these preferences,

$$c_{it}^A = \widetilde{\alpha}_i^A k_{it} = \widetilde{\alpha}_i^A \left(y_{it}^E + y_{it}^W \right),$$

where $\widetilde{\alpha}_i^A = (1 - \beta_i)$, y_{it}^E is the income from enterprise, and y_{it}^W is the labor income; that is,

$$y_{it}^E = \Psi_{t-1}^E \left(\delta_{t-1}^E + \varepsilon_{it-1}^E \right) k_{it-1} s_{it-1}$$
$$y_{it}^W = \Psi_{t-1}^W \left(\delta_{t-1}^W + \varepsilon_{it-1}^W \right) k_{it-1} s_{it-1}.$$

Therefore, and because by definition $\beta_i = \bar{\beta} + \theta_i$, we can write the equation describing optimal consumption in autarky (A) as

$$c_{it}^A = (1 - \bar{\beta} - \theta_i) y_{it}$$
$$= \alpha^A y_{it} + \varepsilon_{it}^A,$$

where y_{it} is the sum of all sources of income ($y_{it} = y_{it}^E + y_{it}^W$), $\alpha^A = 1 - \bar{\beta}$, and $\varepsilon_{it}^A = -\theta_i y_{it}$ is the unobserved component.

Participation in the intermediated sector, on the other hand, allows a household to share any idiosyncratic shock and, as in GJ, get perfect advance information on aggregate shocks δ_t^E, δ_t^W.[20] The bank directs all investment as if each household were exchanging shares in its own return stream for shares in a common mutual fund. The law of motion for wealth is then

$$k_{it+1} = s_{it} k_{it} \max \left\{ \delta_t^W, \delta_t^E \right\} (1 - \tau), \qquad (22)$$

where τ is the marginal intermediation transaction cost. The value function V_I for those in the intermediated sector, I, satisfies the Bellman equation,

$$V_I(k_{it}, \theta_i) = \max_{c_{it}, s_{it}} \left[u(c_{it}) + \beta_i E(V_I(k_{it+1}, \theta_i)) \right],$$

subject to (22). Again $V_I(k_{it}, \theta_i)$ is strictly concave in k_{it}. Policy s_{it} might be a nonlinear function of k_{it}, but again under CRRA preferences, s_{it} is linear in k_{it}. Thus,

$$c_{it}^I = \widetilde{\alpha}_i^I A_t,$$

where the aggregate shock A_t is equal to $\max\{\delta_{t-1}^W, \delta_{t-1}^E\}(1 - \tau)$, and $\widetilde{\alpha}_i^I$ is equal to $(1 - \bar{\beta} - \theta_i)$. Following our previous analysis, we can write

$$c_{it}^I = \alpha^I A_t + \varepsilon_{it}^I,$$

where $\alpha^I = 1 - \bar{\beta}$ and $\varepsilon_{it}^I = -\theta_i A_t$ is the unobserved component.

4.1. Once-and-for-All Participation Decisions and Participation Costs as Instruments

In this section we extend the analysis of GJ. In particular, although GJ has endogenous entry determined by the solution to a dynamic programming problem with a period-by-period decision, we consider the special case of a once-and-for-all entry decision at an initial date. For an empirical application of this idea see Alem and Townsend (2008).

Initially, at $t = 0$, given k_{i0}, the household decides whether to participate in the financial sector or not. Once this is decided, there is no going back. Let Z_i denote an individual-specific participation cost. This subtracts from wealth k_{i0}. Again, this cost is meant to capture exogenous variation in the ability to access intermediation, through either policy variation or physical infrastructure. These

can be thought of as household-specific transaction costs (with any correlation across individuals taken into account by other control variables, which is the way we treat wealth below). In the original GJ model, these costs are subtracted upon entry to the financial system. These are also transaction costs models in the finance literature, for example, Vissing-Jorgensen (2002).

Then, with V_I and W_0 strictly concave in k_{it}, the decision to participate depends on participation cost Z_i and wealth k_{i0}. More precisely, if we denote by I_{i0} the participation decision, we can write

$$I_{i0} = 1 \Leftrightarrow V_I (k_{i0} - Z_i, \theta_i) \geq W_0 (k_{i0}, \theta_i).$$

Additionally, we can write observed consumption at t as a function of potential consumption levels (c_{it}^I, c_{it}^A) and the participation decision I_{i0}:

$$
\begin{aligned}
c_{it} &= c_{it}^A(1 - I_{i0}) + c_{it}^I I_{i0} \\
c_{it} &= \alpha^A y_{it} + (\alpha^I A_t - \alpha^A y_{it}) I_{i0} + v_{it},
\end{aligned}
\tag{23}
$$

where $v_{it} = \varepsilon_{it}^A + I_{i0}(\varepsilon_{it}^I - \varepsilon_{it}^A)$. Equation (23) shows how, if intermediation is effective for those who choose it, idiosyncratic income y_{it} should not determine consumption.

Notice that the error term in (24), v_{it}, depends on the decision made at $t = 0$, I_{i0}, so there is a selection bias argument that prevents the researcher from using OLS in the estimation of (23). In this context, an IV strategy becomes an appealing alternative.

The obvious issue is then how to come up with a valid instrument. Interestingly, the economic model delivers a natural instrument, namely Z_i. In order to see this, notice that under autarky and the assumption of CRRA preferences, optimal saving rates and proportion of savings invested in each sector do not depend on k_{it}. As a result of this, potential consumption in the intermediated and autarkic sectors does not depend on the choice of the intermediation other than at $t = 0$ (when the costs are paid). Consequently, although Z_i affects the initial choice of the intermediation sector versus financial autarky, for all time periods $t > 0$ the individual participation cost does not affect the potential levels of consumption, c_{it}^A and c_{it}^I. These two conditions make Z_i a valid instrument for the effect of intermediation on consumption.

Using the instrument Z_i, the researcher can identify LATE, a causal relationship between financial intermediation and consumption.

Estimating the average treatment effect (ATE) or the treatment effect on those treated (TT) is more delicate. Notice that due to the role of θ_i in the model, I_{i0} is correlated with each of the components of v_{it}, namely ε_{it}^A and $I_{i0}(\varepsilon_{it}^I - \varepsilon_{it}^A)$. This structure is similar to the one discussed in the context of the models introduced in Sections 2 and 3. As in those cases, the presence of unobserved components and the endogenous selection of the individuals into sectors (based on the comparison of counterfactual outcomes affected by unobserved variables) produces heterogeneity in treatment effects. In this context, we can show that under the assumption of

a uniform response of I_{i0} to changes in Z_i (for all i), the instrumental variable estimator will indeed identify a causal effect of I_{i0} on c_{it} [see Heckman et al. (2006). Imbens and Angrist (1994)]. But the causal effect identified by IV might be different from, for example, ATE or TT. Only under the special case of no selection on unobserved gains would IV, ATE, and TT be identical. However, the presence of unobserved components and the endogenous selection process make this case unlikely.[21]

4.2. Sequential Participation Decisions

Now suppose the choice of sector takes place each period t, not just initially. Then for those not yet in the intermediated sector at $t \geq 0$, but who may choose so at $t + 1$, the value function satisfies the Bellman equation

$$W_0 (k_{it}, \theta_i) = \max_{\Psi_t^E, \Psi_t^W, c_{it}, s_{it}} \{U(c_{it}) + \beta_i E \max \{W_0(k_{it+1}, \theta_i), V_1(k_{it+1} - Z_i, \theta_i)\}\}$$

subject to $k_{it+1} = s_{it} \times [\Psi_t^E \times (\delta_t^E + \varepsilon_{it}^E) + \Psi_t^W \times (\delta_t^W + \varepsilon_{it}^W)] \times k_{it}$.

There is a critical family of values $k^*(Z_i, \theta_i)$ that define thresholds for participation. Under some regularity conditions entry is permanent. However, saving $s_t (k_{it})$ and investments $\Psi_t^E (k_{it})$, $\Psi_t^W (k_{it})$ are generally functions of wealth k_{it} even with CRRA utility. It can be established, in fact, that savings and investment in risky assets will rise with k_{it} as that wealth approaches the critical entry $k^*(Z_i, \theta_i)$. See Townsend and Ueda (2006).

Thus variation in Z_i determines both k^* and preparticipation outcomes. Therefore, Z_i cannot be considered a potential instrument. Careful researchers do take into account the impacts of anticipated policy when designing experiments. Subjects are not given full information on what is to happen step by step. See Olken (2007).

4.3. The Identification Power of Policies

Interestingly, the existence of unanticipated policies can allow us to identify the effect of financial intermediation on consumption. For example, assume a once-and-never-more policy that shifts at some date t^* the cost of participation Z_i. Then period t^* can be interpreted as period zero and the earlier analysis applies (except that we have preintervention data, and savings and investment are nonlinear in wealth k_{it}). At period t^* we have pre-established positions for those not yet in, and the participation decision for them is

$$I_{it^*} > 0 \Leftrightarrow V_1 (k_{it^*} - Z_i, \theta_i) \geq W_1 (k_{it^*}, \theta_i).$$

In effect, the policy change can be interpreted as a once-and-for-all wealth shock in the event of joining the financial sector. Consumption equations are as before.

For $t > t^*$, we have

$$c_{it} = c_{it}^A \times (1 - I_{it}) + c_{it}^I \times I_{it}.$$

Then, if the agent enters at t^*, induced by the sudden and temporary policy change, we can analyze this decision as if it were a "once for all" decision. In this case, the policy changes the entry decision, but it does not affect the potential outcomes at $t > t^*$.[22]

However, if the policy is permanent, then the policy is subject to the same qualifications as in the case when the choice of sector takes place each period. Subsequently, pre-entry behavior for those not yet entering at period t^* will be altered.

5. A MODEL OF FINANCIAL INTERMEDIATION WITH MORAL HAZARD AND COLLATERAL CONSTRAINTS

5.1. Statics

In this section, we study the consequences for impact evaluation of a model with financial intermediation with moral hazard. Our model is similar to the one discussed in Paulson et al. (2006) estimated using data from Thailand. This follows the tradition of the earlier literature on occupation choice but attempts to estimate the financial regime in place, that is, moral hazard versus limited commitment. Here we focus on moral hazard and the endogeneity of the intermediation decision.

We first introduce the static version of the model, though for simplicity we suppress occupation choice and focus on firms. We also focus our interest on the empirical consequences of randomized contracts. We then go to the dynamics.[23]

We denote by $u(c_i, e_i)$ the utility function associated with individual i. This function is increasing in consumption c_i and decreasing in effort e_i. The technology in the model is described by a stochastic production function $\Pr(q_i \mid e_i, \theta_i)$, where q_i denotes outcome and θ_i represents the individual's talent or type.

The individual as a firm must decide whether or not to participate in a lottery determining who gets intermediated. If participating in the lottery, s/he must pay an amount b_i to the bank and, as a results of this, s/he gets a randomized contract determining if s/he will have to run her/his business in autarky or if the output will depend on a transfer agreement associated with credit and insurance. The entry into randomization has fixed and individual-specific costs Z_i. However, Z_i produces a natural source of variation that can be used to identify and estimate the effect of financial intermediation on consumption. We illustrate this point in our example.

Overall, the timing of the model is as follow. First, wealth b_i is transferred to the bank. Then, the outcome of the lottery is revealed. If the result is autarky, some wealth may be transferred from the bank to the individual before s/he "opens" the business. This amount is such that the on average the individual ends up

with the same wealth level as in autarky. Let w_{Ai} denote the optimal transfer and $\Pi_A(w_{Ai}, q_i, e_i)$ be the joint distribution of the transfer, production, and effort.

Here $\Pi_A(w_{Ai}, q_i, e_i)$ allows nontrivial probabilities, but much of the outcomes can be deterministic. $\Pi_A(w_{Ai}, q_i, e_i)$ makes the problem linear. The following expressions characterize the problem of determining the optimal transferred level w_{Ai} for each θ_i type;

$$\max_{\Pi_A} \sum_{w_{Ai}, q_i, e_i} \Pi_A(w_{Ai}, q_i, e_i) u(q_i + w_{Ai}, e_i)$$

s.t.

$$\sum_{w_{Ai}} \Pi_A(w_{Ai}, \bar{q}_i, \bar{e}_i) = \Pr(\bar{q}_i \mid \bar{e}_i, \theta_i) \sum_{w_{Ai}, q_i} \Pi_A(w_{Ai}, q_i, \bar{e}_i) \; \forall \bar{q}_i, \bar{e}_i \qquad \textbf{(24)}$$

$$\sum_{w_{Ai}, q_i, e_i} \Pi_A(w_{Ai}, q_i, e_i) w_{Ai} = b_i \qquad \textbf{(25)}$$

$$\Pi_A \geq 0 \quad \text{and} \quad \sum_{w_{Ai}, q_i, e_i} \Pi_A(w_{Ai}, q_i, e_i) = 1. \qquad \textbf{(26)}$$

The first constraint in this program implies that, regardless of the initial transfer, the distribution of output given the effort level is consistent with the production function associated with the individual's type θ_i, $\Pr(q_i \mid e_i, \theta_i)$. The second constraint gives back to the agent his wealth in expectation.

On the other hand, if the outcome of the lottery is "intermediation," the contract defines first a recommended level of effort, and then a distribution of consumption conditional on output. These choices are described by the joint distribution of consumption, output, and effort under intermediation $[\Pi_I(c_i, q_i, e_i)]$. Again, e_i may be deterministic and c_i a nontrivial function of q_i. Additionally, we assume the existence of an individual-specific utility cost κ_{Ii} in the case of being intermediated, as otherwise intermediation would always dominate autarky.

Because effort is only known by the individuals, we need to add the following constraint that makes recommended effort e_i weakly dominate any $\bar{e}_i$:

$$\sum_{c_i, q_i} \Pi_I(c_i, q_i, e_i) u(c_i, e_i) \geq \sum_{c_i, q_i} \frac{\Pr(q_i \mid \bar{e}_i, \theta_i)}{\Pr(q_i \mid e_i, \theta_i)} \Pi_I(c_i, q_i, e_i) u(c_i, \bar{e}_i) \; \forall \bar{e}_i, e_i.$$

$$\textbf{(27)}$$

Additionally, the joint distribution of consumption, output, and effort under intermediation must be consistent with the production technology $\Pr(q_i \mid e_i, \theta_i)$. Thus, we require that

$$\sum_{c_i} \Pi_I(c_i, \bar{q}_i, \bar{e}_i) = \Pr(\bar{q}_i \mid \bar{e}_i, \theta_i) \sum_{c_i, q_i} \Pi_I(c_i, q_i, \bar{e}_i) \; \forall \bar{q}_i, \bar{e}_i. \qquad \textbf{(28)}$$

In sum, the contract can be characterized by the joint distribution of transfer, output, and effort under financial autarky $\Pi_A (w_{Ai}, q_i, e_i)$ and by the joint distribution of consumption, output, and effort under intermediation $\Pi_I (c_i, q_i, e_i)$.

We must impose

$$\Pi_I, \Pi_A \geq 0 \tag{29}$$

$$\sum_{c_i, q_i, e_i} \Pi_I(c_i, q_i, e_i) + \sum_{w_{Ai}, q_i, e_i} \Pi_A(w_{Ai}, q_i, e_i) = 1. \tag{30}$$

Finally, we impose a zero–expected profit condition on our bank. Therefore, the following constraint must hold for each θ_i type:

$$\sum_{c_i, q_i, e_i} \Pi_I(c_i, q_i, e_i) (c_i - q_i) + \sum_{w_{Ai}, q_i, e_i} \Pi_A (w_{Ai}, q_i, e_i) w_{Ai} = b_i - Z_i. \tag{31}$$

This constraint implies that the net expected amount of transfers to the individuals (under both regimes) equals the initial transfer received by the bank.

Program 1 describes the efficient arrangements given b_i.

Program 1.

$$U(b_i; \theta_i, Z_i) = \max_{\Pi_I, \Pi_A} \sum_{c_i, q_i, e_i} \Pi_I(c_i, q_i, e_i)[u(c_i, e_i) - \kappa_{Ii}]$$

$$+ \sum_{w_{Ai}, q_i, e_i} \Pi_A(w_{Ai}, q_i, e_i)u(q_i + w_{Ai}, e_i)$$

s.t.   (24)–(31).

The outcome of this program $U(b_i; \theta_i, Z_i)$ is the indirect utility from the contract given a wealth level b_i and the individual's type θ_i and cost Z_i. Notice the important role of k_{Ii}. If $k_{Ii} < 0$ we will have intermediation with probability one. On the contrary, a nonnegative k_{Ii} will make intermediation more attractive for low values of wealth b_i. Therefore, the possibility of randomization will occur only for nonnegative values of k_{Ii}.

Random assignments of wealth can help us to recover instruments at least over specified ranges of ex ante wealth. Figure 3 illustrates this point.

For values of wealth $b_L < b_i - Z_i < b_U$ a lottery puts mass on participation and autarky points in proportion to the utility distance. That is, suppose that an individual with initial wealth b_i in this range forfeits Z_i in wealth and enters the lottery with $b_i - Z_i$. Then the effect of cost Z_i is to shift ex ante wealth to the left and increase the probability of losing the lottery, that is, becoming poor and needing the financial system.

Figure 3 shows that when $b_i < b_L$, intermediation is chosen with probability one, and those agents do not play the lottery (and do not pay costs Z_i).

The point is that in the relevant range of wealth (and only in that range) costs Z_i affect the probability of participation without changing outcomes associated with the participation decision. This logic produces the instrument. Additionally,
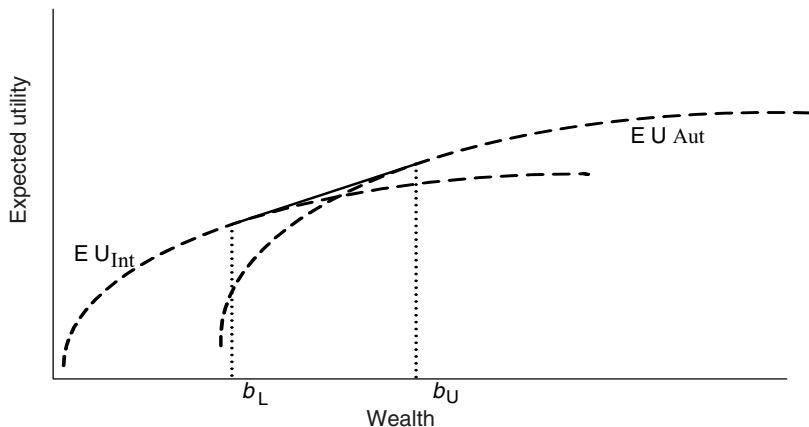
FIGURE 3. Random assignments of wealth as a source of instruments.

changes in the instrument produce uniform or monotonic responses in the chances of getting access to the financial system. Therefore, even under the presence of unobserved talent driving consumption levels and probabilities of intermediation, the IV strategy will identify a causal effect associated with financial intermediation.

We note that ex ante expected utility is a function of the instrument, and we come back to this in a consideration of dynamics.

Example.   To understand the consequences of our analysis for the impact evaluation of financial intermediation, we generate data from our theoretical model and estimate what the effect of financial intermediation would be using different econometric techniques. Specifically, we use our model to generate data on consumption, wealth, and financial intermediation for a sample of approximately 1,800,000 individuals.[24] As previously explained, talent plays a critical role in our theoretical model, but because talent is observed only by the individual, we do not condition on it.

Table 10 presents our parameterization of the theoretical model. In our data, we observe 67.90% of the individuals (endogenously) reporting financial intermediation. Notice that wealth (b) and the instrument (Z) are defined as continuous random variables. However, once we identify the region in which randomization is nontrivial, we solve the model for a set of discrete values of b and Z. Specifically, we work with ten values for both wealth ($b_1, \ldots, b_{10}$) and the instrument ($Z_1, \ldots, Z_{10}$).[25] This allows us not only to make the numerical solution of the theoretical problem feasible but also to mimic what an analyst would face in reality.

First, we estimate the effect of financial intermediation using OLS and IV techniques. We denote by c_i and D_i observed consumption and financial intermediation (dummy variable), respectively. b_i denotes the individual's wealth level.

TABLE 10. Model of financial intermediation with moral hazard and collateral constraints (parameterization)

Utility function	$u(c, e) = -100c^{-1.5} - v(e)$
Disutility of effort	$v(0.06) = 2.9, v(16) = 3$
Probability of high output	$\Pr(q_H \mid e, \theta) = \frac{\theta^{0.5} e^{0.5}}{1 + \theta^{0.5} e^{0.5}}$
Effort grid	$e \in \{0.06, 16\}$
Output grid	$q \in \{0.5, 15\}$
Cost of intermediation	$\kappa_I = 0.1$
Talent (θ) and wealth (b)	$(\theta, b) \sim N \left[(0, 0), \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right]$
Instrument (Z)	$Z \sim U(0, 1)$ with Z independent from θ and b

We carry out the estimation using two different models of consumption,[26]

$$c_i = \alpha + \beta D_i + \delta b_i + \epsilon_i \tag{32}$$

$$c_i = \alpha(b_k) + \beta(b_k) D_i + \epsilon_i \quad \text{for } b_k = b_1, \ldots, b_{10}, \tag{33}$$

where, in (33), the dependency of the coefficients α and β on values of wealth (b_k) indicates that the equation is estimated for each b_k. In this manner, the second equation represents a more flexible functional form than the first (and standard) model. The needs for instruments comes from the fact that intermediation D_i is an endogenous variable. Additionally, given the presence of unobserved talent and the endogenous selection mechanism driving the intermediation decisions, behind expressions (32) and (33) we have a model with heterogeneous treatment effects.

We denote by $\Delta^{\text{OLS}}(b_k)$ and $\Delta^{\text{IV}}(b_k)$ the effect of financial intermediation on consumption (conditional on wealth) obtained from model (33). Table 11 presents our results. The results are ordered increasingly on wealth (i.e., $b_1 < b_2 < \ldots < b_{10}$). The last two rows of Table 11 present the overall effects obtained from (32) and (33) (across wealth levels). The comparison of these columns illustrates the empirical consequences of imposing a priori the restricted functional form (32), i.e., $\alpha(b_k) = \alpha(b'_k)$ and $\beta(b_k) = \beta(b'_k)$ for all (k, k'), and the biased results delivered by OLS.

Additionally, and following Imbens and Angrist (1994), we can decompose the IV estimates ($\Delta^{\text{IV}}(b_k)$) presented in Table 11 into their local components. Specifically, we can write

$$\Delta^{\text{IV}}(b_k) = \sum_{l=1}^{9} \Delta^{\text{IV}}(Z_{l+1}, Z_l; b_k) \lambda_l(b_k) \ \forall b_k \text{ with } k = 1, \ldots, 10,$$

TABLE 11. IV and OLS estimates computed by wealth level—static model

Wealth level	$\Delta^{\text{IV}}(b_k)$ (1)	$\Delta^{\text{OLS}}(b_k)$ (2)	% Difference $\left(\frac{(1)-(2)}{(2)}\right)$
b_1	−0.91	−0.87	−4.8%
b_2	−2.02	−1.85	−8.9%
b_3	−2.92	−2.83	−3.1%
b_4	−3.97	−3.92	−1.4%
b_5	−4.80	−4.91	2.3%
b_6	−5.76	−5.97	3.5%
b_7	−6.91	−7.00	1.3%
b_8	−7.91	−8.04	1.7%
b_9	−9.09	−9.04	−0.5%
b_{10}	−9.75	−10.05	3.0%
Overall effect–full interactions (equation (33))	−5.40	−5.72	5.6%
Overall effect–restricted model (equation (32))	−5.38	−5.45	1.3%

Note: All estimates are statistically significant at the 5% level.

where

$$\Delta^{\text{IV}}(Z_{l+1}, Z_l; b_k)$$

$$= \frac{E(c_i \mid Z = Z_{l+1}, b = b_k) - E(c_i \mid Z = Z_l, b = b_k)}{\Pr(D_i = 1 \mid Z = Z_{l+1}, b = b_k) - \Pr(D_i = 1 \mid Z = Z_l, b = b_k)},$$

and the weights are such that $\lambda_l(b_k) \geq 0$ and $\sum_l \lambda_l(b_k) = 1$ for all b_k with $k = 1, \ldots, 10$. Table 12 presents the estimated $\Delta^{\text{IV}}(Z_{l+1}, Z_l; b_k)$ obtained using data generated from the model, whereas Table 13 presents the associated weights. The variability of $\Delta^{\text{IV}}(Z_{l+1}, Z_l; b_k)$ demonstrates the presence of unobserved heterogeneity in our model across levels of wealth. The IVs presented in Table 11 gives a partial picture of the local effects contained in the data.[27]

Importantly, the local IV estimates presented in Table 12 have a causal interpretation. Specifically, they identify the effects of the treatment for those individuals induced to switch regime as a result of a change in the instrument. In other words, $\Delta^{\text{IV}}(Z_{l+1}, Z_l; b_k)$ identifies $\Delta^{\text{LATE}}(Z_{l+1}, Z_l; b^k) = E(c_i^I - c_i^A \mid D_i(Z_{l+1}) - D_i(Z_l) = 1, b = b_k)$, where c_i^I and c_i^A denote the consumption levels under intermediation and autarky, respectively, and $D_i(Z_l)$ denotes the value for the dummy variable associated with intermediation when individual i faces $Z = Z_l$. This causal interpretation of IV comes from the fact that Z is a valid instrument and from the assumption of a uniform (or monotonic) effect of Z on D (from the lottery). In the next section, we show how this causal interpretation of local IV breaks down in the context of a dynamic model.[28]

TABLE 12. Local IV estimates by wealth level ($\Delta^{\text{IV}}(Z_{l+1}, Z_l; b_k)$, static model)

Wealth level	(Z_2, Z_1)	(Z_3, Z_2)	(Z_4, Z_3)	(Z_5, Z_4)	(Z_6, Z_5)	(Z_7, Z_6)	(Z_8, Z_7)	(Z_9, Z_8)	(Z_{10}, Z_9)	$\Delta^{\text{IV}}(b_k)$
b_1	11.89	−8.95	−0.29	−5.53	7.73	−10.81	5.60	−6.13	9.03	−0.92
b_2	−7.49	−1.31	1.29	−1.32	−2.97	−6.23	−2.69	1.74	1.05	−2.02
b_3	−3.36	−2.79	−3.17	−1.63	−0.93	−6.39	−4.06	0.10	−4.17	−2.92
b_4	−3.76	−3.50	−1.95	−7.04	−5.36	−2.91	−2.22	−3.25	−6.09	−3.97
b_5	−7.05	−3.88	−2.43	−4.84	−6.72	−7.51	−3.65	−1.40	−5.79	−4.80
b_6	−2.84	−5.24	−3.58	−8.85	−5.27	−7.12	−6.86	−3.94	−5.26	−5.76
b_7	−5.58	−7.88	−4.96	−6.42	−8.63	−6.88	−8.22	−6.18	−6.00	−6.91
b_8	−9.72	−6.58	−8.43	−10.29	−10.13	−4.44	−10.11	−3.57	−6.01	−7.91
b_9	−4.17	−8.68	−12.31	−15.37	−1.80	−12.74	1.77	−19.21	−7.99	−9.10
b_{10}	−16.29	−11.53	0.92	−24.90	−12.82	7.29	−22.50	5.74	−17.65	−9.75

TABLE 13. IV weights by wealth level ($\lambda_l(b_k)$) (static model)

Wealth level	(Z_2, Z_1) (1)	(Z_3, Z_2) (2)	(Z_4, Z_3) (3)	(Z_5, Z_4) (4)	(Z_6, Z_5) (5)	(Z_7, Z_6) (6)	(Z_8, Z_7) (7)	(Z_9, Z_8) (8)	(Z_{10}, Z_9) (9)	$\sum_l \lambda_l(b_k)$ $(1) + (2) + \cdots + (9)$
b_1	0.05	0.10	0.13	0.15	0.15	0.14	0.13	0.10	0.05	1.00
b_2	0.05	0.10	0.13	0.15	0.14	0.15	0.13	0.10	0.05	1.00
b_3	0.05	0.10	0.13	0.15	0.15	0.15	0.13	0.10	0.05	1.00
b_4	0.05	0.10	0.13	0.15	0.15	0.15	0.13	0.10	0.05	1.00
b_5	0.05	0.10	0.12	0.15	0.15	0.15	0.13	0.10	0.05	1.00
b_6	0.06	0.10	0.13	0.14	0.15	0.15	0.13	0.10	0.05	1.00
b_7	0.05	0.10	0.13	0.14	0.15	0.15	0.13	0.09	0.05	1.00
b_8	0.05	0.10	0.12	0.16	0.15	0.14	0.13	0.10	0.06	1.00
b_9	0.06	0.10	0.13	0.15	0.14	0.15	0.13	0.09	0.05	1.00
b_{10}	0.06	0.11	0.12	0.13	0.15	0.14	0.14	0.09	0.06	1.00

TABLE 14. Model-generated treatment parameters: ATE, TT, and TUT, by wealth level (static model)

Wealth level	$\Delta^{\text{ATE}}(b_k)$	$\Delta^{\text{TT}}(b_k)$	$\Delta^{\text{TUT}}(b_k)$	Consumption under financial autarky
b_1	-0.81	-0.81	-0.83	22.92
b_2	-1.83	-1.84	-1.82	23.95
b_3	-2.85	-2.87	-2.79	24.97
b_4	-3.87	-3.88	-3.86	25.99
b_5	-4.89	-4.90	-4.87	27.01
b_6	-5.91	-5.92	-5.91	28.03
b_7	-6.93	-6.93	-6.94	29.05
b_8	-7.95	-7.93	-7.99	30.07
b_9	-8.97	-8.97	-8.98	31.09
b_{10}	-9.99	-10.00	-9.99	32.11
Overall	-5.40	-5.17	-5.90	27.52

As previously discussed, in the context of models with unobserved heterogeneity, reduced-form approaches (including IVs) might not give estimates of the average effect of the treatment (ATE) or the treatment effect on those treated (TT). This is because each of these parameters depends in one way or another on counterfactual outcomes, and therefore, their estimation requires additional structure. Fortunately for us, full control of our model allows us to generate these counterfactual states, and consequently, all the treatment parameters. Table 14 presents these treatment parameters. It also presents the average treatment effects for those untreated or TUT. We immediately observe that there are differences among the treatment parameters, which is again a manifestation of the presence of unobserved heterogeneity.

5.2. Dynamic Mechanism Design

Suppose now there are two time periods in our contract model. We continue defining Z_i as a cost of entering the lottery, which is subtracted from wealth b_i.

We denote by b_i' the wealth level in the second period, which is a "decision" variable in the context of the first period. Individuals in our model are allowed to switch from intermediation today to autarky tomorrow, and also the opposite.

The program introduced in Section 5.1 already determined the optimal arrangement in the second period. Importantly, in the first period, not only consumption but also the characteristics of the future arrangement are used to reward individuals. But indirect utility $U(b_i'; \theta_i, Z_i)$ carries all the information from the second period arrangement that is relevant for the characterization of the optimal contract in the first period as only utility matters for incentives. We use this fact in what follows.

When the result of the lottery is autarky in the first period, a particular distribution of transfer to the individual, w_{Ai}, is determined. Then, the individual decides the amount of effort, the output level is obtained, and finally, the individual decides how to split his resources $q_i + w_{Ai}$ between consumption today and wealth level for the second period. Thus, the program determining the optimal policy given the available resources b_i can be written as

$$U_{Ai}(b_i; \theta_i, Z_i)$$
$$= \max_{\Pi_A} \sum_{b'_i, w_{Ai}, q_i, e_i} \Pi_A(b'_i, w_{Ai}, q_i, e_i)[u(q_i + w_{Ai} - b'_i, e_i) + U(b'_i; \theta_i, Z_i)]$$

s.t.

$$\sum_{b'_i, w_{Ai}} \Pi_A(b'_i, w_{Ai}, \bar{q}_i, \bar{e}_i) = \Pr(\bar{q}_i \mid \bar{e}_i, \theta_i) \sum_{b'_i, w_{Ai}, q_i} \Pi_A(b'_i, w_{Ai}, \bar{q}_i, \bar{e}_i) \ \forall \bar{q}_i, \bar{e}_i \quad \textbf{(34)}$$

$$\sum_{b'_i, w_{Ai}, q_i, e_i} \Pi_A(b'_i, w_{Ai}, q_i, e_i) w_{Ai} = b_i \quad \textbf{(35)}$$

$$\Pi_A \geq 0 \quad \text{and} \quad \sum_{b'_i, w_{Ai}, q_i, e_i} \Pi_A(b'_i, w_{Ai}, q_i, e_i) = 1. \quad \textbf{(36)}$$

In this dynamic version of the model, the contract under intermediation can be characterized by the joint distribution of the next period's wealth and the first period levels of consumption, production, and effort, $\Pi_I(b'_i, c_i, q_i, e_i)$.

The incentive constraint under intermediation in the first period is

$$\sum_{b'_i, c_i, q_i} \Pi_I(b'_i, c_i, q_i, e_i)[u(c_i, e_i) + U(b'_i; \theta_i, Z_i)]$$

$$\geq \sum_{b'_i, c_i, q_i} \Pi_I(b'_i, c_i, q_i, e_i) \frac{\Pr(q_i|\bar{e}_i, \theta_i)}{\Pr(q_i|e_i, \theta_i)}[u(c_i, \bar{e}_i) + U(b'_i; \theta_i, Z_i)] \ \forall \bar{e}_i, e_i \quad \textbf{(37)}$$

and the constraint securing that $\Pi_I(b'_i, c_i, q_i, e_i)$ is consistent with the stochastic technology is

$$\sum_{b'_i, c_i} \Pi_I(b'_i, c_i, \bar{q}_i, \bar{e}_i) = \Pr(\bar{q}_i|\bar{e}_i, \theta_i) \sum_{b'_i, c_i, q_i} \Pi_I(b'_i, c_i, q_i, \bar{e}_i) \ \forall \bar{q}_i, \bar{e}_i. \quad \textbf{(38)}$$

Finally, the probability constraints are

$$\Pi_I, \Pi_A \geq 0 \quad \textbf{(39)}$$

$$\sum_{b'_i, c_i, q_i, e_i} \Pi_I(b'_i, c_i, q_i, e_i) + \sum_{b'_i, w_{Ai}, q_i, e_i} \Pi_A(b'_i, w_{Ai}, q_i, e_i) = 1. \quad \textbf{(40)}$$

The bank faces the following zero–expected profit condition:

$$\sum_{b'_i, c_i, q_i, e_i} \Pi_I(b'_i, c_i, q_i, e_i)(c_i + b'_i - q_i)$$

$$+ \sum_{b'_i, w_{Ai}, q_i, e_i} \Pi_A(b'_i, w_{Ai}, q_i, e_i) w_{Ai} = b_i - Z_i. \tag{41}$$

This constraint enssures that the amount initially given to the bank equals the expected amount transferred for the individual under either intermediation or autarky, plus the cost of intermediation Z_i. (A similar constraint is already imposed in the second period).

Therefore, given the initial promise $U(b'_i; \theta_i, Z_i)$, the first period program describing the efficient allocation of resources becomes:

Program 2.

$$\max_{\Pi_I, \Pi_A} \sum_{b'_i, c_i, q_i, e_i} \Pi_I(b'_i, c_i, q_i, e_i)[U(b'_i; \theta_i, Z_i) + u(c_i, e_i) - \kappa_{Ii}]$$

$$+ \sum_{b'_i, w_{Ai}, q_i, e_i} \Pi_A(b'_i, w_{Ai}, q_i, e_i)[U(b'_i; \theta_i, Z_i) + u(q_i + w_{Ai} - b'_i, e_i)]$$

s.t. (34)–(41).

Our main interest in this model is the critical role of ex ante promise utility in the second period, $U(b'_i; \theta_i, Z_i)$. This variable determines the incentives in the first period, and this fact has consequences for the interpretation of Z_i as a valid instrument. Note from our earlier discussion that expected utility in the static problem depends on Z_i. Now the static problem is the second period problem, and so one sees the intuition that varying levels of utility depend on Z through these second period promised utilities. These promises thus impact current period incentives and so vary with Z. In order to see this, notice that the higher Z_i is the less surplus there will be during the second period to maintain a given level of promise utility $U(b'_i; \theta_i, Z_i)$. As a result of this, we might expect $U(\cdot)$ to be a monotone decreasing function of Z_i. On the other hand, assignment to intermediation in the second period, when Z_i is high (low), allows the assignment of a low (high) promise as a threat for bad outcomes in the first period. Therefore, either way the level promise $U(\cdot)$ depends on the instrument Z_i. Thus, we lose the desirable properties of Z_i as a potential instrument. Promised utility in the second period depends on Z_i and promised utility determines current period incentives.

Example.   As in the static case, we study the differences between the model-generated treatment parameters and the estimates obtained by a researcher using observational data on consumption, wealth, and intermediation. We use the same parameterization as in the previous case [see Table 10, and equations (32) and

TABLE 15. IV and OLS estimates computed by wealth level (dynamic model)

Wealth level	$\Delta^{\mathrm{IV}}(b_k)$ (1)	$\Delta^{\mathrm{OLS}}(b_k)$ (2)	% difference ($\frac{(1)-(2)}{(2)}$)
b_1	-1.85	-1.48	-25.0%
b_2	-2.59	-2.16	-19.7%
b_3	-3.08	-2.86	-7.8%
b_4	-3.77	-3.55	-6.1%
b_5	-4.46	-4.24	-5.2%
b_6	-5.14	-4.92	-4.4%
b_7	-5.70	-5.61	-1.6%
b_8	-6.42	-6.30	-1.9%
b_9	-7.16	-6.97	-2.8%
Overall effect—full interactions (equation (33))	-4.74	-4.33	-9.5%
Overall effect—restricted model (equation (32))	-4.70	-4.50	-4.5%

Note: All estimates are statistically significant at the 5% level.

(33)]. In this case, we observe 33.10% of the individuals (endogenously) reporting financial intermediation.[29]

Table 15 presents the IV and OLS estimates (overall and by level of wealth). In general, the differences between IV and OLS are larger than what we computed in Table 11. Table 16 on the other hand, presents the local IV estimates $\Delta^{\mathrm{IV}}(Z_{l+1}, Z_l; b_k)$ for $k = (1, \ldots, 9)$, which—jointly with the weights presented in Table 17—produce the IV estimates reported in Table 15. We again observe a larger variability in IVs (across wealth levels) than for the results in Table 12. This fact reflects the strong selection process driving the decision into intermediation.

Tables 18 and 19 present the model-generated treatment parameters. The numbers in these tables are obtained using the counterfactual consumption levels delivered by the model, which would not be available in observational data (such as those used to generate the numbers in Tables 15, 16, and 17).

The results in Table 18 show important differences between the average treatment effect (ATE), the treatment effect on the treated (TT), and the treatment effect on the untreated (TUT). These differences illustrate how the presence of unobserved talent and the sorting mechanism into financial intermediation generate heterogenous treatment parameters. In this context, the analyst must first state the question she wants to answer, and then use the appropriate empirical approach to identify the treatment parameters of interest.

Table 19, on the other hand, presents the model-generated local average treatment effects $\Delta^{\mathrm{LATE}}(Z_{l+1}, Z_l; b^k) = E(c_i^I - c_i^A | D_i(Z_{l+1}) - D_i(Z_l) = 1, b_i = b_k)$. Given the problematic definition of Z_i as a proper instrument, the model-generated LATE (Table 19) and the estimated local IVs (Table 16) are now different.

TABLE 16. Estimated local IV by wealth level ($\Delta^{IV}(Z_{l+1}, Z_l; b_k)$ dynamic model)

Wealth level	(Z_2, Z_1)	(Z_3, Z_2)	(Z_4, Z_3)	(Z_5, Z_4)	(Z_6, Z_5)	(Z_7, Z_6)	(Z_8, Z_7)	(Z_9, Z_8)	(Z_{10}, Z_9)	$\Delta^{IV}(b_k)$
b_1	3.56	3.93	−1.94	−8.45	1.40	−6.88	−5.11	−1.83	−0.80	−1.85
b_2	−3.80	−3.23	−2.08	−1.64	−1.70	−4.54	−3.19	−4.19	−1.72	−2.59
b_3	−5.14	−0.55	−4.26	−3.37	−2.13	−4.58	−3.60	−4.67	−2.25	−3.08
b_4	−7.10	0.74	−4.96	−3.69	−2.86	−6.26	−5.72	−3.45	−3.45	−3.77
b_5	−5.58	−2.92	−3.32	−6.42	−2.64	−7.46	−6.08	−4.38	−3.83	−4.46
b_6	−2.27	−4.45	−5.32	−4.00	−4.76	−7.03	−5.48	−6.76	−5.06	−5.14
b_7	−9.50	−2.87	−5.88	−6.01	−5.30	−7.78	−5.62	−5.60	−5.81	−5.70
b_8	−7.44	−6.39	−4.18	−7.43	−5.67	−6.46	−7.87	−6.27	−6.54	−6.42
b_9	−8.66	−0.31	−8.88	−8.79	−6.32	−8.90	−6.74	−7.94	−5.80	−7.16

TABLE 17. IV weights by wealth level ($\lambda_l(b_k)$) (dynamic model)

Wealth level	(Z_2, Z_1) (1)	(Z_3, Z_2) (2)	(Z_4, Z_3) (3)	(Z_5, Z_4) (4)	(Z_6, Z_5) (5)	(Z_7, Z_6) (6)	(Z_8, Z_7) (7)	(Z_9, Z_8) (8)	(Z_{10}, Z_9) (9)	$\sum_l \lambda_l(b_k)$ (1) + (2) + $\cdots$ + (9)
b_1	0.03	0.09	0.09	0.12	0.25	0.08	0.13	0.11	0.10	1.00
b_2	0.03	0.09	0.09	0.12	0.26	0.07	0.14	0.11	0.10	1.00
b_3	0.03	0.08	0.09	0.13	0.26	0.06	0.14	0.11	0.10	1.00
b_4	0.02	0.07	0.09	0.13	0.25	0.06	0.15	0.11	0.10	1.00
b_5	0.02	0.07	0.10	0.14	0.25	0.06	0.15	0.11	0.09	1.00
b_6	0.02	0.06	0.10	0.14	0.25	0.06	0.16	0.12	0.09	1.00
b_7	0.02	0.06	0.10	0.15	0.24	0.06	0.16	0.12	0.09	1.00
b_8	0.02	0.05	0.10	0.15	0.23	0.06	0.16	0.12	0.09	1.00
b_9	0.02	0.05	0.10	0.16	0.23	0.07	0.16	0.13	0.09	1.00

TABLE 18. Model generated treatment parameters: ATE, TT, and TUT, by wealth level (dynamic model)

Wealth level	$\Delta^{\mathrm{ATE}}(b_k)$	$\Delta^{\mathrm{TT}}(b_k)$	$\Delta^{\mathrm{TUT}}(b_k)$	Consumption under financial autarky
b_1	−1.49	−1.47	−1.51	22.00
b_2	−2.18	−2.17	−2.19	22.69
b_3	−2.87	−2.84	−2.89	23.38
b_4	−3.56	−3.52	−3.58	24.07
b_5	−4.25	−4.22	−4.26	24.76
b_6	−4.94	−4.92	−4.95	25.46
b_7	−5.63	−5.59	−5.64	26.17
b_8	−6.32	−6.30	−6.32	26.87
b_9	−7.01	−7.05	−7.00	27.57
Overall	−4.51	−4.17	−4.69	25.04

Table 20 summarizes these large differences. In our dynamic model, with dynamic incentives, local IVs would not identify the well-defined causal parameter LATE.

6. CONCLUSIONS

This paper links contract theory models of financial intermediation to econometric policy evaluation. We have discussed a variety of economic models with unobserved heterogeneity and endogenous decisions involving financial intermediation. We have also analyzed econometric techniques and policy evaluations that are appropriate or inappropriate, depending on the vision of the underlying model, the assumptions one is willing to make, and the data at hand.

Even though, under certain assumptions, an IV strategy can accurately recover a true model-generated causal effect (LATE), these are quantitatively different, in order of magnitude and even sign, from other policy impact parameters (e.g., treatment on the treated and the average treatment effect). We also show that laying out clearly alternative models can guide the search for instruments. Mechanism design can deliver natural lotteries of randomization that can be used as sources of identification in empirical analyses. On the other hand, adding more margins of decision, that is, occupation choice and intermediation jointly, or adding more periods with promised utilities as key state variables, as in optimal multiperiod contracts, can cause the misinterpretation of the IV estimates as the causal parameter of interest (e.g., uniformity), so that IV and LATE might no longer coincide.

Our objective is to help researchers and policy makers assess accurately the impact of financial intermediation. In order to identify the impact of financial intermediation, researchers and policy makers need a clear understanding of the role of unobserved heterogeneity (coming from preferences, costs, or talents) and the economic mechanisms driving individuals' endogenous decisions. A limited understanding of the economic fundamentals could result in a misinterpretation of

TABLE 19. Model-generated local average treatment effects by wealth level ($\Delta^{\text{LATE}}(Z_{l+1}, Z_l; b_k)$, dynamic model)

Wealth level	(Z_2, Z_1)	(Z_3, Z_2)	(Z_4, Z_3)	(Z_5, Z_4)	(Z_6, Z_5)	(Z_7, Z_6)	(Z_8, Z_7)	(Z_9, Z_8)	(Z_{10}, Z_9)	$\Delta^{\text{LATE}}(b_k)$
b_1	−1.37	−1.51	−1.77	−1.55	−1.28	−1.76	−1.64	−1.60	−1.45	−1.55
b_2	−2.25	−2.22	−2.01	−2.12	−2.10	−1.92	−2.21	−2.49	−2.26	−2.17
b_3	−2.79	−2.78	−2.78	−2.80	−2.80	−2.67	−2.89	−3.03	−3.06	−2.85
b_4	−3.48	−3.41	−3.45	−3.40	−3.51	−3.49	−3.58	−3.62	−3.75	−3.53
b_5	−4.15	−4.25	−4.20	−4.23	−4.06	−4.26	−4.27	−4.45	−4.45	−4.27
b_6	−4.92	−4.88	−4.88	−4.75	−4.83	−4.87	−4.88	−5.15	−5.04	−4.92
b_7	−5.81	−5.51	−5.58	−5.69	−5.43	−5.39	−5.61	−5.67	−5.89	−5.62
b_8	−6.47	−6.37	−6.16	−6.26	−6.04	−6.08	−6.32	−6.29	−6.47	−6.27
b_9	−6.51	−6.81	−6.96	−6.73	−7.22	−7.00	−6.66	−7.22	−7.15	−7.00

TABLE 20. Model-generated local average treatment effect versus estimated local IVs, by wealth level (dynamic model)

Wealth level	$\Delta^{\text{LATE}}(b_k)$ (1)	$\Delta^{\text{IV}}(b_k)$ (2)	% difference ($\frac{(1)-(2)}{(2)}$)
b_1	−1.55	−1.85	16.2%
b_2	−2.17	−2.59	16.0%
b_3	−2.85	−3.08	7.6%
b_4	−3.53	−3.77	6.4%
b_5	−4.27	−4.46	4.2%
b_6	−4.92	−5.14	4.2%
b_7	−5.62	−5.70	1.4%
b_8	−6.27	−6.42	2.4%
b_9	−7.00	−7.16	2.3%
Overall	−4.50	−4.70	4.3%

policy parameters estimated from observational data. The good news is that there are a wide array of options, so it is a matter of choosing carefully.

NOTES

1. This is easily modified to allow a choice between savings s and consumption c, where $c + s \leq W$ and preferences are determined by a Cobb–Douglas utility function, giving a (myopic) savings rate.

2. Although this wage w is taken as given for each individual choice problem, it is consistent with a market-clearing equilibrium wage.

3. See Rubin (1974) and Heckman and Honoré (1990) for a formal exposition of the Roy model.

4. This is because we assume that the subsidy ψ is not correlated with unobserved talents θ^W and θ^E.

5. If the subsidy takes on a finite number of discrete values, and we order them according to their magnitudes ($\psi_0 < \psi_1 < \ldots < \psi_K$), then Δ^{IV} can be written as a weighted average of $\Delta^{\text{LATE}}(\psi_k, \psi_{k+1})$ with $k = 1, \ldots, K - 1$, where the weights are related to the probability of going into business at the various values of the subsidy [see Yitzhaki (1989); Imbens and Angrist (1994)]. Additionally, if we take the limit as subsidy ψ_k approaches ψ_{k+1}, this delivers the marginal treatment effect (MTE) for those households just indifferent to becoming businesses [see Heckman and Vytlacil (2001)].

6. Formally,

$$\lambda\left(\frac{(\phi_w - 1)w + \phi_b b_i + \psi_i}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}}\right) = E\left(\frac{\theta_i^W - \phi_\theta \theta_i^E}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}} \middle| \frac{\theta_i^W - \phi_\theta \theta_i^E}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}} > \frac{(\phi_w - 1)w + \phi_b b_i + \psi_i}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}}\right)$$

$$= \frac{\phi\left(\frac{(\phi_w - 1)w + \phi_b b_i + \psi_i}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}}\right)}{\Phi\left(\frac{(\phi_w - 1)w + \phi_b b_i + \psi_i}{\sqrt{\sigma_W^2 + \phi_\theta^2 \sigma_E^2}}\right)},$$

where ϕ and Φ represent the probability density and cumulative distribution functions associated with a standard normal distribution, respectively.

7. Notice that this parameter is the limit version of the average local treatment effect. More specifically, $\Delta^{\text{MTE}}(p, b) = \lim_{p' \to p} \Delta^{\text{LATE}}(p, p', b)$.

8. This follows from the fact that

$$\theta_i^W - \phi_\theta \theta_i^E < (\phi_w - 1)w + \phi_b b_i + \psi_i \Leftrightarrow U < p(w, b_i, \psi_i),$$

where $U = F_{\theta_i^W - \phi_\theta \theta_i^E}(\cdot)$ represents the cumulative distribution function associated with the random variable $\theta_i^W - \phi_\theta \theta_i^E$.

9. The empirical implementation of the local instrumental variable estimator involves the nonparametric estimation of the derivative of $E(Y_i | p_i, b_i)$ with respect to p_i. Although the implementation of nonparametric techniques can be considered standard, in small samples they can be infeasible. See Heckman et al. (2006, 2008) for different empirical approaches to implementing the local instrumental variable estimator.

10. Using data from Thailand, Gine and Townsend (2004) estimate a model with characteristics similar to those of the one studied here.

11. Notice that this expression follows directly from the theoretical model [see equation (11)].

12. The average effects defined in Table 3 come from the expression

$$\left. \frac{\Delta E(Y \mid D, b)}{\Delta D} \right|_{b=\bar{b}} = \kappa_2 \bar{b} + \kappa_3,$$

where $\bar{b}$ represents the average wealth in the population. We use $\frac{\Delta Y}{\Delta D}$ to denote a change in Y due to a change in the discrete variable D.

13. Formally, from the model we can generate

$$\Delta^{\text{LATE}}(1, 0; b) = \frac{E(Y_i \mid \psi_i = 1, b_i = b) - E(Y_i \mid \psi = 0, b_i = b)}{E(D_i \mid \psi_i = 1, b_i = b) - E(D_i \mid \psi_i = 0, b_i = b)}$$

and then we compute

$$\Delta^{\text{LATE}}(1, 0) = \int \Delta^{\text{LATE}}(1, 0; t) \, dF_b(t), \tag{42}$$

where $F_b(t)$ represents the cumulative distribution of wealth for those individuals switching occupations as a result of the change in the value of the instrument.

14. Our linear regression model implies the following approximation for LATE (as a function of wealth):

$$\Delta^{\text{IV}}(1, 0; b) = \kappa_2 b + \kappa_3,$$

and consequently, $\Delta^{\text{IV}}(1, 0) = \kappa_2 \bar{b} + \kappa_3$, where $\bar{b}$ denotes the average wealth level. The comparison of $\Delta^{\text{IV}}(1, 0)$ and $\Delta^{\text{LATE}}(1, 0)$ in expression (42) illustrates the source of discrepancies between our estimates.

15. For particular populations in which the occupational decision becomes irrelevant, we can use this method to determine the gains in profits for entrepreneurs induced to use the financial system. Suppose that the random assignment of ψ is such that there exists a population for which the subsidy is so high, ψ^*, that there are only firms, regardless of the assigned values of Q. In this case, we can estimate the local average treatment effect as

$$\frac{E(\xi_i | Q_i = \bar{\bar{Q}}, b_i = b, \psi_i = \psi^*) - E(\xi_i | Q_i = \bar{Q}, b_i = b, \psi_i = \psi^*)}{E(\Upsilon_i | Q_i = \bar{\bar{Q}}, b_i = b, \psi_i = \psi^*) - E(\Upsilon_i | Q_i = \bar{Q}, b_i = b, \psi_i = \psi^*)}$$

which (under uniformity) identifies the income gains associated with intermediation for those who are isolated from the wage sector, or $E(\pi(\theta_i^E, w, r) - \pi(\theta_i^E, b_i, w)|b_i = b, \psi_i = \psi^*, \Upsilon_i(\bar{\bar{Q}}) = 1, \Upsilon_i(\bar{Q}) = 0)$.

16. However, as in the case of financial intermediation, for particular populations we can use the local treatment effect to identify the effect of entrepreneurship for individuals under financial autarky. Specifically, suppose that the random assignment of Q is such that there exists a population for which the costs of using the financial intermediary are too high, Q^*, so that regardless of the assigned values of the subsidy they choose to be in financial autarky. In this case, we can use the instrument ψ to compute

$$\frac{E(\xi_i|\psi_i = \overline{\overline{\psi}}, b_i = b, Q_i = Q^*) - E(\xi_i|\psi_i = \overline{\psi}, b_i = b, Q_i = Q^*)}{E(D_i|\psi_i = \overline{\overline{\psi}}, b_i = b, Q_i = Q^*) - E(D_i|\psi_i = \overline{\psi}, b_i = b, Q_i = Q^*)},$$

which (under uniformity) identifies $E(\pi(\theta_i^E, b_i, w) - (w + \theta_i^W)|b_i = b, Q_i = Q^*, D_i(\overline{\overline{\psi}}) = 1, D_i(\overline{\psi}) = 0)$, which is the income gains associated with entrepreneurship for those individuals isolated from financial intermediation.

17. Formally, suppose individuals decide among J different options. Each option has associated a utility level V_j for $j = 1, \ldots, J$. Let $D_j = 1$ if the individual selects the jth alternative, and 0 otherwise. Furthermore, as in the model of this section, assume $D_j = 1$ if $V_j = \max\{V_1, \ldots, V_J\}$ for $j = 1, \ldots, J$. Let Y_j denote the potential outcome associated with option j. Valid instruments affect choices but are independent from potential outcomes. Let Z_j denote the instrument associated with option j. We present the relationship between instrument and options as $V_j(Z_j)$; that is, instrument Z_j determines the utility level V_j. V_j also depends on unobserved components that can be correlated with potential outcome Y_j. For notational simplicity, we leave this dependence implicit. The observed outcome Y can be written as $Y = \sum_{j=1}^{J} D_j Y_j$. Heckman et al. (2006) show that

$$\Delta^{\text{IV}(Z_j)} = \frac{E(Y|Z = z) - E(Y|Z = z')}{\Pr(D_j = 1|Z = z) - \Pr(D_j = 1|Z = z')},$$

where $z = (z_1, \ldots, z_j, \ldots, z_J)$ and $z = (z_1, \ldots, z_j', \ldots, z_J)$, so that only the variation from Z_j is utilized to compute $\Delta^{\text{IV}(Z_j)}$, identifies the effect on outcome of option j versus the next best option.

18. Notice that when our model is phrased as a model of multiple treatments, intermediation costs Q and subsidy ψ are not valid instruments, in the sense of Z_j entering only V_j (see previous note), for any of the four alternatives.

19. One could present the following regression model for the simultaneous analysis of the effects of occupation and financial intermediation:

$$\Psi_i = \kappa_0 + \kappa_1 b_i + \kappa_2 \Upsilon_i b_i + \kappa_3 D_i (1 - \Upsilon_i) + \kappa_4 b_i D_i (1 - \Upsilon_i) + \kappa_5 \Upsilon_i D_i(r) + \kappa_6 b_i \Upsilon_i D_i(r) + \varepsilon_i.$$

In this case, the information from both instruments (Ψ_i and Q_i) should be used to control for the endogeneity provoked by the selection processes. As previously explained, this model, in which the two margins are simultaneously modeled, has additional complications that go beyond the scope of our analysis in this paper. See Heckman et al. (2006) for an analysis of this case.

20. The risk-sharing role of formal financial institutions is tested in Alem and Townsend (2008).

21. Notice that if the individual does not know her/his unobserved preference parameter θ_i or, alternatively, if s/he knows θ_i but for some reason does not act on it, then the selection process will not be based on unobserved gains. Formally, in this case $E(\varepsilon_{it}^I - \varepsilon_{it}^A|I_{i0}) = 0$, and the model will produce homogeneous treatment effects.

22. A literature on sudden devaluations causing wealth losses from dollar denominated loans is not unrelated.

23. See Karaivanov and Townsend (2009) for further work with the Thai data and the estimation of financial regimes in a dynamic context.

24. It is worth mentioning that experimenting with different sample sizes suggests that reducing the number of observations produces significant losses in the accuracy of the local IV estimates.

25. More precisely, we work with $(b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10}) = (10.5, 10.6, 10.7, 10.88, 10.9, 11.1, 11.2, 11.3, 11.4, 11.5)$, and $(Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8, Z_9, Z_{10}) = (0, 0.03, 0.06, 0.1, 0.13, 0.16, 0.2, 0.23, 0.26, 0.3)$. Given the structure of the model and our ordering, the resulting probabilities associated with the lottery are increasing in Z_j and decreasing in b_k. We also consider a discrete grid for talent θ. Specifically, we solve the model for $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9) = (0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4)$. The distribution of talent and wealth generated using the discrete grids respects the joint distribution associated with these random variables presented in Table 10.

26. A more complicated version would include risk sharing. See Alem and Townsend (2008).

27. The numbers presented in Table 11 are obtained using $\Delta^{\mathrm{IV}}(Z_{l+1}, Z_l; b_k)$ and the IV weights presented in Table 12 and 13, respectively.

28. We do not present the model-generated LATE in this case. This is because, as in the previous examples of a valid instrument satisfying uniformity, they will be close to the estimated local IV estimates, and so we prefer not to repeat the argument.

29. Here we work with the following values for wealth and the instrument: $(b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9) = (18.25, 18.33, 18.41, 18.5, 18.58, 18.66, 18.75, 18.83, 18.91)$ and $(Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8, Z_9, Z_{10}) = (0, 0.03, 0.07, 0.1, 0.14, 0.18, 0.21, 0.25, 0.29, 0.32)$. Given the structure of the model and our ordering, the resulting probabilities associated with the lottery tend to be increasing in Z_j and decreasing in b_j. For talent θ, we solve the model using $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9) = (0.86, 0.88, 0.9, 0.92, 0.93, 0.95, 0.97, 0.98, 1)$.

REFERENCES

Alem, Mauro and Robert Townsend (2008) An Evaluation of Safety Nets and Financial Institutions in Crisis and Growth. Unpublished manuscript, Department of Economics, University of Chicago.

Felkner, John and Robert M. Townsend (2007) Enterprise and the Wealth of Villages. Unpublished manuscript, University of Chicago.

Gine, Xavier and Robert Townsend (2004) Evaluation of financial liberalization: A general equilibrium model with constrained occupation choice. *Journal of Development Economics* 74, 269–307.

Greenwood, Jeremy and Boyan Jovanovic (1999) Financial development, growth, and the distri- bution of income. *Journal of Political Economy* 98, 1076–1107.

Heckman, James J. and Bo E. Honoré (1990) The empirical content of the Roy model. *Econometrica* 58, 1121–1149.

Heckman, James J., Daniel Schmierer and Sergio Urzua (2008) Testing the Correlated Random Coefficient Model. Unpublished manuscript, Department of Economics, University of Chicago; *Journal of Econometrics*, in press.

Heckman, James J., Sergio Urzua and Edward J. Vytlacil (2006) Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics* 88, 389–432.

Heckman, James J. and Edward J. Vytlacil (2001) Local instrumental variables. In Cheng Hsiao, Kimio Morimune, and James L. Powell (eds.), *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, pp. 1–46. New York: Cambridge University Press.

Imbens, Guido W. and Joshua D. Angrist (1994) Identification and estimation of local average treatment effects. *Econometrica* 62, 467–475.

Jeong, Hyeok and Robert Townsend (2008) Growth and inequality: Model evaluation based on an estimation–calibration strategy. *Macroeconomic Dynamics* 12, 231–284.

Kaboski, Joseph P. and Robert M. Townsend (2005) Policies and impact: An analysis of village-level microfinance institutions. *Journal of the European Economic Association* 3, 1–50.

Kaboski, Joseph P. and Robert M. Townsend (2009) The Impact of Credit on Village Economies. Unpublished manuscript, Ohio State University.

Karaivanov, Alexander K. and Robert M. Townsend (2009) Enterprise Dynamics and Finance: Distinguishing Mechanism Design from Exogenously Incomplete Markets Models. Unpublished manuscript, Department of Economics, Simon Fraser University.

Lloyd-Ellis, Huw and Dan Bernhardt (2000) Enterprise, inequality and economic development. *Review of Economic Studies* 67, 147–168.

Matzkin, Rosa L. (1992) Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica* 60, 239–270.

Olken, Benjamin (2007) Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy* 115, 200–249.

Paulson, Anna L., Robert M. Townsend and Alex Karaivanov (2006) Distinguishing limited liability from moral hazard in a model of entrepreneurship. *Journal of Political Economy* 114, 100–144.

Roy, A.D. (1951) Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3, 135–146.

Rubin, Donald B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66, 688–701.

Salop, Steven C. (1979) Strategic entry deterrence. *American Economic Review* 69, 335–338.

Townsend, Robert M. and Kenichi Ueda (2006) Financial deepening, inequality, and growth: A model-based quantitative evaluation. *Review of Economic Studies* 73, 251–293.

Townsend, Robert M. and Kenichi Ueda (in press) Welfare gains from Financial liberalization. *International Economic Review*.

Vissing-Jorgensen, Annette (2002) Towards an Explanation of Household Portfolio Choice Hetero-Geneity: Nonfinancial Income and Participation Cost Structures. Working Paper 8884, National Bureau of Economic Research.

Yitzhaki, Shlomo (1989) On Using Linear Regression in Welfare Economics. Working Paper 217, Department of Economics, Hebrew University.