Supplemental Material for Optimal Contracting, Spatial Competition among Financial Service Providers, and the Impact of Digital Lending

Gustavo Joaquim Robert M. Townsend Victor Zhorin

F Algebra for Section 2

Complete Information. In the complete information case, the problem of the FSP is:

$$\max_{\lambda,t} D(\theta_i,\lambda,t)\left[t-.5\lambda^2\right] \tag{89}$$

taking the FOC:

$$-D^{'}\left[t-\lambda^2\right]+D=0 \tag{90}$$

$$\theta d^{'}\left[t-\lambda^2\right]-D\lambda=0 \tag{91}$$

Dividing Eq. (90) by (91

$$\lambda(\theta_i)=\theta_i \tag{92}$$

To compute the transfers, we replace $\lambda(\theta_i)=\theta_i$ in Eq. (90):

$$t-.5\theta_i^2=D/D^{'}=(\theta_i^2-t-u_{0,\theta_i})\varepsilon^{-1}\Rightarrow t(\theta_i)=\frac{1+.5\varepsilon}{1+\varepsilon}\theta_i^2-\frac{1}{1+\varepsilon}u_0\left(\theta_i\right) \tag{93}$$

Adverse Selection. The problem of the FSP becomes:

$$\max_{\lambda_L,t_L,\lambda_H,t_H} f_L D(\theta_L,\lambda_L,t_L)\left[t_L-.5\lambda_L^2\right]+f_H D(\theta_H,\lambda_H,t_H)\left[t_H-.5_H^2\right] \tag{94}$$

s.t. to the *Truth Telling* constraints:

$$\theta_L\lambda_L-t_L\geq\theta_L\lambda_H-t_H \tag{95}$$

$$\theta_H\lambda_H-t_H\geq\theta_H\lambda_L-t_L \tag{96}$$

We can transform the problem to become:

$$\max_{\lambda_L,u_L,\lambda_H,u_H} D(u_L)\left[\theta_L\lambda_L-.5\lambda_L^2-u_L\right]+D(u_H)\left[\theta H\lambda_H-.5_H^2-u_H\right] \tag{97}$$

s.t.

$$u_L \geq u_H - (\theta_H - \theta_L)\lambda_H \tag{98}$$

$$u_H \geq u_L + (\theta_H - \theta_L)\lambda_L \tag{99}$$

Note that a monotonicity condition ($\lambda_H > \lambda_L$) joint with one truth-telling constraint implies the other, that is

$$u_H = u_L + (\theta_H - \theta_L)\lambda_L \Rightarrow u_L = u_H - (\theta_H - \theta_L)\lambda_L > u_H - (\theta_H - \theta_L)\lambda_H \tag{100}$$

and that a binding TT of the low type is not consistent with the TT of the high type:

$$u_L = u_H - (\theta_H - \theta_L)\lambda_H \Rightarrow u_H = u_L + (\theta_H - \theta_L)\lambda_H > u_H + (\theta_H - \theta_L)\lambda_L \tag{101}$$

Therefore, there is only one potentially binding truth-telling constraint. Contrary to the textbook case, however, we don't know which constraint is binding. As the FSP does not have full market power, it is not trivial that the high type extracts information rents. This is a function of every kind of outside option and shares that each type appears in the population. To see this, consider a case where $f_H \to 0$. In this case, it is better to keep the allocation to the low type undistorted while distorting the allocation of the high type to satisfy the truth-telling constraints.

Assume without loss of generality that it is the one of the high type. The Lagrangian in the problem in Eq.(97) becomes:

$$\mathcal{L} \equiv f_L D(u_L)\left[\theta_L\lambda_L - .5\lambda_L^2 - u_L\right] + f_H D(u_H)\left[\theta_H\lambda_H - .5\theta_H^2 - u_H\right] + \psi(0 - u_L - (\theta_H - \theta_L)\lambda_L + u_H) \tag{102}$$

The FOC system is:

$$\begin{aligned}
f_L D^{'}(u_L)\left[\theta_L\lambda_L - .5\lambda_L^2 - u_L\right] - f_L D(u_L) - \psi = 0 \quad & (u_L) & (103)\\
f_H D^{'}(u_H)\left[\theta H\lambda_H - .5\lambda_H^2 - u_H\right] - f_H D(u_H) + \psi = 0 \quad & (u_H) & (104)\\
f_L D(u_L)\left[\theta_L - \lambda_L\right] + (\theta_H - \theta_L)\psi = 0 \quad & (\lambda_L) & (105)\\
\lambda_H = \theta_H \quad & (\lambda_H) & (106)\\
u_H = u_L + (\theta_H - \theta_L)\lambda_L \quad & (\psi) & (107)
\end{aligned}$$

which shows the no distortion result of Figure 12. In the numerical simulation, we use the following scaling for the demand function:

$$D(u) = \left(\frac{u - u_0}{.5\theta^2 - u0}\right)^{\varepsilon} \tag{108}$$

that is, there is a share between zero and one ($.5\theta^2$ is the perfect competition outcome) that uses intermediation, and the curvature is given by the market power.

F.1 Deriving Eq. (65)

Under the assumption that $\sigma^2 A = 1$, we have that autarky utilities are zero for both types, that is, $U_{A,H} = u_{A,L} = 0$

$$t_H = \frac{1+.5\varepsilon}{1+\varepsilon}\theta_H^2 \text{ and } t_L = \frac{1+.5\varepsilon}{1+\varepsilon}\theta_L^2$$

Therefore, using the full information solution of t_L, t_H as above and $\lambda_H = \theta_H$ and $\lambda_L = \theta_L$, we can re-write the truth-telling constraint of the low type, which is generically given by

$$\theta_L\lambda_L - t_L \geq \theta_L\lambda_H - t_H$$

as

$$\theta_L^2 - \frac{1+.5\varepsilon}{1+\varepsilon}\theta_L^2 \geq \theta_L\theta_H - \frac{1+.5\varepsilon}{1+\varepsilon}\theta_H^2 \Leftrightarrow \theta_L(\theta_L - \theta_H) \geq \frac{1+.5\varepsilon}{1+\varepsilon}(\theta_L - \theta_H)(\theta_L + \theta_H) \tag{109}$$

$$\Leftrightarrow \theta_L(.5\varepsilon) \geq (1+.5\varepsilon)\theta_H \Leftrightarrow \theta_L - \theta_H \geq \frac{2}{\varepsilon}\theta_H \tag{110}$$

Moreover,

$$\omega \equiv \frac{1}{1+\varepsilon} \Rightarrow \varepsilon = \omega^{-1} - 1$$

Substituting and manipulating we arrive at

$$\omega \geq \frac{\theta_L - \theta_H}{\theta_H + \theta_L} \tag{111}$$

which is always satisfied in the first best, so the TT constraint is not binding in the full information contracts. Note, however, that we can simply re-do the analysis for the truth-telling constraint on the high type, and in this case, get exactly Eq.(65). ■

G Proof of Lemma 3.1

Proof. The strategy to show that the equilibrium exists and is unique is to show that the vector of best response functions is a contraction. The Nash Equilibrium is then the unique fixed point of the vector of best response functions. This is useful not only theoretically but also numerically: computing the fixed point of a contraction can be done by an iterative algorithm. The first step of the proof is Lemma G.1, which is a version of Blackwell's sufficient conditions for operators between compact subspaces of $\mathbb{R}^n$, which is our case here.

Lemma G.1. *Let $T: C \to C$, $C \subset \mathbb{R}^n$, C compact. Define $\|x-y\| \equiv \max_i |x_i - y_i|$ and $x \leq y$ if $x_i \leq y_i, i = 1, \ldots, n$. Then, if:*

1. *(Monotonicity) $x \leq y \Rightarrow Tx \leq Ty$, $\forall x, y \in C$.*
2. *(Discount) $T(x + \boldsymbol{e}a) \leq T(x) + \beta \boldsymbol{e}a$, $\forall x \in W, a \in \mathbb{R}_+$, $\boldsymbol{e} = (1, \ldots, 1) \in \mathbb{R}^n_+$ and $x + \boldsymbol{e}a \in C$.*

T is a contraction with modulus β.

Proof. $\forall x, y$: $x - y \leq \boldsymbol{e}\|x-y\|$. This imples that $x \leq y + \boldsymbol{e}\|x-y\|$ By properties 1 and 2: $Tx \leq Ty + \beta \boldsymbol{e}\|x-y\|$. Also, the same is true for x in place of y: $Ty \leq Tx + \beta \boldsymbol{e}\|x-y\|$. Therefore: $Tx - Ty \leq \beta \boldsymbol{e}\|x-y\|$ and $Ty - Tx \leq \beta \boldsymbol{e}\|x-y\|$ which implies $\|Tx - Ty\| \leq \beta \|x-y\|$. ∎

Moreover, we present an auxiliary Lemma G.2 on the argmax of problems of a particular condition - which we then show to hold in our case. This is simply a way to simplify the exposition.

Lemma G.2. *Let $f, g: \mathbb{R} \to \mathbb{R}$ be strictly concave functions, f strictly decreasing in x and g strictly increasing in x. Moreover, Let g be continuously differentiable in x. Let I be a compact interval. Let $\lim_{x \to \max_I} f(x) = -\infty$, and g bounded above. Finally, let δf be the correspondence function of subgradients of f. Define: $x^* \equiv arg\max_{x \in I} f(x) + g(x)$. Then x^* exists, is unique, and is s.t.:*

$$(x^* - \min I) \text{ and } g^{'}(x) < -[\max\{\delta f(x)\}], \forall\, x \in I \tag{112}$$

or

$$g^{'}(x^*) \in -\delta f(x^*) \tag{113}$$

Proof. Existence and uniqueness comes from strict concavity and continuity. The optimum is not at $\max I$ since $\lim_{x \to \max_I} f(x) = -\infty$. Therefore, the optimum must either be at $\min I$ or satisfy the FOC. To derive the FOC, take x_0 s.t. $g^{'}(x_0) \in -\delta f(x_0)$. Due to the strict concavity of f, g, $f(x) - f(x_0) < -g^{'}(x)(x - x_0)$ and $g(x) - g(x_0) < g^{'}(x)(x - x_0), \forall x \neq x_0$, Therefore: $f(x) + g(x) < f(x_0) + g(x_0)$. If the optimum is at $\min I$, however, it must be the case that $g^{'}(x) < -[\max\{\delta f(x_0)\}]$, $\forall\, x \in I$ - otherwise we could find an interior maximum. ∎

Back to our original problem, we do the following steps. First, we do a transformation where we include the outside option as a phantom player. Second, we show that the BR function satisfies the condition of the Lemma G.2 and the implications for the specific case of BR functions. Third, we discuss the boundaries of changes in the BR function. Fourth, we show that the conditions for Lemma 3.1 are satisfied and conclude the proof.

Step 1. The Phantom Player. We assume that there is a phantom player in the game, the *outside option* player. This player represents the outside option: we assume that it is as if it is another FSP, but it always plays the outside option. We denote it by $b = B_p + 1$, i.e., the extra bank in the province. We do this transformation to facilitate the proof of uniqueness and existence. Define the best response function vector (i.e., of all FSPs) by [37]

$$BR(u_1, u_2, \dots, u_n) \equiv \left(BR_1\left(u_{-1}\right), BR_2\left(u_{-2}\right), \dots, BR_{B_p+1}\left(u_{-B_p+1}\right)\right)$$

BR is a function that maps the cartesian products of the strategy spaces in itself: $BR : W^{B_p+1} \to W^{B_p+1}$ maps a set of strategies of all FSPs $\{u_b\}_b \in W_{B_p}$ in the best response of an FSP given the strategies of all other FSPs.

Step 2. Best Response and Auxiliary Lemma. Note that S is decreasing and concave in u. As S is concave, it is continuous everywhere and differentiable almost everywhere (a.e.) in $u \in W$. Also, note that W is a compact subset of $\mathbb{R}$. Moreover, note that:

$$BR_b\left(\{u_b\}_{b=1}^{B_p+1}\right) = \arg\max_{u \in W} S\left(u\right)\mu\left(u, u_{-b}\right) = \arg\max_{u \in \hat{W}} \ln\left[S\left(u\right)\right] + \ln\left[\mu\left(u, u_{-b}\right)\right] \tag{114}$$

where $\hat{W} \equiv W \cap \{u \mid S(u) \geq 0\}$, which is compact. The idea is that if $\exists u \in W s.t. S(u) > 0$, then no utility in equilibrium is played with $S(u) \leq 0$. This implies that the function:

$$f\left(u\right) \equiv \ln\left(S\left(u\right)\right)$$

is: (i) strictly concave, decreasing in u, (ii) $\lim_{u \to \max \hat{W}} \Sigma(u) = -\infty$. Therefore, $f(u)$ plays the role of f in the Lemma G.2. Moreover, let

$$g\left(u, u_{-b}\right) \equiv \ln\left(\mu\left(u, u_{-b}\right)\right)$$

which is: (i) strictly concave, increasing in u, (ii) continuously differentiable at $u \in interior(W)$, (iii) bounded above by zero. Therefore, $g(u, u_{-b})$ plays the role of g in Lemma G.2 for any given value of u_{-b}. Define: $\Sigma : \hat{W} \rightrightarrows \mathbb{R}$ as:

$$\Sigma(u) \equiv -\frac{\delta S(u)}{S(u)} \tag{115}$$

and:

$$\Upsilon(u, u_{-b}) \equiv \frac{\partial_{u_b}\mu\left(u, u_{-b}\right)}{\mu\left(u, u_{-b}\right)} \tag{116}$$

[37] The best responses are also a function of the locations of FSPs and villages, taken as given. We remove it from the notation at this point to facilitate the understanding.

to represent the equivalents of $\delta f, g^{'}$ in Lemma G.2, respectively. From Lemma G.2:

$$BR_b\left(\{u_b\}_{b=1}^{B_p+1}\right) = \begin{cases} \min \hat{W} \text{ and } \Sigma(u) > \Upsilon(u, u_{-b}), \ \forall u \in \hat{W} \\ \text{or } \Upsilon\left(BR_b\left(\{u_b\}_{b=1}^{B_p+1}\right), u_{-b}\right) \in \Sigma\left(BR_b\left(\{u_b\}_{b=1}^{B_p+1}\right)\right) \end{cases}, \forall \ b = 1, \ldots, B$$

For our phantom player - the outside option:

$$BR_b\left(\{u_b\}_{b=1}^{B_p+1}\right) = u_0$$

Step 3. The BR Boundaries. We start this step with two observations: $\Sigma(u)$ is strictly increasing in u and $\Gamma(u, u_{-b})$ is strictly decreasing in u, strictly increasing in u_{-b}. We know that $-\delta S$ is increasing in u. Moreover, S is strictly decreasing in u. The ratio $\Sigma(u)$ is thus increasing, meaning that $u > \hat{u} \Leftrightarrow x < y, \forall x \in \Sigma(u), y \in \Sigma(\hat{u})$. Second, as μ is log-concave in u, Υ must be decreasing in u and increasing in u_{-b} (since Υ is the first derivative of $\ln(\mu)$).

The fact that $\Sigma(u)$ is strictly increasing in u and $\Gamma(u, u_{-b})$ is strictly decreasing in u guarantees that:

$$\left[BR_b\left(\{u_b\}_{b=1}^{B_p+1}\right) - BR_b\left(\{\hat{u}_b\}_{b=1}^{B_p+1}\right)\right]^2 \leq \left[BR_b^{FOC}\left(\{u_b\}_{b=1}^{B_p+1}\right) - BR_b^{FOC}\left(\{\hat{u}_b\}_{b=1}^{B_p+1}\right)\right]^2 \tag{117}$$

where: $BR_b^{FOC}\left(\{u_b\}_{b=1}^{B_p+1}\right)$ is defined as the point that satisfies the equation:

$$\Upsilon\left(BR_b^{FOC}\left(\{u_b\}_{b=1}^{B_p+1}\right), u_{-b}\right) \in \Sigma\left(BR_b^{FOC}\left(\{u_b\}_{b=1}^{B_p+1}\right)\right)$$

even if $BR_b^{FOC} \notin W$. BR_b^{FOC} is picking the utility that solves the FOC if there is no lower bound to possible levels of utility that are offered. Given Eq. (117), it is sufficient to show that

$$\left[BR_b^{FOC}\left(\{u_b\}_{b=1}^{B_p+1}\right) - BR_b^{FOC}\left(\{\hat{u}_b\}_{b=1}^{B_p+1}\right)\right]^2 \lesssim a$$

to guarantee that the same is true for the BR functions. Therefore, we assume in the following step that the condition that the equilibrium utility is in the interior of W never binds. Moreover, note that Eq. (117) is always satisfied for the outside option since the LHS of Eq. (117) is always zero.

Step 4. The contraction.

Given the conditions on the BR function, we now proceed to show that the two conditions in *Lemma 1* hold for BR_b^{FOCs} and, thus, for BR_b. Focus on the FOC of our problem, that is:

$$\Upsilon\left(u_b^*, u_{-b}\right) \in \Sigma\left(u_b^*\right) \tag{118}$$

where $u_b^* \equiv BR_b\left(\{u_b\}_{b=1}^{B_p+1}\right)$. Then:

1. **Monotonicity.** If $\hat{u}_{-b} \geq u_{-b} \Rightarrow \hat{u}_b^* \geq u_b^*$. Assume by contradiction that $\hat{u}_{-b} \geq u_{-b}$ and $\hat{u}_b^* < u_b^*$. We know that:

$$\Upsilon\left(\hat{u}_b^*, \hat{u}_{-b}\right) < \Upsilon\left(\hat{u}_b^*, u_{-b}\right) < \Upsilon\left(u_b^*, u_{-b}\right) \in \Sigma\left(u_b^*\right) < \Sigma\left(\hat{u}_b^*\right) \Rightarrow \Upsilon\left(\hat{u}_b^*, \hat{u}_{-b}\right) < \Sigma\left(\hat{u}_b^*\right) \quad (119)$$

which cannot happen with an interior solution.

2. **Discounting.** If $\tilde{u}_{-b} = u_{-b} + ea$, $a > 0, e = (1, \ldots, 1), u_{-b} + ea \in \hat{W}^B \subset \mathbb{R}^B \Rightarrow \tilde{u}_b^* \in (u_b^*, u_b^* + a)$. We know from monotonicity $\tilde{u}_b^* > u_b^*$. Assume by contradiction that $\tilde{u}_b^* \geq u_b^* + a$.

$$\Upsilon\left(\tilde{u}_b^*, \tilde{u}_{-b}\right) > \Upsilon\left(u_b^* + a, \tilde{u}_{-b}\right) = \Upsilon\left(u_b^*, u_{-b}\right) \in \Sigma\left(u_b^*\right) > \Sigma\left(\tilde{u}_b^*\right) \Rightarrow \Upsilon\left(\tilde{u}_b^*, \tilde{u}_{-b}\right) > \Sigma\left(\tilde{u}_b^*\right) \quad (120)$$

which cannot happen with an interior solution. Note that this is where we use the condition of Eq. (118). The above reasoning guarantees that, $\forall a \in \mathbb{R}_+, \exists \beta_b(a)$ s.t.:

$$\tilde{u}_b^* \leq u_b^* + \beta_b(a) \quad (121)$$

As we know that: $\beta_b(a) < 1$, take the β of the contraction as: $\beta \equiv \max_b \max_a \beta_b(a)$. Note that, as $\hat{W}$ is compact, $\beta_b(a) < 1 \Rightarrow \beta < 1$.

Conclusion. Note that, as the Steps 1-4 above are true for all FSPs and the phantom bank, it must be true that the BR of Eq. (114) is a contraction. The Nash Equilibrium is the unique fixed point of the best responses and, thus, can be found through an iterative procedure (see). See Figure (14) in the main text for a more intuitive approach to the proof. ■

H Spatial Costs and Sufficient Conditions For Lemma 3.1

Lemma H.1 has the conditions for log-concavity when $\psi > 0$. If $\psi \leq \overline{\psi}$ as defined in Eq. (124), Lemma's H.1 condition is satisfied. Moreover, if market shares are always smaller than .5 (i.e., a very segmented market), Lemma's H.1 condition is satisfied.

Lemma H.1. *If for all banks $b, \hat{b}$ and for any two villages, $v, \hat{v}$, the spatial cost ψ and logit variance, σ_L imply that the market share at the village level, $\{\mu_v^b, \mu_{\hat{v}}^b, \mu_v^{\hat{b}}, \mu_{\hat{v}}^{\hat{b}}\}$ satisfies*

$$\sum_{i \in \{v, \hat{v}\}, j \in \{v, \hat{v}\}} \left[N_i N_j \mu_i^b (1 - \mu_i^b) \mu_j^b \left(2\mu_i^b - \mu_j^b\right)\right] > 0 \quad (122)$$

and

$$\sum_{i\in\{v,\hat{v}\},j\in\{v,\hat{v}\}}\left[N_iN_j\mu_i^b\mu_i^{\hat{b}}\mu_j^b\left(2\mu_i^b-\mu_j^b\right)\right]>0 \tag{123}$$

the market share defined by Eqs. (20)-(21) satisfies the conditions of Lemma 3.1. A sufficient condition for Eqs. (122)-(123) is that maximum ($\overline{\mu}$) and minimum ($\underline{\mu}$) market share between villages in the same market at any given level of utilities satisfies $\overline{\mu}\leq\underline{\mu}\left[4+\sqrt{11}\right]$. In terms of (ψ,σ_L), this means

$$\psi\leq\overline{\psi}\equiv\frac{\ln(4+\sqrt{11})\sigma_L}{[\max_{b,v}\|x_b-x_v\|-\min_{b,v}\|x_b-x_v\|]} \tag{124}$$

Proof. **Step 1. Bounded Away from Zero.** With the logit formulation, the minimum market share in a given village is s.t.

$$\mu_v\left(\varphi_b\right)>\frac{e^{\sigma_L^{-1}[\mathbb{V}(u_{min},x_b,x_v)-u_0]}}{1+\sum_{\hat{b}=1}^{B}e^{\sigma_L^{-1}\left[\mathbb{V}(u_{max},x_{\hat{b}},x_v)-u_0\right]}}>0$$

where the inequality comes from replacing u_{max} also for b at the denominator. As $\sigma_L>0$, it is the case that the RHS is larger than zero.

Step 2. Log-concave in u. Taking the derivative of Υ (which corresponds to the second derivative of log-market share)

$$\Upsilon\left(u_b,u_{-b}\right)=\frac{\partial_{u_b}\mu\left(\varphi_b\right)}{\mu\left(\varphi_b\right)}=\frac{\sum_{v=1}^{V}N_v\mu_v\left(\varphi_b\right)\left[1-\mu_v\left(\varphi_b\right)\right]}{\sum_{v=1}^{V}N_v\mu_v\left(\varphi_b\right)}=1-\frac{\sum_{v=1}^{V}N_v\mu_v\left(\varphi_b\right)^2}{\sum_{v=1}^{V}N_v\mu_v\left(\varphi_b\right)} \tag{125}$$

Note that in the case with spatial cost $\psi=0$ (or a single market, i.e., $V=1$), the above condition reads as $\Upsilon\left(u_b,u_{-b}\right)=1-\mu(\varphi_b)$, which is trivially strictly decreasing in u_b. In our problem with $\psi>0,V>1$, however, we need to do some additional steps. Taking the derivative of Eq. (125) and simplifying the notation of $\mu(\varphi_b)$ to μ:

$$\partial_{u_b}\Upsilon\left(u_b,u_{-b}\right)=-\frac{2\left[\sum_{v=1}^{V}N_v\mu_v^2(1-\mu_v)\right]\left[\sum_{\hat{v}=1}^{V}N_{\hat{v}}\mu_{\hat{v}}\right]-\left[\sum_{v=1}^{V}N_v\mu_v(1-\mu_v)\right]\left[\sum_{\hat{v}=1}^{V}N_{\hat{v}}\mu_{\hat{v}}^2\right]}{\left[\sum_{v=1}^{V}N_v\mu_v\right]^2} \tag{126}$$

Selecting the terms on top for any pair $v,\hat{v}$, we recover the equation in Lemma H.1.

Step 3. Υ increasing in u_{-b}. Taking the derivative of Υ w.r.t. $u_{\hat{b}}$ (which corresponds to the cross derivative of log-market share)

$$\partial_{u_{\hat{b}}}\Upsilon\left(u_b,u_{-b}\right)=\frac{2\left[\sum_{v=1}^{V}N_v\left(\mu_v^b\right)^2\mu_v^{\hat{b}}\right]\left[\sum_{\hat{v}=1}^{V}N_{\hat{v}}\mu_{\hat{v}}^b\right]-\left[\sum_{v=1}^{V}N_v\mu_v^b\mu_v^{\hat{b}}\right]\left[\sum_{\hat{v}=1}^{V}N_{\hat{v}}\left(\mu_{\hat{v}}^b\right)^2\right]}{\left[\sum_{v=1}^{V}N_v\mu_v^b\right]^2} \tag{127}$$

Selecting the terms on top for any pair $v,\hat{v}$, we recover the equation in Lemma H.1.

Step 4. The Sufficiency of $\overline{\psi}$. We show here that it guarantees the log-supermodularity condition, but the proof is the same for the log-concavity. As utility in equilibrium is bounded below (since consumption is greater than the lower bound of the grid), whenever there is a level of utility that the bank can offer and make a positive profit:

$$\mu_v^b(\varphi_b)\geq\underline{\mu}>0$$

On the other hand, as there is the outside option:

$$\mu_v^b(\varphi_b)\leq\overline{\mu}<1$$

Therefore: $\mu_v\in[\underline{\mu},\overline{\mu}]$. Note that the signal of the $\partial_{u_{\hat{b}}}\Upsilon\left(u_b,u_{-b}\right)$ is the same as

$$\begin{aligned}
&\sum_v\sum_{\hat{v}}N_vN_{\hat{v}}\mu_v^b\mu_{\hat{v}}^b\mu_{\hat{v}}^{\hat{b}}\left[2\mu_{\hat{v}}^b-\mu_v^b\right]\\
&=\sum_v N_v^2\left[\mu_v^b\right]^3\mu_{\hat{v}}^b+\sum_v\sum_{\hat{v}}N_vN_{\hat{v}}\left\{\mu_v^b\mu_{\hat{v}}^b\mu_{\hat{v}}^{\hat{b}}\left[2\mu_{\hat{v}}^b-\mu_v^b\right]+\mu_{\hat{v}}^b\mu_v^b\mu_v^{\hat{b}}\left[2\mu_v^b-\mu_{\hat{v}}^b\right]\right\}\\
&=\sum_v N_v^2\left[\mu_v^b\right]^3\mu_{\hat{v}}^b+\sum_v\sum_{\hat{v}}N_vN_{\hat{v}}\mu_v^b\mu_{\hat{v}}^b\left\{\mu_{\hat{v}}^{\hat{b}}\left[2\mu_{\hat{v}}^b-\mu_v^b\right]+\mu_v^{\hat{b}}\left[2\mu_v^b-\mu_{\hat{v}}^b\right]\right\}
\end{aligned} \tag{128}$$

Assume that there is a gap of $\mu_{u_{max}}=A+\mu_{u_{min}}$ given a level of utility. In this case:

$$\begin{aligned}
\mu_{\hat{v}}^{\hat{b}}\left[2\mu_{\hat{v}}^b-\mu_v^b\right]+\mu_v^{\hat{b}}\left[2\mu_v^b-\mu_{\hat{v}}^b\right]&\geq\mu_{u_{min}}\left[2\mu_{u_{max}}-\mu_{u_{min}}\right]+\mu_{u_{max}}\left[2\mu_{u_{min}}-\mu_{u_{max}}\right]\\
&=4\mu_{u_{min}}\mu_{u_{max}}-\mu_{u_{min}}^2-\mu_{u_{max}}^2\\
&=4A\mu_{u_{min}}+4\mu_{u_{min}}^2-A^2-\mu_{u_{min}}^2+2A\mu_{u_{min}}-\mu_{u_{min}}^2\\
&=6A\mu_{u_{min}}+2\mu_{u_{min}}^2-A^2\\
&\geq 0\Leftrightarrow\frac{A}{\mu_{u_{min}}}\in\left[3-\sqrt{11},3+\sqrt{11}\right]\Leftrightarrow\mu_{u_{max}}\in\left[0,\mu_{u_{min}}\left(4+\sqrt{11}\right)\right]\\
&\Leftrightarrow\frac{\mu_{u_{max}}}{\mu_{u_{min}}}<7.31(\approx)
\end{aligned} \tag{129}$$

$$\mu_{u_{max}}=\frac{e^{u_b-\psi\min\|x_b-x_v\|}}{e^{u_b-\psi\min\|x_b-x_v\|}+\sum_{\beta\in B/b}e^{u_\beta-\psi\|x_\beta-x_v\|}+e^{u0}}$$

$$\mu_{u_{min}} = \frac{e^{u_b - \psi \max \|x_b - x_v\|}}{e^{u_b - \psi \max \|x_b - x_v\|} + \sum_{\beta \in B/b} e^{u_\beta - \psi \|x_\beta - x_v\|} + e^{u0}}$$

Note then that:

$$\frac{\mu_{u_{max}}}{\mu_{u_{min}}} < e^{[\psi \max_{b,v} \|x_b - x_v\| - \psi \min_{b,v} \|x_b - x_v\|] \sigma_L^{-1}}$$

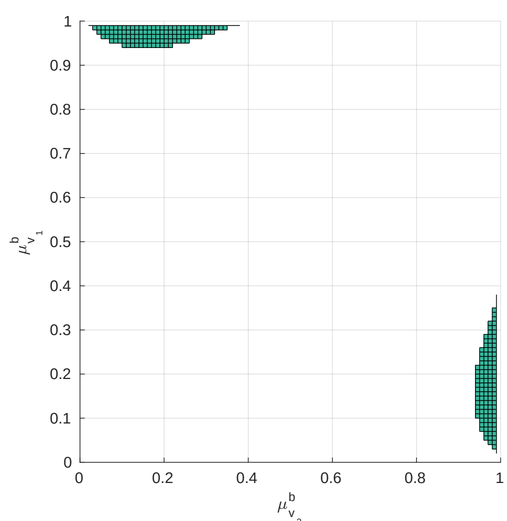
A sufficient condition for log-supermodularity of the game is that:

$$\psi \leq \frac{\log(4 + \sqrt{11}) \sigma_L}{[\max_{b,v} \|x_b - x_v\| - \min_{b,v} \|x_b - x_v\|]}$$

and $\log(4 + \sqrt{11}) \approx 2$, which is easily verifiable. ■

To provide an example, if $N_v = N_{\hat{v}}$, the regions shadowed in Figure 16 represent the combination of market shares in different villages for a given bank that would fail to satisfy the condition of Eq. (122) in Lemma H.1.

Figure 16: Conditions on village level Market Shares to violate Log-Concavity



Note: Combination of market shares in villages v_1, v_2 for bank b, $\mu^b_{v_1}, \mu_{v_2^b}$, that guarantee Lemma H.1 is satisfied assuming $N_{v_1} = N_{v_2}$. Shaded regions represent points where the condition is violated.

I Distance to Nash

Here we propose the following conservative technique to order by rank all possible strategies with metric we call "distance to Nash". We use a more general notation, since this algorithm can be used in other general settings. For simplicity, we illustrate the algorithm with two players: 1 and 2. Let G be a set of strategies by both players and $P_1(G)$ and $P_2(G)$ their payoffs. In the case of adverse selection, this corresponds to Eq.(25). Let G_1 be strategies of player 2 - that is, all that is necessary for 1 to compute its best response (and equivalently G_2).

We can define and compute for any of those deviating strategies the following metrics

$$d(G,G_1) = \max(P_1(G_1) - P_1(G), 0)$$

$$d(G,G_2) = \max(P_2(G_2) - P_2(G), 0)$$

Thus, in the first step of procedure we compute $P_1(G)$ and $P_2(G)$ for a trial strategy set of G. Then, in the second stage we solve

$$\max_{G1} d(G,G_1) \text{ subject to } P_1(G) > 0, \ \forall G_1. \tag{130}$$

$$\max_{G2} d(G,G_2) \text{ subject to } P_2(G) > 0, \ \forall G_2. \tag{131}$$

Let us denote the solution of those maximization problems as $\overline{d(G,G_1)}$ and $\overline{d(G,G_2)}$. Then we compute distance to Nash as

$$d(G,G_1,G_2) = \overline{d(G,G_1)} + \overline{d(G,G_2)}$$

And in the final stage we solve

$$\min_{G} d(G), \ \forall \{G,G_1,G_2\}. \tag{132}$$

At true Nash equilibrium G_{Nash} the solution of this two-step optimization problem

$$\overline{d(G_{Nash})} = 0.$$

At all other strategies, this function is strictly positive and well-defined. All possible strategies can be rank-ordered by their "distance from Nash," even if no true Nash equilibrium exists.

I.0.1 Numerical accuracy of distance to Nash algorithm

When distance to Nash is Lipschitz bounded[38] $d(G) < \lambda * P_{1,2}(G)$ we accept the outcome as an instance of Nash equilibrium. We conduct the same accuracy checks for each case of simultaneous Nash equilibrium we study. Although we don't provide proofs of existence

[38] In this case Lipschitz constant λ specifies a stopping criteria for optimization algorithm with distance to Nash $d(G)$ to act as a "measure" of Nash-closeness in the space of strategies with respect to profit level.

and sufficiency conditions here, those checks serve to filter numerically well-bounded constructively obtained equilibria from outcomes where Nash equilibrium might not exist. The Lipschitz condition λ is set at 10^{-6} value for Nash equilibrium to be considered well-resolved in our numerical examples.

J Relationship Lending

In this section, we show the results of a model of relationship lending. The setting is the same as in Section 4 but with one key difference. In Section 4, we assumed FSPs were heterogeneous in two dimensions. The local bank had an informational advantage but a locational disadvantage with respect to the national bank. This section considers a simpler case where both FSPs are subject to the same spatial costs, but one has an information advantage. We show how equilibrium utilities, profits, and market shares change with changes in both spatial costs ψ and the logit variance σ_L.

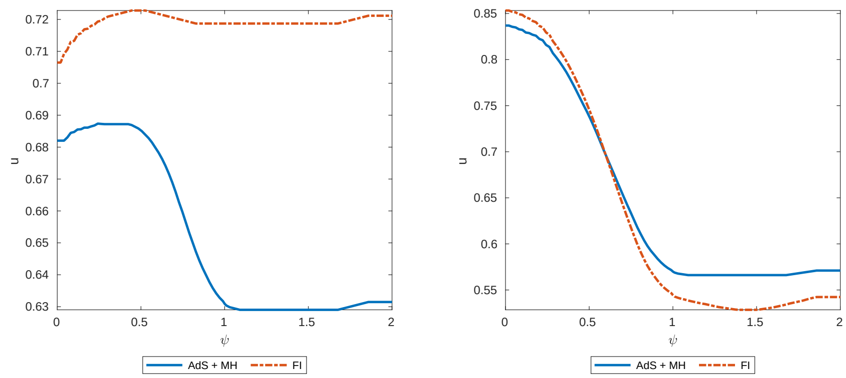
Figure 17 shows equilibrium utilities for each type, and Figure 18 shows the respective market shares and profits of FSP. When the spatial cost ψ increases, the utility for the good type increases to partially offset the higher costs, while it decreases for the bad type (through the local monopoly channel). However, when the spatial cost is sufficiently high, and the local monopoly channel dominates for both types, we have that the FSP subject to MH + ADS must keep utilities somewhat consistent between the two types, as anticipated, while FSP that contracts under FI does not. This generates an asymmetry in the response of utilities for each type when spatial costs are altered. As we assume types are uniformly distributed across villages, this generates regional inequality: villages closer to the FI FSP are better off on average with rising spatial costs. For these villages, however, inequality across types increases within the village.

We repeat the same experiment but for changes in the logit variance σ_L. The results for utilities and market share and profits are, respectively, in Figures 19 and 20. Utilities for the good type, θ_L, are decreasing, while they are hump-shaped for the bad type, θ_H. Note that when SMEs are sensitive to utilities choosing FSPs (σ_L low), the FSP subject to FI has market share advantages in the good type, θ_L, since it can offer a higher utility without also having to increase utilities for the bad type. We observe the opposite as the logit variance σ_L increases, and both FSPs have more market power.

For low values of σ_L, note that the uninformed bank has a very high share of bad clients, which could indicate a worse portfolio (riskier, for instance). In our model, we do not explicitly take this into account since FSPs are risk-neutral and there are no aggregate shocks. The systemic risk this generates (all bad clients with the same FSP) could be relevant to explain macro fluctuations. This is a direction for future research.

Figure 17: Relationship Lending and Spatial Costs: Equilibrium Utilities

(a) θ_L - Good Type (b) θ_H - Bad Type



Note: Equilibrium utilities played by two FSPs in a Hotelling line. One FSP is located at $x = 0$, while the other is at $x = 1$. The FSP at $x = 0$ contracts under FI, while the one at $x = 1$ contracts under AdS + MH. Parameters for estimation are in Table 6. We solve the equilibrium using the distance to Nash algorithm (Appendix I). The x-axis, ψ, is spatial costs. Utilities are normalized such that zero is the autarky and one is the FI, which is the perfect competition level for the bad type.

K Details of Eqs.(27) and (149)

From our market share by village equation, 21, we have that

$$\mu_{v,b} = \frac{e^{\sigma_L^{-1}[\mathbb{V}(u_b,x_b,x_v)-u_0]}}{1+\sum_{\hat{b}=1}^{B} e^{\sigma_L^{-1}[\mathbb{V}(u_{\hat{b}},x_{\hat{b}},x_v)-u_0]}} \tag{133}$$

Taking logs

$$\ln \mu_{v,b} = \sigma_L^{-1}\left[\mathbb{V}(u_b,x_b,x_v) - u_0\right] - \ln\left(1 + \sum_{\hat{b}=1}^{B} e^{\sigma_L^{-1}[\mathbb{V}(u_{\hat{b}},x_{\hat{b}},x_v)-u_0]}\right) \tag{134}$$

For the outside option

$$\ln \mu_{v,0} = -\ln\left(1 + \sum_{\hat{b}=1}^{B} e^{\sigma_L^{-1}[\mathbb{V}(u_{\hat{b}},x_{\hat{b}},x_v)-u_0]}\right) \tag{135}$$

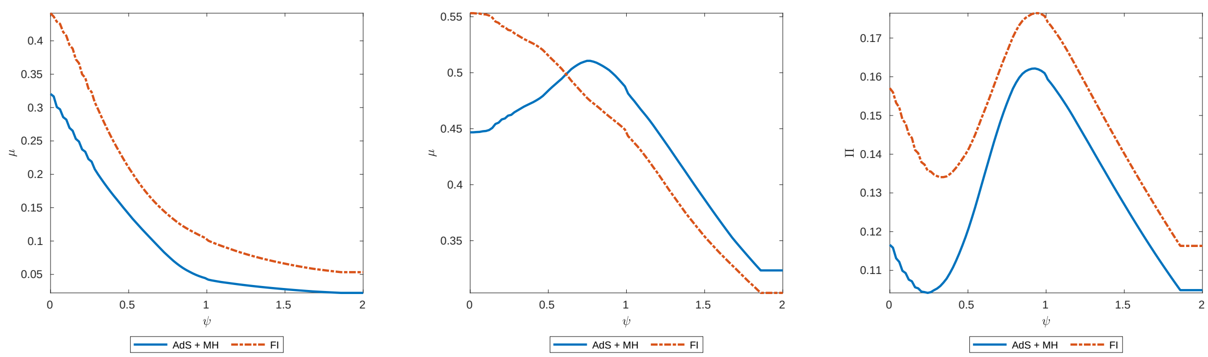
Thus:

$$\ln \mu_{v,b} - \ln \mu_{v,0} = \sigma_L^{-1}\left[\mathbb{V}(u_b,x_b,x_v) - u_0\right] \tag{136}$$

To get (27), we simply add the measurement error in the end. It could also be an error that FSPs do not anticipate when making contract decisions. From our market share by village

Figure 18: Relationship Lending and Spatial Costs: Market Shares and Profits

(a) Market Share, θ_L (b) Market Share, θ_H (c) Profits



Note: Market shares and profits implied by the equilibrium utilities of the game between two FSPs in a Hotelling line. One FSP is located at $x=0$, while the other is at $x=1$. The FSP at $x=0$ contracts under FI, while the one at $x=1$ contracts under AdS + MH. Parameters for estimation are in Table 6. We solve the equilibrium using the distance to Nash algorithm (Appendix I). The x-axis, ψ, is spatial costs.

equation, 21, we have that

$$\begin{aligned}\partial_u \mu_{v,b} &= \sigma_L^{-1} \frac{e^{\sigma_L^{-1}[\mathbb{V}(u_b,x_b,x_v)-u_0]}}{1+\sum_{\hat{b}=1}^{B} e^{\sigma_L^{-1}\left[\mathbb{V}(u_{\hat{b}},x_{\hat{b}},x_v)-u_0\right]}} + \sigma_L^{-1} \frac{e^{\sigma_L^{-1}[\mathbb{V}(u_b,x_b,x_v)-u_0]}}{1+\sum_{\hat{b}=1}^{B} e^{\sigma_L^{-1}\left[\mathbb{V}(u_{\hat{b}},x_{\hat{b}},x_v)-u_0\right]}} \frac{\sigma_L^{-1}\left[\mathbb{V}(u_{\hat{b}},x_{\hat{b}},x_v)-u_0\right]}{1+\sum_{\hat{b}=1}^{B} e^{\sigma_L^{-1}\left[\mathbb{V}(u_{\hat{b}},x_{\hat{b}},x_v)-u_0\right]}} \\ &= \sigma_L^{-1}\left[\mu_{v,b}-\mu_{v,b}^2\right] \qquad (137)\end{aligned}$$

Therefore:

$$\partial_u \mu_b = \sum_{v=1}^{V} N_V \partial_u \mu_{v,b} = \sigma_L^{-1} \sum_{v=1}^{V} N_V \left[\mu_{v,b}-\mu_{v,b}^2\right] \qquad (138)$$

$$= \mu_b - \sigma_L^{-1} \sum_{v=1}^{V} N_V \mu_{v,b}^2 \qquad (139)$$

Dividing by μ

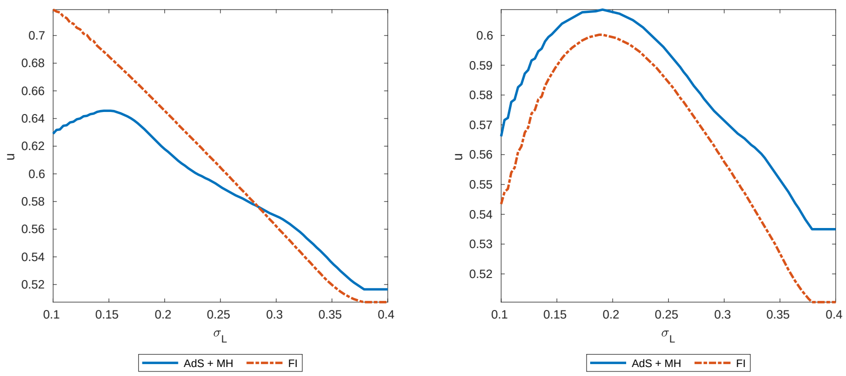
$$\partial_u \mu_b = \sum_{v=1}^{V} N_V \partial_u \mu_{v,b} = \sigma_L^{-1} \sum_{v=1}^{V} N_V \left[\mu_{v,b}-\mu_{v,b}^2\right] \qquad (140)$$

$$= 1 - \sigma_L^{-1} \frac{\sum_{v=1}^{V} N_V \mu_{v,b}^2}{\sum_{v=1}^{V} N_V \mu_{v,b}} \qquad (141)$$

which delivers Eq.(149).

Figure 19: Relationship Lending and Logit Variance: Equilibrium Utilities

(a) θ_L (b) θ_H



Note: Equilibrium utilities of the game between two FSPs in a Hotelling line. One FSP is located at $x = 0$, while the other is at $x = 1$. The FSP at $x = 0$ contracts under FI, while the one at $x = 1$ contracts under AdS + MH. Parameters for estimation are in Table 6. We solve the equilibrium using the distance to Nash algorithm (Appendix I). The x-axis, σ_L, is the logit variance, which changes market share sensitivity to utilities. Utilities are normalized such that zero is the autarky and one is the FI, the perfect competition level for the bad type.

L Other Comparative Statics Results

L.1 Logit Variance

We show in this section the analogous of the results of Section D, but varying the logit variance, σ_L. This variance pins down the scale of utilities. The scale of utility - which here is exactly pinned down by the inverse of σ_L - is important for the equilibrium since it determines how market share changes with an equivalent change in utility. This can be seen in Figure 14. Changes in σ_L affect the downward-sloping curve and, thus, the equilibrium determination.

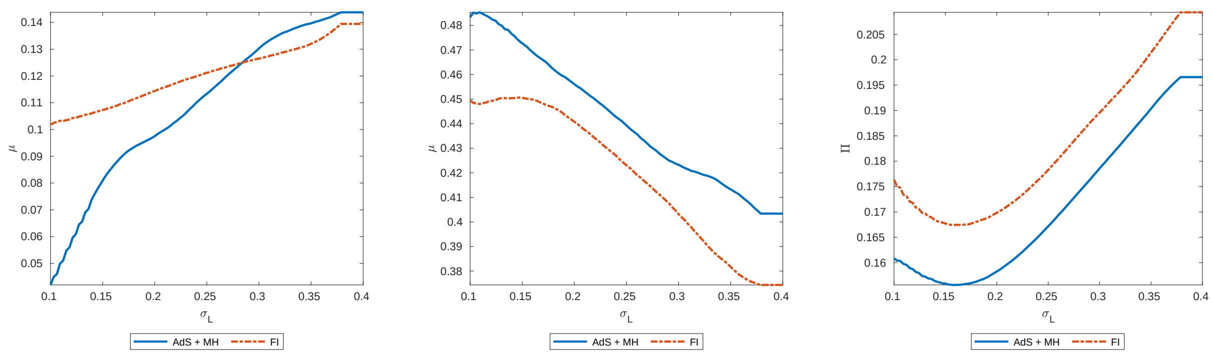
In Figure 21, we show how profits and welfare of the villages in $x \in \{0, .5, 1\}$ vary with σ_L. The welfare is as in Eq.(24).

L.2 Complementarity of Competition

We simultaneously vary spatial costs ψ and logit variance σ_L to understand if changes in σ_L and ψ are substitutes or complements and how this changes with the level of competition in the economy. For simplicity, we change our baseline economy to be symmetric in locations and have one FSP at each location, that is, $b_L = b_R = 1$. We compute the overall welfare in

Figure 20: Relationship Lending and Logit Variance: Market Shares and Profits

(a) Market Share, θ_L - Good Type (b) Market Share, θ_H - Bad Type (c) Profits



Note: Market shares and profits implied by the equilibrium utilities of the game between two FSPs in a Hotelling line. One FSP is located at $x = 0$, while the other is at $x = 1$. The FSP at $x = 0$ contracts under FI, while the one at $x = 1$ contracts under AdS + MH. Parameters for estimation are in Table 6. We solve the equilibrium using the distance to Nash algorithm (Appendix I). The x-axis, σ_L, is the logit variance, which changes market share sensitivity to utilities. Utilities are normalized such that zero is the autarky and one is the FI, the perfect competition level for the bad type.

the economy as in Eq.(142) and plot the results in Figure 22, panel (a).

$$\mathcal{W}(\psi, \sigma_L) \equiv V^{-1} \sum_{v} W_v(\psi, \sigma_L) \tag{142}$$

where $W_v(\psi, \sigma_L)$ is the one defined in Eq.(24).

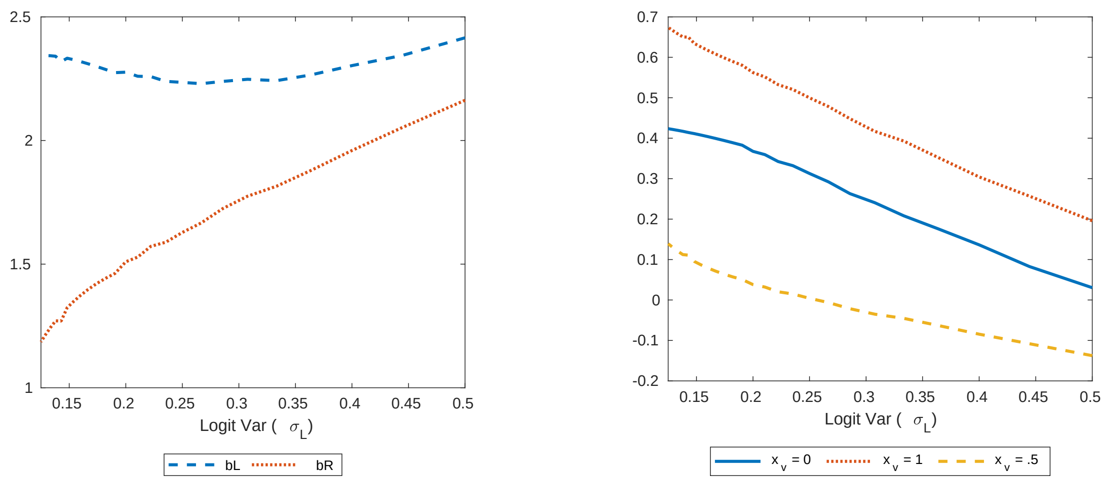
The effects of reducing spatial costs are more pronounced with lower values of σ_L, which indicates that if utility offerings do not sufficiently change market shares, reduction in spatial costs is also less effective in inducing welfare changes. This can be seen in panel (a) of Figure 22. For low values of σ_L, welfare is larger *and* increases by more when spatial costs are reduced than for high values of σ_L. In panel (b), we plot the difference in welfare between a more competitive economy with $b_L = 4 = b_R = 4$ with the welfare plotted in panel (a) with $b_L = b_R = 1$. Note that more FSPs increase levels of utilities since the surface - the welfare differential - is all in positive numbers. This change is not constant across the parametric space of spatial costs ψ and logit variance σ_L. Reduction in spatial costs is passed through more to consumers when competition is higher.

L.3 Local Competition

We consider the effects of introducing additional FSPs in profits and welfare of households in villages at $x \in \{0, .5, 1\}$. Finally, we show how profits and utilities in equilibrium differ between full information and moral hazard and limited commitment in this case.

Figure 21: Profits of FSPs and welfare of villages $x \in \{0, .5, 1\}$ with changes the logit variance

(a) Profits (b) Welfare



Note: Profits of FSPs and Welfare (as in Eq.(24) for three villages - the ones located in $x \in \{0, .5, 1\}$. Equilibrium with the spatial configuration of Figure 4 and parameters of Table 5, changing the logit variance, denoted by σ_L. Contracting frictions are MH + LC.

The introduction of FSPs at $x = 1$ increases the welfare of the village at $x = 1$ by a significant amount - from 20 % to almost 50 % of the perfect competition full information utility. The effect on the village at $x = .5$ is qualitatively similar but quantitatively smaller given the distance of this village to this new, more competitive locale. In our framework, as we have logit market shares that come from idiosyncratic preferences of households within a village, competition in $x = 1$ can decrease the welfare of households at $x = 0$ since some of them prefer to pay spatial costs to visit the FSPs in $x = 1$. This comes from our measure of welfare used. We do not take into account in panel(b) of Figure 23 the idiosyncratic preferences effects (that generate the logit market share), and thus, it may seem that welfare is decreasing when, in fact, it is not. [39]

M Numerical Method

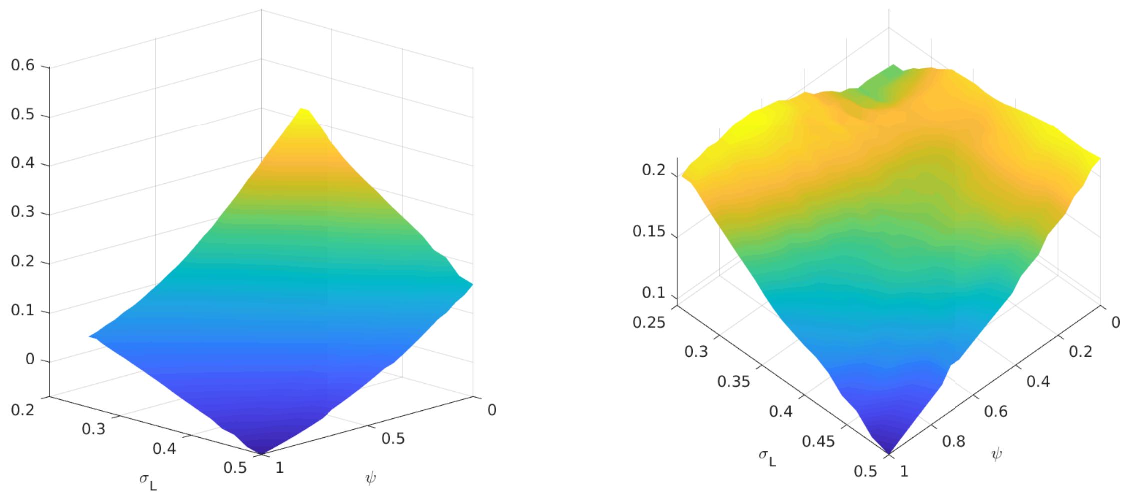
M.0.1 Identification of the Frontier

In this subsection, we assume that the model cannot replicate the data perfectly because FSPs do not fully understand the structure of the model (either the frontier or market shares) and/or there is a measurement error in market shares, as in Eq.(27). In particular, we assume

[39] The idiosyncratic preference shocks imply a few households will travel now larger distances. We capture the large distances in our measure of welfare but not the effect of increased varieties.

Figure 22: Average Welfare varying spatial costs and logit variance

(a) Welfare with $b_L = 1, b_R = 1$ (b) Welfare Gains with $b_L = 4$, $b_R = 4$



Note: Panel(A): Welfare (as in Eq.(142). Equilibrium with spatial configuration of Figure 4 and parameters of Table 5, changing the logit variance, denoted by σ_L, and spatial costs, denoted by ψ. One FSP at $x = 0$ and one at $x = 1$, that is $b_L = b_R = 1$. Panel (B): Welfare difference between the economy with $b_L = 4 = b_R = 4$ and the economy with $b_L = 1 = b_R = 1$. Contracting frictions are MH + LC.

that the profit of an FSP is given by Eq. (143) and that the errors FSPs make are province-specific.

$$\Pi\left(\varphi_b^p\right) \equiv S(u_b^p)\mu(\varphi_b)\,\chi_b^p(u_b^p), \text{ where } \chi_b^p(u_b^p) \equiv e^{\varsigma_b\left[u_b^p - u_0\right]} \tag{143}$$

Without the error $\psi_{b,p}$, we are back at the profit function defined in Eq.(19). The form of the error in Eq. (143) guarantees that the FOC of an FSP, given what other FSPs are doing, is given by Eq.(144). The difference from Eq.(70) (the version without the error) is now that FSPs do not follow that FOC exactly due to the error

$$-\frac{\partial_{u_b} S(u_b^*)}{S(u_b^*)} + \varsigma_b = \frac{\partial_{u_b}\mu\left(u_b^*, u_{-b}\right)}{\mu\left(u_b^*, u_{-b}\right)} \tag{144}$$

For notation purposes, we define the log difference in market shares as $\omega_{v,b-0}^p$, that is

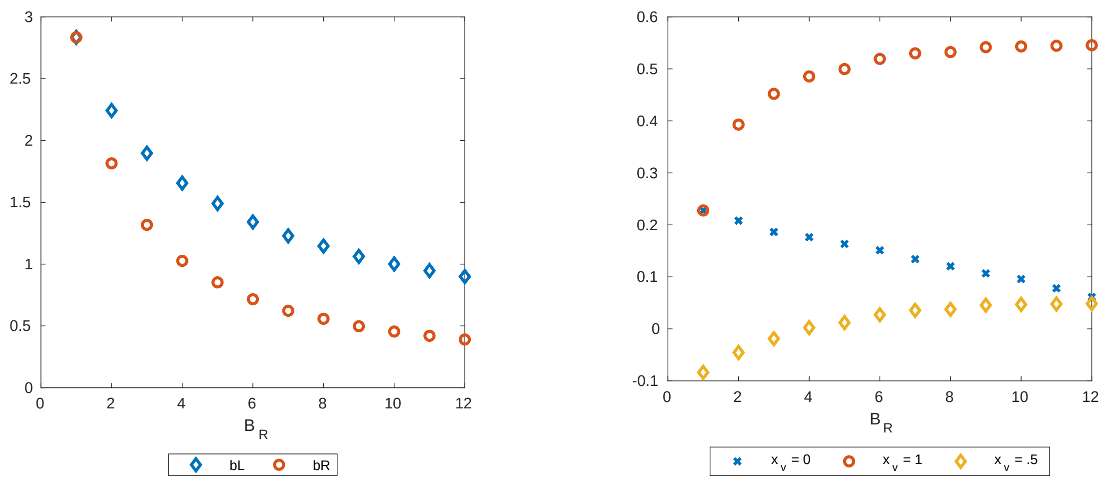
$$\omega_{v,b-0}^p \equiv \ln\left(\mu_{v,b}^p\right) - \ln\left(\mu_{0,v}^p\right) \tag{145}$$

Moreover, let $\bar{x}_b^p$ represent the mean of a given variable $x_{v,b}^p$ over villages. This will be useful for applying the insight of a fixed effects panel model, where we subtract the within FSP-province mean of a variable and use the variation to estimate the spatial cost.

First, we focus on the identification and estimation of ψ. We can re-write the difference of

Figure 23: Profits of FSPs and welfare of villages $x \in \{0, .5, 1\}$ with changes in b_R

(a) Profits　　　　(b) Welfare



Note: Profits of FSPs and Welfare (as in Eq.(24) for three villages - the ones located in $x \in \{0, .5, 1\}$. Equilibrium with the spatial configuration of Figure 4 and parameters of Table 5, changing the number of FSPs in $x = 1$, denoted by b_R. Contracting frictions are MH + LC.

variables with respect to their means across villages in Eq.(27) as

$$\omega^p_{v,b-0} - \bar{\omega}^p_{v,b-0} = \psi\left[t(x^p_b, x^p_v) - \bar{t}(x^p_b, x^p_v)\right] + \vartheta^p_{b,v} - \bar{\vartheta}^p_{b,v} \tag{146}$$

Thus, intra-province variation in market shares allows us to estimate the spatial cost, which we can do in Eq.(146) through OLS. This is not surprising. Within a market, we expect villages to have different market shares in each FSP due to the travel time between them. Let $\hat{\psi}$ the OLS estimator of ψ in Eq.(146).

Second, we focus on the identification of $\Sigma(u)$. If we identify $\Sigma(u)$, S is identified up to a constant.[40] Given a value of ψ, we can define an estimator for the observed utilities, $\hat{u}^p_b$ as a function of $\hat{\psi}$, as in Eq.(148).

$$\hat{u}^p_b \equiv \bar{\omega}^p_{v,b-0} + \hat{\psi} t(x_{v^p}, x_b) \tag{148}$$

From the FOC of the FSPs in Eq.(144), we can estimate a value for $\hat{\Sigma}_{b,p}$, the value of Σ *in*

[40]We can integrate $\Sigma(u)$ to obtain S as in Eq.(147). The constant does not change bank choices at the margin, so it can be ignored here.

$$S(u) = \text{ct} \cdot \exp\left[\int_{u_{min}}^{u} \Sigma(v) dv\right] \tag{147}$$

equilibrium for FSP b in province p given by [41]

$$\hat{\Sigma}_b^p = \hat{\Upsilon}_b^p = 1 - \frac{\sum_{v^p} N_v^p \left(\hat{\mu}_{v,b}^p\right)^2}{\sum_{v^p} N_v^p \hat{\mu}_{v,b}^p} \tag{149}$$

where the second equality comes simply from taking the derivative of market share in Eq. (29).

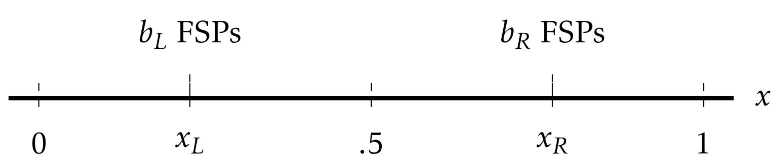
Although there may be more efficient possibilities to estimate $\hat{\Sigma}$ and $\hat{u}_{b,p}$, we focus on the simple approach of Eqs. (148)-(149) of simply computing sample averages. Once we obtain the vectors $\{\hat{u}_b^p, \hat{\Sigma}_b^p\}$, we can estimate a non-parametric function for $\Sigma(u)$. We explore the exact way to do this in Section M.0.2.

M.0.2 Numerical Example

To showcase the power of the methodology, We present a numerical example. We conduct two counterfactuals: the introduction of an additional bank and changes in the spatial cost, and compare the true and estimated welfare effects.

We simulate P markets with a different spatial configuration and number of FSPs. As in Section D, each market has a Hotelling as in Figure 24. We assume that V villages are uniformly distributed in the $(0,1)$ and that FSPs are in two different locations, one to the left of the middle point $x = .5$, denoted by x_L, and one to the right, denoted by X_R For simplicity, we assume that $x_R = 1 - x_L$, i.e., the positions of FSPs are always symmetric. We randomly select a position for FSPs and the number of FSPs in x_L and x_R - between 1 and 5 for each location. Each village has a continuum of SMEs.

Figure 24: A Province p in the simulation



Note: a visual representation of a sample province in our simulation exercise. We assume that there are V villages uniformly distributed between 0 and 1. We choose x_L random between $[0,.5]$ (and use $x_R = 1 - x_L$) and $b_L, b_R \in \{1, \ldots, 5\}$ for each province p .

For this numerical exercise, we compute the equilibrium in utilities assuming that there is only one type of entrepreneur and the frontier is given by Eq.(150). This particular functional form is one of many concave forms from which we could choose. We opt to define the frontier

[41] To derive Eq.(149), simply compute the derivative of the logit market shares over itself. The square term comes from the fact that the derivative of the market share at μ is given by $\sigma_L^{-1}\mu(1-\mu)$.

in terms of parameters instead of microfounding it (as in Section 3.1) to highlight that it does not matter where the frontier comes from.

$$S(u \mid \zeta_S) \equiv 1 - e^{\zeta_S (u-1)} \tag{150}$$

As we cannot identify the scale of utilities, we assume each FSP can choose a utility $u \in [0,1]$. We use an iterative best response to find the equilibrium in utilities (See Lemma 3.1).

To be consistent with the previous section, we assume that FSPs do not fully understand the model and each FSP has the error $\chi_b^p(u_b^p)$ that distorts the FOC and that market shares are observed with an error $\vartheta_{b,v}^p$. Table 15 shows the baseline parameters we use for the estimation. We compute the standard errors using a non-parametric bootstrap by re-sampling provinces.

Table 15: Parameters used for Numerical Simulation

Parameter	Value	Meaning
ζ_S	2	Concavity of Frontier
ψ	1.5	Spatial Cost
V	50	Number of Villages by Province
P	250	Number of Provinces
u	[0,1]	Possible utility values
σ_ϑ	.25	Std. Dev. in $\vartheta_{b,v}^p$, measurement errors in $\mu_{v,b}^p$
σ_ς	.05	Std. Dev. in ς_b^p, error of FOC of FSPs
b_L, b_R	$\{1,\ldots,5\}$	Possible Number of FSPs in Each Location
x_L	[0,.5]	Possible Position of Left Location of FSPs

Note: parameters used to estimate $\Sigma(u)$, the curvature of the frontier, and ψ, the spatial costs.

We start by using Eq. (146) to estimate the spatial cost by OLS. The true spatial cost is $\psi = 1.5$, while our estimate is $\hat{\psi} = 1.49$ (with standard error .006). The heterogeneity in market shares by the same FSP in different villages is what identifies this parameter. We use Eqs. (148)-(149) to recover a dataset of $\{\hat{u}_b^p, \hat{\Sigma}_u^p\}$. Although we could non-parametrically estimate the curvature of the frontier, we opt for the simplicity of fitting a polynomial regression as in Eq. (151) [42]

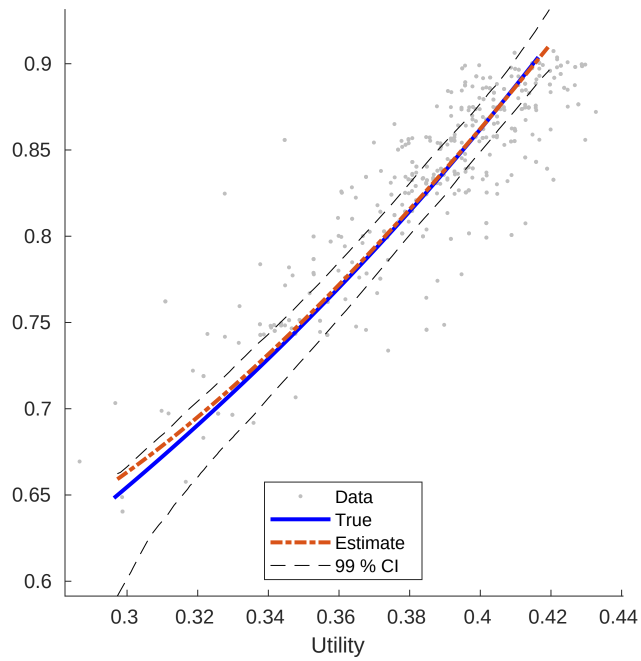
$$\hat{u}_b^p = \beta_0 + \beta_1 \hat{\Sigma}_u^p + \beta_2 \left(\hat{\Sigma}_u^p\right)^2 + \beta_3 \left(\hat{\Sigma}_u^p\right)^{1/2} + \nu_{b,p} \tag{151}$$

The results are in Figure 25. We can approximate the frontier well overall, and the confidence interval is only large at low levels of utility. This is a consequence of errors $+\varsigma_b$ in the

[42] Is is worth noting that given the structure of the model, we opt for running Eq. (151) with $\{\hat{u}_b^p\}$ as the dependent variable to avoid having the error the estimation of $\{\hat{u}_b^p\}$ to be correlated with $\nu_{b,p}$, the true error in the model .

FOC of FSPs, impacting the utility choices in monopolies. In competitive markets, the effect of the errors on utilities is reduced through competition.

Figure 25: Identification of Frontier Through Market Share Data



Note: Estimation of $\hat{\Sigma}$ using market share data. The blue solid line represents the true $\Sigma(u) \equiv \frac{\partial S(u)/\partial u}{S}$. The red dashed-dotted line is estimated. The dashed black lines represent the lower and upper bound of the confidence intervals at 1 %, computed with 1000 bootstrap repetitions re-sampling provinces. Grey dots are observed $\{\hat{u}^p_b, \hat{\Upsilon}^p_b\}$.

We focus now on the counterfactuals. First, we consider the effects of adding FSPs in three different markets. In all markets, FSPs are located at $x_L = 0$ and $x_R = 1$, that is, in the extremes of the Hotelling line. The markets differ, however, in their baseline number of FSPs, which can be 2, 3, or 4 (in each location). The results are in Table 16. As utility is a cardinal concept, and here we assume that there is no model to translate utility gains to consumption gains, we showcase utility changes from the policy over the range observed in the data. This means that we can interpret changes in welfare as a percentage of the variation we observe between the minimum and maximum utilities we recover from market shares. This is informative because it can tell how much a policy can add in terms of making some villages without competition in intermediation closer to the villages with competition in intermediation in the sample. As seen in Table 16, there are gains from competition from introducing FSPs in markets, but this gain is decreasing with the baseline number of FSPs, as expected. Note that our method does a good job of estimating the effect and providing reasonable confidence intervals to it.

Table 16: The Welfare Effect of Additional Banks

Change in FSPs in $\{x_L, x_R\}$	True	Estimated
2 to 3	.1781	.1638
		[.1405, .1868]
3 to 4	.0911	.0935
		[.0870, .1022]
4 to 5	.0434	.0409
		[.0373, .0468]

Note: Welfare effects of including additional banks in each location for three different levels of baseline competition. Each province we analyze has either 2, 3, or 4 FSPs in both $x_L = 0$ and $x_R = 1$ in the baseline, and we add one bank in both locations. Parameters used to estimate $\Sigma(u)$, the curvature of the frontier, and ψ, the spatial costs, are given by Table 15. Confidence intervals were computed with 1000 bootstrap repetitions re-sampling provinces.

Table 17 has the equivalent results for changes in spatial costs. Not only does our model work well, but we can recover the insight of Figure 22 on the Comparative Statics Exercises (Section D), where we discussed that reductions in spatial costs can increase welfare by significantly more in more competitive intermediation environments.

Table 17: The Welfare Effect of Reducing Spatial Costs: from $\psi = 1.5$ to $\psi = .75$

Baseline FSPs in $\{x_L, x_R\}$	True	Estimated
2	.6166	.6168
		[.6158, .6183]
3	.6724	.6728
		[.6719, .6737]
4	.7757	.7746
		[.7728, .7768]

Note: Welfare effects of reducing the spatial costs for three different levels of baseline competition. Each province we analyze has either 2, 3, or 4 FSPs in both $x_L = 0$ and $x_R = 1$ in the baseline. Parameters used to estimate $\Sigma(u)$, the curvature of the frontier, and ψ, the spatial costs, are given by Table 15. Confidence intervals were computed with 1000 bootstrap repetitions re-sampling provinces.

M.0.3 Unobserved Heterogeneity

In the presence of unobserved heterogeneity, our method to identify the frontier can still be used. We consider that there are two types of agents within each village, which are observed by the FSP but not by the researcher. In particular, instead of observing the market share of FSPs across types, the researcher only observes the market share of banks in the village (aggregated across types). For simplicity, we assume that half of the population is of type I, which has the following frontier:

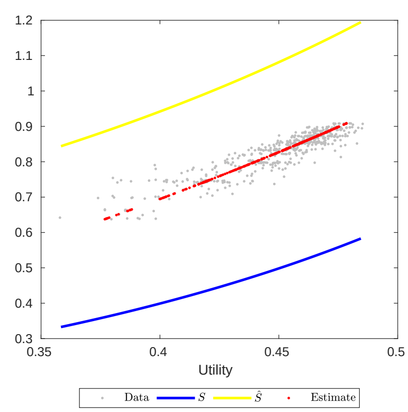
$$S_I(u) = 1 - e^{1.75(u-1)}$$

while the other half is of type II, which has a frontier given by

$$S_{II}(u) = 1 - e^{4(u-1)}$$

These frontiers are sufficiently different: they generate a difference in utility of around 50%. We conduct the same simulation exercise as in Section M.0.2, but assuming only the average market share between types one and two is observed. The curvature of the two frontiers and estimated are in Figure 26. As in the common applied econometrics literature, the estimated Average Treatment Effect of Local Average Treatment Effect is a weighted average of the true economic parameters of each individual. The estimated frontier is in an average of the true frontier for two types, which we can then use to compute counterfactuals as in Section M.0.2.

Figure 26: Identification of Frontier Through Market Share Data: Unobserved Heterogeneity



Note: Estimation of $\hat{\Sigma}$ using average market share data between types I and II, which have frontiers given by $S_I(u) = 1 - e^{1.75(u-1)}$ and $S_{II}(u) = 1 - e^{4(u-1)}$. Parameters are as in Table 15. Blue and yellow solid line represents the true $\Sigma_i(u) \equiv \frac{\partial S_i(u)/\partial u}{S_i}$ for $i \in \{I, II\}$. The red dashed-dotted line is the estimated. Grey dots are observed $\{\hat{u}_b^p, \hat{\Upsilon}_b^p\}$.

M.1 Location of Banks and Household Level Data

In this section, we develop a likelihood estimator based on household-level data (consumption, income, and capital) and the location of banks. We develop this estimator for two reasons. First, we observe this in the Townsend Thai data and use it in our application in Section 6. Second, we want to show how to take contracting models in general to the data.

With a model of contracting in utilities and the market structure, the utilities in equilibrium imply contracts, which themselves imply how the joint distribution of consumption, income, and capital should be in household data.

There are three steps in constructing the likelihood we use here. First, we generalize the problem of banks to include an additive random term, uncorrelated across banks and locations. As in Bresnahan and Reiss (1991), our model will imply the number of banks that *should* be present in each location and, thus, a likelihood function. Second, we extend the analysis in Karaivanov and Townsend (2014) to map contracts played in equilibrium to household income, consumption, and capital data. As seen in Section 3, contracts played in equilibrium are lotteries. We add a measurement error to household-level data (either due to data collection or to the finite grids we define contracting over) and combine the measurement error with the lotteries probabilities to recover what should be observed in household-level data in terms of the joint distribution of consumption, income, and capital if the model was true (i.e., a likelihood). Third, we must combine the first two steps in one unique likelihood to be optimized. We show that we can maximize the log-likelihood of household level data *given* the number of banks we actually observe in the data plus the log-likelihood of observing that number of banks in a given location.

After constructing the likelihood function, we discuss our numerical method. We show that some parameters can be computed indirectly, which speeds up computation significantly. We show that parameters on the market structure (σ_L, ψ) are identified through Monte Carlo experiments. We explore the identification of other parameters and provide pseudo-codes in Appendix N.

M.1.1 Likelihood of FSP Location Data

With the stochastic terms in the profit as in Eq. (32), we can assign a probability for the number of FSPs in the data given the model structure. This will be the likelihood of the supply side. Note, however, that for each $p \in P$, the above system of inequalities is not independent: for a given number of FSPs in a given potential location, the utilities in equilibrium are different in all locations across that province [43]. We must compute a new equilibrium in the *whole* market for a deviation at *each* location. What we can compute for each p given data on the position of banks and assuming that the marginal entrant is such that they will compete afterward is given in Eq. (152). We can write the intersection as a multiplication because we

[43] Provinces are always assumed to be independent of each other

assume that the errors are independent across locations.

$$\mathbb{P}(\mathfrak{S}|\zeta) = \mathbb{P}\left\{\bigcap_{m^p} \mathcal{E}(m^p) = 1 \middle| \zeta\right\} = \prod_{m^p} \mathbb{P}\left\{\mathcal{E}(m^p) = 1 \middle| \zeta\right\} \tag{152}$$

Finally, using the normality assumption for the ι_{m^p}'s in Eq. (32), we can write Eq. (152) as Eq. (153) [44]

$$\ln\{\mathbb{P}(\mathfrak{S}|\zeta)\} = \sum_{m^p} \ln\left(\Phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right] - \Phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{s}\right]\right) \tag{153}$$

where $\Phi(.)$ is the standard normal cdf. Note here that the model is not scale invariant and, thus, potentially informative about the variance s of locations' profit shocks $\iota_{m^p} \sim_{idd} \mathcal{N}(c_E, s^2)$.

M.1.2 Likelihood of Household Level Data

Due to actual measurement error or actual contracting beyond our grids, (c,q,k) data is not limited to be in the small finite grids and, therefore, we associate the probability of observing a triple (c,q,k) given the measurement error as in Eq. (34). Define $\#Y \equiv C \times Q \times K$, i.e., the Cartesian product of the grids. Therefore, for any y implied in the model (pre-measurement error), we have that $y \in \#Y$. Let $l = 1,\ldots,L$ represent the different elements of $\hat{y}_j$ - (c_j, q_j, k_j) here. Then, for a given household j, the likelihood of observing a given $\hat{y}_j$ is given by Eq. (154). In others, for each possible point in the grid, we compute the function $f_v(.)$, the probability that a point is the outcome of contracting. Given a contracting outcome, we sum all probabilities of the actual observed $\hat{y}_j$ given the measurement error (for each element of $\hat{y}_j$), if y was the actual contracting outcome.

$$F_v\left(\hat{y}_j;\zeta,\gamma_{ME}\right) \equiv \sum_{r=1}^{\#Y} f_v(c,q,k|\zeta) \prod_{l=1}^{L} \Phi\left(\hat{y}_j^l \mid y_r^l, \varsigma^l(\gamma_{ME})\right) \tag{154}$$

where $\Phi(.|\mu,\varsigma)$ stands for the CDF of a normal distribution with mean μ and variance ς^2.

We now need to sum F_v over all households in all villages to get the full sample likelihood. Let the village v in province p of household j^p be given by $\hat{v}_j^p$. Denote the demand side data as $\mathfrak{D} = \left\{(\hat{y}_j^p, \hat{v}_j^p)\right\}_{j,p}$, that is, consumption, income, capital and village for each household j, in each province p. The likelihood of demand $\mathfrak{D}$ *given* the position of banks observed in the data, is given by Eq.(155) (re-introducing province superscripts). To obtain the log-likelihood for the overall sample, we sum the log of F_v for all households in all villages and provinces.

$$\ln\{\mathbb{P}(\mathfrak{D}|\mathfrak{S},\zeta)\} = \sum_{p}\sum_{v^p}\sum_{j^p} \mathbb{1}_{\hat{v}_j^p = v^p} \ln F_v^p\left(\hat{y}_j^p,\zeta\right) \tag{155}$$

[44]For notation purposes, we define $\Pi^E(0|.) = \infty$, since we want $\Phi\left[\frac{\Pi^E(0|.)}{s}\right] = 1$.

Eq. (155) is the likelihood of demand $\mathcal{D}$ *given* supply because we use the actual number of observed banks in each potential location m^p to compute the utilities in equilibrium in Eq.(36). The next subsection shows that this is sufficient to combine the likelihoods.

Unobserved Heterogeneity and A reduced Form of Competition. Our structural method is also flexible to deal with unobserved heterogeneity. This can be a relevant state variable that is not observed, such as types θ or initial asset positions or the equilibrium utilities themselves. For that, assume that we have a set s_0 of unobserved states relevant to determine the distribution of s_0.

Let $h_v^{k,s_0}(k,s_0)$ is the joint distribution of capital k and the unobserved state, s_0, in village v. We can write the joint distribution as a function of the marginal distribution of k, observed, and the conditional distribution of s_0 conditional on k, given by $h_v^{s_0|k}$, not observed as in Eq.(156). We assume that this distribution is parametrized by parameters ζ^u, which we can estimate in the likelihood.

$$h_v^{k,s_0}(k,s_0) = h_v^k(k)h_v^{s_0|k}(s_0|k,\zeta^u) \tag{156}$$

We can re-write Eq.(37) as Eq.(157)

$$f_v(c,q,k|\zeta,\zeta^u) = \sum_{s_0} g_v(c,q|k,\zeta)h_v^k(k)h_v^{s_0|k}(s_0|k,\zeta^u) \tag{157}$$

From Eq.(157), we can simply modify the likelihood computation to include the parameters ζ^u. This methodology is used in Karaivanov and Townsend (2014) to estimate initial asset holdings in simply borrowing/savings contracts.

This likelihood formulation in Eq.(157) also provides a useful result to understand the effects of competition in utilities. Assume that we do not know or don't want to assume the market structure of the model. We can define s_0 to be utilities in equilibrium for the level of capital k in a given province. In particular, we can parametrize it as a normal distribution with different means/standard deviation for each level of capital as $s_0 \sim \mathcal{N}(\kappa_k, \sigma_u^k)$. This would allow a researcher to understand the effects of competition *without* a model of how the equilibrium is determined. The downside is that it is not possible to conduct counterfactuals in this way since we do not take a stand on how the equilibrium utilities are determined.

M.1.3 Combining The Two Datasets

We explore now how to combine the likelihood of Section M.1.1 and Section M.1.2. We show that we can compute the demand log-likelihood (based on household-level data) *given* the observed location of banks and sum with the log-likelihood of the supply (the number of banks and potential locations). We also provide an additional result, [45] which is how to combine the number of banks for all provinces and household level data for the subset of provinces and villages that are in fact observed. This is useful for a researcher who observes the numbers of banks across several provinces but only has detailed household data for a few specific provinces.

[45] In this version of this paper we do not use this result, but we do intend to use it in future versions.

These two results are mathematically stated in Lemma M.1. Intuitively, Lemma M.1 states that we do not have to re-compute the likelihood of potential deviations from the number of banks that are not observed, which speeds up computation significantly. In other words, we can decompose the competition structure and its results from the demand side data.

Lemma M.1. *Let ζ be the set of structural parameters and $\mathcal{L}$ to be the log-likelihood of supply and demand data for province p, denoted by, respectively, $\mathfrak{S}_p$, and $\mathfrak{D}_p$. Let P be the provinces we observe both, and $\overline{P}$ the set we observe only $\mathfrak{S}_P$. Then*

$$\mathcal{L}\left(\zeta|\mathfrak{S}_{p\in\overline{P}},\{\mathfrak{D}_p\}_{p\in P}\right)=\sum_{p\in\overline{P}}\ln\left[\mathbb{P}\left(\mathfrak{S}_p|\zeta\right)\right]+\sum_{p\in P}\ln\left[\mathbb{P}\left(\mathfrak{D}_p|\zeta,\mathfrak{S}_p\right)\right] \tag{158}$$

Proof. See Appendix M.2 ■

Moreover, Lemma M.1 is useful in our numerical method, which we explore in Section M.1.4. The idea is that there is a subset of structural parameters ζ that is relevant for supply and one that is relevant for demand. Some parameters, like spatial costs ψ, are important for both. The measurement error variance γ_{ME}, on the other hand, determines only demand-side likelihood. This will be valuable when we discuss the empirical method in Section M.1.4.

M.1.4 Numerical Method

We now explore the numerical method given the result of Lemma M.1 and Eqs. (153) and (155). The parameters in our model are

$$\zeta\equiv\{\zeta_S,\psi,\sigma_L,\gamma_{ME},c_E,s\} \tag{159}$$

where ζ_S are the parameters that change the frontier. Our objective is to solve the optimization problem in Eq.(160)

$$\max_{\zeta}\sum_{p\in P}\ln\left[\mathbb{P}\left(\mathfrak{S}_p|\zeta\right)\right]+\ln\left[\mathbb{P}\left(\mathfrak{D}_p|\zeta,\mathfrak{S}_p\right)\right] \tag{160}$$

where $\ln\left[\mathbb{P}\left(\mathfrak{S}_p|\zeta\right)\right]$ is defined in Eq.(153) $\ln\left[\mathbb{P}\left(\mathfrak{D}_p|\zeta,\mathfrak{S}_p\right)\right]$ from Eq.(155). We could simply numerically solve the problem above in Eq.(160). However, Eq.(160) has several characteristics that allow for a more efficient solution. First, not all parameters are entered in both terms. Moreover, we can separate the optimization problem into two parts - for any values of ζ_S,ψ,σ_L, we solve for the optimal values of the optimal parameters and then optimize over ζ_S,ψ,σ_L. In particular, we have that the problem in Eq.(160) is equivalent to Eq.(161).

$$\max_{\zeta_S,\psi,\sigma_L}\left(\max_{c_E,s}\sum_{p\in P}\ln\left[\mathbb{P}\left(\mathfrak{S}_p|\{\zeta_S,\psi,\sigma_L,c_E,s\}\right)\right]+\max_{\gamma_{ME}}\sum_{p\in P}\ln\left[\mathbb{P}\left(\mathfrak{D}_p|\{\zeta_S,\psi,\sigma_L,\gamma_{ME}\},\mathfrak{S}_p\right)\right]\right) \tag{161}$$

First, we explore how to estimate γ_{ME},c_E and s given ζ_S,ψ and σ_L, that is, the inner maximization problems in Eq.(161). Second, we discuss how to estimate ψ and σ_L and show that it is numerically identified from Monte-Carlo experiments. We do not focus on the estimation of ζ_S in this paper, given that our innovation is on the market structure side given contracting. We do discuss the estimation and identification of ζ_S in Appendix N. So, we simplify the notation and exclude the dependence from it.

Estimating $\{c_E, s, \gamma_{ME}\}$. For now, assume that ψ and σ_L are as if known. We show how to solve inner maximization problems in Eq.(161) given these values.

Note that the variance of the measurement error, denoted by γ_{ME}, does not affect (i) the frontier and (ii) Equilibrium utilities, and (iii) the likelihood of the supply side in Eq.(153). It only changes the likelihood at the computation of F_v in Eq. (154). This does not mean that in maximizing the likelihood there is no interaction between the parameters. Still, it means that *given* ψ, σ_L and ζ_S we can easily compute the estimator $\hat{\gamma}_{ME}$, the argmax of Eq.(162) as a function of ψ and σ_L, without having to recompute the frontier or the equilibrium.

$$\hat{\gamma}_{ME} \equiv \arg\max_{\gamma_{ME}} \sum_{p \in P} \ln\left[\mathbb{P}\left(\mathfrak{D}_p | \{\gamma_{ME}, \psi, \sigma_L\}, \mathfrak{S}_p\right)\right] \tag{162}$$

Furthermore, note that the entry cost, c_E, and standard deviation of location-specific shocks, s, do not affect (i) the frontier and (ii) Equilibrium utilities, and (iii) the likelihood of the demand side in Eq.(153). Thus, we can easily compute the estimators $\hat{c}_E, \hat{s}$, the argmax of Eq.(163) as a function of ψ and σ_L .

$$(\hat{c}_E, \hat{s}) \equiv \arg\max_{c_E, s} \sum_{p \in P} \ln\left[\mathbb{P}\left(\mathfrak{S}_p | \{c_E, s, \psi, \sigma_L, \zeta_S\}\right)\right] \tag{163}$$

As γ_{ME} is the std. dev. of the normal distribution, F_v in Eq. (154) is differentiable in γ_{ME}. As we know what the analytical derivative is, it is straightforward to compute an optimal value for γ_{ME} using a grid search method, which is computationally fast. See Appendix M.3 for specific equations and details.

Analogously, we can take the FOC of Eq.(153) with respect to c_E, s and easily solve for it through a grid search (although now is a FOC system with two equations). See Appendix M.4 for specific equations and details.

Estimating ψ and σ_L. We have to estimate ψ (spatial cost) and σ_L (logit variance) numerically. For each value used of ψ and σ_L, we use the method above to compute $\{c_E, s, \gamma_{ME}\}$ The numerical optimization of ψ and σ_L is done through a mix of a grid search and the Matlab built-in *patternsearch*. See Appendix M.5 for more details and a pseudo-code.

Numerical Identification. Although we have given indications in Section D that we can identify the parameters from the microdata, we show that it is the case numerically. The intuition is that the overall levels of utility, which imply consumption, capital, and income dynamics, identify σ_L. In contrast, the variation between these levels across villages identifies ψ, as seen in Section D. To validate this intuition, we conduct a Monte-Carlo experiment. We generate model-simulated data and use it to estimate the parameters in question. We use only consumption, production, and capital data in this exercise. [46]. As we are ultimately interested in estimating the spatial costs ψ and logit variance,

[46] As we assume that the observed data corresponds to an equilibrium in terms of bank entry (as in Bresnahan

σ_L, we mainly focus on the maximization of Eq. (155) on these two parameters. The numerical results show that our method identifies $\{\psi, \sigma_L\}$ from the data. The details and results are in Appendix N.

M.2 Proof of Lemma M.1

Proof. Note that

$$\prod_{p\in P}\mathbb{P}\left(\mathfrak{S}_p,\mathfrak{D}_p|\zeta\right)=\prod_{p\in P}\mathbb{P}\left(\mathfrak{S}_p,\mathfrak{D}_p|\zeta\right)\prod_{p\in\overline{P}-P}\mathbb{P}\left(\mathfrak{S}_p,\mathfrak{D}_p|\zeta\right)=\prod_{p\in P}\mathbb{P}\left(\mathfrak{S}_p,\mathfrak{D}_p|\zeta\right)\prod_{p\in\overline{P}-P}\mathbb{P}\left(\mathfrak{S}_p|\zeta\right)\mathbb{P}\left(\mathfrak{D}_p|\mathfrak{S}_p,\zeta\right)$$
$$\propto\prod_{p\in P}\mathbb{P}\left(\mathfrak{S}_p,\mathfrak{D}_p|\zeta\right)\prod_{p\in\overline{P}-P}\mathbb{P}\left(\mathfrak{S}_p|\zeta\right)\prod_{p\in P}\mathbb{P}\left(\mathfrak{D}_p|\mathfrak{S}_p,\zeta\right)\prod_{p\in\overline{P}}\mathbb{P}\left(\mathfrak{S}_p|\zeta\right) \tag{164}$$

Taking logs and re-arranging delivers the expected result.

■

M.3 Estimator for (γ_{ME})

The partial derivative of the likelihood of demand given supply in Eq. (155) to γ_{ME} is given by

$$\begin{aligned}
\frac{\partial\ln\{\mathbb{P}\left(\mathfrak{D}|\mathfrak{S},\zeta\right)\}}{\partial\gamma_{ME}} &= -L\sum_{p}\sum_{v^p}\sum_{j^p}\mathbb{1}_{\hat{v}_j^p=v^p}\,\frac{\sum_{r=1}^{\#Y}f_v(c,q,k|\zeta)\gamma_{ME}^{-L-1}\exp\left\{\sum_{l=1}^{L}-\frac{\left(\hat{y}_j^l-y_r^l\right)^2}{2\chi_l^2\gamma_{ME}^2}\right\}}{F(\hat{y}_j,\zeta)}\\
&+\sum_{p}\sum_{v^p}\sum_{j^p}\mathbb{1}_{\hat{v}_j^p=v^p}\,\frac{\sum_{r=1}^{\#Y}f_v(c,q,k|\zeta)\gamma_{ME}^{-L}\exp\left\{\sum_{l=1}^{L}-\frac{\left(\hat{y}_j^l-y_r^l\right)^2}{2\chi_l^2\gamma_{ME}^2}\right\}\left[\sum_{l=1}^{L}\frac{\left(\hat{y}_j^l-y_r^l\right)^2}{\chi_l^2\gamma_{ME}^3}\right]}{F(\hat{y}_j,\zeta)}\\
&=\gamma_{ME}^{-1}\sum_{p}\sum_{v^p}\sum_{j^p}\mathbb{1}_{\hat{v}_j^p=v^p}\,\mathbb{1}_{j\in v}\,\frac{\sum_{r=1}^{\#Y}f_v(c,q,k|\zeta)\exp\left\{\sum_{l=1}^{L}-\frac{\left(\hat{y}_j^l-y_r^l\right)^2}{2\chi_l^2\gamma_{ME}^2}\right\}\left[\sum_{l=1}^{L}\frac{\left(\hat{y}_j^l-y_r^l\right)^2}{\chi_l^2\gamma_{ME}^2}-L\right]}{\sum_{r=1}^{\#Y}f_v(c,q,k|\zeta)\exp\left\{\sum_{l=1}^{L}-\frac{\left(\hat{y}_j^l-y_r^l\right)^2}{2\chi_l^2\gamma_{ME}^2}\right\}}
\end{aligned} \tag{165}$$

From Eq.(165), we have that

$$\lim_{\gamma_{ME}\to 0}\frac{\partial\ln\{\mathbb{P}\left(\mathfrak{D}|\mathfrak{S},\zeta\right)\}}{\partial\gamma_{ME}}>0\quad\text{and}\quad\lim_{\gamma_{ME}\to 1}\frac{\partial\ln\{\mathbb{P}\left(\mathfrak{D}|\mathfrak{S},\zeta\right)\}}{\partial\gamma_{ME}}<0 \tag{166}$$

---

and Reiss (1991)), we do not know the entry/exit process nor its dynamics, such that it is very challenging to simulate an equilibrium in the position of each bank .

In the optimum, $\hat{\gamma}_{ME}$:

$$\hat{\gamma}_{ME}^2 = L^{-1} \sum_p \sum_{v^p} \sum_{j^p} \mathbb{1}_{\hat{v}_j^p = v^p} \frac{\sum_{r=1}^{\#Y} f_v(c,q,k|\zeta) \exp\left\{\sum_{l=1}^{L} -\frac{\left(\hat{y}_j^l - y_r^l\right)^2}{2\chi_l^2 \hat{\gamma}_{ME}^2}\right\}\left[\sum_{l=1}^{L} \frac{\left(\hat{y}_j^l - y_r^l\right)^2}{\chi_l^2}\right]}{\sum_{r=1}^{\#Y} f_v(c,q,k|\zeta) \exp\left\{\sum_{l=1}^{L} -\frac{\left(\hat{y}_j^l - y_r^l\right)^2}{2\chi_l^2 \hat{\gamma}_{ME}^2}\right\}} \tag{167}$$

We can re-write it as

$$\sum_p \sum_{v^p} \sum_{j^p} \mathbb{1}_{\hat{v}_j^p = v^p} \frac{\sum_{r=1}^{\#Y} f_v(c,q,k|\zeta) \exp\left\{\sum_{l=1}^{L} -\frac{\left(\hat{y}_j^l - y_r^l\right)^2}{2\chi_l^2 \hat{\gamma}_{ME}^2}\right\}\left[\sum_{l=1}^{L} \frac{\left(\hat{y}_j^l - y_r^l\right)^2}{\chi_l^2 \hat{\gamma}_{ME}^2}\right]}{\sum_{r=1}^{\#Y} f_v(c,q,k|\zeta) \exp\left\{\sum_{l=1}^{L} -\frac{\left(\hat{y}_j^l - y_r^l\right)^2}{2\chi_l^2 \hat{\gamma}_{ME}^2}\right\}} = 1 + L \tag{168}$$

the LHS of Eq.(168) is constant and the RHS is a weighted average. When γ_{ME} increases, we increase the relative weight of high $\left(\hat{y}_j^l - y_r^l\right)^2$ terms and decrease all terms. Therefore, it is not trivial to state if the LHS is decreasing or increasing. Therefore, there is no general proof that the function is concave [47]. Still, we know from Eq. (166) that a zero partial derivative is a necessary condition, which translates to $\hat{\gamma}_{ME}$ satisfying Eq. (167).

M.4 Estimators for c_E, s

As $\{c_E, s\}$ are the mean and std. dev. of the normal distributions, $\ln\{\mathbb{P}(\mathfrak{S}|\zeta)\}$ in Eq. (153) is differentiable in $\{c_E, s\}$. As we know its analytical derivative, it is straightforward to compute an optimal value for $\{c_E, s\}$ using a grid search method, which is computationally fast. In particular, analogous to what did in Section M.3, $\{\hat{c}_E, \hat{s}\}$ are the solution to the non-linear system in Eq. (169)-(170) (as we show later on).

$$\sum_{m^p} \frac{\phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{\hat{s}}\right] - \phi\left[\frac{\Pi^E(B_{m^p}|.)}{\hat{s}}\right]}{\Phi\left[\frac{\Pi^E(B_{m^p}|.)}{\hat{s}}\right] - \Phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{\hat{s}}\right]} = 0 \tag{169}$$

$$\sum_{m^p} \frac{\phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{\hat{s}}\right]\Pi^E\left(B_{m^p}+1|.\right) - \phi\left[\frac{\Pi^E(B_{m^p}|.)}{\hat{s}}\right]\Pi^E\left(B_{m^p}|.\right)}{\Phi\left[\frac{\Pi^E(B_{m^p}|.)}{\hat{s}}\right] - \Phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{\hat{s}}\right]} = 0 \tag{170}$$

where the dependence of the system of c_E comes implicitly from its effect on profits, i.e.: profit = revenue $- c_E$ in Eq. (32). We now move to show that Eq. (169)-(170) determine the optimal value of $\{\hat{c}_E, \hat{s}\}$.

Fixed Cost. the partial derivative of the supply likelihood (Eq. 153) w.r.t. to the fixed cost c_E is given by Eq. (171).

[47] In all numerical runs, the likelihood was concave in γ_{ME}. As shown above, it is not trivial to guarantee this analytically

$$\frac{\partial \ln\{\mathbb{P}(\mathfrak{S}|\zeta)\}}{\partial c_E} = s^{-1}\sum_{m^p} \frac{\phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{s}\right]-\phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]}{\Phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]-\Phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{s}\right]} \tag{171}$$

Note that at $c_E \leq 0$, as we have that the profit Π^E is non-increasing in the number of intermediaries in a location m^p and $\Pi^E \geq 0$ (since there is no cost and the state space is limited to points with positive frontier without loss of generality), we have that:

$$\phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right] \leq \phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{s}\right] \tag{172}$$

And:

$$\phi\left[\frac{\Pi^E(1|.)}{s}\right] > 0 \tag{173}$$

Together, (172)-(173) imply that (171) is positive at $c_E \leq 0$, i.e.:

$$\left.\frac{\partial \ln\{\mathbb{P}(\mathfrak{S}|\zeta)\}}{\partial c_E}\right|_{c_E\leq 0} > 0 \tag{174}$$

Moreover, at $c_E \to \infty$:

$$\begin{aligned}
\sum_{m^p}\mathbb{1}_{m^p>0}\lim_{c_E\to\infty}\frac{\phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{s}\right]-\phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]}{\Phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]-\Phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{s}\right]} &\underset{L'Hospital}{=} \sum_{m^p}\mathbb{1}_{m^p>0}\lim_{c_E\to\infty}\frac{\phi'\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]-\phi'\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]}{\phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]-\phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]} \\
&= \sum_{m^p}\mathbb{1}_{m^p>0}\lim_{c_E\to\infty}\frac{\phi'\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]-\phi'\left[\frac{\Pi^E(B_{m^p}+1|.)}{s}\right]}{\phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{s}\right]-\phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]} \\
&= \sum_{m^p}\mathbb{1}_{m^p>0}s^{-1}\lim_{c_E\to\infty}\frac{\Pi^E(B_{m^p}+1|.)\,\phi\left[\frac{\Pi^{BR}(N_m+1|.)}{s}\right]-\Pi^E(B_{m^p}|.)\,\phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]}{\phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{s}\right]-\phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]} \\
&= -\infty
\end{aligned} \tag{175}$$

and:

$$\lim_{c_E\to\infty}\frac{\phi\left[\frac{\Pi^E(1|.)}{s}\right]}{1-\Phi\left[\frac{\Pi^E(1|.)}{s}\right]} = 0 \tag{176}$$

Therefore, it must be the case that:

$$\partial_{c_E}\left.\frac{\partial \ln\{\mathbb{P}(\mathfrak{S}|\zeta)\}}{\partial c_E}\right|_{\hat{c}_E} = 0 \tag{177}$$

where $\hat{c}_E$ is the argmax of the likelihood given the data (i.e., the estimator). The idea here is that as the function is differentiable, increasing at zero and decreasing at $c_E \to \infty$, there is a global max, and it must satisfy the necessary condition in Eq. (177).

Standard Deviation s. the partial derivative of the supply likelihood (Eq. 153) w.r.t. to the variance of location specific profit shocks s is given by Eq. (178).

$$\frac{\partial \ln\{\mathbb{P}(\mathfrak{S}|\zeta)\}}{\partial s} = s^{-2}\sum_{m^p} \mathbb{1}_{m^p>0} \frac{\phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{s}\right]\Pi^E(B_{m^p}+1|.) - \phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right]\Pi^E(B_{m^p}|.)}{\Phi\left[\frac{\Pi^E(B_{m^p}|.)}{s}\right] - \Phi\left[\frac{\Pi^E(B_{m^p}+1|.)}{s}\right]}$$
$$+ s^{-2}\sum_{m^p} \mathbb{1}_{m^p=0} \frac{\phi\left[\frac{\Pi^E(1|.)}{s}\right]\Pi^{BR}(1|.)}{1-\Phi\left[\frac{\Pi^E(1|.)}{s}\right]} \tag{178}$$

Using the same arguments as in Eq. (175), one can show that

$$\lim_{s\to 0} \frac{\partial \ln\{\mathbb{P}(\mathfrak{S}|\zeta)\}}{\partial s} > 0 \text{ and } \lim_{s\to\infty} \frac{\partial \ln\{\mathbb{P}(\mathfrak{S}|\zeta)\}}{\partial s} < 0 \tag{179}$$

Therefore, it must be the case that:

$$\left.\frac{\partial \ln\{\mathbb{P}(\mathfrak{S}|\zeta)\}}{\partial s}\right|_{\hat{s}} = 0 \tag{180}$$

where $\hat{s}$ is the argmax of the likelihood given the data (i.e., the estimator). The idea here is that as the function is differentiable, increasing at zero and decreasing at $s \to \infty$, there is a global max, and it must satisfy the necessary condition in Eq. (180).

M.5 Details on Numerical Maximization

First, we discuss the pseudo-code we use for numerical maximization. We discuss first how we compute the likelihood for fixed values of $\{\psi, \sigma_L\}$.

One likelihood computation. Given $\{\psi, \sigma_L\}$.

1. As a function of $\{\sigma, \theta\}$ only, compute the frontier of Section 3.1, S .
   - Don't have to redo this step if we are calibrating $\{\sigma, \theta\}$ instead of estimating it (as we are here).
   - Use the LP formulation with the *gurobi* [48] linear solver.
2. Given Step 1 (the frontier), compute the equilibrium utilities and the resulting contracts.
   - uses iterative procedure based on supermodularity of Lemma 3.1.
3. Likelihood:

[48] Available for free for academic use at:

- Demand Given Supply. Given Steps 1 and 2, compute the likelihood of the demand using the adaptation of Karaivanov and Townsend (2014) method presented (Eq. 155).
  - Use a grid search to find $\hat{\gamma}_{ME}$ that satisfies Eq. (167) and compute Eq. (155) already at the optimum $\hat{\gamma}_{ME}$.
- Supply. Given Steps 1 and 2 - i.e., it can be computed parallel to the demand - compute the likelihood of the supply using the entry model and the normality assumption (Eq. 153).
  - Use a grid search to find $\{\hat{c}_E, \hat{s}\}$ that satisfie Eqs.(177) -(180) and compute Eq. (153) already at the optimum $\{\hat{c}_E, \hat{s}\}$.
- Sum Demand Given Supply and Supply. as in Eq.()158).

Global Optimum. We optimize over ψ, σ_L by first doing a grid search and then using the *patternsearch* command in Matlab from the optimal point in the grid search. We guess $\psi = 1$, $\sigma_L = .33$, and use 25 point grids between .1 and 5 times the original values for both.

N Identification

N.1 Numerical Identification

Although we have given indications in Section D that we can identify the parameters from the micro-data, we now show that it is the case numerically. For that, we conduct a Monte-Carlo experiment. We generate model-simulated data and use it to estimate the parameters in question. We use only consumption, production, and capital data in this exercise. As we assume that the observed data corresponds to an equilibrium in terms of bank entry (as in Bresnahan and Reiss (1991)), our methodology makes it much harder to discuss identification from it. [49]. As we are ultimately interested in estimating the spatial costs ψ and logit variance, σ_L, we focus on this section on the maximization of Eq. (155) on these two parameters but present results for risk aversion σ and cost of exerting effort φ.

To be closer to our actual application, we use a spatial configuration here with two dimensions, in a 'Manhattan' style as in Figure 27. We assume each intersection has a random number of FSPs (which can be zero), and villages are uniformly distributed within roads. [50]

We simulate the data for four provinces, with 10 villages in each road, each with $N = 75$ households each, with the same parameters of the comparative statics exercises (Section D). [51]. In particular, a pseudo-code for this identification experiment is as follows:

Pseudo-Code for identification:

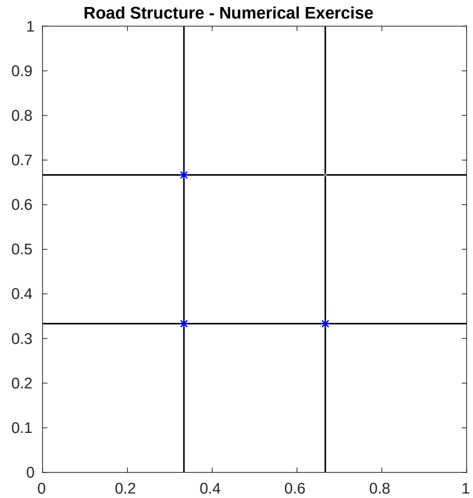
1. Draw a number of banks for each position as in Figure (27)).

[49] We are assuming that we do not know the entry/exit process and its dynamics, such that it is very challenging to simulate an equilibrium in the position of each bank.

[50] We have the equivalent results using the actual spatial configuration in the Townsend Thai Data upon request. We prefer to showcase this version since we can understand better the dynamics of competition and contracting .

[51] Frontier parameters given by Table 4 and the true value of ψ and σ_L as in Table 5

Figure 27: Simplified Map 'Manhattan' Style



A circular dot in one of the intersections means that there is one bank there, and a star means that there are two. Villages are uniformly distributed throughout the roads.

2. Given the map configuration of Step 1, solve for the equilibrium in utilities and compute the optimal contracts.

3. Produce probability distributions of (c, q, k) assuming that the distribution of k in the simulated data is the same as in the sample.

4. Draw N observations from Step 3 and then add the measurement errors to (c, q, k) obtained (belonging to the grid) to generate the dataset.

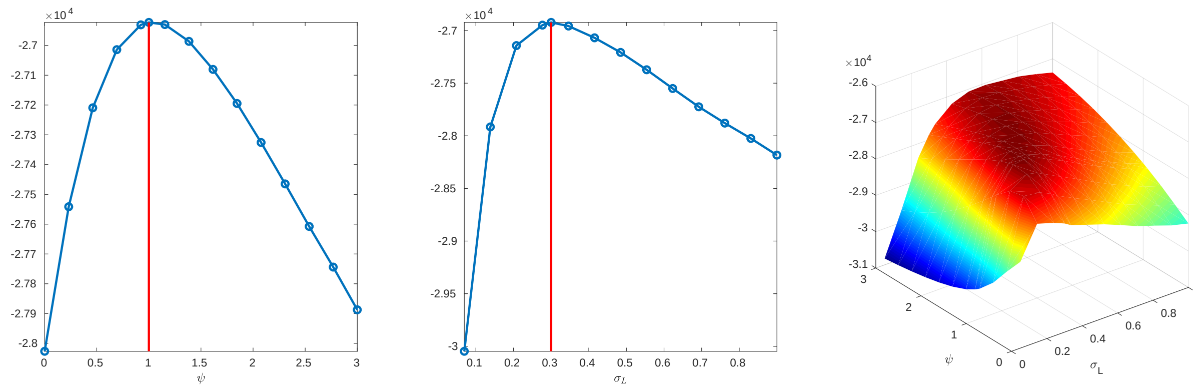
5. Fix all the other parameters and vary the parameters we plot.

Results are in Figure 28, where we can see our method successfully identifies the market structure parameters: spatial cost ψ and logit variance σ_L. We first vary one parameter at a time (panels (a) and (b)) and then both parameters jointly. The overall levels of utility, which imply consumption, capital, and income dynamics, identify σ_L, while the variation between these levels across villages identifies ψ, as seen in Section D. Note that although ψ and σ_L are jointly identified, the higher differences from the likelihood appear for different ratios of ψ/σ_L, which is the overall spatial cost in terms of utilities.

We repeat the experiment above for risk aversion σ and disutility of effort φ. The results are in Figure 29. Again, our results show that the model successfully identified this parameter. To identify them, the model uses the joint variation in consumption/income and the implicit distribution of all variables given the utilities and effort levels implied by competition.

Figure 28: Log-Likelihood of Household Level Data as a Function of Spatial Cost (ψ) and Logit Var (σ_L) for Simulated data

(a) Spatial Cost, ψ (b) Logit Var, σ_L (c) Joint



Note: Likelihood of household level data (Eq. 155). The red line is the true value. The dotted blue line is the log-likelihood. The map here is 'Manhattan' style (Figure 27). Data was simulated for four provinces, with 10 villages on each road, each with $N = 75$ households. Frontier parameters are given by Table 4 and the true value of ψ and σ_L as in Table 5.

O From Utilities to Consumption

Consumption and Distance. With the utility given by Eq.(16), we want to solve for Δ in Eq.(181). The value Δ is the % growth in consumption corresponding to moving from zero to the median distance of intermediaries and villages in the sample. In this case:

$$u((1+\Delta)c) - z^{\varphi} - \psi \, \text{med}\left(t(x_v, x_b)\right) = u(c) - -z^{\varphi} \tag{181}$$

Eq.(181) implies Eq.(182)

$$\left\{(1+\Delta)^{1-\sigma} - 1\right\} u(c) = \psi \, \text{med}\left(t(x_v, x_b)\right) \tag{182}$$

and Eq.(182) yields:

$$\Delta = \left[\psi \, \frac{\text{med}\left(t(x_v, x_b)\right)}{u(c)} + 1\right]^{\frac{1}{1-\sigma}} - 1 \tag{183}$$

which delivers Eq. (38).

Utilities and Consumption. With the utility given by Eq.(16), we now want to solve for Δ in Eq.(183). The value $\tilde{\Delta}$ is the % growth in consumption that corresponds to moving Δu. In this case:

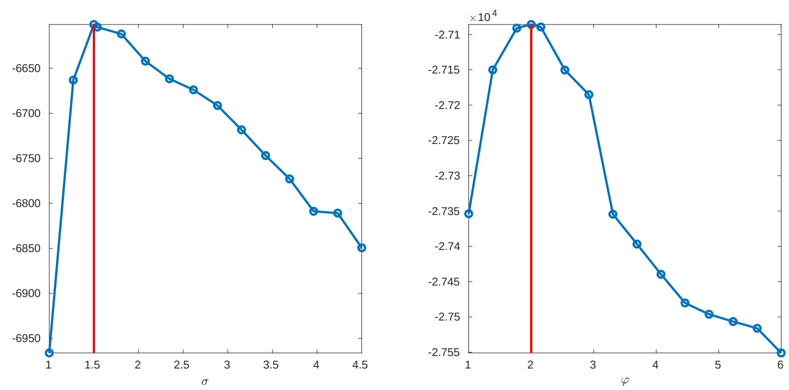
$$\Delta u = u((1+\tilde{\Delta})c) - z^{\varphi} - \psi t(x_v, x_b) - [u(c) - z^{\varphi} - \psi t(x_v, x_b)] = (1-\sigma)^{-1}\left[(1+\tilde{\Delta})^{1-\sigma} - 1\right] c^{1-\sigma} \tag{184}$$

For $\sigma = 2$

$$\tilde{\Delta} = \left[\frac{\Delta u}{u(c)} + 1\right]^{\frac{1}{1-\sigma}} - 1 \tag{185}$$

Figure 29: Log-Likelihood of Household Level Data as a Function of Risk Aversion (σ) and Disutility of Effort (φ) for Simulated data

(a) Risk Aversion (σ) (b) Disutility of Effort (φ)



Note: Likelihood of household level data (Eq. 155). The red line is the true value. The dotted blue line is the log-likelihood. The map here is 'Manhattan' style (Figure 27). Data was simulated for four provinces, with 10 villages on each road, each with $N = 75$ households. Frontier parameters are given by Table 4 and the true value of ψ and σ_L as in Table 5.