Spatial Competition among Financial Service Providers and Optimal Contract Design*

Robert M. Townsend and Victor V. Zhorin

December 28, 2014

Abstract

We present a contract-based model of industrial organization that allows us to consider in a unified way both different information frictions (moral hazard, adverse selection, both) and a variety of market structures (monopoly, imperfect competition, various strategic interactions). We show how this method can be applied to the spread of the banking industry in emerging market countries, emphasizing observed transitions, namely the geographic locations of branches. Local collusive monopoly organizations and Bertrand-like competitive environments in location and utility space are considered alongside with frictions affecting the outcome, namely provincial spatial costs and the information structure. Mixed environments with fully informed local incumbents and entrants facing adverse selection are analyzed. Our larger goal, beyond calibrated numerical examples, is to develop a framework with an operational toolkit for empirical work.

1 Introduction

This paper provides an industrial organization, contract theoretic, model of the spread of financial service providers in emerging market countries, focusing on the geography of branch expansion as well as on the terms of the actual loan/insurance contracts which are offered. The environment features heterogeneous households running running small and medium enterprises (SME), on their own lacking credit and insurance instruments and with a nontrivial spatial costs of traveling to the location of bank branches, as in a Hotelling (1929), D'Aspremont et al. (1979), Prescott and Visscher (1977)[1]. A priori we place no restrictions on the location of branches nor the contracts which can be offered, though we allow for and compare outcomes as we vary

*We thank Varadarajan Chari, Glenn Ellison, Amy Finkelstein, Thomas Holmes, Ariel Pakes, Christopher Phelan, Michael Whinston, seminar participants at Minneapolis Fed, MIT, Harvard, University of California/Berkeley-Haas for very useful comments. We gratefully acknowledge research support from the National Institute of Child Health and Human Development (NICHD), the research initiative "Private Enterprise Development in Low-Income Countries" (funded jointly by the CEPR and the DFID), the John Templeton Foundation, and the CFSP at the University of Chicago (funded by Bill & Melinda Gates Foundation).

[1]Agarwal and Hauswald (2010) study the effects of physical distance on the acquisition and use of private information in credit markets. Rajan and Petersen (2002) document that the distance between small firms and lenders is increasing. Alessandrini et al. (2009b) show show that greater functional distance stiffened financing constraints, especially for small firms. Butler (2008) suggests that investment banks with a local presence are

financial information regimes. Here we also drawing on a theory literature using promised utilities as a key state variable (Green (1987), Spear and Srivastava (1987)), and also building on the computation methods of Prescott and Townsend (1984), Karaivanov and Townsend (2014). The financial regimes we consider: full information (complete insurance/perfect credit); unobserved effort (moral hazard) but with complete information on types; unobserved types hence adverse selection but with perfect information and contract-defined effort schedules; and finally combinations of adverse selection and moral hazard. Our methods in this paper allow for multiple types with combinations of what is observed or unobserved by the banks and discrimination across types (unconstrained or constrained).

We can allow for comparison imposed, a priori contracts (incomplete markets), so that we can more fully explore existing regulations and make welfare comparisons. But we choose to emphasize the endogenous, unrestricted contracts. Indeed, for that matter, the actual structure of observed bank contracts (credit and insurance arrangements for households and SME) is not simple, i.e. it does not fit the stylized contracts of theory, of borrowing at interest with collateral and fixed-term payments, with presumed repayment but allowing for default. Instead, typical contracts offered by banks represent a blend of credit and insurance, e.g, loans are rolled over, some interest is forgiven, and indeed there are well known and explicit contingencies under which an effective indemnity is paid and some or all of principal is written off (as if paid with the indemnity).

Our motivation for this research is both positive and normative. On the positive side we seek to understand better the industrial organization of financial service providers in terms of both the geography of branches and expansion over time as well as in terms of the actual loan/insurance contracts which are offered. On the normative side, we seek to answers policy questions such as the coexistence of local and national banks and the role of information and competition ((Petersen and Rajan (1995)); the impact of deregulation which alleviates artificial geographic or policy/segmentation boundaries ((Brook et al. (1998), Demyanyuk et al. (2007)); and the welfare and distributional consequences of dfferent market structures, different obstacles to trade (information, trade costs) (Koijen and Yogo (2012), Martin and Taddei (2012))), and the interaction of these obstacles with market structure. Our goal is to develop a toolkit to answer those questions.

Picking a particular provincial market with specified spatial costs, picking a particular financial information regime, and specifying parameters of utility functions and production technology (consistent with those estimated in other studies), the various model of market structure considered make predictions about contracts which are offered; the implications for households/firms in terms effort, borrowing, capitalization, output, and the level and variation of consumption with output; and the location, market share, and profits of the banks by branch. We can then do various comparative static exercises. The principle one we feature is variation in the spatial cost as we move from province to province, with observed variation in spatial costs, as measured for example with a geographical information system GIS which computes travel times on road networks from each village to each branch; see Assuncao et al. (2012).

better able to assess private information and place difficult bond issues. Degryse and Ongena (2005) report the comprehensive evidence on the occurrence of spatial price discrimination in bank lending. Alessandrini et al. (2009a) show that small and medium enterprises (SMEs) located in provinces where the local banking system is functionally distant are less inclined to introduce process and product innovations, while the market share of large banks is only slightly correlated with firms propensity to introduce new products.

Though we can allow for multiple, endogenous number of banks each with multiple branches, by specifying costs of entry and branch establishment, instead we simplify and provide illustrative, key examples with two bank each with one branch. That said, observed provincial markets can be quite thin in the economies and time periods for which we have observations: Thailand, Brazil, Bangladesh.

Our first market structure imagines the banks are implicitly colluding with each other (explicitly a two branch monopoly). We compare the full information and moral hazard regime for the case of one homogenous type. The basic building block or primitive of the models here and below is the profit/surplus of the branch per customer as a function of utility promised to the household/firm (not allowing spatial discrimination). The profit maximizing contract for a given spatial cost as utility is varied satisfies a familiar elasticity condition: the elasticity of the profit/surplus frontier with respect to utility, on the intensive margin, equals the elasticity of increased market share on the extensive margin; the tradeoff between lower profits per customer and more customers is balanced and determines a particular operational point on the frontier. But since these surplus frontiers vary with information regime, with the frontier of the moral hazard regime having a lower level but also a lower slope, there are corresponding implications for comparative statics. Throughout the range of spatial costs, profits are lower and market shares are higher in the moral hazard regime relative to the full information regime. Further, as spatial costs move from low to high, labor effort, capitalization/borrowing and expected output are monotonically decreasing. But finally, and a bit more novel, the moral hazard regime initially specifies higher and then lower levels of these variables, with the crossing point varying by the particular variable considered. Though expected and actual consumption are constant in the full information regime, the degree of insurance (the variation of consumption with output) varies in the moral hazard regime, naturally, but varies with the provincial spatial costs, depending on inducement of low or higher levels of effort.

We then consider a competitive Nash equilibrium in contracts, keeping the branches of the two banks separated at the same locations as the two branch monopoly, but allowing heterogeneity in types and comparing the range of financial/information regimes. As spatial costs increase across provinces, profits first increase and then decrease, and utilities of households/firms decrease and then increase. Low spatial costs imply intense competition and hence low profit. At intermediate costs, banks struggle to retain customers in their respective hinterlands. At high spatial costs active market segments do not overlap, and we move toward and obtain local branch monopolies. Yet the transition points are different across the different information regimes. In sum, though higher spatial costs is a worse physical environment over all, competing banks can actually gain from this in certain ranges, and the extent of this gain depends on the information structure. Related effort, capitalization (borrowing), and expected output are now each non-monotonic with spatial costs, that is, rise and then fall as spatial costs increase. Further, the range and peaks of nontrivial values are different for different variables and different types, and again vary with the financial information regime. For example, capital is most responsive for low types in the adverse selection plus moral hazard regime but most responsive for the high type in the moral hazard regime.

An interesting policy experiment compares the two-branch monopoly to this competitive outcome. Here we illustrate by comparing the full information regime to the adverse selection regime, again as we move across provinces that vary in spatial costs. At relatively low spatial costs the switch from monopoly to competition naturally increases the household utility, but with some twists. With full information the biggest gain is for the risky type. With adverse selection it is much harder to distinguish across types, so the overall gain from liberalization/competition is similar for both types. Plus safe type gains more from liberalization in the adverse selection regime than in the full information regime. At yet higher spatial costs there is no gain for either type, though the transition point from gain to no gain happens earlier, at lower spatial costs in the adverse selection regime.

More generally, we make endogenous not only contracts but also locations, and then the timing of entry begins to matter. Here we can distinguish various degrees of commitment or inflexibility. For example, suppose the first entrant into a province chooses a branch location in a first stage, before a second entrant does so in a second stage, and before both compete on contracts in a third stage. In this setting and at relatively low spatial costs, the first bank has an incumbency, first-mover advantage, picking a more central location and being able to set lower utility, thus gaining higher market share and higher profits. However, for provinces with intermediate and higher spatial costs, this first-mover advantage can be lost sometimes, but not always. Again it is non-monotonic. Further, and related, locations, utilities, market shares, and profits are all non-monotonic in spatial costs. The point is that there is little that is straightforward about moving toward economically advantageous environments with lower spatial costs, as with the emergence of electronic rather than brick and mortar banking.

When the first entrant has to commit not only to its location but also to its contracts, to its complete business model, then the second entrant has an inherent advantage in choosing its contracts to attract customers[2].

In our last setting we postulate that the first-mover has an information advantage, knowing firm types whereas the second entrant suffers an adverse selection problem[3]. Here even without logits we get salient and consistent patterns. At a low range of spatial costs, the informed incumbent ends up dealing exclusively with the safer type, and the entrant with the risky type. That said, the incumbent, anticipating the ultimate competition, has moved toward the hinterland and makes lower profits in the end. For provinces with intermediate spatial costs, there is less specialization, but now it is the incumbent taking more of the risky types. At yet higher spatial costs the two banks begin to mirror each other; the incumbent information advantage disappears.

Our larger goal in this paper is to allow us and other researchers to move beyond these calibrated numerical examples and to develop a tool kit, an operational empirical framework. That is, our ultimate goal is to do for industrial organization and contract theory with transition dynamics what Doraszelski and Pakes (2007) did for industrial organization and steady state dynamics. So we do not shy away from reporting what we know about computation.

As has been noted implicitly, we are gaining empirical identification by the assumed observability of spatial costs and the fact that these costs enter into one side of equilibrium elasticity conditions on the extensive margin side, that is how quickly a monopoly can pick up customers when increasing utility offers. Again, suppose the financial information regime and the parameters of utility functions and production technologies are fixed. This fixes the Pareto/surplus frontier, the tradeoff between profits and utilities. Then, as spatial costs increase, this demand

[2]Here the outcomes are not only non-monotonic but can be discontinuous; this is not a numerical computation issue but rather has to do with the underlying economics of competition. Still, for display purposes, we smooth things out both with logits in the model and here imagine imperfectly-observed spatial costs.

[3]We do not model the explicit dynamics that would generate this but take it as given.

side moves, and the equilibrium is in effect tracing out points along the surplus frontier, decreasing profit and increasing utility. Of course as utility varies, so does the underlying contract. This logic is less transparent when there is competition, though Armstrong and Vickers (2001) derive an analytic solution for what the equilibrium is maximizing when there are logistic demands and zero spatial costs. Of course when we also allow endogenous branching and varying degrees of commitment, we loose these analytic results, but the intuition of identification remains. That is, we still expect solutions to vary systematically as we move across provinces with varying spatial costs, as we have seen in the above calibrated examples. The comparative statics tell us that there are restrictions on data.

A key tool for us, given our determination to move beyond fixed contracts is our ability to compute solutions for arbitrary information regimes (for the moment, conceptually, still holding parameters fixed). The contracting problem is one of maximizing surplus, that is, maximizing profit subject to a fixed utility of the household/firm for each type, subject to technological, incentive (moral hazard) and selection (truth telling) conditions. Approximating the space of underlying choice variables by a fine grid over the continuous probability distribution, this contracting problem becomes a linear programming problem, granted with a potentially large number of variables and constraints, but computable with a high degree of reliability even for strongly non-linear information problems. Then varying promised utility for the household/firm as in a cross section, one gets implications from the model for what should be observed in cross-sectional data: histograms of lending/capitalization, output, consumption. Indeed, taking a partial equilibrium approach and parameterizing that unobserved distribution of promised utility as a mixture of normal distributions, one can now allow those and the other underlying parameters of preferences and technology to vary and estimate them. Indeed, ones does not have to restrict the model environment to pre-specified structural equations for preferences and technology. Using maximum likelihood to estimate parameters for a given financial/information regime, and Vuong test to choose the financial regime that best fits the data, the fundamental underlying obstacles to trade can be investigates. Though we do not have an analytic, non parametric identification argument, extensive Monte Carlo simulations show that true parameters and the actual financial regime can be recovered well[4].

The current paper makes endogenous the distribution of promised utilities faced by households/firms as an outcome along with branch locations, profits, and market shares from the equilibrium industrial organization of financial service providers. The point is that the entire structure can thus be estimated with full information maximum likelihood techniques, comparing both information regimes and market structures varying with the degree of commitment.

We make clear throughout the paper what we know about existence of equilibria, multiplicity, the continuity of comparative static exercises, and when one can run into problems. Existence and comparative statics are straightforward for the monopoly. With fixed locations and no commitment as well as for full commitment sequential entry regimes, one may need to restrict ranges of variables, parameters values, and make assumptions on endogenous objects. For example the fixed point theorem of Glicksberg has a sufficient continuity and quasi concavity conditions, and these may be shown to hold at least locally. Those assumptions are standard in literature. In general these are hard problems, as is evident in the Hotelling-D'Aspremont exchange; see also Dasgupta and Maskin (1986). In any event we have developed and use an

[4]subject to ties across regimes depending on the degree of estimated measurement error, sample size, and number of variables being fit, see Karaivanov and Townsend (2014) for actual empirical applications

algorithm that computes the distance to a Nash equilibrium and terminates with an apparent fixed point up of zero distance up to machine accuracy - in effect these are approximate epsilon Nash equilibrium with very small epsilons. We do not display figures that fail this criteria.

We keep our environments relatively simple in order to compute for the full scan of potential spatial costs. We would like to include in the adverse selection environments with more than two types, but maintaining menus of contracts subject to truth telling is challenging, given the hierarchical two step process that we use for two types. In future work we may have to compromise a bit on this dimension and allow a parametric structure with some bunching. On the other hand, we can easily allow more than two branches and many time periods for transition dynamics if we make use of the data; for example Assuncao et al. (2012) restrict location candidates to the actual locations that are eventually filled with branches.

Unlike much if not all of applied empirical literature on selection markets, we draw on a theory literature using promised utilities as state variables. Thus we allow an infinite number of contract characteristics and do our analysis directly in that space of type-dependent constrained utilities. See also Armstrong and Vickers (2001) who also make this point clearly. In particular we do not assume an additively separate cash good, out of which prices are paid. In our credit application, even simple interest rates are not really prices, given default option and the the equity-like amounts repaid under project success. We can also make the point that if the world is like out model, then using observed contracts in an equilibrium as a basis for restricting the contract space can be misleading. The equilibrium is determined by out-of-equilibrium, counterfactual contract possibilities (at least in the absence of regulation).

Our welfare metrics also differ from some, but not all, of the applied literature. Even in oligopoly, financial firms operate on obstacle-constrained Pareto frontiers, without price distortions or loss of social surplus. The equilibrium outcomes on these frontiers, the contracts, and the division of gains between households and firms, are determined by obstacles and the degree and nature of market competition. The two branch monopoly equilibrium and the Bertrand competitive outcome are each Pareto optimal and the comparison between the two is Pareto non-comparable. In particular, we do not employ the welfare metric of a utilitarian social planner but rather respect the information incentive and truth telling constraints and the timing in most competitive market structures. That is, there is no contracting on interim draws of types, so using a utilitarian planner that is averaging over types seems an odd contrast. See Mas-Colell et al. (1995) who forcefully make this point in the context of Akerlof lemons problem, yielding a constrained efficient outcome. On the other hand, dynamic limited commitment issues would create inefficiency. We return to these and other generalizations in the concluding section.

The paper proceeds as follows. Section 2 introduces the basic model of competition in utility space with micro foundations as the common building block defining the contract characteristics space. Section 3 develops and compares the models of market competition, including single-branch monopoly, two-branch monopoly or colluding banks, and finally, no commitment, partial commitment and full commitment Nash competition with choice of location and contracts. Section 4 presents results on competition under asymmetric information, in particular, for relationship banking trying to hold its ground against the global uninformed challenger. In Section 5 we discuss dynamic contract considerations and potential extensions of our framework.

2 Hotelling competition in utility and location space

2.1 Demand side

The agents are distributed spatially inside a unit ball (for simplicity here we limit to linear competition case in $\mathbb{R}^1 : [0;1]$[5]. Total market mass is set to one for all types. The densities $p(\theta)$ of agents with type θ, though explicit in this section, are assumed in the subsequent sections to be equal and spatially independent. Location x of the agent defines the cost to access financial services at location x_i of an intermediary. Let's assume that cost is linear $\overline{L} * |x - x_i|$ with a scale factor $\overline{L}$[6]. Different local environments can have specific values of cost parameter $\overline{L}$, representing a variety of regional or historical features. We imagine we see the entire cross section of economies and, for now, that cost $\overline{L}$ is fully observable. We thus look at solutions as $\overline{L}$ is varied for the entire range.

The agents with type θ at location x could potentially choose to go to bank i if the offer of utility from bank i at least satisfies the participation constraint

$$u_{i\theta} - \overline{L} * |x - x_i| \geq \hat{u}_{0\theta}$$

where $\hat{u}_{0\theta}$ is reservation utility from agent's problem solution (autarky value).

The agents at location x choose to go to bank i at location x_i if the offer of utility from bank i satisfies both the above participation constraint and bank i local value $V_i(x)$ offers a strong domination constraint against local value offered by bank i'

$$V_i(x) = u_{i\theta} - \overline{L} * |x - x_i| > u_{i'\theta} - \overline{L} * |x - x_{i'}| = V_{i'}(x)$$

when contract utility $u_{i'\theta}$ is posted by bank i'. With equal value offered at the point x the market is split equally between banks i and i'.

The spatial distribution of agents can be thought of literally in the geographical sense as in our preferred application. However, one can also imagine that households have preferences for a certain bank, following a smooth distribution as in discrete choice. Indeed, we add a logit extreme value distribution in section 3.5, in addition to the separation in space. The agents can also have spatially different characteristics. All we have to do in such a case is to integrate over densities and is easily incorporated.

2.2 Micro foundations: contracts for household enterprise

We consider an economy populated by spatially distributed SME's, that is output-producing agents, and by financial intermediaries. By default all agents are in an autarky regime, as a reservation strategy. They produce a single commodity output using a stochastic technology and that output is fully consumed by the agents.
Agents have preferences

$$u(c, a|\theta),$$

[5]Restriction to finite nodes as potential bank locations is a trivial simplification. Allowing location choices in $\mathbb{R}^2$ with real geographical restrictions is straightforward and feasible.

[6]We also tried other functional specifications for costs, including quadratic and square root, with no qualitative difference in major results.

where c is consumption and a is an action that can be either observable or hidden.

We chose the following standard functional form for utility

$$u(c,a|\theta) = \frac{c^{(1-\sigma(\theta))}}{(1-\sigma(\theta))} + \chi(\theta)\frac{(1-a)^{\gamma(\theta)}}{\gamma(\theta)}$$

Utility is strictly concave in consumption featuring risk aversion $\sigma(\theta)$, hence optimal contracts offer insurance, not simply credit. In principle, we can use any parameterized utility function, such as CARA here. The disutility of labor is specified same way as in Moll et al. (2013). We do not rule out wealth effects. Our formulation allows preference to vary in the population.

The agents also can choose to contract with an intermediary. The bank keeps all surplus left from output q. The output is fully observable, and the contract can be made conditional on output. Intermediated agents can augment their effort by borrowing capital k in standard Cobb-Douglas functional form[7].

There is a set of stochastic production technologies $p(q|a)$ available to agents

$$P(q = high|k,a,\theta) = p(q = high|a)f(\theta)k^{\alpha}; P(q = low|k,a,\theta) = 1 - P(q = high|k,a,\theta) \quad (1)$$

where $P(q|k,a,\theta)$ is a probability to reach the output q that depends on agent's type θ and the effort a exercised by an agent. We can also use non parametrically estimated production function if relevant empirical data is available. Here θ stands for observed (and potentially unobserved) characteristics of the household/SME. Depending on informational frictions we consider, thus either the type or the effort could be unobservable to the intermediary leading to either moral hazard or adverse selection or both. We can allow correlation of θ-types in preferences and in production. More specific interpretation of θ-heterogeneity will be discussed later. Heterogeneity in wealth endowments is possible as well but we don not include that here.

We begin with our basic building block, as if there is one monopoly lender. The optimal contract offered to the agents must be cost minimizing or surplus maximizing for each potential utility offered to agent[8]:

$$S(\overline{u_\theta}) = \max_{\pi(q,c,k,a|\theta)} \left[\sum_{q,c,k,a,\theta} \pi(q,c,k,a|\theta)\,[q - c - k] \right] \quad (2)$$

where $\pi(q,c,k,a|\theta)$ is a probability distribution[9] over the vector (q,c,k,a) given the agent's type θ and $\overline{u_\theta}$ is the specific utility offered to the agents by the bank.

We put the following constraints in place.

[7] To give intermediation an extra advantage, output is higher in the intermediated sector. On net it does make the utility gain over autarky larger, making it easier to display solutions. There is also risk-sharing available through the intermediaries mitigating potential downside risk of failure. However, the output is effectively under control of the intermediaries who structure the contract in a way to maximize their own profit. Note that the cost of capital is the level k, as the intermediary borrows at zero net interest in an outside market. The output is capital-neutral in autarky production function, i.e. $k = 1$. This gives autarky a slight advantage but it is outweighed by the above disadvantages.

[8] As it stands, the monopoly would pick the utility of autarky as the reward, but see further discussions of market shares in Section 3.

[9] We use Prescott-Townsend lotteries over discrete grids to find solution, it is also possible to write down the problem in a general integral form typically used in contract theory

- Participation Constraints (participation must be voluntary):
  $\forall\ \theta \in \Theta$

$$\sum_{Q,C,K,A} \pi\left(q,c,k,a|\theta\right) u\left(c,a|\theta\right) \geq u_{0\theta}$$

  where $u_{0\theta}$ is the autarky utility.

- Utility Assignment Constraints (UAC):
  $\forall\ \theta \in \Theta$

$$\sum_{q,c,k,a} \pi(q,c,k,a|\theta)u(c,a|\theta) = \overline{u_\theta} \tag{3}$$

- Mother Nature/Technology Constraints[10]:
  $\forall \left\{\overline{q},\overline{k},\overline{a}\right\} \in Q \times K \times A$ and $\forall\ \theta \in \Theta$

$$\sum_{c} \pi(\overline{q},c,\overline{k},\overline{a}|\theta) \quad = \quad P\left(\overline{q}|\overline{k},\overline{a},\theta\right) \sum_{q,c} \pi(q,c,\overline{k},\overline{a}|\theta) \tag{4}$$

- Incentive Compatibility Constraints (ICC) for action variables[11]
  (Moral Hazard problem on unobserved effort):
  $\forall\ a,\hat{a} \in A \times A$ and $\forall\ k \in K$ and $\forall\ \theta \in \Theta$:

$$\sum_{q,c} \pi\left(q,c,k,a|\theta\right) u\left(c,a|\theta\right) \quad \geq \quad \sum_{q,c} \pi\left(q,c,k,a|\theta\right) \frac{P(q,|k,\hat{a},\theta)}{P(q,|k,a,\theta)} u(c,\hat{a}|\theta) \tag{5}$$

- Truth Telling Constraints (TTC) for unobservable types (Adverse Selection problem: type θ must be prevented from pretending to be of type $\theta', \theta \neq \theta'$) :
  $\forall\ \theta,\theta' \in \Theta \times \Theta$:

$$\sum_{q,c,k,a} \pi\left(q,c,k,a|\theta\right) u\left(c,a|\theta\right) \quad \geq \quad \sum_{q,c,k,a} \pi\left(q,c,k,a|\theta'\right) \frac{P(q,|k,a,\theta)}{P(q,|k,a,\theta')} u(c,a|\theta) \tag{6}$$

Please note that under explicitly dynamic contracts we could have an additional state dimension - a continuation utility w and discount rate β, so that UAC (3) could be modified to: $\forall\ \theta \in \Theta$

$$\sum_{q,c,k,a} \pi(q,c,k,a|\theta)[u(c,a|\theta) + \beta w] = \overline{u_\theta} \tag{7}$$

[10] Prescott and Townsend (1984)

[11] We act as if either there is moral hazard or, below, adverse selection. The key term is the "likelihood ratio", $\frac{P(q,|k,\hat{a},\theta)}{P(q,|k,a,\theta)}$ or $\frac{P(q,|k,a,\theta)}{P(q,|k,a,\theta')}$, which reflects the fact that by deviating on effort or lying on its type, the agent changes the probability distribution of output.

structural parameter	σ	χ	γ	α
value	0.5	1.4	1.2	1/3

a	$p(q = low)$	$p(q = high)$
0	.9	.1
.2	.75	.25
.4	.6	.4
.6	.4	.6
.8	.25	.75
1	.1	.9

Table 1: Specifications and technology relating action to probability of each output

We limit the exposition to static contracts in this paper.

The optimal contract:

$$\overline{\pi(q, c, k, a|\theta)} = \underset{\pi(q,c,k,a|\theta)}{\operatorname{argmax}} \left[\sum_{q,c,k,a,\theta} \pi\left(q, c, k, a|\theta\right) [q - c - k] \right] \tag{8}$$

results in bank surplus $S(\overline{u_\theta})$, where

$$\overline{u_\theta} = \sum_{q,c,k,a} \overline{\pi(q, c, k, a|\theta)} u(c, a|\theta)$$

The surplus maximization problem is a standard linear program. The densities of all grids are chosen as to minimize sensitivity to further increase in density. The effort grid is set in [0 1] range. The outcome space for autarky lies in [0, 2] range and for bank-mediated sector it lies in [1, 4] range. Thus the bank-mediated sector first-order stochastically dominates autarky.

For our baseline numerical example (see specifications in Table 1 that are similar to Phelan and Townsend (1991)) we get the following graph (Figure 1)[12] for Pareto frontier for first-best (always full insurance) case. There is the usual trade off between profits of the principal and utility of the agent. In many settings, as with a seller, the price the seller gets is what the buyer surrenders. Our set up is similar, but with risk aversion the frontier is concave. Type dependent technology function is set to be linear here: $f(\theta) = \theta$, it can be made non-parametric. With moral hazard constraints added, the frontier per each type becomes flatter, though to save space we do not display that here; the surplus in moral hazard case is strictly interior to the surplus in full information case when the incentive constraint on effort is binding.

The surplus frontier starts at autarky utility that bounds the utility offer space from below. The first point on all surplus frontiers pictured is always an autarky utility, the last one is the utility at zero surplus for the principal.

A higher ability type within the range of possible utility offers always generates a higher surplus frontier compared to lower ability types. The higher ability types have higher probability of reaching high output; the lower ability requires higher effort to reach the same output. Autarky utilities are different for types, and that causes the surplus frontier to start at different

[12] We have generated many such figures for different parameters values. What we present in this paper is illustrative.

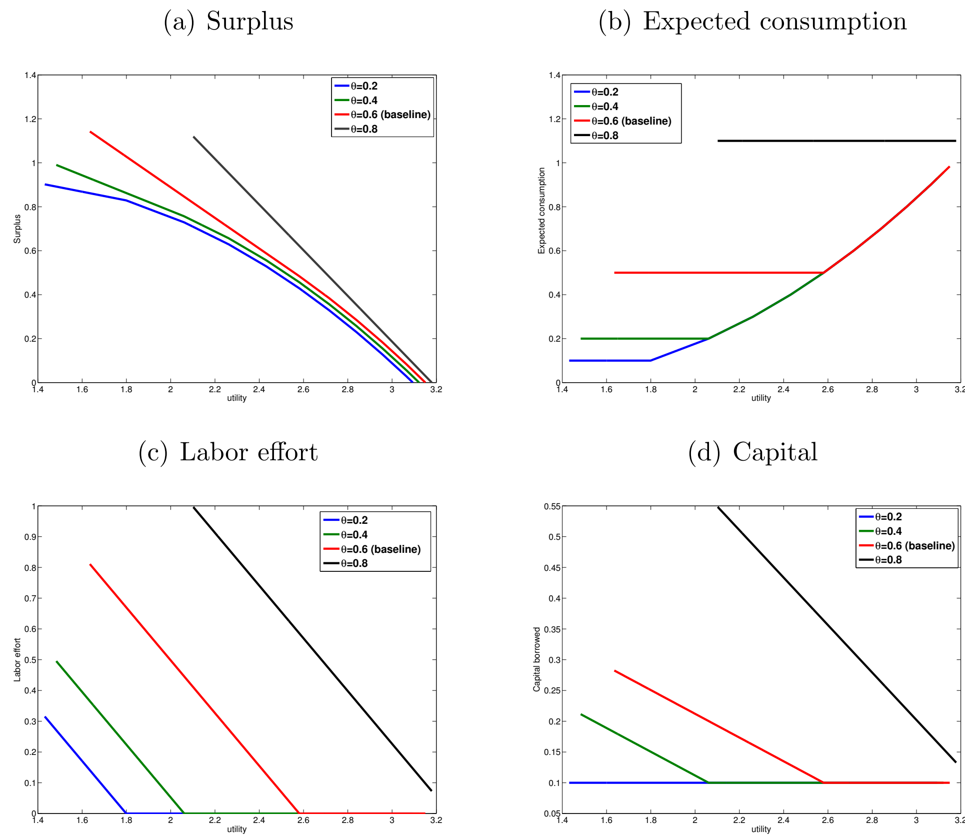


Figure 1: Principal surplus $S(u_\theta)$ and optimal contracts for the agents type θ under full information (FI) regime with full output insurance.

minimum utility offer since participation constraints cut the curve at different values. Note that the slope of surplus for agents carrying high production risk is lower in absolute value than for safer types because the effort of those types doesn't have as much influence on the output. The lower production risk types with higher θ have stricter larger expected output produced at all effort levels than the riskier types with lower θ[13].

The properties of the optimal contract are shown at Figs.1. Notice that the borrowing k is higher for higher ability agents. The costless borrowing is complementary to the effort in this setup. We can easily accommodate costly borrowing, with positive net interest, as well.

The first-best compensation contract is not conditional on output at any utility offers. When utility goes up in full insurance case, the agent first gets more leisure then consumption starts to rise[14]. When moral hazard is added and the incentive constraint is binding, Fig.2, the contract

[13] In the range of productivity shocks theta and actions that we consider, low theta types have a higher variance of output. In any event low theta types are more likely to fail, other things equal.

[14] Financial service provider minimizes the loss in surplus for a given positive change in utility. The impact of consumption change on utility is the marginal utility of consumption times consumption necessary and cost in surplus is one to one to consumption change. The same applies to cost measured in leisure but the impact in surplus is through the production function. Either consumption or leisure is at a corner solution depending on specifications costs, leading to different preferences for a provider to "purchase" an additional unit of input from agents either by giving more leisure or by paying more in consumption.

offers a conditional reward structure as part of the agent's compensation contract.

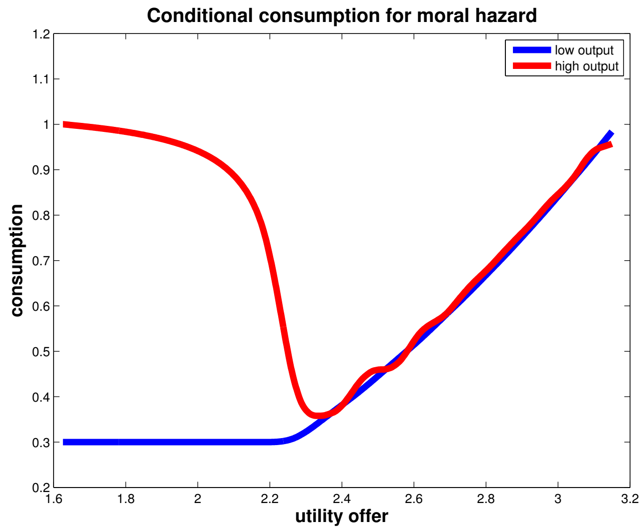


Figure 2: Consumption contracts for the agents of baseline type under moral hazard (MH) regime.

In case of adverse selection (AdS) contracts for two types

$$\theta \in \{\theta_1, \theta_2\}$$

UAC (3) (given offer of utility per type) can be substituted into truth-telling constraints (TTC) to result in explicit bounds for TTC:

$$\overline{u(\theta_2)} \geq \sum_{q,c,k,a} \left[\pi\left(q, c, k, a|\theta_1\right) \frac{P(q, |k, a, \theta_2)}{P(q, |k, a, \theta_1)} u(c, a|\theta_1) \right]$$

$$\overline{u(\theta_1)} \geq \sum_{q,c,k,a} \left[\pi\left(q, c, k, a|\theta_2\right) \frac{P(q, |k, a, \theta_1)}{P(q, |k, a, \theta_2)} u(c, a|\theta_2) \right]$$

This makes utility offers for types close. Though these can sometimes could be undistinguishable in the graphs, the contracts nevertheless are still different in consumption/effort/borrowing schedules for types, even at seemingly close utility levels; see Fig.3, for an example.

We can reduce the space of infeasible contracts under adverse selection while scanning the space of all possible contracts. The utility of one type is fixed (assigned) in UAC by the principal, the utility of another type is obtained through surplus maximization subject to truth telling constraints. It is either not optimal or infeasible to use a different value of utility for another, linked type. Then, we can search over all promises for the first, initial type forming the connected Pareto frontiers. That is, we do not need to scan independently across Pareto frontiers for two types, and we also guarantee that we always stay in contract feasible range

Under all information frictions and all utility offers, the principal offers a type-specific surplus optimal contract. Thus there is no inefficiency in production per each agent's type at a given utility offer under specific information-constrained conditions. The multi-dimensional

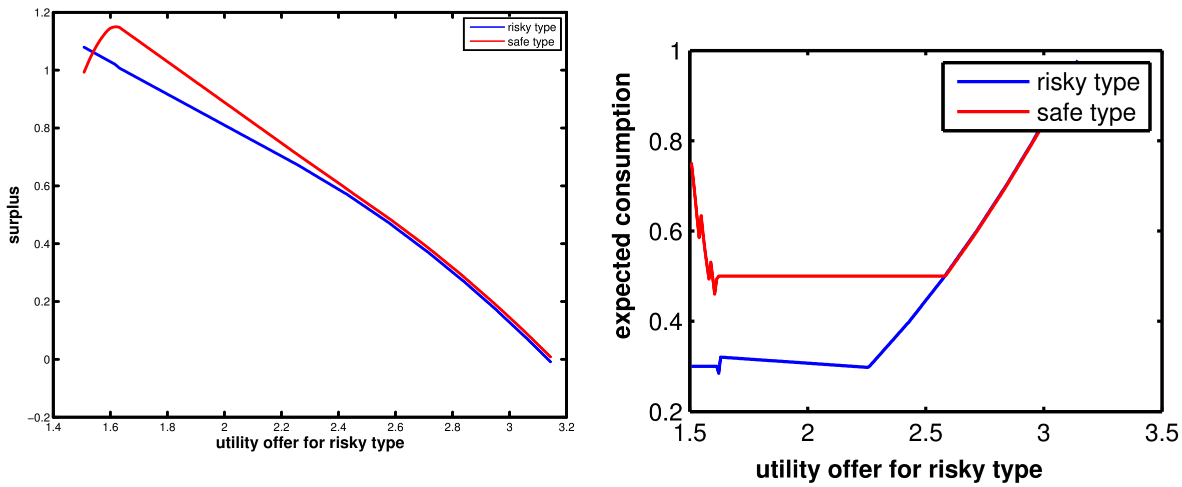


Figure 3: Adverse selection (AdS) contract and surplus for $\theta_1 = 0.5, \theta_2 = 0.6$

characteristics of optimal contracts with utility u are completely encoded in a contract problem generating surplus frontier $S(u)$.

To conclude, in this section we introduced the space of contracts defined by a range of surplus-optimal utility offers from financial intermediaries that depend on agents type and underlying information frictions. We can amend this structure to allow only a fraction of the surplus as written to enter as profits of the bank, the rest covering real intermediation costs. Likewise we can add shocks as random variables into the surplus function, even arguably as a function of locations that we describe below. Though that would move us toward more standard industrial organization setups, it would only cloud the picture here as we try to focus on basics, first.

In the next section we consider several types of market structure that define Pareto optimal contracts with specific utility offers that depend on competition (or lack of such) among financial service providers and their strategic positioning.

3 Financial market structure and equilibrium contracts

All contracts we consider are micro-economic optimal in a sense that all contracts are principal surplus maximizing contacts derived from solving a principal-agent contract problem. Now we proceed to specify the **equilibrium contract** resulting from the market interactions of supply-side (intermediaries) and demand-side (agents) in the spirit of the classical Hotelling setup.

3.1 Supply side

3.1.1 Single branch monopoly

Let's consider a single branch monopoly that selects location x_m (in the middle at 1/2) and utility offers u_θ simultaneously. The share of agent θ market captured by a single branch monopoly is

$$\mu(u_\theta, x_m|\theta)$$

attracting agents at any x where value offered by a contract with an intermediary net costs weakly dominates autarky

$$\mu(u_\theta, x_m|\theta) := \int_0^1 \mathbb{1}_{V(x|\theta)\geq \hat{u_0}}\, dx; V(x|\theta) = u_\theta - \overline{L} * |x - x_m|$$

where $\mathbb{1}_{V(x|\theta)\geq \hat{u_0}}$ is an indicator function equal to one where underlying constraint is satisfied and zero elsewhere. The market share multiplied by the value on surplus frontier $S(u_\theta|\theta)$ set by the utility offers u_θ determines a total profit for monopoly to maximize[15]

$$P(\overline{u_\theta}, \overline{x_m}) = \max_{\{u_\theta, x_m\}} \sum_\theta p(\theta) S(u_\theta|\theta)\mu(u_\theta, x_m|\theta);$$

With no binding corners and the market not fully covered, as when cost L exceeds a critical value L^*, then the FOC on profit allows us to obtain relation between surplus elasticity and market share elasticity:

$$\frac{S'(u_\theta|\theta)}{S(u|\theta)} = -\frac{\mu'(u_\theta|\theta)}{\mu(u_\theta|\theta)}, \forall\theta \tag{9}$$

It can be shown that the monopolist offers utility value u^*_θ, independent of the spatial cost $\overline{L}$:

$$\frac{S'(u_\theta|\theta)|_{u_\theta=u^*_\theta}}{S(u^*_\theta|\theta)} = -\frac{1}{u^*_\theta - u_{0\theta}}, \forall\theta \tag{10}$$

When the entire market is covered, at $\overline{L} < L^*$ the monopolist will offer utility $u^*_\theta(\overline{L})$ such that the most distant household is indifferent to autarky:

$$u^*_\theta(\overline{L}) = \frac{\overline{L}}{2} + u_{0\theta}, \forall\theta \tag{11}$$

Finally, we have the following definition of Pareto optimality:

[15] Existence is guaranteed for continuous profit function here defined on compact set.

Definition A contract with utility $\overline{u(\theta)}$ offered by financial service provider at location $\overline{x_m}$ is *Pareto optimal* if there is no feasible contracts u^*_θ and location x^*_m such that $P(u^*_\theta, x^*_m) > P(\overline{u_\theta}, \overline{x_m})$ and $V^*(x|\theta) = u^*_\theta - \overline{L} * |x - x^*_m| \geq \overline{u_\theta} - \overline{L} * |x - \overline{x_m}| = V(x|\theta) > u_{0\theta}$ for the agents at $\forall x$ that joined financial system under contract $\overline{u_\theta}$ offered at location $\overline{x_m}$ and $V^*(x|\theta) \geq V(x|\theta) > u_{0\theta}$ for those who were in autarky.

Appendix A contains proofs for those results and for Pareto optimality of single-branch monopoly eqilibrium contract. A single-branch monopoly is the simplest market structure we consider. Analytical results are available in this case to verify the accuracy of numerical methods we develop later for more complex models of competition and to provide bounds on possible outcomes for comparative statics.

For example, Eq.(10) for a single-branch monopoly allows us to specify a unique contract value[16] that is going to be common at the high spatial cost bound among all types of market competition we consider.

A monopolist could potentially offer a spatially discriminating contract. In this case the monopolist locates at $\overline{x}$ and offers a location specific utility $u_\theta(x) = u_{0\theta} + \overline{L} * |x - \overline{x}|$ to leave households at all x indifferent to autarky $V(x|\theta) = u_{0\theta}$ while extracting surplus along $S(u_\theta(x)|\theta)$ frontier from all intermediated agents. The market is covered up to the point $x = \hat{x}$ such as $S(u_\theta(\hat{x})|\theta) = 0$. The agents are weakly worse off under such spatially discriminating contract at all spatial costs[17].

3.1.2 Two branch monopoly (collusion among banks that are supposed to be competitors)

In this section we start with the simplification of one type, thus dropping dependence on θ temporarily. In the numerical example we focus on moral hazard versus full information. Let's consider a monopoly that selects locations x_1 and x_2 simultaneously for two branches or two banks that collude to share total surplus with no competition. This is a benchmark that we use later to compare against the results of competition with two competing banks. This monopoly also offers utilities u_1, u_2 simultaneously at each of its branches. Thus for each branch we can write the the share of market captured as

$$\mu_1(u_1, x_1, u_2, x_2) := \int_0^1 \mathbb{1}_{\substack{u_1 - \overline{L}*|x-x_1| \geq u_2 - \overline{L}*|x-x_2| \\ u_1 - \overline{L}*|x-x_1| \geq \hat{u_0}}} \, dx;$$

$$\mu_2(u_1, x_1, u_2, x_2) := \int_0^1 \mathbb{1}_{\substack{u_2 - \overline{L}*|x-x_2| > u_1 - \overline{L}*|x-x_1| \\ u_2 - \overline{L}*|x-x_2| \geq \hat{u_0}}} \, dx;$$

The total profit for monopoly is defined as

$$\Pi(\{\overline{u_2}, \overline{x_2}, \overline{u_1}, \overline{x_1}\}) = \max_{\{u_2,x_2,u_1,x_1\}} [S(u_1)\mu_1(u_2, x_2, u_1, x_1) + S(u_2)\mu_2(u_2, x_2, u_1, x_1)];$$

$$\{\overline{u_2}, \overline{x_2}, \overline{u_1}, \overline{x_1}\} = \underset{\{u_2,x_2,u_1,x_1\}}{\operatorname{argmax}} [S(u_1)\mu_1(u_2, x_2, u_1, x_1) + S(u_2)\mu_2(u_2, x_2, u_1, x_1)]$$

[16] Local monopoly equilibrium contract has numerical value of $u^* = 2.4$ in our baseline type specification for full information agent problem.

[17] We do not consider spatial discrimination in more detail in this paper. The incomplete nature of the contract, not contingent on location, makes the problems more interesting (and arguably more realistic).

Proposition 3.1. *Let* $\Pi(\{\overline{u_2}, \overline{x_2}, \overline{u_1}, \overline{x_1}\})$ *be a maximum profit for two branch monopoly with branches at locations* $\overline{x_1}$ *and* $\overline{x_2}$ *offering utility* $\overline{u_1}$ *and* $\overline{u_2}$ *correspondingly. Then monopoly equilibrium is Pareto optimal.*

This follows as in Propositions A.1-A.2 earlier, with one branch, since this problem is strictly concave. Intuitively, without loss of generality, monopoly branches are located at $\overline{x_1} = 1/4, \overline{x_2} = 3/4$. Thus, each branch can be considered as a local single branch monopolist serving half of the total active market[18]. Each of those monopolists offers a contract that is surplus maximizing according to Propositions A.1-A.2 for its half of the market. See the discussion which follows.

To visualize the effects of supply side action on demand side reaction we introduce the concept of real value for a household at location x from the financial contracts with utilities u_1 and u_2 offered at two locations x_1 and x_2 defined as

$$V^{u_1,u_2,x_1,x_2}(x) = \max(u_1 - L * |x - x_1|, u_2 - L * |x - x_2|, u_0), \tag{12}$$

where utility promises and locations are for two competing banks or two branches of monopoly (or a central planner choices), u_0 is the level of utility the household gets from staying in autarky and not incurring any costs to join financial system.

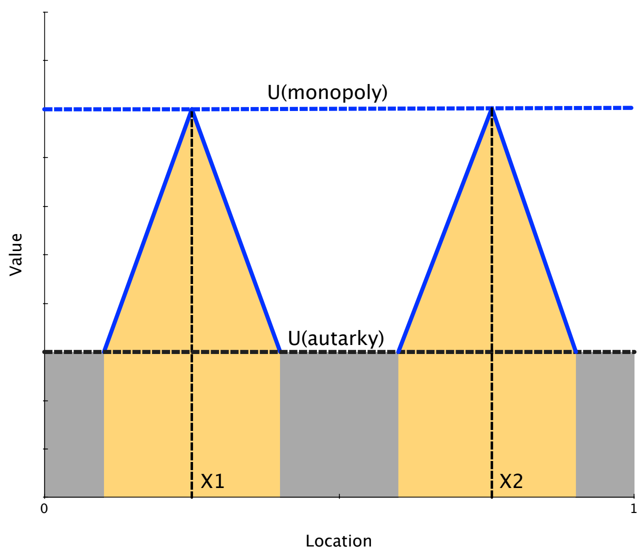


Figure 4: Autarky islands and two-branch monopoly

The real value from contracts to households is plotted on Fig.4.

Those tent-shaped graphs represent the net benefit for households located at various x values. The area above an autarky value is a net gain for households from joining financial system.

The apex of each tent corresponds to financial service provider branch location and utility offer (location on the x axis and promise u on the y axis). Households that are located in the immediate vicinity get the largest value since utility offered is not spatially discriminating.

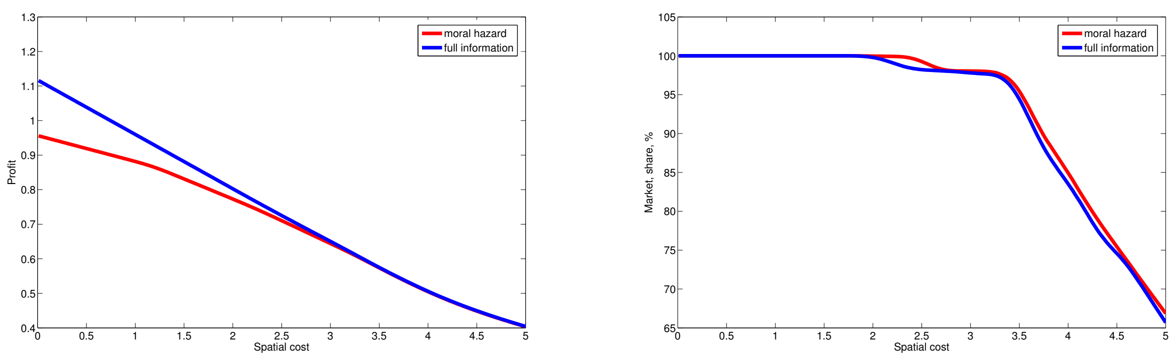
[18]If there are autarky islands then branch locations are indetermined as long as the market for each branch is separated by an island of autarky

The marginal customer in this case is the customer who gets exactly the autarky value from contract when spatial costs are subtracted from utility offered. There are islands of autarky in the middle and at the margins in this particular case. The spatial costs are high enough so as to make a monopoly to be a profit optimal while servicing $< 100\%$ of the market. The high level of surplus extracted from the agents at relatively lower utility offer overweighs the benefits of adding additional agents that could be attracted by higher utility offers at lower level of surplus for the intermediary.

The slope of tents gets steeper with increasing spatial costs, the level of utility at the apex first goes up; but then it stays constant at local monopoly value.

We solve the optimization problem for the equilibrium contract numerically as we scan the range of spatial costs from zero spatial cost case of $L = 0$ to $L = 5$. The profit monotonically goes down as we increase the spatial costs. Locations of branches for monopoly are fixed symmetrical at $x_1 = 1/4; x_2 = 3/4$; our baseline type has $\theta = 0.6$.

Monopoly utility offer at zero spatial costs starts with autarky utility level $u \approx u_0$. As we increase the spatial costs, the profitable profit extraction by a monopoly is possible only from a smaller share of the market. The autarky islands appear, less people get access to the financial services.



(a) Profit

(b) Market share

Figure 5: Profit and market shares for two-branch monopoly

In case of both full information and moral hazard, the equilibrium monopoly offer from banks has to beat the autarky at the margin. The profits under moral hazard are strictly lower than under full information at all spatial costs; see Fig.5.

Since the autarky value is the same for full information and moral hazard, we have identical comparative values for utilities for both regimes when market is fully covered. The value of spatial cost at which monopoly becomes a local monopoly according to Eq.(10) with spatial cost independent offer u^* is at $\overline{L} = 4(u^* - u_0) \approx 3.3$. It is quantitatively different for MH regime, but here, not by much.

However, the contract schedules are different in the full information compared to the moral hazard regime throughout the whole range of costs scanned; see Fig.6.

While consumption schedule is the same over all spatial costs in full-information regime, the moral hazard compensation is sharply and non-monotonically spatial cost dependent. The capital used in production, the labor effort, and the output decrease faster with higher spatial

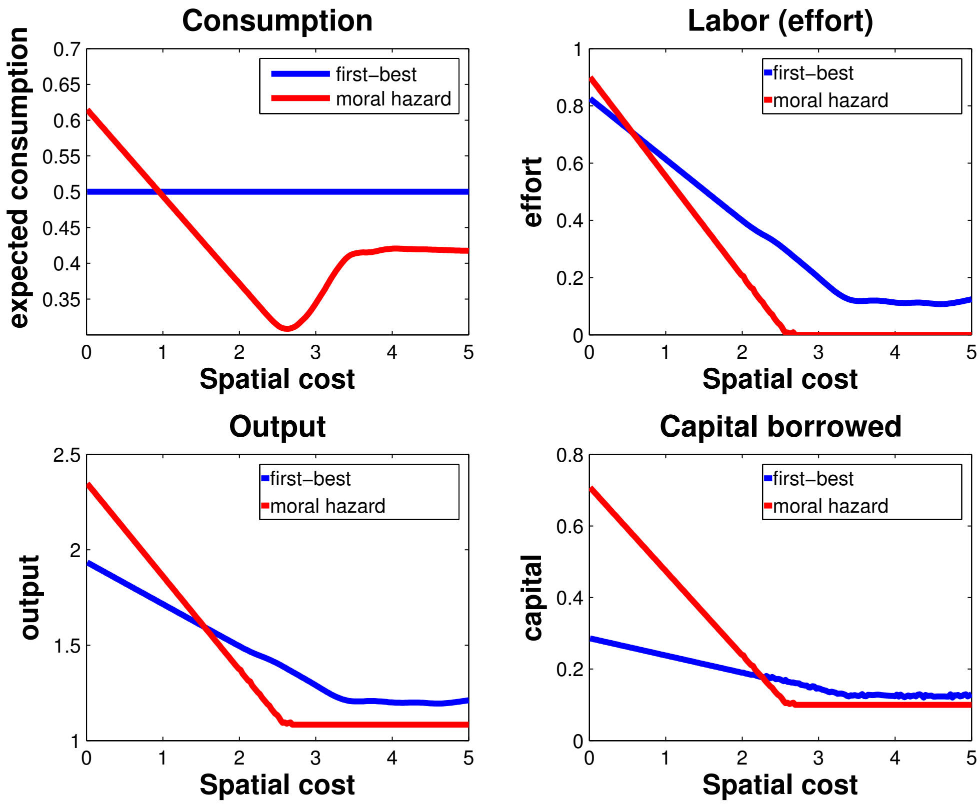


Figure 6: Equilibrium contracts for two-branch monopoly

costs under the moral hazard in the range of spatial costs where market is fully covered. A monopoly can't extract the surplus from the agents under moral hazard in the same way as under full information at the same contract value posted; it has to reallocate relatively larger part of the surplus to agent's compensation. Since inducing effort is costly for a monopoly and the assigned effort is a cost for an agent too, the capital under moral hazard is used relatively more even though the capital is complementary to the effort in production.

The optimal utility (contract) choice as a function of spatial cost is well-resolved, unique and it moves continuously across varying spatial costs[19]. This will not necessarily be true for competitive market structures which we consider next.

The profit maximizing contract for a given spatial cost as utility is varied satisfies a familiar elasticity condition: the elasticity of the profit/surplus frontier with respect to utility, on the intensive margin, equals the elasticity of increased market share on the extensive margin; the tradeoff between lower profits per customer and more customers is balanced and determines a particular operational point on the frontier. But since these surplus frontiers vary with information regime, with the frontier of the moral hazard regime having a lower level but also a lower slope, there are corresponding implications for comparative statics. Throughout the range of spatial costs, profits are lower and market shares are higher in the moral hazard regime relative to the full information regime. Further, as spatial costs move from low to high, labor effort, capitalization/borrowing and expected output are monotonically decreasing. But

[19]Berge's Maximum Theorem applies here.

finally, and a bit more novel, the moral hazard regime initially specifies higher and then lower levels of these variables, with the crossing point varying by the particular variable considered. Though expected and actual consumption are constant in the full information regime, the degree of insurance (the variation of consumption with output) varies in the moral hazard regime, naturally, but varies with the provincial spatial costs, depending on inducement of low or higher levels of effort.

3.2 Competition with no commitment: simultaneous Nash competition on contracts and locations with heterogeneous agents

Nash equilibrium in case of two market entrants is defined by $\{u^*_{1\theta}, x^*_1, u^*_{2\theta}, x^*_2\}$ that satisfy

$$\sum_\theta S(u^*_{1\theta}|\theta)\mu_1(u^*_{2\theta}, x^*_2, u^*_{1\theta}, x^*_1) \geq \sum_\theta S(u_{1\theta}|\theta)\mu_1(u^*_{2\theta}, x^*_2, u_{1\theta}, x_1) \tag{13}$$

$$\sum_\theta S(u^*_{2\theta}|\theta)\mu_2(u^*_{2\theta}, x^*_2, u^*_{1\theta}, x^*_1) \geq \sum_\theta S(u_{2\theta})|\theta)\mu_2(u_{2\theta}, x_2, u^*_{1\theta}, x^*_1) \tag{14}$$

$\forall\{u_{1\theta}, x_1, u_{2\theta}, x_2\}, \forall\theta$

3.2.1 Competition with no commitment: comparative statics of FI vs MH vs AdS vs AdS+MH as spatial cost L is varied

We present computational results for no commitment Nash with locations of competitors fixed at $x_1 = 0.25; x_2 = 0.75$ on Figs.7-11. To solve for optimal policies in the general case we use min-max algorithm we call "minimizing distance to Nash", described in Appendix B. Results are symmetric; both banks have the same profit, offer the same utility and have equal market share.

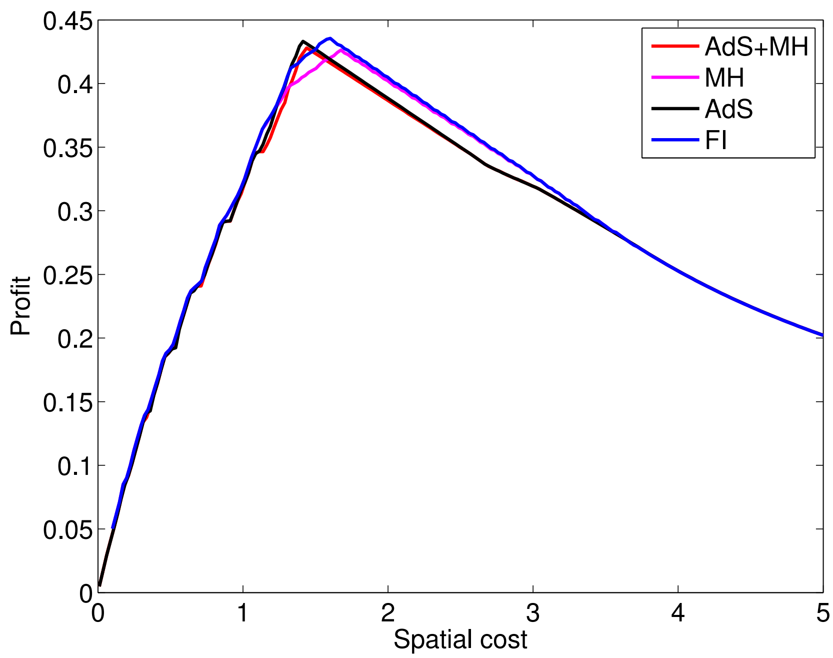
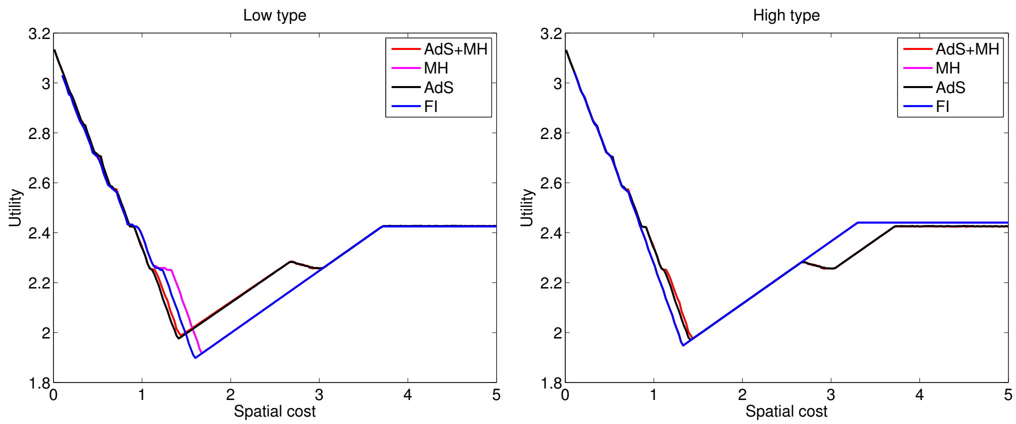


Figure 7: Profits, fixed location Nash, $x_1 = 0.25; x_2 = 0.75$

The levels of profits and utilities on Figs.7-8 are similar, although clearly distinguishable at certain spatial costs, across the various financial frictions[20].

[20]The inflection point with competitive regime switching to monopoly solution is different across regimes, in

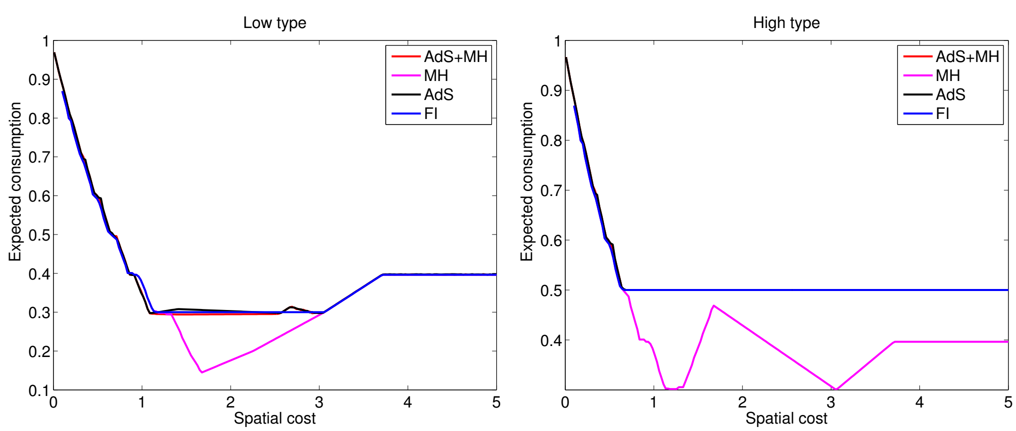


(a) Low ability type (b) High ability type

Figure 8: Utility, fixed location Nash, $x_1 = 0.25; x_2 = 0.75$, $\theta_1 = 0.5$ (low type), $\theta_2 = 0.6$ (high type)

The range of optimal utility values spans most of the underlying domain when spatial costs are varied. This has implications for empirical estimations, as in Karaivanov and Townsend (2014), for example, the spatial information can be used to add additional restrictions on utility distributions to distinguish between regimes with more precision.

By reverse mapping of equilibrium utility back to contract characteristics we can see the difference resulting from financial frictions; see Figs.9-11.

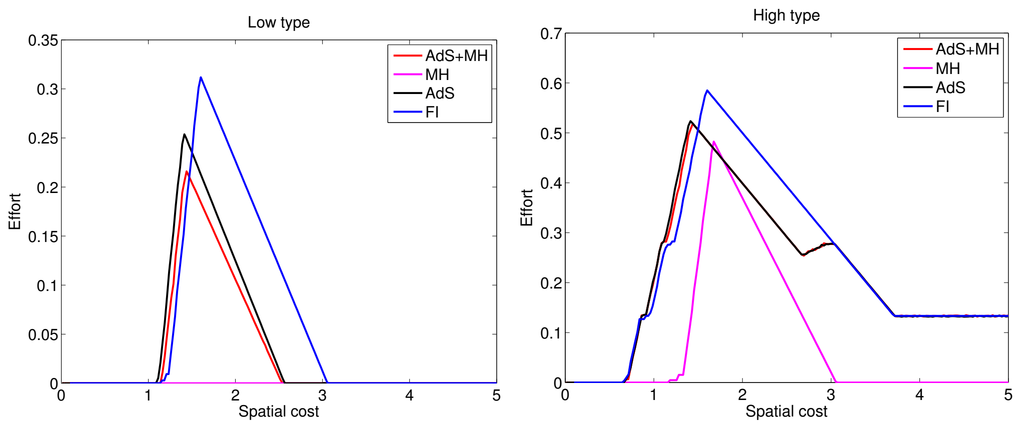


(a) Low ability type (b) High ability type

Figure 9: Consumption, fixed location Nash, $x_1 = 0.25; x_2 = 0.75$, $\theta_1 = 0.5$ (low type), $\theta_2 = 0.6$ (high type)

Due to ICC, the expected consumption part of the equilibrium contract in comparative statics best distinguishes moral hazard regime and also the type of the customer, from the other regimes.

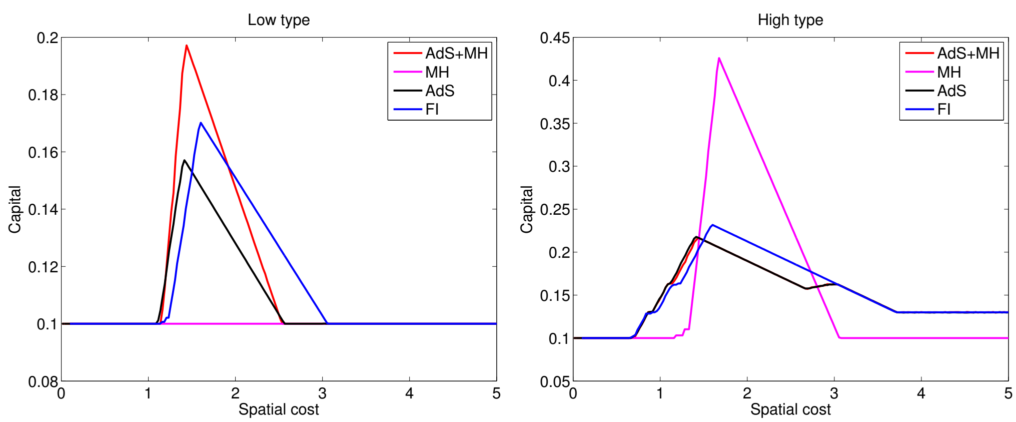
particular, it is most pronounced for regimes with adverse selection because both types have to switch at the approximately same time due to intertwined utilities.



(a) Low ability type (b) High ability type

Figure 10: Effort, fixed location Nash, $x_1 = 0.25; x_2 = 0.75$, $\theta_1 = 0.5$ (low type), $\theta_2 = 0.6$ (high type)

The effort moment is clearly different across all regimes and types. The combination of moral hazard with adverse selection moves the optimal effort closer to the first-best contract compared with pure MH. Pure adverse selection is the closet to FI.



(a) Low ability type (b) High ability type

Figure 11: Capital, fixed location Nash, $x_1 = 0.25; x_2 = 0.75$, $\theta_1 = 0.5$ (low type), $\theta_2 = 0.6$ (high type)

The level of capital borrowed for production is generally higher across all frictions for the high ability type (note the change in scale). There is substantially higher level of borrowing for the low ability type under combined MH and AdS frictions compared with pure MH regime when the type is low. But the high type capital is highest in pure MH.

We consider a competitive Nash equilibrium in contracts, keeping the branches of the two banks separated at the same locations as the two branch monopoly, but allowing heterogeneity in types and comparing the range of financial/information regimes. As spatial costs increase across provinces, profits first increase and then decrease, and utilities of households/firms decrease and then increase. Low spatial costs imply intense competition and hence low profit.

At intermediate costs, banks struggle to retain customers in their respective hinterlands. At high spatial costs active market segments do not overlap, and we move toward and obtain local branch monopolies. Yet the transition points are different across the different information regimes. In sum, though higher spatial costs is a worse physical environment over all, competing banks can actually gain from this in certain ranges, and the extent of this gain depends on the information structure. Related effort, capitalization (borrowing), and expected output are now each non-monotonic with spatial costs, that is, rise and then fall as spatial costs increase. Further, the range and peaks of nontrivial values are different for different variables and different types, and again vary with the financial information regime. For example, capital is most responsive for low types in the adverse selection plus moral hazard regime but most responsive for the high type in the moral hazard regime.

3.2.2 Mechanics of competition among banks: splitting the market

In this section we overview the basic features of competitive behavior that lie behind the figures considered but in some sense these are general for all market structures we consider.

We assume that both banks already made choice of locations and utilities at the end of competitive market process and we take their choices as given here.

The outcome of that choice in competitive regime is illustrated at Fig.12 for two competing entrants. See Fig.4 for comparison. In the case shown both banks have non-zero profit from a fully covered market.

The market share of bank 1 is determined by the critical z_θ^1 where the t-pees cross, namely:

$$z_\theta^1 = \mu_1(u_{1\theta}, x_1, u_{2\theta}, x_2|\theta) = \frac{x_1 + x_2}{2} + \frac{u_{1\theta} - u_{2\theta}}{2L}$$

The market share of bank 1 is thus z_θ^1 and bank 2 is the residual:

$$z_\theta^2 = 1 - z_\theta^1$$

Profits of bank 1 and 2 from market types θ:

$$P_1^\theta(u_{1\theta}, x_1|u_{2\theta}, x_2) = S(u_{1\theta}|\theta)z_\theta^1; P_2^\theta(u_{2\theta}, x_2|u_{1\theta}, x_1) = S(u_{2\theta}|\theta)(1 - z_\theta^1); \tag{15}$$

Note, that as in D'Aspremont et al. (1979) there is potential discontinuity in payoff functions as one t-pee overlaps and then encompasses the other. This forces a decision on whether to fight it out with higher utility offer or to separate spatially with lower intensity of competition with possible discontinuous jumps in policies. There are two related points to keep in mind. One is the discontinuity, the second is a potentially non-concave profit defined by two disjoint strategies, i.e. centrally located competition vs. spatial separation. Appendix D illustrates this non-concavity for the full commitment regime we consider later and this stays true even when logit formulation is used eliminating the discontinuity.

From FOC on profits in Eq.(15) with market fully covered for fixed symmetric bank locations at $x_1 = \frac{1}{4}, x_2 = \frac{3}{4}$ we get:

$$\begin{aligned} S'(u_{1\theta}|\theta))\,[L + u_{1\theta} - u_{2\theta}] + S(u_{1\theta}|\theta)) = 0 \\ S'(u_{2\theta}|\theta))\,[L + u_{2\theta} - u_{1\theta}] + S(u_{2\theta}|\theta)) = 0 \end{aligned} \quad \forall\theta \tag{16}$$

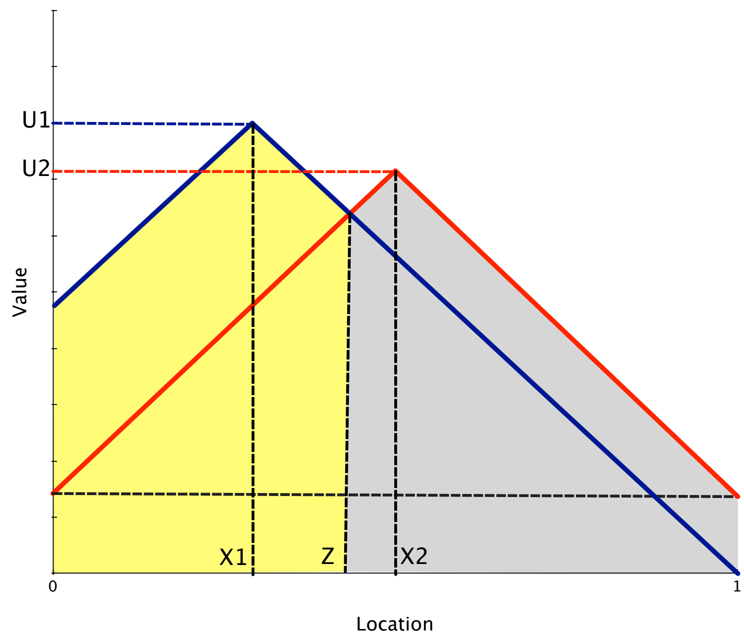


Figure 12: Splitting the market under given choice of locations and utilities.

And finally we obtain the following simple conditions for optimal policies u_θ^* under assumption that both banks are following standard symmetric Nash best-response logic[21]:

$$\frac{S'(u_\theta^*|\theta)}{S(u_\theta^*|\theta)} = -\frac{1}{L}, \forall\theta \tag{17}$$

At low spatial costs $L \to 0$ this results in pure Bertrand outcome with utility at the maximum possible value and surplus at zero limit.

As noted in the case of monopoly, in general, two most important factors defining optimal level of utility offers chosen by financial service providers are the level and the slope of surplus but in combination. The surplus elasticity $S'(u_\theta|\theta)/S(u_\theta|\theta)$ defines how much of surplus is lost when utility is raised to get extra market share and that in turn interacts with the market share elasticity $-\mu'(u_\theta|\theta)/\mu(u_\theta|\theta)$.

The value of optimal utility (equilibrium point u_θ^*) can be higher for high-risk (bad) type compared with low-risk (good type), as shown in Fig.13 at certain spatial costs under our two types specification. The point of intersection of downward sloping surplus frontier with its slope scaled by the spatial costs is an equilibrium point as in Eq.(17), defining the optimal choice of utility.

Of considerable interest, the absolute value of surplus elasticity for high production risk types is lower at low utility promises and higher at high promised utilities compared with low risk types (figure not shown here). That means relatively less surplus is lost at low utility promises for risky types when utility is raised at the margin than for safe types and vice versa for high promised utilities; see Fig.8 with risky type getting higher utility at certain spatial

[21] This applies for type-separating (full information and moral hazard) regimes.

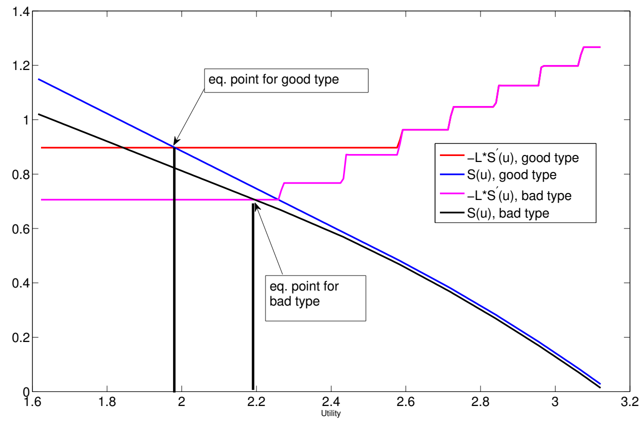


Figure 13: Equilibrium point, $\theta_1 = 0.5$ (bad type), $\theta_2 = 0.6$ (good type), $L = 1.3$

costs. So the attractive type so to speak varies with the form of overall competition, as that determines the range of utilities in play.

More generally, as the spatial cost increases, competition in utility offer decreases according to Eq.(17) until the level of utility hits a local monopoly solution with two separate branches. A monopoly solution is defined by the participation constraint, which means that the marginal market share is now attracted from the autarky rather than from the competitor. With rising spatial costs, it becomes relatively more important to protect the outermost borders, a "hinterland" of the market, from falling into autarky than to attempt to steal a customer from a competitor. Thus, the optimal solution with market still fully covered becomes as follows:

$$u_\theta^m = \frac{L}{4} + u_{0\theta},\ \forall\theta$$

when $u_\theta^* < u_\theta^m$ and $x_1 = \frac{1}{4}, x_2 = \frac{3}{4}$.

Finally, at very large spatial costs L with market not fully covered we have monopolies with autarky islands offering unique value of utility defined previously in Eq.(10).

To conclude, we introduce several characteristic features for competitive regime; see Fig.8. At low spatial costs, the entities compete most intensely for the customers in the middle of the market; indeed, close to $L = 0$ we are at the Bertrand-like solution but as cost L increases, the competition intensity decreases. Eventually at medium spatial cost the effect of autarky constraints kicks in. The market is still entirely covered, but the market tents only intersect on at the base, the point in space where the utility of the household is at autarky. In effect the entities are bound by the requirement to protect the "hinterland" as in Prescott and Visscher (1977). Finally, at yet higher spatial costs the tents no longer have a common point, and the competing entities become fully spatially separated local monopolies with autarky islands both in between them and at the "hinterland" margins as in Hotelling (1929). In this region the utility value is pinned as argued previously, but the extent of the market is shrinking as L continues to increase.

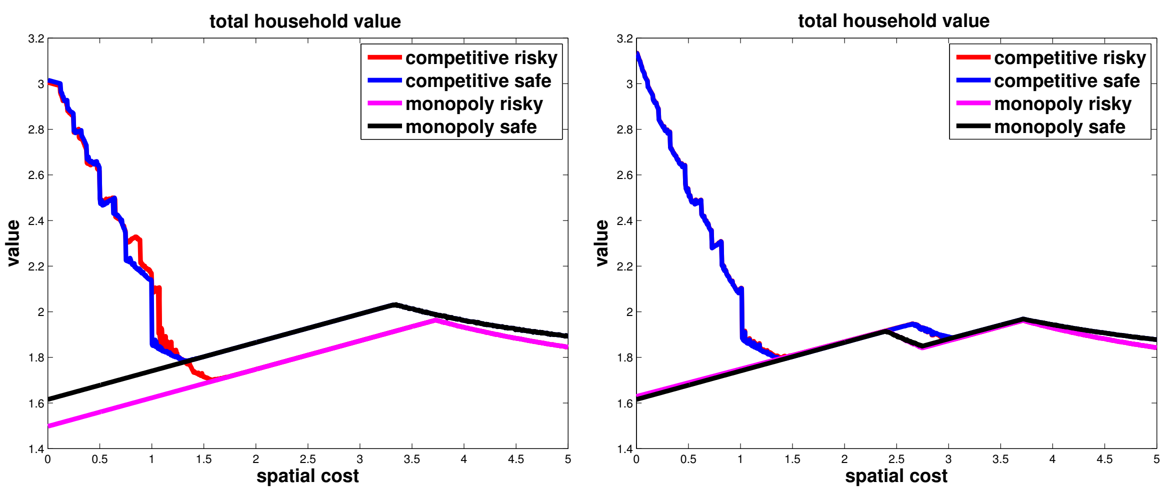
3.2.3 Welfare implications of financial liberalization experiment

We perform the following experiment to describe a possible impact of financial liberalization, such as the interstate branching reform of 1994 in the US that eliminated restrictions on interstate bank acquisitions, or more simply, we break up a monopoly bank into pieces.

First, we compute the two-branch monopoly solution with branch locations at [1/4; 3/4] as the pre-existing condition prior to the reform. Then, interstate banking is allowed under full information competition on contracts and adverse selection competition on contracts (in contrast to previous section we drop MH).

Specifically, each of monopoly's branches at fixed locations can be bought by one of competing banks out of state or, more simply, the two preexisting branches are forced to compete with each other and collusion can be thwarted. We assume that opening a new additional branch is either too costly or restricted by regulation. We compute welfare implication using a comparison of total real value received by the households.

Results are shown in Fig.14 for FI monopoly versus simultaneous Nash competing FI entities and also for the same entities under adverse selection. Types are $\theta_1 = 0.5$ and $\theta_2 = 0.6$. The utilities for risky and safe types in adverse selection are close to each other; they merge into single line on the scale of this graph. The largest welfare gains from competition are at low spatial cost. Under full information the largest gains are for the risky type.



(a) Full Information

(b) Adverse Selection

Figure 14: Welfare, liberalization experiment, $\theta_1 = 0.5$ (risky type), $\theta_2 = 0.6$ (safe type),

At higher spatial costs, the gains from liberalization are smaller. In Mahoney and Weyl (2014), competition under adverse selection results in production-distorted oversupply of services and market power is shown to be socially beneficial for some markets (subprime auto lending). This is not the case here since banks operate with constrained optimal contracts that are not distorting production decisions, even under adverse selection. In our framework, liberalization produces strictly higher welfare for both FI and AdS regimes.

The adverse selection regime is relatively less "extractive" under monopoly compared to full information regime because the gap in optimal level of utility offers for two types is minimal, while full information monopoly offers significantly lower utility for risky type. Information frictions prevent a monopoly from extracting the surplus optimally for both types.

At relatively low spatial costs the switch from monopoly to competition naturally increases the household utility, but with some twists. With full information the biggest gain is for the risky type. With adverse selection it is much harder to distinguish across types, so the overall gain from liberalization/competition is similar for both types. Plus safe type gains more from liberalization in the adverse selection regime than in the full information regime. At yet higher spatial costs there is no gain for either type, though the transition point from gain to no gain happens earlier, at lower spatial costs in the adverse selection regime.

3.3 Competition with partial commitment: full commitment on location choice, followed by simultaneous Nash competition in contracts

Here and subsequently we drop contract dependence on types to concentrate on location choice effects. This is done for clarity of exposition, and results can easily be extended for heterogeneous types. We return to heterogeneous types in Section 4.

Let's consider a sequential game with the first bank coming to location $\overline{x_1}$ and second bank coming later to location $\overline{x_2}$.

The first bank anticipates the entry of the second bank and it chooses its location with respect to the best possible response by the second bank. Both banks anticipate subsequent simultaneous Nash competition in contracts as in Section 3.3. This is a common assumption in IO literature: for example, ex-ante investment with ex-post price competition, similar to Doraszelski and Pakes (2007).

We use hierarchical programming formulation for this sequential leader-follower problem, with full commitment to location choice and ex-post simultaneous Nash competition in contracts.

Simultaneous Nash equilibrium (ex-post competition in utilities for arbitrary locations) in case of two market entrants would be defined by strategy $G^N = \{u_1^*, u_2^*\}$ that satisfy

$$\begin{aligned} P_1(G^N) = S(u_1^*)\mu_1(u_1^*, u_2^*, x_1, x_2) \geq \\ S(u_1)\mu_2(u_1, u_2^*, x_1, x_2) \end{aligned} \tag{18}$$

$$\begin{aligned} P_2(G^N) = S(u_2^*)\mu_2(u_2^*, u_1^*, x_1, x_2) \geq \\ S(u_2)\mu_2(u_2, u_1^*, x_1, x_2) \end{aligned} \tag{19}$$

$\forall\{u_1, u_2, x_1, x_2\}$

We define the second bank problem (the lower level location choice problem) as follows:

$$\max_{\{x_2\}} \{P_2(x_1, x_2) : \mu_2(x_1, x_2) \geq 0\} \tag{20}$$

Let $x_2^*(x_1)$ denote the solution set of problem Eq.(20) for a fixed x_1.

Then, the bilevel programming problem (or the upper level location choice problem) can be stated as:

$$\max_{\{x_1\}} \{P_1(x_1, x_2(x_1)) : \mu_1(x_1, x_2(x_1) \geq 0, x_2(x_1) \in x_2^*(x_1)\} \tag{21}$$

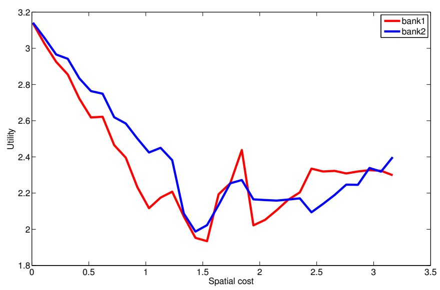
Definition A strategy $G^{SNE} \otimes G^N = \{\overline{x_1}, \overline{x_2}(\overline{x_1})\} \otimes \{u_1^*, u_2^*\}$ in utilities and locations constitutes sequential Nash eqilibrium (SNE) on location with ex-post simultaneous Nash competition in utilities if it solves Eq.(20)-(21) and it satisfies Eq.(18)-(19).

If locations and utilities are discretized, then the standard extensive games eqilibrium existence conditions apply. We don't concentrate on analysis of such option, instead we directly compute Nash equilibrium relying on algorithm described in Appendix B for computing a minimum distance to Nash equilibrium[22].

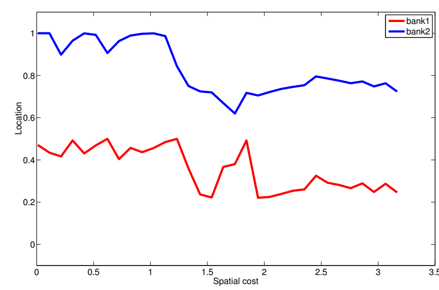
Results of numerical experiments are in Fig.15. With higher spatial costs solution converges to local monopoly solution we discussed before, and not shown here.

Figure 15: Competition: full insurance, no commitment

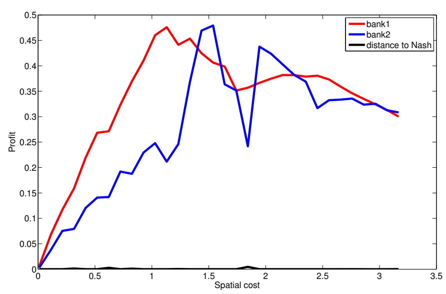
(a) Utility offer choice



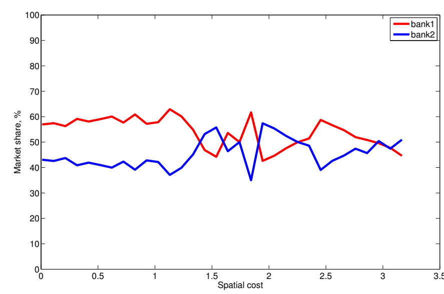
(b) Location choice



(c) Profit



(d) Market share



The first entrant at low spatial costs tries to use his first mover advantage by choosing a location in the middle, and the second entrant tries to get his market share by staying at the margin, $\overline{x_2} \approx 1$, while raising utility offer in subsequent simultaneous move above the first entrant. Thus, for the first entrant the optimal business strategy is to get larger market share at relatively larger surplus per agent utilizing the better central location. At higher spatial cost bank 2 moves toward the center, the first mover advantage disappears and market shares become more equal. Indeed, bank 2 can sometimes do better than bank 1.

The endogenously separating equilibrium for banks with ex post competition in utility promises is different from pooling eqilibrium for banks under no commitment in contracts (see Fig.7-Fig.11) and as we will see it will be different from full commitment outcome below. Obvious but worth repeating, if utilities are different, contracts are different. Note in particular the non-monotonic dynamics in movements under a spatial cost scan.

We make endogenous not only contracts but also locations, and then the timing of entry begins to matter. Here we can distinguish various degrees of commitment or inflexibility. For example, suppose the first entrant into a province chooses a branch location in a first stage, before a second entrant does so in a second stage, and before both compete on contracts in a third stage. In this setting and at relatively low spatial costs, the first bank has an incumbency,

[22]We have a proof of existence with a different market in the next section.

first-mover advantage, picking a more central location and being able to set lower utility, thus gaining higher market share and higher profits. However, for provinces with intermediate and higher spatial costs, this first-mover advantage can be lost sometimes, but not always. Again it is non-monotonic. Further, and related, locations, utilities, market shares, and profits are all non-monotonic in spatial costs. The point is that there is little that is straightforward about moving toward economically advantageous environments with lower spatial costs, as with the emergence of electronic rather than brick and mortar banking.

Here we introduced asymmetry in location choice resulting in asymmetric Nash outcome; next we are going to consider asymmetry in both location and utility choice under a full commitment regime.

3.4 Competition with full commitment on location choice and contracts

As in previous section, we use bilevel hierarchical programming formulation for sequential leader-follower competition with full commitment to location choice and contract.

We define the second bank problem (the lower level problem) as follows:

$$\max_{\{G_2\}} \{P_2(G_1, G_2) : \mu_2(G_1, G_2) \geq 0\} \tag{22}$$

where $G_1 = \{x_1, u_1\}$ and $G_2 = \{x_2, u_2\}$ are the strategy sets of bank 1 and 2.

Let $G_2^*(G_1)$ denote the solution set of problem Eq.(22) for a fixed G_1.

Then, the bilevel programming problem (or the upper level problem) can be stated as:

$$\max_{\{G_1\}} \{P_1(G_1, G_2(G_1)) : \mu_1(G_1, G_2(G_1)) \geq 0, G_2(G_1) \in G_2^*(G_1)\} \tag{23}$$

Definition A strategy $G = \{\overline{x_1}, \overline{u_1}, \overline{x_2}(\overline{x_1}, \overline{u_1}), \overline{u_2}(\overline{x_1}, \overline{u_1})\}$ constitutes sequential Nash eqilibrium (SNE) if it solves Eq. (22)-(23).

3.4.1 Mapping with logistic probabilities, attenuation function

We consider a case where characteristics of the agent, in particular, his locations are seen with error by the econometrician. In addition, there is a distribution of preferences for a given bank; the agents joining a service provider are split between the banks according to logistic function, as is typical for demand estimation models in IO.

The real value function, a preference for bank 1, for households at locations x under a given sequential game strategy

$$G = \{x_1, u_1, x_2(x_1, u_1), u_2(x_1, u_1)\}$$

is denoted

$$V(x|G) = u_1 - \overline{L} * |x - x_1| - (u_2(x_1, u_1) - \overline{L} * |x - x_2(x_1, u_1))|)$$

The simplest distributional assumption on bank specific preferences is Type I extreme value distribution, as in the aggregate logit model (Nevo (2000)).

Total market shares are then

$$M_1(G) = \int_0^1 \mathfrak{L}\left(\frac{V(x|G)}{\mathfrak{M}}\right) dx - \int_0^1 \mathbb{1}_{u_1(x_2,u_2) - \overline{L}*|x - x_1(x_2,u_2)| < \hat{u_0}} \, dx; \tag{24}$$

$$M_2(G) = \int_0^1 \mathfrak{L}\left(\frac{-V(x|G)}{\mathfrak{M}}\right) dx - \int_0^1 \mathbb{1}_{u_2 - \overline{L}*|x - x_2| < \hat{u_0}} \, dx; \tag{25}$$

where

$$\mathfrak{L}(\mathfrak{x}) = \frac{1}{1 + \exp^{-\mathfrak{x}}}$$

With specific logit tail mass $\mathfrak{M}$ we integrate over all agents. The mass of agents in autarky is computed precisely and is subtracted from the total possible share, hence the indicator function in Eq.(24)-(25).

The total profits to maximize for each bank are then convex in strategies:

$$P_1(G) = M_1(G) * S(u_1(x_2, u_2))$$

$$P_2(G) = M_2(G) * S(u_2)$$

Nash equilibrium exists per the usual Berge's maximum theorem. The second bank maximizes a continuous function over a compact set. By Berge's theorem, the maximizer is an upper semicontinuous function of bank 1's choice. In turn, bank 1 is maximizing an upper semicontinuous function over compact domain. Though maximizing solutions and the sequential Nash equilibrium exists, the overall solution, in general, is not upper semicontinuous in parameters. Graphs in parameter L, for example, will not be necessarily smooth.

We also suggest an "attenuation function" in an aggregate logit model for the relative bank value offer advantage at location x. This function is defined based on the distance of averaged bank's offers from the maximum possible utility as follows:

$$\mathfrak{M}(u_1(x_2, u_2), u_2) = \frac{\Delta}{exp(u_{max} - \overline{L}/2 - (u_1(x_2, u_2) + u_2)/2))}$$

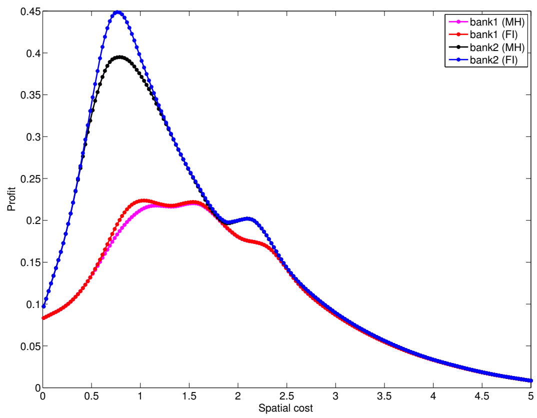
where Δ is a scaling parameter for logit tail mass $\mathfrak{M}$ and u_{max} is a maximum utility possible for specific contract problem. When both banks post offers that are close to be surplus zero, the tail mass automatically increases making sudden movement less critical for survival of the bank. This tends to smooth comparative statics as cost L is varied. When we put relatively more mass in the tail, the change in utility offer or location of the bank doesn't affect the market share much. It prevents banks from engaging in price war with Bertrand outcome, trying to jump over each other's location and utility offer, destroying both entities in the process. We call this regime a viscous competition with policies; the agents tend to stick with a particular bank nor matter what it does. This tends to help competitive statics as cost L is varied.

When intensity of competition is low, on the other hand we put less mass in the tail, allowing for sharper movements in policies since that would prevent banks from becoming local monopolists. We call this regime a "frozen ice" competition with policies that are sensitive to an underlying specifications. Even the slightest change of offer or bank location can cause fast movements in payoff functions .

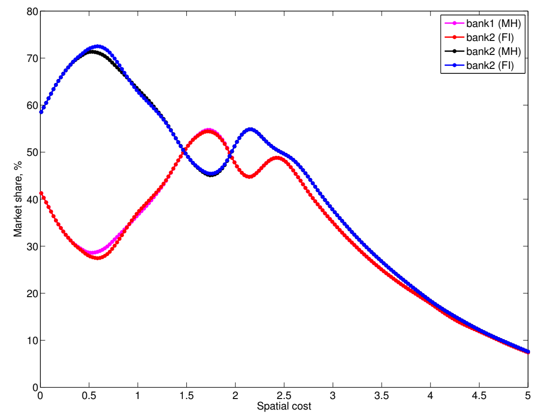
We proceed with numerical examples showcasing the developed logit approach.

3.4.2 Competition with full commitment and attenuated logits: comparative statics of MH vs FI as spatial cost L is varied

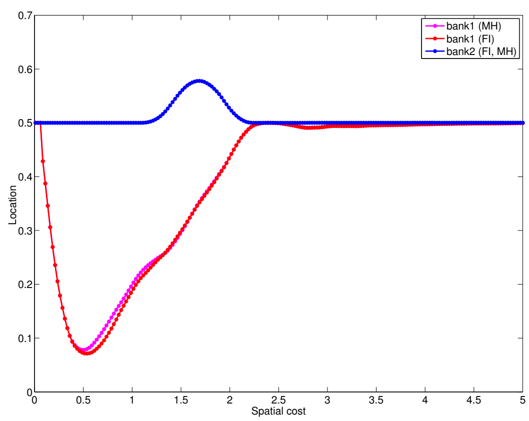
Results for sequential Nash equilibrium computations are presented at Fig.16, these results are smoothed by projection pursuit regression (PPR). This non-parametric regression method allows us to get the mean response to slight jumps in underlying spatial costs, imitating a real data scenario with mismeasured spatial costs without assuming any particular structure for the errors.



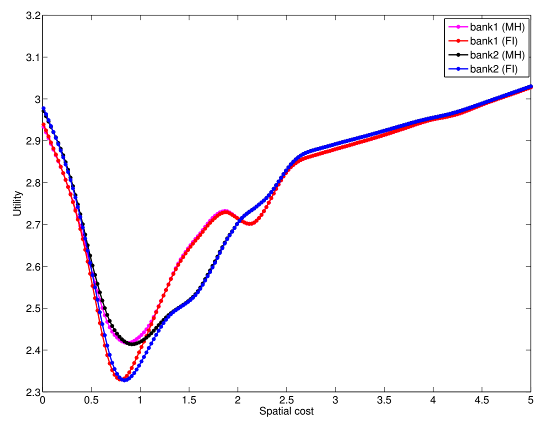
(a) Profit



(b) Market share



(c) Location



(d) Utility

Figure 16: Logit attenuating formulation, $\Delta = 1/10$, FI and MH, full commitment: profits, market shares, location and utility choices

The FI and MH choices are attenuated with endogenously derived tail mass function shown at Fig.17. It monotonically increases at $L > 1.5$ to prevent banks from becoming local monopolists. An attenuated competition takes place over the whole range of spatial costs with no extremes of Betrand competition or spatially separated competitors who become local monopolists.

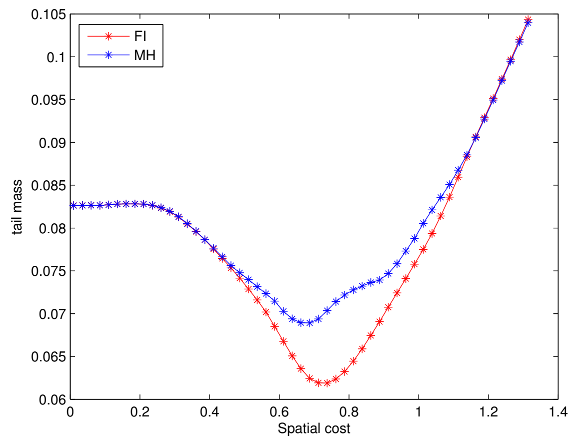


Figure 17: Logit attenuating function (LAF), $\Delta = 1/10$, FI and MH, full commitment

The profits are asymmetric, the second bank has larger profits compared to bank 1 - this is the first case where we observe a clear "second mover" advantage at low and intermediate spatial costs. The second entrant has relatively larger advantage in profit under full information compared to moral hazard. The presence of moral hazard smooths out the profit difference over the spatial cost range due to a flatter surplus frontier. The preempting spatial separation to the margin of the market at low spatial costs prevents the incumbent from being outcompeted completely. The central location is preferred by both banks at high spatial costs, hence we observe no local spatially separated monopoly class of solutions in this formulation.

This case presents both the richest dynamics in moments and in policy choices with both location and utility offers spanning the most parts of underlying compact domain under spatial cost scans. We consider it empirically relevant, and it also establish direct link from our formulation to methods widely used in IO literature.

4 Relationship banking

Here we consider another case of full commitment on contracts and locations choice with banks operating in different information regimes. There are two types (riskier and safer) among the agents. No logit mapping is used. Rather the information asymmetry is shown to be an important force for comparative static stability.

The first entrant (a local bank or an incumbent) comes to the area and studies his customers long enough to gain full information about their true types, as if establishing relationships. The first entrant anticipates that there will be another player in the area (a global bank or a challenger), and it commits to location and contract menu based on information and established relationships, trying to prevent the challenger from taking over his market share. The global bank doesn't have the information advantage that the first entrant possesses, and he operates in adverse selection regime with truth-telling constraints restricting the optimal menu of contracts for types. The global bank does have the second mover advantage in a sense that he already knows both location and contracts offered by the local bank. It is capable of undercutting the competitor on utility offer if such strategy were profit maximizing.

(a) Market shares of bank1 (FI) (b) Market shares of bank2 (AdS)

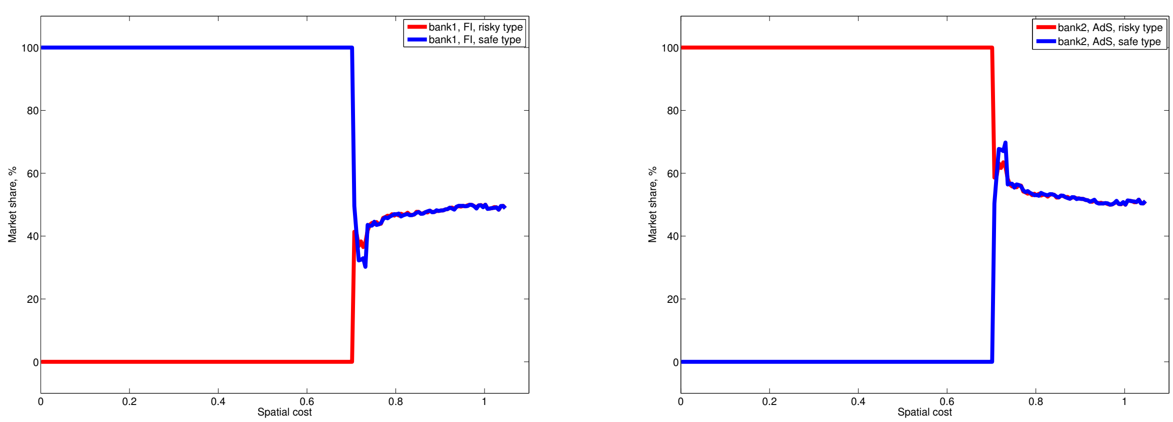


Figure 18:

In Fig. 18 we observe the remarkable result that at low spatial costs under the assumed information asymmetry, the incumbent gets exactly 100% of good (safer) types and the uninformed challenger gets exactly 100% of bad (riskier) types. The global bank specializes in what can be called "subprime lending", the local bank keeps relationships established with the better clients with no subprime activity on the books. The global bank makes higher profits and it is located in the middle of market; the incumbent gives up on more preferable central location by preemptively moving away from the center. In this sense the incumbent can be viewed as a small regional bank.

When spatial costs are low enough, the local bank increases the gap between contracts for riskier and safer types. The local bank stops caring about attracting bad types, and it raises the utility offer for safer clients in such way so as to make the global bank incapable of taking over by offering high utility for them. Since the global bank is constrained by truth-telling conditions, it can't customize the contracts in the way that full information bank does. If the global bank sets his menu of contract so as to compete on safer type segment, he would overpay the riskier type that are intertwined with safer types. As a result, the global bank outcompetes the local bank on riskier types by offering utility for riskier types just above the one posted by first entrant. It loses completely, however, on good (safer) types. The global bank is more than compensated by intermediating riskier types with relatively higher profits as the local bank chooses not to compete for this segment of market.

The local bank posts a nominal offer for riskier type (say, at autarky level) without entering into serious competition with global bank. For example, this offer can also be set at previous local monopoly value posted at time period before global players enter. The local bank indirectly sets the level of profits for the global player by posting a nominal offer for riskier types which serves as a reservation utility when those types are intermediated by a global player. The local bank, however, should be careful not to post an offer for riskier types that is high enough to become a real offer with new non-optimal equilibrium resulting in lower profits both both players[23]. If the local (or global) regulator is to insist on fully competitive price posted by the local player for riskier types, then the global player is likely not to survive, so after a transient welfare improvement the local bank would be left as a local monopolist.

With raising spatial cost the equilibrium changes from type-separating to a pooling with market shares of each banks containing a mix of types. The profit of the incumbent rises since he can more efficiently set the menu of contracts in autarky bounding regime while the challenger loses surplus by overpaying for bad types.

Note that such an outcome is possible only in full commitment market structure, where it is mutually beneficial for both banks to use a price discrimination strategy across market segments. With partial or no commitment, the full information incumbent can outbid the adverse selection entrant by posting a better structured menu of contracts, that is not constrained by truth-telling conditions, while staying at strictly profits positive level. Thus, the information-constrained entrant can not survive against full information incumbent at low spatial cost unless, as in full commitment case, the full information entrant pre-commits to mutually profitable market segmentation.

In our last setting we postulate that the first-mover has an information advantage, knowing firm types whereas the second entrant suffers an adverse selection problem. Here even without logits we get salient and consistent patterns. At a low range of spatial costs, the informed incumbent ends up dealing exclusively with the safer type, and the entrant with the risky type. That said, the incumbent, anticipating the ultimate competition, has moved toward the hinterland and makes lower profits in the end. For provinces with intermediate spatial costs, there is less specialization, but now it is the incumbent taking more of the risky types. At yet higher spatial costs the two banks begin to mirror each other; the incumbent information advantage disappears.

[23]For example, the local bank can post autarky level offer for the risky type with no intention to capture a market share. It also can offer higher utility up to the point where it actually can capture a market share leading the second bank to start competing for both risky and safe types. There is potential multiplicity issue here for the first player in the level of utility choice for the risky type, but not for the second player, since each risky-type offer with the same total profit for the local bank causes well-defined and unique response from the second bank affecting his level of profit. The local bank can either accommodate the global player by setting his offer at autarky or indirectly hurt the global player by choosing the highest possible profit-indifferent offer for risky types

5 Extensions and Conclusions

As noted in the introduction, our goal, ultimately, is to make this framework with its associated toolkit empirically operational. This provokes a number of remarks here in this conclusion on the generality of what we have done, and next steps.

In terms of the underlying environment, we have focused the analysis largely on two firms, each with one branch. In the appendix we present some numerical calculations where we allow more potential entrants and fixed costs to enter. This is a more typical IO framework, but examples make the point that the market may sustain relatively few firms in equilibrium, sometimes only two, perhaps three. Related, two financial firms each with two branches, established sequentially, may mean the second entrant established only one as in case of full commitment example with in the appendix. The implicit point is that we can continue to use our methods to compute equilibria with more than two firms.

The underlying environment also assumes the stereo-typical Hotelling line. Here three comments are in order. First, we did not assume the Salop's circle since we do want to consider the location of branches in the hinterland, i.e. centrally located or on the fringe, though granted, this can bring with it the existence problem of D'Aspremont et al. (1979), Shaked and Sutton (1982) and others have recognized. As a consequence we are careful to check on this as part of the algorithm, so to speak. Second, we can do the analysis in R^2, a more realistic geography. However, there is of course discretion in move in multiple directions, rather than one, and computations are more delicate. Third, and more realistic for empirical applications, we can restrict branches to a potentially small number of locations, as for example actual location of branches over the period of time observed, initially and then expansion, as in the paper of Assuncao et al. (2012). This still allows households in villages in discrete locations to travel to the bank (or for the bank to come and transact or monitor them at home, so to speak). This speeds up computations enormously, and frequently resolves non existence issues. With these advantages one might ask why we did not go this route. The answer is that we are intent in the paper in establishing the tradeoff of movements in homogenous space vs movement in contract terms, so we can see how equilibria vary as we do comparative statics, varying travel costs and not concealing non monotonicities and jumps.

In turn, we can ask empirically how spatial locations and contracts vary as we move across counties/provinces, regions with have good and bad infrastructure/transport and communication costs. Indeed this is how one can begin to think about identification, since that cost parameter determines the extensive margin market elasticity, and in varying it, we are moving along the surplus frontier. This most clear in the case of monopoly, but the logic extends to Nash competition. Bertrand-like forces make utility to the households higher the lower is the spatial cost, as financial firms compete harder. Thus the equilibrium distribution of promised utilities is endogenous, though a given, fixed distribution has implications for contracts and observables. Indeed, holding spatial costs fixed, equilibrium promised utilities will vary with information obstacles and the nature of assumed competition (sequential, full Nash etc). One can compare graphically the figures in the paper which have spatial cost on the x axis and consumption, effort, investment, insurance, credit, branch location, profits, and market shares as dependent variable outcomes, and begin to think about identification. In contrast, but closely related, Karaivanov and Townsend (2014) are silent about the industrial organization of financial service providers and take the distribution of promised utilities as an unobserved initial condition, which needs to be estimated. Parameters of that distribution, along with parameters of technology and preferences are recovered well, along with the underlying true financial regime (obstacle) that generated the data, even with nontrivial measurement errors in the observables, according to Monte Carlo simulations.

We have in this paper focused consistently on exclusive contracts; a household chooses one bank or another, though the latter are in competition with each other. There is a potential multiplicity of equilibrium issue. This turns up in the adverse selection model with menus and latent contracts, but there we understand the economics of what is going on and how to impose reasonable criteria for what we might expect to observe. In other settings the multiplicity is either obvious, as in local monopoly having continuum of location choices, or non-detectable.

Moving to common agency models as in Bernheim and Whinston (1986) is a huge step we do not embark on in this paper, though our methods do help to think about the problem.

We have not in this paper focused much on true dynamics. Virtually all of this paper could be extended to include explicit dynamics, i.e. multi-period contracts for borrowers, not simply dynamic IO considerations. On the household side, one need only put next period's promised utility as an additional control variable. Karaivanov and Townsend (2014) is about likelihood estimation with this structure. Equilibrium outcome will be information-constrained Pareto efficient, if there is perfect costless long term commitment on all sides. This allows both moral hazard and a new kind of adverse selection (ex-post hidden types) simultaneously. However, in the context of sequential competition among a local incumbent and new entrant, it seems more likely that households cannot commit to stay with the local bank, especially if there is a limit in the extent to which bank can front-load the contract. This would be true either if the new entrant comes in as a surprise move, unanticipated, or if the incumbent anticipates. Under the latter, the outcome might not be information-constrained efficient, as the first bank cannot take full advantage of long term contracts which are otherwise beneficial. The surplus Pareto frontier over a range of contracts might have spots where the previously optimal contract dominated. This is the first instance then in the paper where the IO equilibrium can be inefficient and may require some regulation, fine tuned, to deal with this problem.

Acknowledgements

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575. We acknowledge the University of Chicago Research Computing Center for support of this work.

References

Agarwal, S. and R. Hauswald (2010). Distance and Private Information in Lending. *Review of Financial Studies 23*, 2757–2788.

Alessandrini, P., F. F. Presbitero, and A. Zazzaro (2009a). Bank size or distance: what hampers innovation adoption by SMEs? *Journal of Economic Geography 10*, 845–881.

Alessandrini, P., F. F. Presbitero, and A. Zazzaro (2009b). Banks, Distances and Firms' Financing Constraints. *Review of Finance 13*, 261–307.

Armstrong, M. and J. Vickers (2001). Competitive price discrimination. *RAND Journal of Economics 32*(4), 579–603.

Assuncao, J., S. Mityakov, and R. Townsend (2012). Ownership Matters: the Geographical Dynamics of BAAC and Commercial Banks in Thailand. Working paper.

Bernheim, B. D. and M. D. Whinston (1986). Common Agency. *Econometrica 54*(4), 923–942.

Brook, Y., R. Hendershott, and D. Lee (1998). The Gains from Takeover Deregulation: Evidence from the End of Interstate Banking Restrictions. *The Journal of Finance 53*(6), 2185–2204.

Butler, A. W. (2008). Distance Still Matters: Evidence from Municipal Bond Underwriting. *Review of Financial Studies 21*(1), 763–784.

Dasgupta, P. and E. Maskin (1986). The Existence of Equilibrium in Discontinuous Economic Games. *Review of Economic Studies 53*(1), 1–26.

D'Aspremont, C., J. J. Gabszewicz, and J.-F. Thisse (1979). On Hotelling 'Stability in Competition". *Econometrica 47*(5), 1145–1150.

Degryse, H. and S. Ongena (2005). Distance, Lending Relationships, and Competition. *Journal of Finance 60*, 2757–2788.

Demyanyuk, Y., C. Ostergaard, and B. E. Sørensen (2007). U.S. Banking Deregulation, Small Businesses, and Interstate Insurance of Personal Income. *The Journal of Finance 62*(6), 2763–2801.

Doraszelski, U. and A. Pakes (2007). A Framework for Applied Dynamic Analysis in IO. In *Handbook of Industrial Organization*, Volume 3. Elsevier.

Green, E. J. (1987). Lending and the Smoothing of Uninsurable Income. In E. C.Prescott and N. Wallace (Eds.), *Contractual Arrangements for Intertemporal Trade*. University of Minnesota Press, Minneapolis.

Hotelling, H. (1929). Stability in Competition. *The Economic Journal 39*(153), 41–57.

Karaivanov, A. and R. Townsend (2014). Dynamic Financial Constraints: Distinguishing Mechanism Design from Exogenously Incomplete Regimes. *Econometrica 82*(3), 630–49.

Koijen, R. S. and M. Yogo (2012). The Cost of Financial Frictions for Life Insurers. *NBER Working Paper No. 18321*.

Mahoney, N. and E. G. Weyl (2014). Imperfect Competition in Selection Markets. SSRN 2372661.

Martin, A. and F. Taddei (2012). International capital flows and credit market imperfections: A tale of two frictions. *Journal of International Economics*, 441–452.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995). *Microeconomic Theory.* Oxford University Press.

Moll, B., R. M. Townsend, and V. Zhorin (2013). Financial Obstacles and Inter-Regional Flow of Funds. *NBER Working Paper No. 19618*.

Nevo, A. (2000). A Practitioners Guide to Estimation of Random-Coefficients Logit Models of Demand. *Journal of Economics and Management Strategy 9*(4), 513–548.

Petersen, M. A. and R. G. Rajan (1995). The Effect of Credit Market Competition on Lending Relationships. *The Quarterly Journal of Economics 110*(2), 407–443.

Phelan, C. and R. M. Townsend (1991). Computing Multi-Period, Information-Constrained Optima. *The Review of Economic Studies 58*, 853–881.

Prescott, E. C. and R. M. Townsend (1984). Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard. *Econometrica 52*, 21–46.

Prescott, E. C. and M. Visscher (1977). Financial Obstacles and Inter-Regional Flow of Funds. *The Bell Journal of Economics 8*(2), 378–393.

Rajan, R. G. and M. A. Petersen (2002). Does Distance Still Matter? The Information Revolution in Small Business Lending. *Journal of Finance 57*(6), 2533–2570.

Shaked, A. and J. Sutton (1982). Relaxing Price Competition Through Product Differentiation. *Review of Economic Studies 49*(1), 3–13.

Spear, S. E. and S. Srivastava (1987). On Repeated Moral Hazard with Discounting. *Review of Economic Studies 54*(4), 599–617.

Appendix

A Pareto optimality for single-branch monopoly

For all regimes with the exception of adverse selection the problem separates per types θ, we will suppress θ notation here for reading convenience, all results hold for all types θ independently in case of full information and moral hazard regimes.

Proposition A.1. *Let $S(u) * \mu(u)$ be the total profit for one branch monopoly at location x_m. Suppose that monopoly profit is at maximum and that monopoly doesn't cover the whole market. Then any movement in utility offer u will decrease monopoly profit. So, by the Pareto principle we can not make someone better of without making someone else worse off.*

Proof. Since the monopoly does not cover the entire market, there are numerous locations that maximize profit. For example, location $x_m = 1/2$ is fine without loss of generality. Hence there is only one argument to choose, w. From FOC wrt to utility promise

$$\frac{\partial\left[S(u) * \mu(u)\right]}{\partial u} = S'(u) * \mu(u) + S(u) * \mu'(u) = 0$$

From here we derive useful equality for surplus and market share elasticities

$$\frac{S'(u)}{S(u)} = -\frac{\mu'(u)}{\mu(u)} \tag{26}$$

The second-order condition is written as

$$S''(u)\mu(u) + 2S'(u)\mu'(u) + S(u)\mu''(u)$$

Let's consider terms in this expression:

$$S''(u) \leq 0 \wedge \mu(u) \geq 0 \Rightarrow S''(u)\mu(u) \leq 0$$

since surplus function is concave and market share is non-negative by definition.

$$S'(u) \leq 0 \wedge \mu'(u) \geq 0 \Rightarrow 2S'(u)\mu'(u) \leq 0$$

since surplus function is monotonically decreasing and market share is a linearly increasing function of u (follows from definition of market share) in case of monopoly.

$$\mu''(u) = 0 \Rightarrow S(u)\mu''(u) = 0$$

Thus second-order condition is negative. □

The surplus elasticity $\epsilon_S(u) = \frac{S'(u)}{S(u)}$ is downward sloping (see Fig.19), it asymptotically goes to $-\infty$ at $S(u) \to 0$.

The market share elasticity $\epsilon_\mu(u) = -\frac{\mu'(u)}{\mu(u)}$ is upward sloping or flat at zero level (when utility offer is high enough to cover the whole market, $\mu'(u) = 0$), it asymptotically goes to $-\infty$ at $u \to \hat{u_0}$. Thus, equation (26) is a condition where surplus elasticity function intersects with market share elasticity function at optimal (profit maximizing) utility level. The surplus elasticity is independent of spatial costs $\overline{L}$, so the increase in $\overline{L}$ moves market share elasticity to the right as will be shown in examples later. To anticipate, as $\overline{L}$ moves we trace out the market share elasticity function for this monopoly problem.

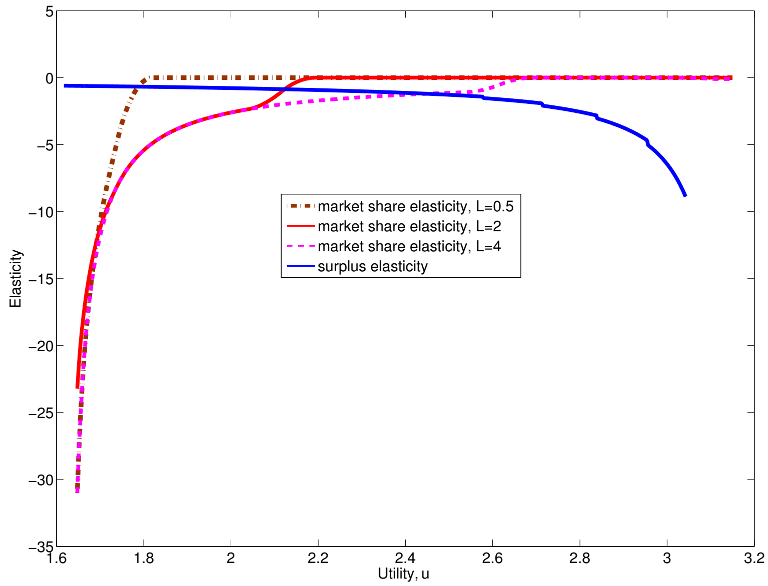


Figure 19: Surplus elasticity vs. market share elasticities at different spatial costs L

Proposition A.2. *If monopoly covers the whole market and monopoly profit is at maximum. Then no change in location x_m or utility u is possible without hurting optimal monopoly profit at optimal monopoly utility and location. However, when the spatial cost $\overline{L}$ rises, if the monopolist is to maintain 100% of the market, then utility will have to be increased to retain this marginal customer at his autarky value.*

Proof. Since in this case monopoly still wants the marginal customer who is located at maximum distance $x = 0$ from monopoly location $\overline{x_m} = 1/2$ (without loss of generality) it offers minimum utility $u = \hat{u_0} + L/2$ to attract that borderline customer. Any higher offer would hurt monopoly surplus without increasing market share any further. Any deviation $\delta(x_m)$ from the central location to the left will cause pure loss in value for the right segment of the market and vice versa.

□

Theorem A.3. *(Equlibrium contract for local monopoly)*
For a given contract problem the eqilibrium contracts for all market structures at high spatial costs converge at spatial cost independent utility value u^* equal to local monopoly eqilibrium contract with autarky islands (market is not fully covered).

Proof. With downward sloping surplus defined on compact set and unbounded spatial cost disutility there exist a spatial cost value at which the market is not fully covered.

First, we derive the optimal utility value offered by monopoly when market is not fully covered. Since there are islands of agents (see Fig.4) in autarky the marginal profit from attracting customers from autarky equals the average costs of paying existing customers extra utility du . If existing customers are located within range x from monopoly location then

$$u = u_0 + L * x/2$$

or

$$x = \frac{2(u - u_0)}{L} \tag{27}$$

and monopoly total profit

$$S(u) * x = \pi^*$$

Since monopoly is neutral between raising utility offer and staying with profit π^*

$$S(u)dx + S'_u|_{u=u^*} x du = 0$$

We use Eq. 27 for x and substitute $dx = \frac{2du}{L}$ to get the condition from which the optimal spatial contract with utility u^* can be obtained. It does not depend on L.

$$-\frac{S'_u|_{u=u^*}}{S(u^*)} = \frac{1}{u^* - u_0} \tag{28}$$

□

Corollary A.4. *(Pareto Optimality of single-branch monopoly contract)*
Let $S(u) * \mu(u, x_m)$ be a total profit for one branch monopoly. Then monopoly equilibrium is Pareto optimal.

Proof. Immediately follows from Propositions A.1 - A.2 and Theorem A.3 Either the monopoly profit or the agent's utility will be harmed at any deviation from equilibrium contract or location. □

B Distance to Nash

Here we propose the following conservative technique to order by rank all possible strategies with metric we call "distance to Nash". Both banks enter the market simultaneously and use a strategy set of $G = \{u_1^*, u_2^*, x_1^*, x_2^*\}$ with payoff $P_1(G) = S(u_x^*)\mu_1(u_2^*, x_2^*, u_1^*, x_1^*)$ and $P_2(G) = S(u_2^*)\mu_2(u_2^*, x_2^*, u_1^*, x_1^*)$.

If a strategy set G is a true Nash equilibrium then there exist no deviating strategy sets $G_1 = \{u_1, u_2^*, x_1, x_2^*\}$, $G_2 = \{u_1^*, u_2, x_1^*, x_2\}$ that would provide higher profit for any of the correspondingly deviating banks taking other bank strategy as given. We can define and compute for any of those deviating strategies the following metrics

$$d(G, G_1) = \max(P_1(G_1) - P_1(G), 0), P_1(G_1) = S(u_1)\mu_1(u_2^*, x_2^*, u_1, x_1)$$

$$d(G, G_2) = \max(P_2(G_2) - P_2(G), 0), P_2(G_2) = S(u_2)\mu_2(u_2, x_2, u_1^*, x_1^*)$$

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

$$\underset{G1}{\text{maximize}}\ d(G, G_1) \text{ subject to } P_1(G) > 0,\ \forall G_1. \tag{29}$$

$$\underset{G2}{\text{maximize}}\ d(G, G_2) \text{ subject to } P_2(G) > 0,\ \forall G_2. \tag{30}$$

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

$$\underset{G}{\text{minimize}}\ d(G),\ \forall\{G, G_1, G_2\}. \tag{31}$$

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.

B.0.3 Numerical accuracy of distance to Nash algorithm

When distance to Nash is Lipschitz bounded[24] $d(G) < \lambda * P_{1,2}(G)$ we accept the outcome as an instance of Nash eqilibrium. We conduct the same accuracy checks for each case of simultaneous Nash equilibrium we study. Although we don't provide proofs of existence and sufficiency conditions here, those checks serve to filter numerically well-bounded constructively obtained equilibria from outcomes where Nash eqilibrium 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.

C Costly entry, multiple entry, multi-stage entry

In Table 2 we report results of sequential (full commitment) Nash equilibrium with entry cost. At low spatial cost, banks are in Bertrand competition making essentially zero net profits (the gross profit is just enough to cover the entry cost). The entrants can't move away from a central position to increase profits by avoiding price competition because profits are not going to be high enough to cover the entry cost. With larger spatial cost, the first entrant can prevent a new entry by posting utility that is high enough to prevent a new entrant from absorbing the entry costs. The market mass is too low at high spatial costs to support two entry costs because possible surplus level is bounded by participation constraints (autarky level), the level of utility has to be relatively high.

In Table 3 we report results of sequential Nash equilibrium for three entrants. The second entrant is "sandwiched" between the first and the last. This is not enough to support another entrant.

Let's consider a sequential entry of two banks that at entry select locations $\overline{x_1^a}, \overline{x_1^b}$ and x_2^a, x_2^b simultaneously for two branches. Bank 1 enters first, then bank 2 enters.

Each bank also offers utilities $\overline{u_1^a}, \overline{u_1^b}, u_2^a, u_2^b$ simultaneously at each of its branches.

Results are in Table 4. It can be seen that even with two branches per each entrant the market is saturated already, not all branches are profit positive.

[24]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.

Spatial Cost $L \in [0,1)$ (entry cost absorbed)				
bank	location	utility offer	market share	profit
1	center	max to cover entry	50%	0
2	center	max to cover entry	50%	0
Spatial Cost $L = 2$ (deterrence of entry)				
bank	location	utility offer	market share	profit
1	center	2.4 (to prevent entry)	83%	0.18
2	border	2.4 (doesn't enter)	0%	0

Table 2: SNE with entry costs, entry cost $c_E = 0.3$

Spatial Cost $\overline{L} = 2$				
bank	location	utility offer	market share	profit
1	0.6375	2.3555	49%	0.3132
2	0.3750	2.3655	51%	0.3224
bank	location	utility offer	market share	profit
1	0.6250	2.3746	43.5%	0.2723
2	0.5000	2.3759	6.5%	0.0406
3	0.3750	2.3759	44%	0.2750

Table 3: SNE with multiple entrants

Spatial cost = 1

	Bank 1 branch 1	Bank 1 branch 2	Bank 2 branch 1	Bank 2 branch 2
Utility offer	1.8500	2.3463	2.3056	2.3463
Share of market	0%	40%	56.5%	0%
Location	0.0694	0.6538	0.5125	0.5500
Profit	0	0.2536	0.3717	0

Spatial cost = 2

	Bank 1 branch 1	Bank 1 branch 2	Bank 2 branch 1	Bank 2 branch 2
Utility offer	2.3457	2.1568	2.4045	2.3000
Share of market	27.5%	26%	46.5%	0%
Location	0.8428	0.0678	0.5750	0.5500
Profit	0.1744	0.1922	0.2788	0

Table 4: SNE with simultaneous branch opening, full committment

D Full and no commitment with counterfactuals

D.1 Convex out of equilibrium response, no logits, full commitment

Let's consider what happens if the first entrant makes an error in his choice of location x_1. We compute the best responses over all out of equilibrium range of locations, shown at Fig. 20. Out of equilibrium profit functions are convex for $L \in (\underline{L}, \infty)$. That provides perfect identification for locations and utilities. At higher spatial costs the optimal location lies in continuous interval $x_1 \in (\underline{x}, \frac{1}{2})$, the profit surface becomes less steep in location choice and eventually bank1 becomes a local monopoly

(a) Profit at $L = 1.8$ (b) Profit at $L = .5$

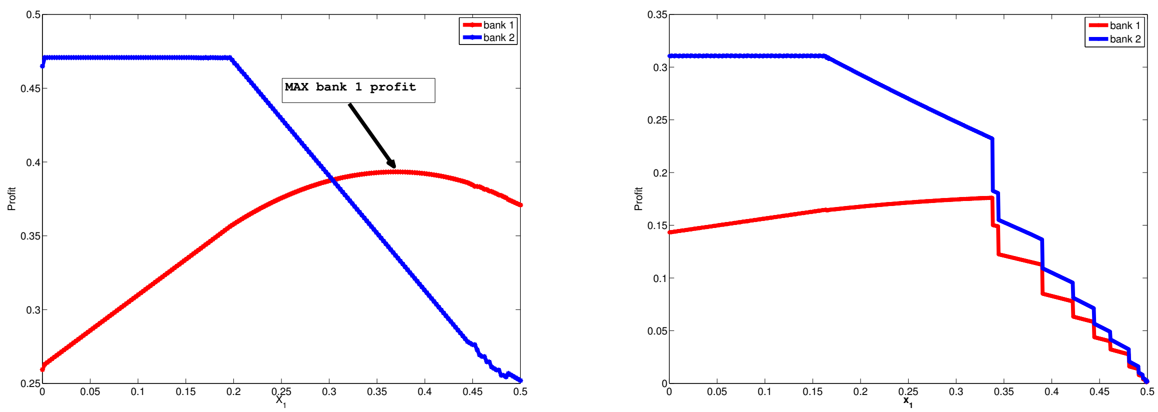


Figure 20: Out of equilibrium responses for full commitment

D.2 No commitment with counterfactuals: Nash surface with logits

We compute the no commitment Nash solution for fixed tail mass logits. In Fig.21 we graph the out-of-equilibrium Nash surface at all possible strategies of bank 1 deviating from Nash. Those surfaces are fully convex at non-zero tail mass at low to medium spatial costs. At larger spatial cost the spatial separation strategy becomes optimal. Fig.21(a) illustrates typical Nash equilibrium with middle locations at equilibrium which is a stable strategy over a large range of spatial costs and tail mass values $\mathfrak{M}$, Fig.21(b) shows a ridge in the surface that causes spatial separation under larger spatial costs, results are generally applicable in $\mathfrak{M} > 0, L > 0$ domain

(a) $L = 0.5, \mathfrak{M} = 0.1$ (b) $L = 2, \mathfrak{M} = 0.3$

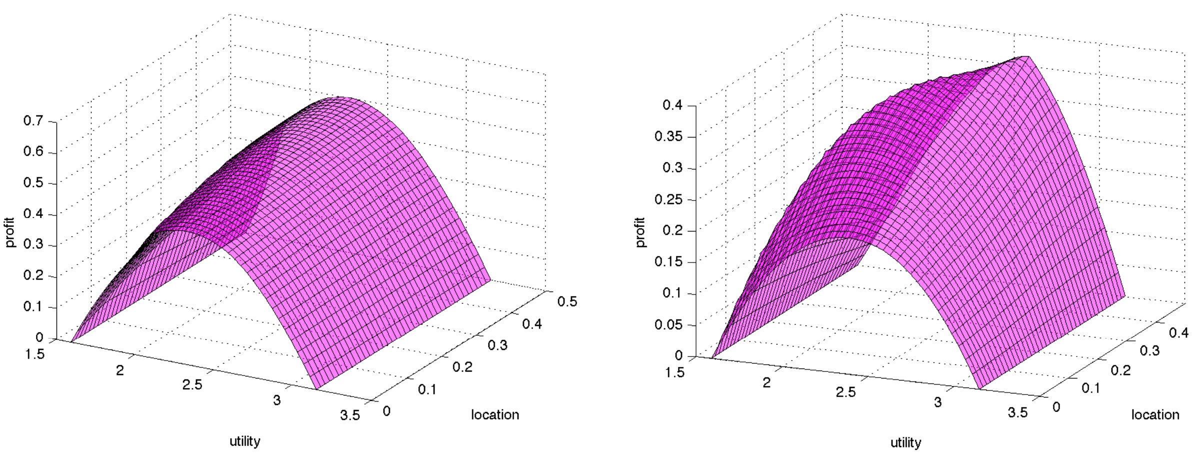


Figure 21: Out of equilibrium responses for no commitment