Informal land markets and ethnic kinship in sub-Saharan African cities

Lucie Letrouit[*], Harris Selod[†]

September 13, 2023

Abstract

We present an urban land use model with land tenure insecurity and information asymmetry regarding risks of contested land ownership, a very common issue in sub-Saharan African cities. A market failure emerges as sellers do not internalize the impact of their market participation decision on the average quality of traded plots, which in turn affects other sellers and buyers' decisions: The equilibrium is suboptimal and has too many transactions of insecure plots and too few transactions of secure plots. This market failure can be addressed when agents trade along trusted kinship lines that discourage undisclosed sales of insecure plots. Such kinship matching is an important feature of West African societies, including on the market for informal land, as illustrated by a unique survey administered in Bamako, Mali. In the model, the extent to which the market failure is addressed increases with the intensity of kinship ties. When sellers also have the possibility of registering their property right in a cadastre, this not only further attenuates information asymmetry but also helps reduce risk. We find complementarity between kinship matching and registration: As transactions between kins tend to involve plots that are more secure on average, kinship matching makes registration better targeted at insecure plots traded outside kinship ties. In this context, a fully-funded partial subsidy to cover the registration fee can bring the economy to the social optimum.[1]

Keywords: Land markets, property rights, information asymmetry, informal land use, land registration, ethnic kinship

JEL classification: P14, Q15, R14

[*]Gustave Eiffel University, 16 Bd Newton, 77420 Champs-sur-Marne, France. Email: lucie.letrouit@univ-eiffel.fr

[†]The World Bank, 1818 H Street NW, Washington, DC 20433. Email: hselod@worldbank.org.

[1] We are grateful to two peer reviewers for comments that helped us substantively improve the paper as well as to Amadou Cissé for the many discussions about social structures in West Africa that initially motivated this paper, to Demba Karagnara who coordinated field surveys in Bamako in 2012 and 2022, and to Pierre M. Picard, Jan Brueckner and Tony Yezer for suggestions and technical comments on the model. We are also grateful to Eliana la Ferrara and Karen Macours for useful insights on kinship and land markets in developing countries, as well as to Thierry Verdier, Miren Lafourcade and the participants to the Labor and Public Economics Seminar at the Paris School of Economics and to the International Conference of Development Economics for various suggestions. We gratefully acknowledge the funding of the World Bank and of the Paris School of Economics Mobility Grant, a public grant overseen by the French National Research Agency (ANR-18-CE22-0013-01), the Labex OSE and the Labor and Public Economics Group at the Paris School of Economics.

1 Introduction

In developing countries, informally holding land is more often the norm than the exception. In sub-Saharan African cities in particular, a large fraction of landowners—in some cases, a large majority—do not hold a formal property title. To a great extent, high levels of tenure informality tend to mirror the deficiencies of land registration systems in the region, which often remain prohibitively costly and unaffordable to most households (see, e.g., Ali et al., 2016, who show that registration costs deter registration in Dar es Salaam, Tanzania). The economic literature stresses that informal tenure is generally not desirable, as it comes with many ills and can have large private and socioeconomic costs. The risk of eviction associated with informal land can reduce investment in land (Besley, 1995) or discourage labor market participation due to the necessity of spending time guarding one's land plot (Field, 2007). Households residing on informal plots—which are often found in slums—are also exposed to a wide range of externalities, including crime, poor health from low housing quality, and associated negative human capital externalities (Galiani and Schargrodsky, 2010, Galiani et al., 2017, Nakamura, 2017). Additionally, informal land tenure also hinders the tradability of land, possibly leading to land misallocation and loss of economic efficiency (Chen et al., 2022, Gottlieb and Grobovsek, 2019).

In developing country cities, informal land markets are mainly characterized by risks of contested ownership and information asymmetry regarding those risks. This is modeled by Lanjouw and Levy (2002), who show that buyers and sellers of urban land may respond to weak property rights by transacting among "family and friends" who share information on risks. The authors find empirical validation of their theory in the cities of Quito and Guayaquil, Ecuador. Similarly, in rural tenancy markets in the Dominican Republic and in Guatemala, Macours et al. (2010) and Macours (2014) find that households resort to ethnic matching strategies in response to risks associated with informal property rights. Recent land market studies on Bamako, Mali, and Yaounde, Cameroon, also report qualitative evidence of land transactions occurring between trusted parties as a way to address information asymmetry regarding tenure risks (see Durand-Lasserve et al., 2015, and World Bank, 2020).

The objective of this paper is to study how matching along ethnic kinship lines endogenously emerges in response to tenure insecurity in contexts of incomplete property right systems, a salient feature of sub-Saharan African cities.[2] For this, we develop a theoretical framework according to which purchasing informal land is risky for buyers, as their ownership of the purchased plot might be contested in the future in the absence of a property title. The risk of contested land ownership can be eliminated through registration of ownership in a registry or cadastre, leading to the issuance of a fully secure property title. This solution, however, is costly and may only be chosen by a fraction of the population. The rest of the population will be acquiring non-registered land but without having information on the plot's intrinsic risk that is known only to the seller. Among these households, some may decide to transact along trusted ethnic relationships, which reduces the information asymmetry between buyers and sellers and decreases the likelihood that buyers end up unknowingly purchasing insecure land plots. We model this idea in an urban land use framework with both tenure insecurity and information asymmetry, where we study equilibrium land market transactions and associated inefficiencies. In our framework, informal plots are of two types: risky (insecure) plots, whose ownership might be contested in the future, and risk-free (secure) plots, whose ownership cannot be contested. When selling a plot, sellers do not inform buyers of the intrinsic risk associated with the plot. However, buyers and sellers of land plots may have reciprocal duties based on trusted ethnic kinship: If a risky plot is exchanged between individuals who are linked by ethnic kinship, the seller is considered to have violated his duty and a social penalty will be imposed on him. Such social penalties are at the core of interethnic relations and have been observed in informal settlements as a means to deter land conflicts (see, e.g., Adam 2014, in the case of peri-urban areas in Ethiopia). In our context, a buyer will be ready to pay a premium when transacting with a seller he is ethnically connected with. This is because the buyer will expect the seller to be more likely to sell him a secure plot rather than an insecure one, due to the threat of the social penalty. Knowing this, sellers may decide whether to transact with ethnically or non-ethnically related buyers, depending on the expected risk on

[2] In the paper, we indifferently use the terms kinship matching or ethnic matching to refer to matching between sellers or buyers of ethnic groups who claim to be related by kinship. See, e.g., Dunning and Harrison (2010) who use the term "ethnic cousins".

the plot, the exogenous social penalty, and the market-determined price premium for informal transactions along ethnic lines. This mechanism differs from those previously presented in the literature: With the introduction of a social penalty, we do not need to make the unrealistic assumption that groups of agents share the same information regarding risks as in Lanjouw and Levy (2002). Contrary to Macours et al. (2010) who focus on the risk that tenants in rental markets could squat on the agricultural land they rent from someone else, we focus on sales markets of land for residential use in urban and peri-urban areas and, in our model, the risk of losing the plot to another party is borne by buyers of land rather than by landlords renting out land. To our knowledge, our paper is the first to present an equilibrium theory of ethnic matching in informal urban land markets. It is also the first land use model with interpersonal transactions, an important feature that is largely missing in the theoretical literature on land markets in developing countries. The introduction of ethnic matching allows us to analyze the respective advantages of transactions sanctioned by the registration of property rights and of those made under ethnic matching, and to study the coexistence of the two practices within a single city, as commonly observed in sub-Saharan African urban contexts.

The paper is organized as follows. We start by presenting the literature our model relates to in Section 2, before presenting stylized facts on informal urban land markets and on the city structure of Bamako, Mali – a city that is representative of the sub-Saharan African context – in Section 3. We then present a benchmark monocentric urban economics model with tenure insecurity in Section 4. In Section 5, we introduce the possibility of buyers and sellers matching according to a trusted ethnic relationship. In the following section, we further add the possibility of registering property rights and study the impact of a registration subsidy. The final section concludes.

2 Literature review

Our paper is at the intersection of two main strands of literature. The first is the mainly anthropological literature on ethnic ties and reciprocal duties which studies the links among groups in a wide range of societies. Individuals from groups linked in such a way are referred to as “allies”, “kins” or “cousins”—in a figurative sense—and exhibit codified reciprocal duties along those links (Mauss,1923). These duties may take various forms, including the requirement to treat one another fairly or to exchange gifts such as food or shelter. Such links are very commonly found in sub-Saharan Africa, with academic publications covering to our knowledge the contexts of Burkina Faso, Burundi, the Gambia, Guinea, Mali, Rwanda, Senegal, Tanzania and Zambia (see Freedman, 1977, Ndiaye, 1992, Fouéré, 2004, Smith, 2004 and 2006, Diallo, 2006, Dunning and Harrison, 2010). In West and Central Africa, the social institution underpinning those links is referred to under the generic French term of “cousinage” (also translated as cousinage in English) and designates the social links between groups of so-called “cousins”, a term that we will use throughout the paper to designate matching along a trusted ethnic relationship.[3] These relationships often correspond to alliances between pairs of social groups defined by ethnicity, patronyms and/or the professions traditionally practiced by members of these groups.[4] These alliances “are set in stone by blood pacts or inviolable words of honor under penalty of discredit and banishment” (Attino, 2021).[5] Although cousinage relationships come from a very old tradition,[6] they are still widely used nowadays. In Senegal, for instance, it was assessed that 46 percent of Senegalese practice cousinage everyday and an additional 30 percent practice it occasionally (Smith, 2004). Although the anthropological literature is mostly focused on reciprocal social relationships, several authors mention the role of cousinage in markets as revealed by price bargaining along ethnic lines (see Hagberg 2006 for Burkina Faso, and Birkeland 2007 and Jones 2007 for Mali).

The second strand of literature that our model builds on involves the emerging urban economics literature on land tenure insecurity in developing countries. This literature focuses on analyzing the causes of informal tenure in developing country cities and its implications

[3] An alternative term for cousinage is “joking relationships” (in French, “cousinage à plaisanterie”), which refers to the codified jokes that individuals exchange upon their first encounter. Joking according to ritualized mocking allows to identify the nature and intensity of bilateral relationships and stresses reciprocal duties before parties engage in social interactions.

[4] For example, the Sérère and Poular are two ethnic groups that are considered to be ethnic cousins in Senegal and the Gambia. Ba and Diallo are two family names linked with a cousinage relationship in Senegal. In Mali, groups that traditionally practiced the profession of blacksmith are considered to be the “cousins” of all other traditional professions.

[5] Translation from French by the authors.

[6] In Mali, cousinage (known as “sanankuya”) is believed to have been ordained by the 13th century ruler Sundiata Keita, as part of the oral constitution of the Mali Empire.

for the functioning of urban land markets or its impact on welfare. The literature began with Jimenez's (1985) seminal model of urban squatting in which informal dwellers coordinate land invasions to protect themselves from evictions. Brueckner and Selod (2009) further studied the emergence of a city's squatter settlements in a general equilibrium with inelastic land supply. In their model, squatting "squeezes" the formal land sector, explaining the high price of formal land in an equilibrium configuration where formal and informal settlements coexist.[7]

Our paper, however, does not involve squatting whereby land is occupied without being purchased or rented out from its rightful owner. Instead, it focuses on the broader context of tenure insecurity whereby the occupant of a land plot may be its legitimate owner and yet not have a formal property right, exposing him to the possibility of competing claims and conflicts. A small number of recent models account for these issues in extensions of the standard monocentric land use model of urban economics initially developed by Alonso (1964), Mills (1967) and Muth (1969). In Selod and Tobin (2018), urban households compete for land and simultaneously decide the type of property right to purchase from a land administration among a menu of rights that provide various degrees of tenure security. The model leads to an equilibrium with formal and more secure property rights at the proximity of the city center, a prediction that also holds in our model. Cai et al. (2018) adapt the Selod and Tobin model to a calibrated dynamic stochastic model with internal migration that allows them to study the long term trajectory of formal and informal land uses and the persistence of informal settlements over time when agglomeration effects are not strong enough. Picard and Selod (2020) study the conversion of agricultural land into urban residences and the associated changes in land tenure. They introduce information asymmetry between buyers and sellers of risky plots—a feature that is also present in our model—and find that information asymmetry deters land market participation and hinders the land use conversion process at the periphery of the city. They also show empirically that information asymmetry is present in informal land markets in Bamako, Mali. Other spatial papers study specific types of informal housing. This is the case of Brueckner et al. (2019) who develop a theory explaining the emergence of a

[7]Brueckner (2013) further extended the model with the introduction of a rent-seeking organizer. Shah (2014) modified the model to account for squatting on public land. Turnbull (2008) proposed a non-spatial but dynamic model of the landowner-squatter relationship that focuses on the timing of evictions.

rental market for backyard structures in South African cities, and of Pfeiffer et al. (2019) who propose a dynamic land-use model with formal and informal housing, including traditional informal settlements as well as backyard structures. Using various modeling approaches, other recent studies have focused on the determinants of informal housing and urban slums, stressing the role of migration and of the relative elasticities of formal and informal housing supply in determining the amount of informal housing (Alves, 2021, Henderson et al., 2016, Henderson et al., 2021, Cavalcanti et al., 2019). Bird and Venables (2020) use an urban simulation model to show how formalization of traditional tenure can lead to a more efficient land use in the case of Kampala, Uganda—a result reminiscent of the welfare improving impact of formalization demonstrated in Brueckner and Selod (2009). It is important to note that, overwhelmingly, the existing literature focuses on impersonal markets (Arruñada, 2012). To our knowledge, the idea of interpersonal relationships in urban land transactions was only previously explored by Lanjouw and Levy (2002) who contrast land transactions between family members and land transactions between non-related parties and by Marx et al. (2019) who find that ethnicity affects the bargaining power of Kenyan slum dwellers over rents but who do not explicitly focus on issues of tenure insecurity and risk. To our knowledge, none of the papers with interpersonal transactions adopt a theoretical spatial setting as we do.

3 Stylized facts

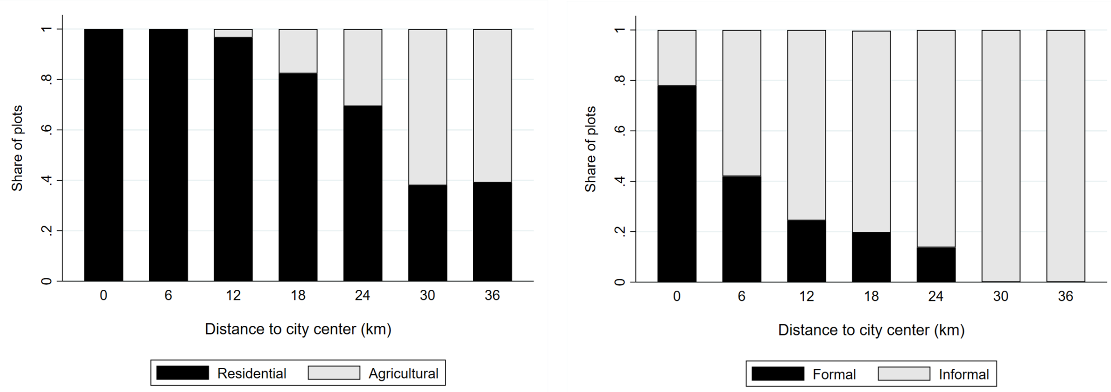
3.1 Spatial patterns of land tenure

Because our model predicts the spatial distribution and the prices of formal and informal land, we present evidence on the location and pricing of formal and informal land in Bamako, Mali, a representative city in sub-Saharan Africa for which such data is uniquely available. The data comes from a World Bank survey of a representative selection of unbuilt land plots transacted between 2009 and 2012 in the greater Bamako area (see Durand-Lasserve et al., 2015, and Appendix A of this paper for a detailed presentation of the survey).[8]

[8] Along with Durand-Lasserve et al. (2015), we define the greater Bamako area as the space comprising the six central municipalities of the Bamako District and eight peri-urban municipalities surrounding the District (see Appendix Figure A1). Observations were sampled at regular intervals around road corridors extending

Georeferenced information is available on price, land tenure, intended land use (i.e., residential or agricultural), area, presence of infrastructure services, distance to road and river. As regards land tenure, we distinguish formal and informal tenure: Formal plots are those that are held with a property title or a permit to occupy. Informal plots are those with no documentation or only a sales document or an administrative document which does not provide a legal right (see Durand-Lasserve et al., 2015, for more details on the typology of tenure situations in Bamako). The spatial distribution of destined land uses (i.e., whether plots are intended for a residential or an agricultural use) is represented on the left-hand side panel of Figure 1 below, which shows that the plots closest to the city center are residential, and that after a certain distance (12 km), agricultural and residential plots coexist, with the share of agricultural plots rising with distance to the city center. The right-hand side panel of Figure 1 shows that the share of formal plots decreases with distance to the city center.[9]

Figure 1: Intended land use (left panel) and tenure status (right panel) by distance to the city center



Note: The left-hand side panel in this figure represents intended land use (residential or agricultural) of plots by bins of distance to the city center. The right-hand side panel represents tenure status for residential plots by bins of distance to the city center.

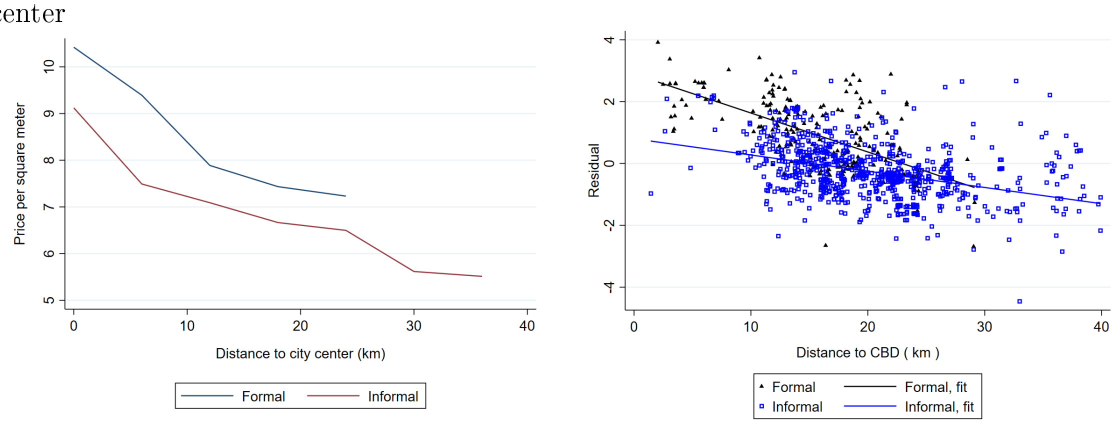
The left-hand side panel of Figure 2 below represents land prices per square meter by

outward from the city center. The sample comprises over a thousand observations (represented as dots on Appendix Figure A1).

[9]In line with our data, Bertrand (2016 and 2019) notes the lower prevalence of formal plots in the peri-urban communes of the greater Bamako area than in the central communes of the Bamako district. It is a common feature of sub-Saharan African cities (see for instance Agegnehu et al. (2016) for Ethiopia).

distance to the city center. We see that prices decrease with distance to the city center and that formal plots sell at a premium compared to informal plots in similarly distant locations. The right-hand side panel of the same figure plots the residuals obtained in a hedonic regression of land plot prices on plot characteristics against distance to the city center. This provides a representation of how prices vary with distance to the city center, controlling for all other price determinants (see Appendix A and Appendix Table A1 for details of the regression). The figure confirms that, all else being equal, both formal and informal land markets value proximity to the city center. Importantly, the slope is steeper for formal than for informal plots, indicating that the increment in land value for a location marginally closer to the city center is greater on the formal market than on the informal market. Our analysis in Section 4 will shed light on what causes these patterns.

Figure 2: Price gradients (left panel) and tenure status (right panel) by distance to the city center



Note: The left-hand side panel in this figure displays the logarithm of the price of land plots per square meter (in CFA) against distance to the city center by tenure status. The right-hand side panel displays the residuals from hedonic regressions of land plot prices (excluding distance to the city center) depending on distance to the city center. Separate regressions are run for formal plots (black triangles) and informal plots (blue squares). See Appendix A for details.

3.2 Ethnic matching and tenure insecurity

We report here the analysis of a 2022 survey also carried out by the World Bank in which over a thousand individuals were randomly selected throughout the Greater Bamako Area within street discussion groups known as "grins" (See Appendix A for a more detailed description of the sampling approach). These discussion groups are a tradition in Malian society and are present in every neighborhood, with each neighborhood having several such discussion groups. They usually consist of a group of men—although women participate as well—and are open to all. All topics can be publicly discussed and all views freely defended. Randomly selected respondents attending those grins were asked to provide basic demographic characteristics (age, gender, occupation and ethnicity). They were then presented with three fictive land purchase situations in which they were asked to assess the risk of purchasing a plot from fictive individuals whom they were randomly matched with. These fictive matches were randomly drawn to be either ethnic cousins or non-cousins of the respondent, with the ethnicity of the fictive seller clearly mentioned. For instance, a respondent identified as belonging to the Bozo ethnic group could be presented with a fictive member of the Dogon ethnic group (with which the Bozo group has a cousinage link) or with a fictive member of the Soninké ethnic group (with which the Bozo group has no cousinage link). The three land purchase situations corresponded to three different tenure situations and associated levels of tenure risk, with the fictive land plot being either a formal plot (registered property title), an informal plot, or a customary plot.[10] Respondents were also asked about attempts they might make at formalizing the plot if they were to purchase it from these fictive individuals in these various land tenure situations. This allowed to have variation in the sample, with some individuals paired with an ethnic cousin and others with an ethnic non-cousin for the purchase of a plot with similar tenure risk. In addition, respondents were also presented with three fictive sale situations with various degrees of tenure risk, in which they were asked to choose among fictive potential buyers of land whose ethnicity was explicitly mentioned.[11] The list of fictive buyers was conditioned on the respondent's ethnicity to ensure that it included an ethnic cousin and two non-cousins.

[10]In the survey, a customary plot is defined as being sold by a customary owner (i.e., by a member of a customary group who initially obtained the plot according to customary allocation rules but decided to sell it; See Picard and Selod (2020) for more details on this). Customary plots can be considered informal since they are not held with a property title or a permit to occupy. In the analysis, however, we treat customary plots as a separate category to allow for differences in the perception of risks over informal and customary plots.

[11]The different levels of tenure risks were conveyed to the respondents by mentioning in some instances that the fictive plot had a registered property title (which is commonly known to be risk-free), that it did not have a registered property title but ownership was not contested (an intermediate level of risk) or that it did not have a registered property title and ownership was contested (high level of risk).

Towards the end of the questionnaire, the collected data also included opinions regarding cousinage practices, past experience of land sales and purchases, land tenure documentation held on their actual plot, as well as experience of land conflicts.

Appendix Table A2 presents the sample's descriptive statistics, distinguishing between individuals surveyed in one of the six municipalities of the Bamako district and those residing in one of the eight peripheral municipalities. The sample includes a majority of men but also a significant share of women (20 percent). There is a lot of ethnic variation, both in the center and in the periphery of the greater Bamako area, with a majority of Bambaras, followed by Malinkés and Peuhls, in proportions that are consistent with the latest census information available (INSTAT, 2009). Opinions expressed by the respondents show that social structures are deemed important: Respondents believe that cousinage relationships need to be abided by, with an average score of 3.4 on a scale of 1 (not important) to 4 (very important). These opinions are homogeneous across space, which reflects the universality of cousinage norms. 28 percent of the surveyed individuals have already purchased a land plot and 14 percent have already sold one. Although the share of individuals who previously sold plots is greater in the periphery (possibly reflecting the dynamism of land markets in the periphery in relation with urban expansion), there is no spatial variation in the share of individuals who purchased plots. Interestingly, a large share of acquired plots (41 percent) and a large share of sold plots (49 percent) were transacted with family members or ethnic cousins. The share of sales to family or ethnic cousins, however, is lower for individuals surveyed in the peripheral municipalities (42 percent) than for those surveyed in the Bamako District (63 percent).[12] Both past experiences of sales and purchases overwhelmingly involve informal plots (i.e., plots for which respondents had at best a weak form of documentation such as an allocation letter issued by authorities or a sales certificate). For purchased plots, slightly less than one third have a formal property right (11 percent have a property title and 20 percent have a permit to occupy). Consistently with the stylized facts presented in the previous subsection, the share of plots purchased with formal

[12]These figures do not imply that the share of transactions along ethnic lines is lower in peri-urban areas than in city centers because the survey did not collect information on where the transacted plots were located. In fact, many households residing in urban centers tend to acquire plots in peri-urban areas as investments (see Durand-Lasserve et al., 2015).

property rights is significantly greater in the city center (27 percent) than in the periphery (5 percent). Interestingly, the share of respondents who personally experienced a land conflict or who know someone in their inner circle who experienced a land conflict is very high, at 38 percent.[13]

We first run a regression that estimates whether respondents find it risky or not to purchase the land plot, depending on whether the respondent was matched with a fictive ethnic cousin seller or not, controlling for age, occupation, municipality, gender, dummies indicating previous purchase and sale experience, and whether the respondent faced or knew someone who faced a land conflict in the past.[14] This was estimated using a logit model where the explained variable is equal to 1 if the respondent deems that it is "risky" or "very risky" to purchase the plot and 0 otherwise, and the regression is run separately for the different fictive tenure situations of the plot (formal plots, informal plots, and customary plots). We report the results in Table 1 below. For formal plots (column 1), which are a priori secure, we see that cousinage does not play any significant role in the buyer's assessment of the risk. In contrast, for informal plots and customary plots which both carry intrinsic tenure risk, respondents feel the transaction is significantly less risky when in a situation of purchasing from a randomly assigned fictive ethnic cousin seller than from a fictive non-cousin seller. For informal plots (resp. customary plots) (column 2, resp. 3), the marginal effect is a reduction of 6.9 (resp. 5.4) percent in the perception of risk when presented with an ethnic cousin seller.

In Table 2 , we then investigate whether respondents would be more likely to undertake steps to formalize an informal or a customary plot after purchasing it from a fictive ethnic non-cousin seller instead of a fictive ethnic cousin seller. For this, we regress the willingness of respondents to undertake steps to formalize the plot depending on whether they were presented with a fictive ethnic cousin seller or with a fictive non-cousin seller. For informal plots, respondents are 4.7 percent less likely to take any step to formalize if the fictive seller is an ethnic cousin. The result is significant at the 10 percent level. For customary plots, respondents are also less likely to formalize if the seller is an ethnic cousin although the effect is not statistically

[13] Our finding confirms that of Neimark et al. (2018) who report high levels of tenure insecurity in the Bamako area.

[14] We use the same controls in all the regressions that follow in this section.

Table 1: Perception of risks depending on cousinage with seller (logit)

	(1) Formal plot	(2) Informal plot	(3) Customary plot
Cousin seller	0.174	-0.428***	-0.327**
	(0.192)	(0.152)	(0.149)
Observations	948	1,106	1,106
Pseudo R^2	0.170	0.166	0.195

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Logit regressions include controls for respondent's age, occupation, municipality, gender, dummies indicating previous purchase and sale experience, and whether the respondent was faced with or knew someone who was faced with a land conflict, and a constant. Column (1) has a smaller number of observations due to observations in two municipalities being droppped as their fixed effects deterministically predict the outcome. Dropping those two municipalities or merging them with neighborhood municipalities does not qualitatively changes the results but slightly increases the marginal effects (Tables available upon request).

significant. Both results from Table 1 and Table 2 support the stronger willingness to purchase from ethnic cousins and illustrate the lower perceived risk when purchasing from an ethnic cousin rather than from an ethnic non-cousin.

Table 2: Decision to formalize depending on cousinage with seller (logit)

	(1) Informal plot	(2) Customary plot
Cousin seller	-0.228*	-0.0408
	(0.134)	(0.133)
Observations	1,106	1,105
Pseudo R^2	0.102	0.100

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Logit regressions include controls for respondent's age, occupation, municipality, gender, dummies indicating previous purchase and sale experience, and whether the respondent was faced with or knew someone who was faced with a land conflict, and a constant. Regression (2) only has 1,105 observations due to a non-response.

Next, we focus on fictive sale situations. Table 3 shows estimates of whether respondents decide to sell to an ethnic cousin or to an ethnic non-cousin depending on the fictive tenure risk of the plot. Respondents are consecutively presented with three sales situations with different tenure risk levels (a formal plot which is fully secure, an uncontested informal plot which bears a low risk, and a contested informal plot which bears a high risk). We consider the formal plot sale situation as the benchmark and assess how sellers' ethnic matching decisions vary for informal plots depending on the risk level. For each sale situation, the respondent is presented with three fictive potential buyers, at least one of whom is an ethnic cousin, and asked which buyer they would choose to sell their plot to, with the possibility of stating indifference between potential buyers. Respondents can choose an ethnic cousin as their buyer (32 percent of respondents), a non-cousin (27 percent), or mention that they are indifferent between the potential buyers (41 percent). We run a multinomial logit regression that assesses the impact of tenure risk on the sellers' ethnic matching decisions, the reference decision being indifference between any of the potential buyers. Column (1) shows the regression considering only the impact of selling an informal plot (irrespective of its risk level) on the ethnic matching decision. We find that selling an informal plot instead of a formal plot increases the probability to choose a non-cousin buyer by 88 percent as compared to choosing any potential buyer. This result is consistent with the idea that sellers are reluctant to sell to their cousins plots that they know are risky. There is no statistical effect on the relative probability of choosing a cousin relative to being indifferent between potential buyers. In column (2), we distinguish the effects of informal plots that have low and high risks. Both significantly increase the probability of choosing a non-cousin buyer relative to being indifferent between potential buyers at the 1 percent level. Unexpectedly, the point estimate seems greater for low-risk plots than for high-risk plots, but the difference is not statistically significant. In Appendix Table A3 , we reproduce Table 3 for a subsample of men over 40 years old who already experienced a land market transaction (i.e, people who are more likely to be active in land markets) and find a greater point estimate for high-risk plots than for low-risk plots (although still not statistically different), which is consistent with the idea that greater risks incentivize sellers to sell to non-cousins.

The survey has shown that land markets in Bamako are far from being impersonal given the omnipresence of codified bilateral ethnic relationships which affect all aspects of life in Mali.

Table 3: Decision to sell to an ethnic cousin or a non-cousin depending on tenure risk (multinomial logit)

		(1)	(2)
Cousin	Informal (low risk)		-0.0243
			(0.105)
	Informal (high risk)		-0.0116
			(0.104)
	Informal (both risks)	-0.0179	
		(0.0898)	
Non-cousin	Informal (low risk)		0.674***
			(0.115)
	Informal (high risk)		0.595***
			(0.115)
	Informal (both risks)	0.635***	
		(0.102)	
Observations		3,318	3,318
Pseudo R^2		0.0917	0.0919

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Multinomial logit regressions include controls for respondent's age, occupation, municipality, gender, dummies indicating previous purchase and sale experience, and whether the respondent was faced with or knew someone who was faced with a land conflict, and a constant.

Since the survey did not focus on georeferenced land market transactions, it is difficult to say whether ethnic matching in land markets present geographic patterns. This said, the survey of plot transactions presented in the previous subsection showed that the share of informal and customary plots (i.e., risky plots) is greater in the periphery than in the city center. This would imply that ethnic matching could be more prevalent in peripheral locations. Although we do not have a direct measure of this for Bamako, another study on Dakar, Senegal—the capital of a neighboring country that was previously part of the Mali empire and which shares similar cousinage social structures—reports clear spatial patterns of cousinage: Analyzing a sample of 324 randomly selected individuals, Smith (2004) finds that the frequency of individuals practicing cousinage daily increases with distance to the city center (42 percent in the city center, 50 percent in the suburbs, and 57 percent in peri-urban areas). Only 7 percent of peri-urban respondents declare that they never or rarely practice cousinage whereas this percentage reaches 30 percent for city-center respondents.

4 Urban land-use model with tenure insecurity and information asymmetry (benchmark model)

In this section, we present a benchmark urban land-use model in which we introduce tenure insecurity and information asymmetry and for which we derive the equilibrium city structure and surplus. Presenting this benchmark model is a useful stepping stone to highlight the market failure in our core setting and to derive optimality properties before introducing matching along trusted ethnic relationships in the next section.

4.1 Main assumptions

The urban space is represented by a line segment at the extremity of which lies a Central Business District (CBD) where all jobs are located. Each location on this segment (denoted by its distance x to the city center) has a unit mass of absentee landowners, each endowed with one land plot.[15] Each landowner decides whether or not to sell his land plot to a potential migrant coming to the city, thereby extracting the migrant's willingness to pay to reside in that particular location. Because migrants will be working in the CBD, they value proximity to the city center. As migrant buyers are competing with one another, sellers sell their plots to the highest bidder. We consider an open-city model, in which buyers migrate to the city until the expected utility in the city is equalized with the rural utility level u.[16]

In our model, land tenure is insecure for some plots in the sense that a buyer can lose his plot in the future with a non-zero probability. As discussed in the previous sections, this probability reflects contested ownership of land, which is prevalent in many developing country cities.[17] Because not all plots are insecure, we assume that there are two possible levels of tenure security $Q \in \{q, 1\}$, with $q < 1$. Only insecure plots may be contested and have a probability q of remaining in the hands of their buyer in the future, whereas secure plots are

[15] Our assumption of a unit of land in each location makes our framework akin to Alonso (1964).

[16] We assume a linear utility function and a price of the composite good normalized to 1, so that the utility in the city—defined as the consumption of the composite good—is exactly equal to the expected disposable income.

[17] Typically, conflicts over land ownership may oppose heirs, customary owners and investors, private parties and public authorities. See Durand-Lasserve et al. (2015) for a full typology of land conflicts.

uncontested and have a probability 1 of remaining in the hands of their buyer. We denote π the exogenous initial proportion of secure plots and assume it is constant across all locations.

The key assumption in our model is the existence of an information asymmetry between sellers (i.e., initial landowners) and buyers. Whereas sellers know the tenure security level of their plots (i.e., they know if a competing claim might emerge following the sale), migrant buyers cannot observe this characteristic before the transaction takes place. In what follows, we will refer to the initial owners of secure plots as 1-owners and to the initial owners of insecure plots as q-owners. We assume that tenure insecurity only emerges after the sale so that if landowners decide not to sell their plot to a migrant, they simply keep it for agricultural use and obtain a fixed revenue equal to the agricultural land rent R_a.[18] We assume that $u \leq R_a$, reflecting the fact that migrants are rural laborers who, by definition, cannot be paid above the agricultural land rent.

We present below the market behavior of buyers and sellers in each location and derive the spatial extent of the urban land market.

4.2 Sellers and buyers' decisions

The owners' decision to sell is modeled with a binary choice variable $P \in \{0, 1\}$, with $P = 1$ if the owner transacts with a migrant—in which case we refer to the owner as a seller—and $P = 0$ if the owner does not participate in the land market.[19] We denote $\pi(x)$ the share of sellers in location x who are 1-owners. Note that $\pi(x)$ generally differs from π, because some landowners (1-owners or q-owners) decide not to sell their land. We consider that buyers have rational expectations and can fully anticipate the value $\pi(x)$. In location x, the buyer of a plot can thus expect to purchase a secure plot with probability $\pi(x)$ and an insecure plot with probability $1-\pi(x)$. If the plot is insecure, it is lost with probability $1-q$ in the future. For the buyer, the expected probability of keeping the plot in the future is thus $\pi(x)+(1-\pi(x))q$ and

[18] The assumption that there is no "before sale" risk in the model is consistent with findings from Wehrmann (2008) or Owusu and Chigba (2020) who note that conflicts emerge after the sale due to issues such as multiple sales and disposition of rights by traditional leaders without consulting other group members.

[19] Note that our model departs from the standard adverse selection setting (see Akerlof 1978) where it is buyers who may opt out of the market.

the expected probability of losing it is $(1-\pi(x))(1-q)$. Furthermore, the buyer knows that, if he is not evicted and is thus able to remain in the city, he will have a utility of $y-xt-R(x)$, corresponding to his urban income net of commuting costs and net of the price paid for the land plot (denoted $R(x)$), which will be endogenously determined. If evicted and having to return to the rural area, the buyer gets utility $u-R(x)$ because the purchase of the plot is a sunk cost.[20] It follows that the expected utility of a buyer purchasing a plot in x is:

$$(\pi(x)+(1-\pi(x))q)(y-xt-R(x))+(1-\pi(x))(1-q)(u-R(x)) \tag{1}$$

We are now ready to derive the bidding behavior of buyers. Equating (1) with the rural utility level u (given our open city assumption) and solving the resulting equality for the land price, we obtain the bid rent of a buyer of a plot located in x:

$$\psi(x,u)=\{\pi(x)+q(1-\pi(x))\}(y-xt-u) \tag{2}$$

The bid rent measures the buyer's willingness to pay for a plot in location x to exactly attain equilibrium utility u. Note that the bid-rent function depends on the buyer's expectation regarding tenure insecurity, so that the buyer's willingness to pay increases with $\pi(x)$, i.e., with the fraction of 1-owners among sellers in x and with q, i.e., with the level of tenure security of risky plots .

As for owners, their decision whether or not to participate in the market will depend on the plot's location x, its intrinsic tenure security level Q and the market price $R(x)$, which we capture with the generic notation $P(x,Q,R)$. In turn, because the benefit B of a landowner will depend on his market participation decision P and on the market prevailing price $R(x)$ in location x, we express it as $B(P,x,Q,R)\equiv PR(x)+(1-P)R_a$. The formula expresses *gross* profit and simply states that landowners who do not participate in the land market ($P=0$) obtain a benefit of $B=R_a$, whereas landowners who participate in the market ($P=1$) obtain a benefit of $B=R(x)$.[21]

[20] Movements from rural to urban areas as well as from urban to rural areas are well documented (see Cattaneo and Robinson, 2020).

[21] Considering that a seller gives up on agricultural production, the *net* profit from a sale is $PR(x)+(1-P)R_a-R_a$.

4.3 The competitive equilibrium

Having characterized the behaviors of both sellers and buyers, we can now define the spatial equilibrium. To do this, however, we first need to introduce the additional notations $L_q(x)$ and $L_1(x)$ for the respective quantities of insecure land and secure land that are transacted in x. With these notations, the total quantity of land transacted in x, can be decomposed as $L(x) = L_q(x) + L_1(x)$.

The set of equilibrium conditions that define the equilibrium are as follows: First, the equilibrium quantity of land that is sold in each location x must be smaller than the initial unit land endowment in that location, which yields the following *land use constraint*:

$$L_q(x) + L_1(x) \leq 1 \quad \text{for any } x \tag{3}$$

Then, in equilibrium, the market participation decision of a landowner, $P^*(x, Q, R)$, conditional on his location, plot tenure security level, and price, maximizes his gross profit, which leads to the *profit maximization condition*:

$$P^*(x, Q, R) \in ArgMax_{P \in \{0,1\}} B(P, x, Q, R) \quad \text{for any } x \tag{4}$$

Additionally, in equilibrium, the *land market clearing condition* implies that prices equalize supply and demand in each location, with demand being defined by the envelope of equilibrium bid-rent functions as standard in open city models (see Fujita, 1989). In other words, due to the infinite potential pool of migrants to the city, sellers are able to perfectly extract buyers' willingness to pay, so that the land price is equal to the bid rent in each location x, taken at the rural utility level u:

$$R(x) = \psi(x, u) \quad \text{for any } x \text{ where } L(x) > 0 \tag{5}$$

Finally, we can directly express the *city boundary* denoted x_b as the location closest to the CBD such that the gross benefits of both 1-owners and q-owners accounting for their optimal market participation decisions are equal to the agricultural land rent R_a. This can be written as:

$$x_b = \min x \text{ s.t. } B^*(x,1,R) = B^*(x,q,R) = R_a \tag{6}$$

where $B^*(x,Q,R) \equiv B(P^*(x,Q,R),x,Q,R)$ is the optimal payoff (i.e. the gross profit evaluated at the optimal participation decision) of an owner of a plot of security level Q located in location x and facing the sale price R. Since there is only one price for land irrespective of the tenure security level (given that buyers cannot distinguish between insecure and secure plots), the condition boils down to $R(x_b) = \psi(x_b,u) = R_a$, which is the standard city fringe condition in urban economics. We have the following equilibrium definition:

Definition 1: *A competitive equilibrium is a set of market participation decisions (P), prices in each location ($R(x)$), and a city fringe (x_b) that satisfies the system of equilibrium conditions (3), (4), (5) and (6).*

Note that our equilibrium is defined "ex-ante" in the sense that agents make choices depending on their expectation that a conflict may materialize, but before the existence of a conflict can be observed. It also assumes rational expectations, in the sense that buyers and sellers know the model of the economy and are able to correctly assess the equilibrium proportion of secure plots among transacted plots in each location.

To solve the system, we identify all possible combinations of owners' participation decisions that are Nash equilibria (i.e., the sets of 1-owners' and q-owners' decisions in which no landowner would gain from revising his participation decision given the participation decisions of all other landowners). We show in Appendix B that although a continuum of equilibria is possible, only one equilibrium is stable. This is the equilibrium we retain in the analysis.

We have the following proposition:

Proposition 1: *There is a single stable equilibrium. In the stable equilibrium, all landowners between the city center and the city border* $x_b = x_a \equiv \frac{1}{t}\left[y - \frac{R_a}{\pi+q(1-\pi)} - u\right]$ *participate in the* *land market. The equilibrium price curve is* $R(x) = [\pi + q(1-\pi)]\,(y - tx - u)$.[22]

Proof: See Appendix B.1.[23]

Observe that the land price curve is linear with slope $\frac{\partial R}{\partial x} = -\left[\pi + q(1-\pi)\right] t$, where $\pi + q(1-\pi)$ is the probability for a buyer to keep a purchased plot. In equilibrium, when marginally moving outwards from the city center, the reduction in land price exactly compensates the increase in *expected* transport costs (given the probability of keeping the plot and commuting to the city center rather than losing the plot and moving back to the rural area without any need to commute anymore). The equilibrium price slope becomes flatter if plots become more insecure (smaller q) or if the fraction of sellers with secure plots is smaller (smaller π). As expected, when there is no tenure insecurity in the model (either because $q = 1$ or $\pi = 1$), the slope is equal to the *certain* marginal transport cost, as in the standard version of the urban economics model with unit land consumption (a variation of the Alonso-Muth-Mills condition). Figure 3 below represents the equilibrium city structure and land price as a function of distance to the CBD (with the value of the slope written in blue below the land price curve).[24]

Observe that, under information asymmetry, both 1-plots and q-plots are transacted over the same zone $[0, x_a]$. In this model, there is no adverse selection that would have 1-plots not transacted in equilibrium. This is due to the risk materializing only after the sale, implying that the intrinsic values of 1-plots and q-plots are the same for an owner who decides not to participate in the market. Since the sale prices of 1-plots and q-plots are also the same under information asymmetry, owners of 1-plots and q-plots thus face the same incentives to sell or

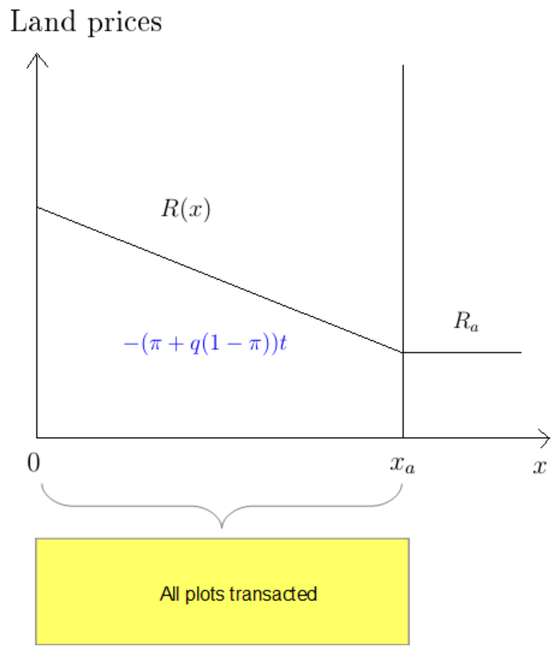
[22] Note that to avoid the degenerate case of an empty city, the model requires the exogenous parameters to verify $y - \frac{R_a}{\pi + q(1-\pi)} - u > 0$. From now on, we assume that this condition is satisfied.

[23] For the intuition of the proof provided in Appendix B, note that there is a continuum of participation decisions that are compatible with equilibrium conditions (3)-(6). In these equilibria, all owners in the segment $[0, x_a]$ participate in the market. These multiple equilibria only differ with respect to market participation decisions over the segment $]x_a, x_a^*[$ with $x_a^* \equiv \frac{1}{t}(y - R_a - u)$ where any transacted land plot is sold at a price exactly equal to the agricultural land rent R_a. Each equilibrium in this continuum of equilibria is characterized by the number of 1- and q-sellers in each location $x \in]x_a, x_a^*[$, which must verify $(L_q(x), L_1(x)) \in [0, 1-\pi[\times [0, \pi[$ and $B(P, x, Q, R) = R_a$. Denoting $\mathcal{E}_{L_q, L_1}$ the equilibrium characterized by functions L_q and L_1 over the interval $]x_a, x_a^*[$, Appendix B shows that $\mathcal{E}_{L_q, L_1}$ is unstable if there exists $x \in]x_a, x_a^*[$ such that either $L_q(x) > 0$ or $L_1(x) > 0$ so that any small deviation in participation decisions from that equilibrium will always trigger a transition towards $\mathcal{E}_{0,0}$, the unique stable equilibrium.

[24] In Appendix B.2, we also present a figure that plots the payoffs of owners and their underlying participation decisions (see Figure B1).

not sell. Inspection of the equilibrium city fringe formula for x_a (see Proposition 1) shows that city size decreases with the proportion of insecure plots $(1-\pi)$ and the level of tenure insecurity $(1-q)$. It is easy to understand that, when either one of these values marginally increases, the "last" seller at the city fringe prefers to keep his plot under agricultural use, since buyers' willingness to pay is reduced.[25]

Figure 3: City structure and land prices (benchmark model)



Note: This figure represents the equilibrium land price and market participation as a function of distance to the city center. The slope of the land price curve is indicated in blue.

4.4 Surplus analysis

Following Fujita (1989), we define the surplus as the city production (sum of wages) minus the costs to organize the city (transport costs, composite good consumption, and foregone

[25]Considering a variant of the model with risk aversion, it can be shown that city size would also decrease with risk aversion, a feature absent from our model. To see this, assume a Von-Neumann Morgenstern context with a CCRA Bernoulli function, $z \to \frac{z^{1-\rho}-1}{1-\rho}$, where ρ is risk aversion. The bid-rent $\psi(x,u)$ is now implicitly defined as the rent satisfying the equation $u = (1-q)(1-\pi(x))(\frac{(u-R)^{1-\rho}-1}{1-\rho}) + (\pi(x)+q(1-\pi(x)))((\frac{(y-tx-R)^{1-\rho}-1}{1-\rho}))$. Applying the implicit function theorem, we show that $\frac{\partial\psi(x,u)}{\partial\rho} < 0$, which further implies that $\frac{\partial x_a}{\partial\rho} < 0$. In other words, the greater the risk aversion, the smaller the city. Introducing risk aversion is, however, not necessary to derive the main results of our model and would make it much less tractable. We therefore abstain from considering risk aversion in the model.

agricultural production).[26] The surplus can be written as a function Γ that depends on parameters q, π and u. For the sake of simplicity, we assume that following an eviction, plots can no longer be used.[27]

In mathematical terms, we have:

$$\Gamma(q,\pi,u) = \int_0^{x_a} \pi(y - xt - u - R_a) + (1-\pi)(q(y - xt - u) - R_a)dx$$

As expected, the surplus increases with the share of 1-plot sellers (π) and decreases with the level of tenure insecurity on risky plots (1-q). This surplus under information asymmetry can be compared to the surplus Γ^{sym} that would be obtained in the same model without information asymmetry (i.e., when buyers have perfect information on risks). It can be shown that removing information asymmetry would result in q-plots and 1-plots being sold over different zones, with all 1-plots being sold between the CBD and location $x_a^* = \frac{1}{t}\left[y - R_a - u\right]$ and all q-plots until location $x_a^q \equiv \frac{1}{t}\left[y - \frac{R_a}{q} - u\right]$. [28] It is easy to see that $x_a^q < x_a < x_a^*$, implying that information asymmetry decreases the zone over which 1-plots are sold and increases that over which q-plots are sold. This effect is magnified by risk as x_a^q decreases with risk.

We have:

$$\Gamma^{sym}(q,\pi,u) = \pi \int_0^{x_a^*} y - xt - u - R_a dx + (1-\pi)\int_0^{x_a^q} q(y - xt - u) - R_a dx$$

It is straightforward to show that $\Gamma(q,\pi,u) < \Gamma^{sym}(q,\pi,u)$ for all $q < 1$, $\pi < 1$, and u, which reflects the surplus-reducing impact of information asymmetry. This points to a

[26] This is mathematically equivalent to another definition of surplus that would consider the utility increment from migration to the city net of the opportunity cost of land use.

[27] This assumption is consistent with our equilibrium being defined ex-ante (i.e. before evictions take place). It also removes a dynamic externality that would arise in the model if plots were to be occupied by other workers or for agriculture following eviction, as buyers would not account for the future use of the land in their decisions. With this simplifying assumption, we can focus on the core mechanism of the model regarding the role of ethnic kinship ties in response to information asymmetry.

[28] Indeed, the respective payoffs of q-plot sellers ($q(y - xt - u)$) and 1-plot sellers ($y - xt - u$) are decreasing in x and equalize the agricultural land rent R_a in x_a^q and x_a^*.

market failure caused by an externality under information asymmetry whereby agents do not internalize the effect of their market participation decision on the composition of the pool of transacted plots, which in turn affects other agents' expectations about risk and their decisions to participate in the market. We show in Appendix B.3 that removing information asymmetry brings the equilibrium to its optimum so that the optimal surplus $\Gamma^{opt}(q, \pi, u)$ is the same as $\Gamma^{sym}(q, \pi, u)$.

Finally, note that if risk is completely removed from the model ($\pi = 1$ or $q = 1$), the issue of information asymmetry becomes irrelevant and the market equilibrium configuration coincides with the optimum city structure which extends until x_a^* (with $x_a^q = x_a = x_a^*$). The risk-free social optimum is thus $\Gamma^* = \Gamma^{opt}(1, 1, u)$.

5 Adding ethnic matching to the model

We now introduce a norm that governs trust between specific ethnic groups and analyze the equilibrium response to information asymmetry and derive the implications in terms of city structure and surplus.

The social norm we introduce in the model corresponds to a set of trusted relationships between specific ethnic groups in line with the cousinage institution described in Sections 2 and 3. A land owner may choose whether to transact with a potential migrant with whom he has an ethnic relationship that involves some amount of trust or with a potential migrant with whom he has no link. Borrowing the language of the anthropological literature presented in Sections 2 and 3, we refer to this behavior as transacting with an ethnic cousin (as opposed to transacting with an ethnic non-cousin).[29] In our setting, there is no need to define ethnic groups and specify their numbers, as we just focus on whether landowners transact with an ethnic cousin or not, with an infinite pool of ethnic cousins potentially supplied by migration to the city. Cousinage relationships are known and observable by all agents. Our only assumption here is that tenure insecurity is an intrinsic characteristic of plots that does not depend on

[29] Although we refer to cousinage relationships, our model is also relevant for any society where trusted relationships exist within or across groups of individuals.

the ethnicity of landowners, implying that, in each location, the proportions of insecure and secure plots do not depend on the owner's group, which thus does not need to be specified.

Conditionally on participating in the market, we denote $C \in \{c, nc\}$, the landowner's decision to sell to an ethnic cousin ($C = c$) or to a non-cousin ($C = nc$). As in the benchmark model, buyers do not know the risk associated with the plot they are purchasing ($Q \in \{q, 1\}$). Mirroring the literature on ethnic groups and social sanctions (see Fearon and Laitin 1996, La Ferrara 2003, Habyarimana et al. 2007), the key assumption in this setting is that selling an insecure plot to an ethnic cousin will always be punished with penalty $J > 0$.[30] In line with the stylized facts presented in Section 3, this ensures that sellers of risky plots have an incentive to sell their plot to an ethnic non-cousin and will be reluctant to sell their plot to an ethnic cousin. The social penalty reflects the ostracism or disapproval faced by individuals who betray trust among ethnic cousins as codified in the cousinage institution. We denote $\pi^c(x)$ the proportion of 1-sellers who transact with an ethnic cousin in x, expressed as a fraction of all sellers who transact with an ethnic cousin. Similarly, $\pi^{nc}(x)$ is the proportion of 1-sellers who transact with an ethnic non-cousin in x, expressed as a fraction of all sellers who transact with an ethnic non-cousin.

Observe that informal land markets are now interpersonal as opposed to the impersonal land markets presented in the benchmark model. Also note that because different levels of trust exist between ethnic cousins and between non-cousins, there are now two different prices for informal land plots, depending on whether the transaction involves ethnic cousins or non-cousins. We denote these prices for informal interpersonal transactions $R^c(x)$ and $R^{nc}(x)$ when the transaction involves ethnic cousins and non-cousins respectively.

5.1 Sellers and buyers' behavior

A land plot owner now has two choices to make. He chooses whether to participate in the market (decision $P \in \{0, 1\}$) and then, conditional on participating in the market, whether to sell to an ethnic cousin buyer or not (decision $C \in \{c, nc\}$).

[30] For simplicity, J is the same among all pairs of ethnic cousin groups.

The respective expected utilities of a buyer purchasing land from an ethnic cousin seller or from a non-cousin seller are:

$$\begin{cases} u_c(x) = \{\pi^c(x) + q(1-\pi^c(x))\}(y - tx) + (1-\pi^c(x))(1-q)u - R^c(x) \quad \text{and} \\ u_{nc}(x) = \{\pi^{nc}(x) + q(1-\pi^{nc}(x))\}(y - tx) + (1-\pi^{nc}(x))(1-q)u - R^{nc}(x) \end{cases}$$

Because we have an open city model, migration will occur until buyers obtain the same utility level u as in the rural area. Inverting the above utility functions in the land price gives the bid-rent functions in each location of the city for transactions of land among ethnic cousins and among non-cousins:

$$\begin{cases} \psi(x, u|C = c) = \{\pi^c(x) + q(1-\pi^c(x))\}(y - tx - u) \\ \psi(x, u|C = nc) = \{\pi^{nc}(x) + q(1-\pi^{nc}(x))\}(y - tx - u) \end{cases}$$

Let us now detail the owners' profit associated with each decision. If an owner decides not to participate in the land market, he receives the agricultural rent R_a. An owner of a secure informal plot selling to an ethnic cousin buyer $(C = c, Q = 1)$ receives a payment $R^c(x)$. An owner of an insecure informal plot selling to an ethnic cousin buyer $(C = c, Q = q)$ receives a payment $R^c(x)$, but faces the social penalty J, which reduces his benefit to $R^c(x) - J$.[31] Finally, an owner selling a plot to a non-cousin buyer receives a payment $R^{nc}(x)$ and there is no social penalty if the transacted plot is insecure as the two parties are not bound by any alliance.

5.2 The competitive equilibrium

We adapt the equilibrium definition to this new setting with matching along trusted ethnic relationships. The decision to participate in the market is now a function of the interpersonal prices of land, R^c and R^{nc}. It can be denoted $P(x, Q, R^c, R^{nc})$. The decision to sell to an ethnic cousin or to a non-cousin, $C(x, Q, R^c\ , R^{nc})$ is also a function of the same arguments. The profit of an owner can now be generically expressed as $B(P, C, x, Q, R^c, R^{nc})$. In this ex-

[31] J captures the disutility experienced by landowners when they are punished. Although the penalty can be non-monetary in nature, landowners behave as if their monetary benefit were reduced by J.

tended version of the model, we decompose transacted land not only according to its intrinsic tenure insecurity but also according to the possibility of ethnic matching between buyers and sellers. This requires the introduction of the notations $L^c(x)$ and $L^{nc}(x)$ for land transacted between ethnic cousins and land transacted between ethnic non-cousins respectively. With these additional notations, we have the following equilibrium definition:

Definition 2: *An equilibrium is a set of market participation and ethnic matching decisions, prices in each location x and a city border that satisfies the following equilibrium conditions:*

$$\begin{cases} L(x) = L_q(x) + L_1(x) \leq 1 \quad \textit{for any } x & (7) \\ (P^*(x,Q,R^c,R^{nc}), C^*(x,Q,R^c,R^{nc})) & \\ \qquad \in \underset{(P,C)\in\{0,1\}\times\{c,nc\}}{ArgMax} \quad B(P,C,x,Q,R^c,R^{nc}) \quad \textit{for any } x & (8) \\ R^c(x) = \psi(x,u|C=c) \quad \textit{for any } x \textit{ where } L^c(x) > 0 & (9) \\ R^{nc}(x) = \psi(x,u|C=nc) \quad \textit{for any } x \textit{ where } L^{nc}(x) > 0 & (10) \\ x_b = \min x \ s.t. \ B^*(x_b,1,R^c,R^{nc}) = B^*(x_b,q,R^c,R^{nc}) = R_a & (11) \end{cases}$$

where $B^*(x,Q,R^c,R^{nc}) \equiv B(P^*,C^*,x,Q,R^c,R^{nc})$ is the optimized payoff (i.e., after taking into account the optimal participation and ethnic matching decisions) of an owner of a plot of security level Q in location x facing prices R^c if the buyer is an ethnic cousin, and R^{nc} if the buyer is not an ethnic cousin.

We now have five equilibrium conditions. As in the benchmark version of the model, condition (7) says that the quantity of land sold must be smaller than the initial endowment in location x. Condition (8) characterizes the optimal market participation and ethnic matching decisions of landowners. Conditions (9)-(10) reflect sellers' extraction of buyers' willingness to pay under the different ethnic matching configurations, where $L^c(x)$ and $L^{nc}(x)$ are the quantities of land transacted with an ethnic cousin or a non-cousin buyer respectively.[32] Condition (11) characterizes the city fringe.

We detail the resolution of this extended model in Appendix C. As with the benchmark

[32] Observe that $L_q(x) + L_1(x) = L^c(x) + L^{nc}(x)$

model, there will be a continuum of equilibria, but only one equilibrium is stable for each value of the social penalty J and Pareto-dominates all the other equilibria. These are the equilibria we retain for the analysis. In Proposition 2 below, for illustrative purposes, we present the case of a small J ($J < \underline{J} = \pi R_a \frac{1-q}{q}$), which is both realistic and sufficient to derive the implications of introducing ethnic matching in the model. Other (similar) equilibria are presented in Appendix C. We have the following proposition:

Proposition 2: *Denoting the boundary zone thresholds* $\underline{x}(J) = \frac{1}{t}\left(y - \frac{R_a+J}{\pi(1-q)+q} - u\right)$ *and* $\bar{x}(J) = \frac{1}{t}(y - (R_a + J) - u)$, *the equilibrium city is organized in three zones:*

- ***The most central zone (Zone 1, for*** $x \in [0, \underline{x}(J)]$***) is fully residential****: All owners (q- and* 1*-owners) sell their plot exclusively to ethnic cousins. In each location* x*, the price for these sales is* $R^c(x) = (\pi(1-q)+q)(y - xt - u)$.

- ***The "close periphery" (Zone 2, for*** $x \in]\underline{x}(J), \bar{x}(J)]$***) is a mix of residential and agricultural uses, with all*** 1***-owners selling their land****: All* 1*-owners and some q-owners sell their plots exclusively to ethnic cousins. Whereas all* 1-owners *participate in the market, some q-owners drop out of the market. The price of land in each location* x *is* $R^c(x) = R_a + J$.

- ***The "far periphery" (Zone 3, for*** $x \in]\bar{x}(J), x_a^*]$***) is a mix of residential and agricultural uses, with all*** q***-owners dropping out of the market****: All* 1*-owners sell their plot exclusively to ethnic cousins and all q-owners drop out of the market. The price in each location* x *is* $R^c(x) = y - xt - u$.

- *The city boundary is at* $x_b = x_a^* = \frac{1}{t}[y - R_a - u]$.

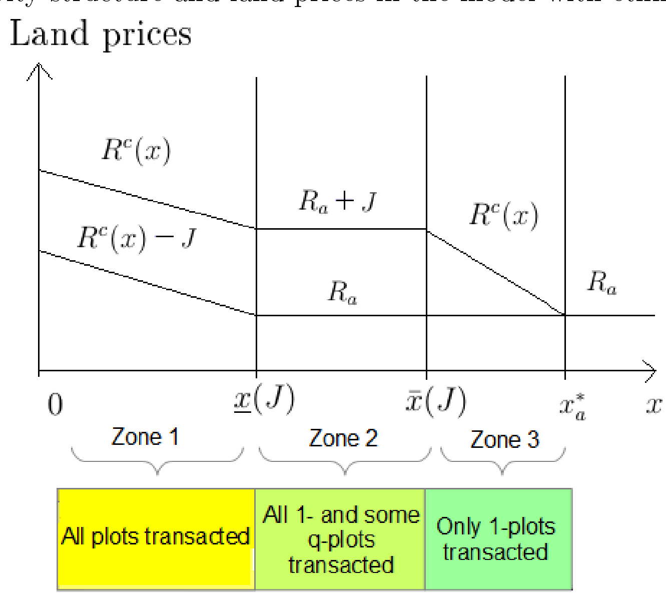
Proof: See Appendix Sections C.1 and C.2.

The structure of the city and the corresponding equilibrium land prices are represented in Figure 4 below.[33] Figure 4 corresponds to the case where the social penalty is relatively small

[33] In Appendix Section C.3, we present a figure that plots the payoffs of owners and their underlying participation and ethnic matching decisions (see Figure C3).

(i.e., with $J < \underline{J}$, where $\underline{J}$ is derived in Appendix C). In the central residential zone (Zone 1), the slope of the land price curve is $-(\pi(1-q)+q)t$, reflecting the trade-off between the land price and expected transport costs as discussed in the benchmark version of the model. Zones 2 and 3 represent the close and far peripheries of the city, where residential and agricultural land uses coexist. Note that, since $\bar{x}(J)$ tends towards x_a^* when the social penalty tends towards zero (see the formula in Proposition 2), Zone 3 tends to disappear for low values of J and the greater periphery is then Zone 2, where the land price tends towards the agricultural land rent. Observe that in this specific case of a low social penalty, all transactions involve ethnic cousins. This is not necessarily the case when the social penalty is larger and q-owners face a stronger disincentive to trade with ethnic cousins.

Figure 4: City structure and land prices in the model with ethnic matching



Note: This figure represents the equilibrium land prices, market participation and ethnic matching as a function of distance to the city center when $J < \underline{J}$. The slopes of the land price curves are indicated in blue.

5.3 Surplus analysis

The city surplus corresponding to the equilibrium city described in Proposition 2 can be expressed as the sum of each zone's contribution to the surplus as follows:

$$\begin{aligned}\Xi(J) = \int_0^{\underline{x}(J)} & \pi(y - xt - u - R_a) + (1-\pi)(q(y - xt - u) - R_a)\, dx \\ & + \int_{\underline{x}(J)}^{\bar{x}(J)} \pi(y - xt - u - R_a) + L_q^c(x, J)(q(y - xt - u) - R_a)\, dx \\ & \qquad + \int_{\bar{x}(J)}^{x_a^*} (y - xt - u - R_a)\pi\, dx\end{aligned}$$

where L_q^c is the mass of q-owners selling to an ethnic cousin.[34] The three integrals correspond to the respective surplus contributions in Zones 1, 2 and 3. For other values of J, the surplus formulas are very similar.[35]

We have the following general proposition:

Proposition 3: *Cousinage in the presence of information asymmetry always increases the surplus compared to the benchmark model with only information asymmetry. For large values of the social penalty ($J \geqslant \overline{J}$), the optimal city structure is obtained in equilibrium and the optimal social surplus is reached. For small values of the social penalty ($J < \overline{J}$), the equilibrium is not optimal.*[36]

Proof: See Appendix section C.4.

Proposition 3 illustrates that cousinage allows a separating mechanism to operate, which can fully or partially address information asymmetry and the associated externality described in Section 4.[37] In this extended model with cousinage, the externality involves individuals not

[34] Here, we explicitly express L_{qc} as a function of both x and J as we are interested in studying the surplus for different contexts of cousinage intensity.

[35] Only the boundaries of the integrals need to be changed to correspond to the boundaries of the different zones represented on Appendix graph C2.

[36] The formula of $\bar{J}$ is derived in Appendix C2. We have $\bar{J} > \underline{J}$.

[37] Note that, when we introduce the possibility of ethnic matching, selection may now arise due to the possibility of signaling secure plots through trade between cousins (which was not the case in the benchmark model). More precisely, selection arises in sections of the city where either only 1-plots are transacted (complete selection) or all 1-plots and some q-plots are transacted (partial selection). This selection is "positive" in the

taking into account the impact of their participation and matching decisions on the ratio of secure plots among informal plots transacted between cousins (we show in Appendix C that transactions between non-cousins always involve q-plots only). When the social penalty is large ($J \geqslant \overline{J}$), q-plot owners are discouraged from trading with cousins, leading to complete sorting with 1-plots exclusively sold among cousins and q-plots exclusively sold among non-cousins. This fully eliminates information asymmetry and leads to the optimal city structure (see Appendix Figure C2) and optimal social surplus $\Gamma^{opt}(q, \pi, u)$. When the penalty is low ($J < \overline{J}$), however, sorting is incomplete and the market failure partially remains: The zone over which q-plots are traded may extend to locations that are even further to the right of x_a^q, implying that the additional q-plot transactions due to ethnic matching negatively contribute to the surplus (see section 4.3). In total, the equilibrium surplus is greater than that under the benchmark model with information asymmetry only, but lower than $\Gamma^{opt}(q, \pi, u)$ (i.e., when information asymmetry is removed).[38]

Finally, note that although cousinage can help address information asymmetry, it does not address the fundamental issue of tenure insecurity. We now introduce into the model the possibility of reducing risk (as well as information asymmetry) through property registration.

6 Adding registration to the model

We now introduce a formal registration system that coexists with the cousinage institution. We then derive the implications of this coexistence in terms of city structure and surplus. Registration of ownership in a cadastre totally extinguishes competing claims on a land plot and makes it fully secure.[39] Registration is observable by all and thus allows buyers to identify a fraction of secure plots (i.e., those which are registered) with certainty. Sellers have the possibility to register their land before entering a transaction, anticipating that a registered

sense that it leads to secure plots being transacted relatively more than insecure plots.

[38]Note that under the polar case where the social penalty is very low ($J \to 0^+$), there is no surplus change associated with cousinage compared to the benchmark model with information asymmetry.

[39]Land registration systems in Western Africa are inspired by the Torrens system where registration (in french "immatriculation") leads to the issuance of an indefeasible property title. Titles cannot be revoked, even if in practice it can be the case that conflicts are not properly resolved or adjudicated before ownership is registered and the property title is issued.

plot will sell at a higher price $R_f(x)$ which capitalizes both the increase in tenure security and the removal of information asymmetry. We assume that there is a registration cost, k, which is the same for all plots, irrespective of the initial tenure security level.[40] Since all plots may not be registered, we will henceforth distinguish between formal (registered) and informal plots.

6.1 Sellers and buyers' behavior

The behavior of buyers and sellers of informal plots is the same as described in Section 5. Conditional on participating in the market, a landowner may now choose whether to formalize his plot by registering it in the cadastre (decision $F \in \{0, 1\}$).[41] If choosing $F = 1$, the plot becomes formal and the tenure security level of the plot is reset at value 1.[42] The expected utility of the buyer of a formal land plot is $u_f(x) = y - tx - R_f(x)$. The bid-rent function for formal land is given by $\psi(x, u|F = 1) = y - tx - u$. The profit of a landowner selling a formal plot is $R_f(x) - k$.

6.2 The competitive equilibrium

The equilibrium definition presented in Section 5 is adapted to account for formal land sales. The decision to register a land plot or not before selling it is a function $F(x, Q, R_f, R^c, R^{nc})$. The participation and cousinage decisions and the profit function presented in section 5 are now also a function of the price of formal land. We denote $L_f(x)$ the mass of formal land in location x. With these additional notations and assumptions, the equilibrium definition becomes:

[40]The registration fee to obtain a formal title might be proportional to the market value of the plot. In Mali for instance, the former Doing Business database reported a registration cost of 11 percent of the plot value. In practice, however, in the absence of proper market valuation and in the presence of much larger informal costs borne by owners (in terms of bribes and time as described in Durand-Lasserve et al., 2015), the total registration cost is not likely to be proportional to the plot value. In the model, we treat the registration cost as constant, an assumption which does not qualitatively affect our findings but simplifies the analysis. We also do not differentiate between the registration cost of secure and insecure plots, an assumption which also does not qualitatively affect our findings while simplifying the analysis.

[41]This decision is motivated by agents' lower risk of eviction on formal plots compared to informal plots (see empirical Section 3 on risk perceptions).

[42]Note that, for registered plots, there is no need to model the choice of an ethnic cousin or non-cousin buyer by the seller because, as no risk remains, there is no social penalty.

Definition 3: *An equilibrium is a set of market participation, ethnic matching, and registration decisions, prices in each location x and a city border that satisfies the following equilibrium conditions:*

$$
\begin{cases}
L(x) = L_f(x) + L_q(x) + L_1(x) \leq 1 \quad \textit{for any } x & (12) \\
(P^*(x,Q,R_f,R^c,R^{nc}), F^*(x,Q,R_f,R^c,R^{nc}), C^*(x,Q,R_f,R^c,R^{nc})) & \\
\qquad \in \underset{(P,F,C)\in\{0,1\}\times\{0,1\}\times\{c,nc\}}{ArgMax} B(P,F,C,x,Q,R_f,R^c,R^{nc}) \quad \textit{for any } x & (13) \\
R_f(x) = \psi(x,u|F=1) \quad \textit{for any } x \textit{ where } L_f(x) > 0 & (14) \\
R^c(x) = \psi(x,u|F=0,C=c) \quad \textit{for any } x \textit{ where } L^c(x) > 0 & (15) \\
R^{nc}(x) = \psi(x,u|F=0,C=nc) \quad \textit{for any } x \textit{ where } L^{nc}(x) > 0 & (16) \\
x_b = \min x \ \textit{s.t.}\ B^*(x_b,1,R_f,R^c,R^{nc}) = B^*(x_b,q,R_f,R^c,R^{nc}) = R_a & (17)
\end{cases}
$$

where $B^*(x,Q,R_f,R^c,R^{nc}) \equiv B(P^*,F^*,C^*,x,Q,R_f,R^c,R^{nc})$ is the new expression for the optimized payoff. Compared to the equilibrium without registration, we have an additional condition (14) which reflects sellers' extraction of buyers' willingness to pay for formal land. The other conditions are adapted to account for $R_f(x)$ and $L_f(x)$. Conditions (15)-(16) are unchanged. The resolution of this version of the model is presented in Appendix D.

Although all cases are presented in the Appendix, for simplicity of presentation, we consider in Proposition 4 the cases where the social penalty is small and k is relatively high.[43] As shown in Appendix figure D4 to D7, these cases correspond to situations where some but not all plots are registered and where transactions only occur among ethnic cousins. We present the other cases in the Appendix.

We have the following proposition:

Proposition 4: *Denoting the new boundary zone threshold* $\mathring{x}(k,J) = \frac{1}{t}\left(y - \frac{k-J}{(1-q)(1-\pi)} - u\right)$, *the equilibrium city is organized in four zones:*

[43] Mathematically, this corresponds to the cases where $0 < J \leq \underline{J} = \pi R_a \frac{1-q}{q}$ and $R_a \frac{1-q}{q} < k < \bar{k}$, or where $0 < J \leq k(q+(1-q)\pi) - R_a(1-q)(1-\pi)$ and $\underline{k} < k < R_a \frac{1-q}{q}$] with $\underline{k} = R_a \frac{(1-q)(1-\pi)}{\pi(1-q)+q}$ and $\bar{k} = (1-q)(1-\pi)(y-u)$. We also assume that $y - u > \max(\frac{k}{(1-q)}, R_a \frac{1}{\pi q(1-\pi)})$. This hypothesis ensures the existence of the city, the possibility of having informal plots, and that $\underline{k} < \bar{k}$.

- ***The most central zone (Zone 1, for $x \in [0, \mathring{x}(k, J)]$) is fully residential with a mix of informal and formal land uses****. All owners (irrespective of the initial tenure security level of their plot) participate in the market. Some q-owners register their plot before the sale. The other q-owners do not register their plot and sell exclusively to their ethnic cousins.* 1*-owners do not register their plots and sell them exclusively to their ethnic cousins. The informal price in each location x is $R^c(x) = y - xt - u - k + J$ and the formal price for registered plots is $R_f(x) = y - xt - u$.*

- ***The next zone (Zone 2, for $x \in]\mathring{x}(k, J), \underline{x}(J)]$) is fully residential and fully informal****: All owners (q- and* 1*-owners) sell their plot informally and exclusively to ethnic cousins. In each location x, the price for these informal sales is $R^c(x) = (\pi(1 - q) + q)(y - xt - u)$.*

- ***The "close periphery" (Zone 3, for $x \in]\underline{x}(J), \bar{x}(J)]$) is a mix of informal residential and agricultural uses, with all* 1*-owners selling their land****: All* 1*-owners and some q-owners sell their plots exclusively to ethnic cousins. Whereas all* 1-owners *participate in the market, some q-owners drop out of the market. The price of land in each location x is $R^c(x) = R_a + J$.*

- ***The "far periphery" (Zone 4, for $x \in]\bar{x}(J), x_a^*]$) is a mix of informal residential and agricultural uses, with all q-owners dropping out of the market****: All* 1*-owners sell their plot exclusively to ethnic cousins and all q-owners drop out of the market. The price in each location x is $R^c(x) = y - xt - u$.*

- *The city boundary is at $x_b = x_a^* = \frac{1}{t}[y - R_a - u]$.*

Proof: See Appendix Section D.1.

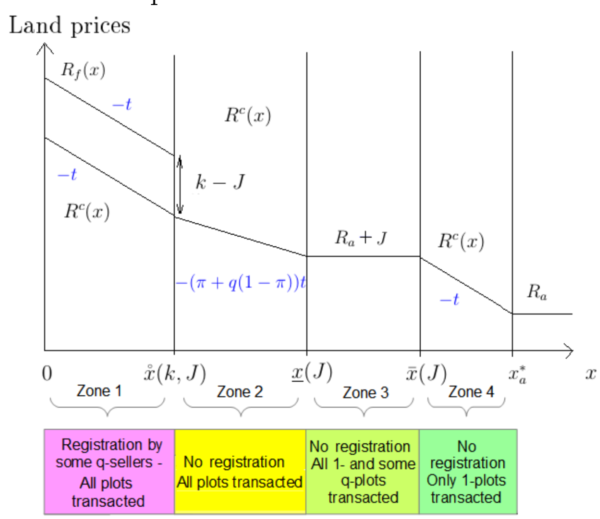
The structure of the city and the corresponding equilibrium land prices are represented in Figure 5 below.[44] The only difference with the equilibrium without registration is that the

[44] In Appendix Section D.2, we present a figure that plots the payoffs of owners and their underlying participation, ethnic matching, and registration decisions (see Figure D8).

former central residential zone is split into two zones: a new central residential zone (Zone 1) and a peripheral residential zone (Zone 2). In Zone 1, which consists of a mix of formal and informal plots, the formal price curve has a slope of $-t$, reflecting the standard trade-off between proximity to the center and land prices under full tenure security. There is a constant markup between the formal price curve and informal price curve equal to $k - J$. This equilibrium markup ensures that q-sellers are indifferent between registering at cost k and selling informally to ethnic cousins while incurring the social penalty J. In the peripheral residential zone (Zone 2, where no plots are registered), the slope of the land price curve is $-(\pi(1-q)+q)t$ as in the benchmark version of the model. New Zones 3 and 4 are the exact same zones as Zones 2 and 3 in the model without registration presented in the previous section.

Figure 5 which represents the equilibrium city structure is well aligned with the actual price and tenure patterns presented in Figure 2 in the stylized facts section.

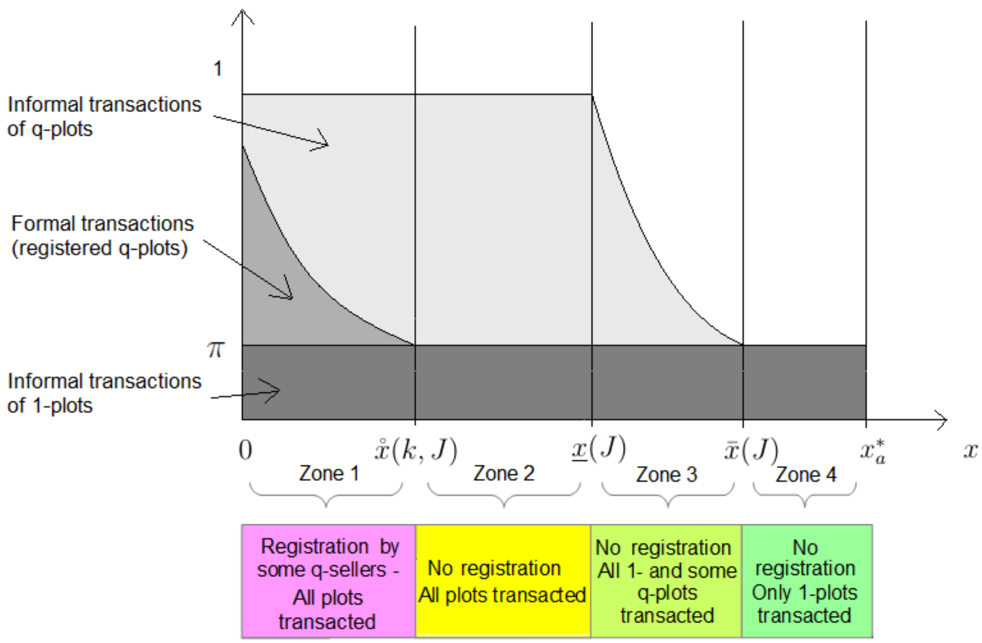
Figure 5: City structure and land prices in the model with ethnic matching and registration



Note: This figure represents the equilibrium land prices, market participation, ethnic matching and registration as a function of distance to the city center when $k > \underline{k}$ and $J < \underline{J}$. The slopes of the land price curves are indicated in blue.

We also represent in Figure 6, the number of transactions in each location and the corresponding breakdown by tenure type. In line with the stylized fact from Section 3, we show in the Appendix that registration, which is limited to Zone 1, decreases with distance to the CBD.The share of secure plots among informally transacted plots, $\pi^c(x)$, is non-monotonic. It decreases with distance to the CBD over Zone 1, is constant over Zone 2, and increases over Zone 3, as informal q-plot transactions are phased out. This indicates that the quality of cousinage transactions increases with distance to the CBD outside the registration zone.

Figure 6: Number of transactions by tenure type in equilibrium



Note: This figure represents the number of transactions by tenure type in equilibrium as a function of distance to the city center when $k > \underline{k}$ and $J < \underline{J}$.

We show that registration and ethnic matching can substitute for one another, as stated in the following proposition:

Proposition 5: *Registration and ethnic matching are substitutes: If the registration cost decreases, some landowners will shift from ethnic matching to registering their plots. If the social penalty decreases, some landowners will shift from registering their plots to ethnic matching.*

Proof: See Appendix Section D.3.

Intuitively, it is easy to see that a decrease in the registration cost increases both the extent of the registration zone and the number of landowners registering their plot in each location of the registration zone. A decrease in the social penalty reduces the registration zone and expands the fully informal sales zone (Zone 2), both toward the city center and the city periphery. The proposition illustrates how, in a context of costly registration, social relationships can partially address information asymmetry issues at the periphery of the city. This is in line with the persistence of trusted ethnic relationships in the periphery of sub-Saharan African cities (as documented by Smith 2004) where registration is also less resorted to. Proposition 5 also predicts that, if registration becomes more affordable over time, the role of ethnic relationships in governing land transactions could be phased out.

6.3 Surplus analysis

The city surplus corresponding to the equilibrium city described in Proposition 4 can be expressed as the sum of each zone's contribution to the surplus as follows:

$$
\begin{aligned}
\Sigma(J,k) = & \int_0^{\mathring{x}(k,J)} (y - xt - u - k - R_a) L_{fq}(x,J,k)\, dx \\
& + \int_0^{\mathring{x}(k,J)} \pi(y - xt - u - R_a) + (1 - \pi - L_{fq}(x,J,k))(q(y - xt - u) - R_a))\, dx \\
& + \int_{\mathring{x}(k,J)}^{\underline{x}(J)} \pi(y - xt - u - R_a) + (1-\pi)(q(y - xt - u) - R_a)\, dx \\
& + \int_{\underline{x}(J)}^{\bar{x}(J)} \pi(y - xt - u - R_a) + L_q^c(x,J)(q(y - xt - u) - R_a)\, dx \\
& + \int_{\bar{x}(J)}^{x_a^*} \pi(y - xt - u - R_a)\, dx
\end{aligned}
$$

where L_{fq} (see Appendix section D.1) is the mass of q-owners registering their plot, and L_q^c is the mass of q-owners informally selling to an ethnic cousin. The first two integrals are the surplus associated with the registration zone (Zone 1), and the next three integrals correspond to the respective surplus contributions in Zones 2, 3 and 4.[45]

We have the following proposition:

[45] For other values of J and k, the surplus formulas are very similar. Only the boundaries of the integrals need to be changed to correspond to the boundaries of the different zones represented on Appendix graphs (D4-D7).

Proposition 6: *Adding registration to ethnic matching in the model always increases the surplus. The surplus gain is greater for a large social penalty J and a small registration cost k. The equilibrium is optimal if and only if $J \geq k$. When $J < k$, too little registration occurs.*

Proof: See Appendix section D.4.

As expected, the introduction of a second institution to address risk and information asymmetry is socially beneficial. Also as expected, registration is more efficient when registration costs are low, as more owners can register their plots, thereby removing both risk and information asymmetry. Note that with the introduction of registration, the externality in our model now revolves around participation, ethnic matching and, additionally here, registration decisions, which are not internalized by agents. We show in Appendix D.4 that the optimal city requires that all q-plots in a central zone of the city are registered (this allows reducing the risk on plots that have a strong contribution to the surplus). However, this does not happen in the competitive equilibrium when the social penalty is too small and the registration cost too high so that some q-sellers in the registration zone prefer not register their plot and exploit the information asymmetry by selling informally. In that case, a market failure still occurs. The innovative finding mentioned in Proposition 6 is that optimality is reached when the social penalty is sufficiently large compared to the registration fee, which incentivizes all q-sellers in the central zone to register their plots (see Appendix graphs D4-D7). Proposition 6 also points to a complementarity between the two institutions. This is because ethnic matching allows for more efficient registration decisions by incentivizing more q-sellers to formalize as J increases.[46] Finally, observe that, when the registration fee is zero, we have $J \geq k$, so that the equilibrium is optimal. Because all agents may formalize at a zero cost, all risk is removed from the model and the surplus corresponds to the risk-free optimal surplus calculated in Section 4 ($\Sigma(0, J) = \Gamma^*$).

[46]In Appendix D.4, we also show that the equilibrium with information asymmetry, ethnic matching and registration dominates the equilibrium in a sheer registration model with information asymmetry but without ethnic matching. This is because although ethnic matching leads to a reduced registration of both 1-plots and q-plots, the positive contribution of the former is greater than the negative contribution of the latter.

6.4 Registration subsidy

We now analyze the effects on surplus of a registration subsidy, which allows to reduce registration costs for all registering landowners by a same fixed amount (i.e., k is reduced to $k - s$, where s is the individual subsidy received by each landowner who decides to register). This registration subsidy is financed through a lump-sum tax on both rural and urban residents (a funding scheme which has no impact on the extent of the city and only affects registration decisions). We already showed that in strong cousinage contexts, where $J \geqslant k$, the equilibrium is optimal so that a registration subsidy would reduce the surplus by increasing the registration zone beyond its optimal limit. In contexts where $J < k$, we are only able to identify sufficient conditions showing that a registration subsidy improves the surplus.

We have the following proposition:

Proposition 7:

- When the social penalty is relatively large ($J \geqslant k$), a registration subsidy will decrease the surplus.
- When the registration cost is sufficiently large ($k > \underline{k}$) and the social penalty is relatively small enough ($J < min(\pi k, (k + R_a)(\pi(1 - q) + q) - R_a) < k$), a partial registration subsidy is desirable.

Proof: See Appendix D.5.

The intuition underpinning the second part of Proposition 7 is as follows: Because the equilibrium has too few q-owners that formalize and a registration zone that is too small, the subsidy increases both the incentives for the owners of risky plots to formalize and the registration zone, thereby improving the surplus. Note, however, that under a full subsidy, the registration zone would become too large and lead to "overformalization". We show in Appendix D.5. that the optimal subsidy s^* needs to be lower than the registration fee net of the social penalty ($s^* < k - J$).

7 Conclusion

The economic development literature has argued that informal institutions of various kinds stand in when formal institutions do not work very well (see e.g., Platteau, 1996 and 2000, Braselle et al., 2002, Munshi, 2004, Panman and Lozano Gracia, 2022, Williamson and Kerekes, 2011). This is certainly true in the case of land property rights in sub-Saharan African cities where many households cannot afford the cost of registration. The model proposed in this paper helped analyze how ethnic matching along trusted relationships—which is extremely common in the whole region—plays such a role when applied to informal transactions of urban land. We empirically documented this feature using a unique survey in Bamako, Mali, that showed that potential buyers of land prefer to buy informal plots from their ethnic cousins and that potential sellers of (insecure) informal plots prefer to transact with non-cousins. Our model showed that tenure insecurity and information asymmetry about contested ownership result in a market failure in the form of insufficient urban development, a result that was first derived by Picard and Selod (2020) in a related but different theoretical setting. The main contribution of our paper is to show how ethnic matching and registration are able to address this market failure.

We showed that the possibility to transact along trusted ethnic relationships—which involves a social penalty incurred by sellers if they deceive buyers about the risk of contested ownership on transacted plots—is always beneficial, as it alleviates information asymmetry. When the social penalty is very high, ethnic matching may even fully remove information asymmetry. We also showed that adding the possibility of registering land in a cadastre further improves the surplus, as registration addresses information asymmetry and reduces risk at the same time. There is also an additional gain from registration in the presence of ethnic matching, as ethnic matching tends to separate risks (with transactions between cousins favoring secure plots) allowing registration to be resorted to exclusively by owners of insecure plots. This complementarity also makes it possible for a registration subsidy to target insecure plots, allowing the city to reach its social optimum.

Finally, our paper sheds light on an ongoing debate in the policy world where it has been argued that the promotion of freehold titles as the unique acceptable solution to hold land could have been misguided (Barry and Augustinus, 2016). Our results lend some credit to that position by recognizing the second-best role of social ties in reducing uncertainty in land transactions. As long as informal transactions continue—potentially because of high registration fees—informal institutions such as cousinage are likely to persist and have a beneficial impact on social welfare. This said, although codified ethnic alliances have long been a source of social cohesion and stability in sub-Saharan Africa, some authors note that informal institutions might in the long run give rise to ethnic tensions (Keefer and Knack, 2002, Letrouit, 2021). Cultural norms could also be weakening over time, especially in conflict-affected economies. Given the challenge of rapid urbanization, it will be all the more important to ensure that an affordable formal land registration system is in place to accompany urban population growth. We leave the study of underlying factors governing these changes and the speed at which they could happen for future research.

References

[1] Adam, A.G. (2014). Informal settlements in the peri-urban areas of Bahir Dar, Ethiopia: An institutional analysis. *Habitat International*, 43, pp.90-97.

[2] Agegnehu, S.K., Fuchs, H., Navratil, G., Stokowski, P., Vuolo, F. and Mansberger, R. (2016). Spatial urban expansion and land tenure security in Ethiopia: Case studies from Bahir Dar and Debre Markos peri-urban areas. *Society & Natural Resources*, 29(3), pp.311-328.

[3] Akerlof, G. A. (1978). The market for "lemons": Quality uncertainty and the market mechanism. In *Uncertainty in economics* (pp. 235-251). Academic Press.

[4] Ali, D.A., Collin, M., Deininger, K., Dercon, S., Sandefur, J. and Zeitlin, A. (2016). Small price incentives increase women's access to land titles in Tanzania. *Journal of Development Economics*, 123, pp.107-122.

[5] Alonso, W. (1964) *Location and land use. Toward a general theory of land rent*, Cambridge, Mass.: Harvard Univ. Press.

[6] Alves, G. (2021) Slum Growth in Brazilian Cities., *Journal of Urban Economics, 122, p.103327.*

[7] Attino, M. (2021) Parenté à Plaisanterie ou Cousinage Entre les Dogons et les Songhays au Mali. *European Scientific Journal,* 17(13), 20 pages.

[8] Arruñada, B. (2012). *Institutional foundations of impersonal exchange: Theory and policy of contractual registries*, University of Chicago Press.

[9] Barry, M., and Augustinus, C. (2016) *Framework for Evaluating Continuum of Land Rights Scenarios.* UN-Habitat GLTN Report 4, 65 pages.

[10] Bertrand, M. (2016). *Analyse des données de la conservation foncière au Mali, rapport final d'étude.* Ministère des domaines de l'Etat et des affaires foncières.

[11] Bertrand, M. (2019). A cadastre for Mali? The production of land titles and the challenge of property data on the periphery of Bamako. *Land Use Policy*, 81, pp.371-381.

[12] Besley, T. (1995) Property Rights and Investment Incentives: Theory and Evidence from Ghana. *Journal of Political Economy*, 103, 903-37.

[13] Bird, J. and Venables, A.J. (2020). Land tenure and land-use in a developing city: A quantitative spatial model applied to Kampala, Uganda. *Journal of Urban Economics*, 119, p.103268.

[14] Birkeland, A. (2007). Managing Fulbe identity: exchange relations and ethnic discourses in a village in Central Mali, *Master's thesis*.

[15] Brasselle, A.S., Gaspart, F. and Platteau, J.P., 2002. Land tenure security and investment incentives: puzzling evidence from Burkina Faso. *Journal of Development Economics*, 67(2), pp.373-418.

[16] Brueckner, J. (2013) Urban Squatting with Rent-Seeking Organizers. *Regional Science and Urban Economics*, 43, 561-569.

[17] Brueckner, J. and H. Selod (2009) A Theory of Urban Squatting and Land-tenure Formalization in Developing Countries. *American Economic Journal: Economic Policy*, 1, 28-51.

[18] Brueckner, J., Rabe C. and H. Selod (2019) Backyarding: Theory and Evidence for South Africa. *Regional Science and Urban Economics*, Volume 79, November 2019.

[19] Cai, Y., Selod H. and J. Steinbuks (2018) Urbanization and property rights. *Regional Science and Urban Economics*, Elsevier, vol. 70(C), pages 246-257.

[20] Cattaneo, A., & Robinson, S. (2020). Multiple moves and return migration within developing countries: A comparative analysis. *Population, Space and Place*, 26(7), e2335.

[21] Cavalcanti, T., D. Da Mata and M. Santos (2019) On the Determinants of Slum Formation. *Economic Journal*, 129(621), 1971-1991.

[22] Chen, C., Restuccia, D., & Santaeulàlia-Llopis, R. (2022). The effects of land markets on resource allocation and agricultural productivity. *Review of Economic Dynamics*, 45, pp.41-54.

[23] Diallo, Y (2006) Joking relationships in western Burkina Faso. *Zeitschrift für Ethnologie*, pages 183-196.

[24] Dunning, T. and L. Harrison (2010) Cross-cutting cleavages and ethnic voting: An experimental study of cousinage in Mali. *American Political Science Review*, 104(01), 2139.

[25] Durand-Lasserve, A., Durand-Lasserve, M. and H. Selod (2015) *Land Delivery Systems in West-African Cities. The Example of Bamako, Mali*, Washington, DC: World Bank and Agence Française de Développement, 106 pages.

[26] Fearon, J. and D. Laitin (1996) Explaining Interethnic Cooperation. *American Political Science Review*, 90, 4, 715-735.

[27] Field, E. (2007) Entitled to Work: Urban Tenure Security and Labor Supply in Peru. *Quarterly Journal of Economics*, 122, 1561-1602.

[28] Fouéré, M-A. (2004) *L'objet ethnologique "Relations à plaisanteries" dans l'espace est-africain (Tanzanie) : De la construction savante d'une coutume à la restitution des situations sociales de l'Utani, PhD dissertation*, Ecole des Hautes Etudes en Sciences Sociales, 366 pages.

[29] Freedman, J. (1977) Joking, affinity and the exchange of ritual services among the Kiga of northern Rwanda: an essay on joking relationship theory. *Man*, New Series, Vol. 12, No. 1, pages 154-165.

[30] Fujita, M. (1989). *Urban Economic Theory: Land Use and City Size.* Cambridge University Press.

[31] Galiani, S., P. Gertler, R. Cooper, S. Martinez, A. Ross and R. Undurraga (2017) Shelter from the Storm: Upgrading Housing Infrastructure in Latin American Slums. *Journal of Urban Economics*, 98, 187-213.

[32] Galiani, S., and Schargrodsky, E. (2010) Property rights for the poor: Effects of land titling. *Journal of Public Economics*, 94(9-10), 700-29.

[33] Gottlieb, C., & Grobovšek, J. (2019). Communal land and agricultural productivity. *Journal of Development Economic*s, 138, 135-152.

[34] Habyarimana, J., M. Humphreys, D. Posner and J. Weinstein (2007) Why Does Ethnic Diversity Undermine Public Good Provision. *American Political Science Review*, 101, 4, 709-725.

[35] Hagberg, S. (2006) The politics of joking relationships in Burkina Faso. *Zeitschrift für Ethnologie*, pages 197-214.

[36] Henderson, J. V., Venables, A. J., Regan, T., & Samsonov, I. (2016). Building functional cities. *Science*, 352(6288), 946-947.

[37] Henderson, J. V., Regan, T., & Venables, A. J. (2021). Building the city: from slums to a modern metropolis. *Review of Economic Studies*, 88(3), 1157-1192.

[38] INSTAT / Institut National de la Statistique, Mali (2009). Recensement General de la Population et de l'Habitat.

[39] Jimenez, E. (1985) Urban Squatting and Community Organization in Developing Countries. *Journal of Public Economics*, 27 (1), 69–92.

[40] Jones, R. (2007) "You eat beans!": Kin-based joking relationships, obligations, and identity in urban Mali, *Digital Commons, Malacester College*.

[41] Keefer, P. and Knack, S., 2002. Polarization, politics and property rights: Links between inequality and growth. *Public Choice,* 111(1), pp.127-154.

[42] La Ferrara, E. (2003) Kin Group and Reciprocity: A Model of Credit Transactions in Ghana. *American Economic Review*, 93, 5, 1730-1751.

[43] Lanjouw, J. O., and Levy, P. I. (2002). Untitled: A study of formal and informal property rights in urban Ecuador*. Economic Journal*, 112(482), 986-1019.

[44] Letrouit, L. (2021). Why can't we be friends? An evolutionary approach to ethno-cultural hierarchies. HAL working paper.

[45] Macours, K. (2014). Ethnic divisions, contract choice, and search costs in the Guatemalan land rental market. *Journal of Comparative Economics*, 42(1), 1-18.

[46] Macours, K., De Janvry, A., and Sadoulet, E. (2010). Insecurity of property rights and social matching in the tenancy market. *European Economic Review*, 54(7), 880-899.

[47] Marx, B., Stoker, T. M., and Suri, T. (2019). There is no free house: Ethnic patronage in a Kenyan slum. *American Economic Journal: Applied Economics*, 11(4), 36-70.

[48] Mauss, M. (1923) Essai sur le don: forme et raison de l'échange dans les sociétés archaïques. *L'Année sociologique* (1896/1897-1924/1925), 1:30-186.

[49] Mills, E. (1967) An aggregative model of resource allocation in a metropolitan area. *American Economic Review*, 57(2):197-210.

[50] Munshi, K., 2004. Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Journal of development Economics*, 73(1), pp.185-213.

[51] Muth, R. (1969) *Cities and housing; the spatial pattern of urban residential land use*. University of Chicago Press.

[52] Nakamura, S. (2017) Does Slum Formalisation Without Title Provision Stimulate Housing Improvement? A Case of Slum Declaration in Pune, India. *Urban Studies*, 54(7), 1715-1735.

[53] Neimark, B., Toulmin, C. and Batterbury, S., 2018. Peri-urban land grabbing? Dilemmas of formalising tenure and land acquisitions around the cities of Bamako and Ségou, Mali. *Journal of Land Use Science*, 13(3), pp.319-324.

[54] Ndiaye, A (1992) Correspondances ethno-patronymiques et parenté plaisante: une problématique d'intégration à large échelle. *Environnement africain*, (32):34.

[55] Pande, R., & Udry, C. (2005). Institutions and development: A view from below. Yale University Economic Growth Center Discussion Paper, (928).

[56] Panman, A. and Lozano Gracia, N., 2021. Titling and Beyond: Evidence from Dar es Salaam, Tanzania. *Land Use Policy*, 2022, 117, C,Article 105905.

[57] Picard, P. and H. Selod (2020) Customary land conversion and city formation in sub-Saharan Africa, *World Bank Policy Research Working Paper WPS9192*.

[58] Pfeiffer, B., Rabe, C., Selod, H., and Viguié, V. (2019). Assessing Urban Policies Using a Simulation Model with Formal and Informal Housing: Application to Cape Town, South Africa*, World Bank Policy Research Working Paper WPS8921*.

[59] Platteau, J.P., 1996. The evolutionary theory of land rights as applied to sub-Saharan Africa: a critical assessment. *Development and Change*, 27(1), pp.29-86.

[60] Platteau, J.P., 2000. Allocating and enforcing property rights in land: informal versus formal mechanisms in Subsaharan Africa. *Nordic Journal of Political Economy*, 26(1), pp.55-81.

[61] Selod, H. and L. Tobin (2018) The Informal City, *World Bank Policy Research Working Paper WPS8482*, 57 pages.

[62] Shah, N. (2014) Squatting on Government Land. *Journal of Regional Science*, 54, 114-136.

[63] Smith, E. (2004) Les cousinages de plaisanterie en Afrique de l'ouest, entre particularismes et universalismes. *Raisons politiques*, (1):157-169.

[64] Smith, E. (2006) La nation par le côté. le récit des cousinages au Sénégal. *Cahiers d'études africaines*, 46(184):907-965.

[65] Turnbull, G. (2008) Squatting, Eviction and Development. *Regional Science and Urban Economics*, 38, 1-15.

[48] Williamson, C. R., & Kerekes, C. B. (2011). Securing private property: formal versus informal institutions. *Journal of Law and Economics*, 54(3), 537-572.

[66] World Bank (2020) *Système d'Approvisionnement en Terrains Résidentiels de la Ville de Yaoundé*. Unpublished document. 73 pages.

[67] World Bank Doing Business 2004-2020 (databse): https://archive.doingbusiness.org/en/doingbusiness

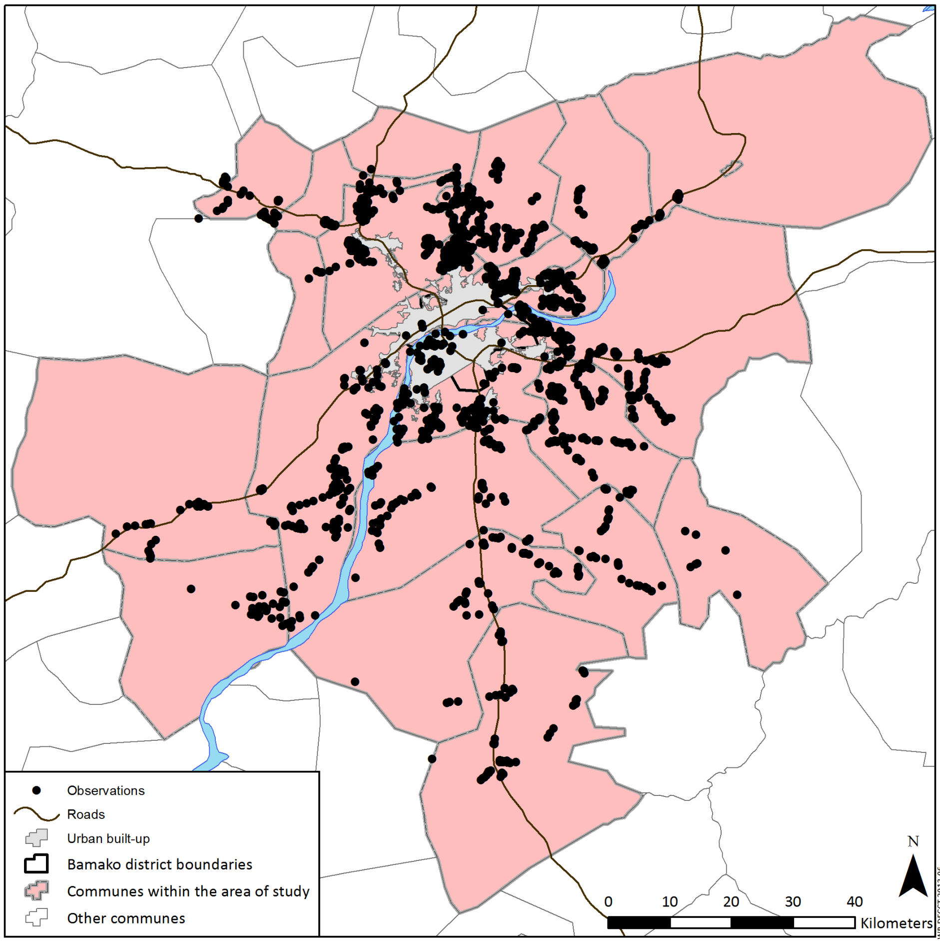
Appendix A - Data and empirical analysis

In the stylized fact section of this paper, we make use of two data sets collected by the World Bank. The first data set is a survey of unbuilt land plots that was carried out in 2012 in the greater Bamako area (i.e., the six municipalities of the Bamako District and eight surrounding municipalities: Kalabancoro, Mandé, Dogodouman, Dialakorodji, Sangarébougou, N'Gabacoro-droit, Baguinéda-Camp et Sanankoroba). Although a strict random selection of plots was not possible in the absence of a sampling frame (due to the absence of an exhaustive cadastre), information on unbuilt land plots that were transacted during the period 2009-2012 was collected at regular intervals around roads extending outward from Bamako, ensuring a uniform coverage of the urban area that is sufficient to shed light on spatial patterns of tenure in the city. The data set includes information on plot characteristics (whether the plot is destined to a residential or an agricultural use, the plot area, the presence of water and electricity, the distance to the closest paved road, the GPS coordinates of the plot, its tenure status, and the transacted price in CFA Francs for plots that were subject to a monetary transaction). Map A1 below represents the study area and the locations of unbuilt plots included in the survey.

The data set includes 1,007 plots with monetary transactions and complete information on price, location, area and the presence of infrastructure services which we used to generate Figures 1 and 2 in Section 3.1. For the right-hand side panel of Figure 2, we ran a hedonic regression of the log square meter land price on whether the plot is located to the south of the river (reflecting market segmentation), the log plot area (to capture diminishing returns), the distance to the road (to capture local accessibility), water and electricity (to capture the capitalization of infrastructure), and transaction year dummies (to account for inflation). We intentionally excluded distance to the city center, which is thus included in the error term. The results from the hedonic regression are presented below in Table A1 for all plots (column 1), for formal plots that have a permit to occupy or a property title (column 2), and for informal plots (column 3). The right-hand side panel of Figure 2 in Section 3.1 is then generated by plotting the residuals of regressions (2) and (3) on distance to the city center of Bamako.

The second data set was also collected by the World Bank in March 2022 on a sample of 1,106 individuals in the greater Bamako area. Stratified random sampling was applied to ensure that all 14 municipalities were covered, and within each municipality, 2 or 3 villages (for rural municipalities) or neighborhoods (for urban municipalities) were randomly selected. In each village or neighborhood, about 25 respondents were surveyed. The respondents were randomly selected from "grins", which are places where the population meets to discuss on a daily basis. The survey was rolled out by a field

Figure A1: Map of observations in the 2012 survey of the Greater Bamako Area



Note: This maps shows the locations of the unbuilt plots surveyed by the World Bank in the District of Bamako and surrounding municipalities. Source: Durand-Lasserve et al. (2015)

coordinator and 10 qualified enumerators over a period of 9 days following an initial pilot. The questionnaire was administered with smartphones running the Survey Monkey application. The collected data includes information on the demographic and ethnic characteristics of respondents, opinions regarding cousinage practices, experience of land sales and purchases and land tenure documentation held, as well as experience of land conflicts. Respondents were asked to choose whether they would find risky the purchase of a plot from randomly matched fictive individuals for fictive purchase situations of varying risks (formal plot, informal plot, and customary plot) and attempts they might make at formalizing the plot after a purchase from these fictive individuals. They were also asked to choose among fictive buyers of land randomly matched with them in contexts of fictive land sales exhibiting various risks (i.e., with or without a competing claim on the plot). Fictive matching was randomly drawn to ensure that the names and ethnic groups of the fictive buyers or sellers would correspond to ethnic cousins or non-cousins of the respondents.

Table A1: Hedonic regressions

	(1) All plots Log(land price) (CFA/m2)	(2) Formal plots Log(land price) (CFA/m2)	(3) Informal plots Log(land price) (CFA/m2)
South bank dummy	0.495***	0.555***	0.506***
	(0.072)	(0.168)	(0.068)
Log(area) (m2)	-0.740***	-0.510***	-0.778***
	(0.040)	(0.099)	(0.035)
Distance to road (km)	-0.100***	-0.161***	-0.074***
	(0.008)	(0.025)	(0.007)
Water dummy	1.107***	0.614*	0.927***
	(0.201)	(0.322)	(0.221)
Electricity dummy	1.308***	0.819	1.189*
	(0.433)	(0.529)	(0.663)
Sale year dummy 2010	0.223**	0.063	0.098
	(0.096)	(0.260)	(0.086)
Sale year dummy 2011	0.256***	-0.344	0.213**
	(0.094)	(0.254)	(0.085)
Sale year dummy 2012	0.686***	0.551*	0.356**
	(0.154)	(0.332)	(0.151)
Constant	11.559***	11.625***	11.485***
	(0.273)	(0.705)	(0.245)
Observations	1,007	228	779
R-squared	0.429	0.450	0.505

Note: This table shows the results of a regression of land prices expressed in logarithms on plot characteristics. Column (1) is for the full sample, whereas columns (2) and (3) are for formal and informal plots respectively. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A3 replicates Table 3 on a smaller sample comprising all men 40 years or older who were already involved in a housing transaction.

Table A2: Descriptive statistics

	(1) Central municipality		(2) Peripheral municipality		(3) Difference	(4) All	
	Mean	S.d.	Mean	S.d.		Mean	S.d.
Age	44.88	14.46	42.89	14.61	-1.99*	43.77	14.57
Women	0.23	0.42	0.18	0.39	-0.05	0.20	0.40
Ethnic group							
Bambaras	0.35	0.48	0.41	0.49	0.06	0.38	0.49
Malinkés	0.12	0.33	0.18	0.38	0.06**	0.15	0.36
Peuhls	0.14	0.34	0.12	0.33	-0.01	0.13	0.33
Other ethnic group	0.42	0.49	0.31	0.46	-0.10***	0.36	0.48
Opinions							
Importance of family	3.66	0.74	3.74	0.60	0.08*	3.71	0.66
Importance of social relationships	3.51	0.77	3.55	0.70	0.05	3.53	0.73
Importance to abide by cousinage rules	3.48	0.94	3.39	1.08	-0.10	3.43	1.02
Contravening cousinage should be punished	2.73	1.17	2.69	1.32	-0.04	2.71	1.25
Land purchase experience							
Previously bought land	0.30	0.46	0.26	0.44	-0.03	0.28	0.45
Among individuals who previously bought land:							
Bought from family or cousin	0.42	0.50	0.41	0.49	-0.02	0.41	0.49
Land documentation:							
Ownership title	0.17	0.38	0.05	0.22	-0.12***	0.11	0.31
Precarious title	0.27	0.45	0.14	0.35	-0.13**	0.20	0.40
Allocation letter	0.60	0.49	0.64	0.48	0.04	0.62	0.49
Authenticated sales certificate	0.40	0.49	0.36	0.48	-0.04	0.38	0.49
Non-authenticated sales certificate	0.09	0.29	0.06	0.24	-0.03	0.07	0.26
No document	0.01	0.12	0.11	0.31	0.10***	0.06	0.25
Strongest land documentation:							
Ownership title	0.17	0.38	0.05	0.22	-0.12***	0.11	0.31
Precarious title	0.21	0.41	0.14	0.35	-0.07	0.17	0.38
Allocation letter	0.43	0.50	0.54	0.50	0.11*	0.49	0.50
Authenticated sales certificate	0.15	0.36	0.14	0.35	-0.01	0.15	0.35
Non-authenticated sales certificate	0.03	0.16	0.04	0.19	0.01	0.03	0.18
No document	0.01	0.12	0.10	0.30	0.09***	0.06	0.24
Land sale experience							
Previously sold land	0.11	0.31	0.17	0.38	0.06**	0.14	0.35
Among individuals who previously sold land:							
Sold to family or cousin	0.63	0.49	0.42	0.50	-0.22*	0.49	0.50
Land documentation:							
Ownership title	0.12	0.32	0.06	0.23	-0.06	0.08	0.27
Precarious title	0.17	0.38	0.07	0.25	-0.11	0.10	0.30
Allocation letter	0.60	0.50	0.69	0.47	0.09	0.66	0.48
Authenticated sales certificate	0.63	0.49	0.38	0.49	-0.25**	0.46	0.50
Non-authenticated sales certificate	0.12	0.32	0.06	0.23	-0.06	0.08	0.27
No document	0.06	0.24	0.09	0.29	0.04	0.08	0.28
Strongest land documentation:							
Ownership title	0.12	0.32	0.06	0.23	-0.06	0.08	0.27
Precarious title	0.12	0.32	0.07	0.25	-0.05	0.08	0.28
Allocation letter	0.42	0.50	0.63	0.49	0.21*	0.56	0.50
Authenticated sales certificate	0.23	0.43	0.15	0.36	-0.08	0.18	0.38
Non-authenticated sales certificate	0.06	0.24	0.02	0.14	-0.04	0.03	0.18
No document	0.06	0.24	0.08	0.28	0.03	0.08	0.27
Land conflict experience							
Land conflict expericence in their inner social circle	0.37	0.48	0.39	0.49	0.01	0.38	0.49
Observations	487		619			1106	

Note: Descriptive statistics from a survey of grins participants in the greater Bamako area (2022). Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Replication of Table 3 on a subsample consisting of all men of 40 years of age or older with previous housing transaction experience (multinomial logit)

		(1)	(2)
Cousin	Informal (low risk)		-0.0316
			(0.220)
	Informal (high risk)		-0.0483
			(0.220)
	Informal (both risks)	-0.0398	
		(0.187)	
Non-cousin	Informal (low risk)		1.153***
			(0.308)
	Informal (high risk)		1.226***
			(0.303)
	Informal (both risks)	1.190***	
		(0.281)	
Observations		723	723
Pseudo R^2		0.169	0.169

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Multinomial logit regressions include controls for respondent's age, occupation, municipality, gender, dummies indicating previous purchase and sale experience, and whether the respondent was faced with or knew someone who was faced with a land conflict, and a constant.

Appendix B - The benchmark model

B.1. Proof of Proposition 1 - Competitive equilibrium in the benchmark model

To derive the competitive equilibrium, we study, in each x, all possible combinations of participation decisions that 1- and q-owners may take to satisfy (4) subject to (5). We then study the stability of these configurations and retain only the stable one. We finally check that this stable configuration verifies the equilibrium conditions (3)-(6).

Sellers' participation decisions and spatial city configuration We consider the three possible cases in terms of 1-sellers' participation decisions, i.e. $L_1(x) = \pi$, $L_1(x) \in]0, \pi[$ or $L_1(x) = 0$. For each case, we then derive the implications for the participation decisions of q-sellers and find the set of compatible city locations for these participation decisions.

- Let's start with the case $L_1(x) = \pi$, which means that all 1-owners participate in the market. Because q-owners face the same payoff function as 1-owners (since they cannot be distinguished from one another), all of them also participate in the market so that $L_q(x) = 1 - \pi$. Therefore, the proportion of 1-sellers among all sellers in x, $\pi(x) \equiv \frac{L_1(x)}{L_1(x)+L_q(x)}$, is equal to π. Plugging this expression into the participation constraint of 1-sellers $B(P = 1, x, Q = 1, R) \geq R_a$ simplifies to $x \leq \frac{1}{t}(y - \frac{R_a}{\pi+q(1-\pi)} - u) = x_a$ as defined in Proposition 1. We have shown that:

$$L_1(x) = \pi \Rightarrow (L_1(x), L_q(x)) = (\pi, 1 - \pi) \Rightarrow x \leq x_a$$

- Let's consider the second case $L_1(x) \in]0, \pi[$, which means that only a fraction of 1-owners located in x sell their land, requiring indifference between participation and non-participation with $B(P = 1, x, Q = 1, R) = \{\pi(x) + q(1 - \pi(x))\} (y - xt - u) = R_a$. Because q-owners face the same payoff as 1-owners, they are also indifferent between participation and non-participation in the market. The above indifference condition provides an explicit formula for $\pi(x)$ and thus for the ratio $\frac{L_q(x)}{L_1(x)} = \frac{(y-xt-u-R_a)}{R_a-q(y-xt-u)}$. Observe that the numerator in this ratio represents the maximum net gain that a seller can obtain from a sale in location x (since the buyer of a plot would be willing to pay $y - xt - u$ if he knew for sure that the plot is secure). Given that plots are transacted in x under information asymmetry, the numerator of $\frac{L_q(x)}{L_1(x)}$ is necessarily positive, which implies $x < \frac{1}{t}(y - R_a - u) \equiv x_a^*$. Furthermore, in order to have $\frac{L_q(x)}{L_1(x)} > 0$, we must also have $R_a - q(y - xt - u) > 0$, which requires that $x > \frac{1}{t}(y - \frac{R_a}{q} - u \equiv x_a^q)$. We have shown that:

$$L_1(x) \in]0,\pi[\Rightarrow (L_1(x), L_q(x)) \in]0,\pi[\times]0,1-\pi[\Rightarrow x \in]x_a^q, x_a^*[$$

- In the third case, $L_1(x) = 0$, which means that 1-owners prefer not to sell. As q-owners have the same payoffs as 1-owners, they also prefer not to sell, so that $L_1(x) = L_q(x) = 0$. Observe that we are in a polar case where the function $\pi(x)$ is actually not defined. From a buyer's perspective, given the shares of secure and insecure plots in location x, if a plot were to be offered on the market, it would be a secure plot with probability π. Non-participation thus requires $R_a \geq (\pi + (1-\pi)q)(y - xt - u)$, where the RHS is the willingness to pay of a buyer in x. The latter inequality boils down to $x \geq x_a$. We have shown that:

$$L_1(x) = 0 \Rightarrow (L_1(x), L_q(x)) = (0,0) \Rightarrow x \geq x_a$$

We have derived necessary conditions for the three above cases. Since it can easily be checked that $x_a^q < x_a < x_a^*$, this implies the following spatial configuration:

- For any $x \leq x_a^q$, we have $(L_1(x), L_q(x)) = (\pi, 1-\pi)$.
- For any $x \geq x_a^*$, we have $(L_1(x), L_q(x)) = (0,0)$.
- On the interval $x \in]x_a^q, x_a[$, one may encounter any $(L_1(x), L_q(x)) \in]0,\pi[\times]0,1-\pi[$ (second case) or $(L_1(x), L_q(x)) = (\pi, 1-\pi)$ (first case).
- For $x = x_a$, one may encounter any $(L_1(x), L_q(x)) \in]0,\pi[\times]0,1-\pi[$ (second case) or $(L_1(x), L_q(x)) = (\pi, 1-\pi)$ (first case) or $(L_1(x), L_q(x)) = (0,0)$ (third case).
- On the interval $x \in]x_a, x_a^*[$, one may encounter any $(L_1(x), L_q(x)) \in]0,\pi[\times]0,1-\pi[$ (second case) or $(L_1(x), L_q(x)) = (0,0)$ (third case).

Stability of the different configurations The multiplicity of solutions for $L_1(x)$ and $L_q(x)$ on $x \in]x_a^q, x_a^*[$ implies that we potentially have a continuum of equilibria. To study the stability of each of the possible combinations identified, we look at whether each combination is robust to a small deviation in the participation decisions made by sellers in x. We have three cases, depending on whether we have full-participation, no-participation, or partial participation in the market. We have the following results for each one of these cases:

- The full-participation case $((L_1(x), L_q(x)) = (\pi, 1-\pi))$ is stable on $[0, x_a]$. To show this, observe that the participation constraint of owners is $\{\pi + q(1-\pi)\}\ (y - xt - u) > R_a$. If a mass ϵ of owners stops selling, owners will now compare $\{\pi(x) + q(1 - \pi(x))\}\ (y - xt - u)$ and R_a to decide whether to participate in the market. As buyers will expect deviant owners to include $\epsilon\pi$ 1-owners and $\epsilon(1-\pi)$ q-owners, $\pi(x)$ remains unchanged and equal to π. It follows that the market participation constraint is unchanged so that deviant sellers will come back to their initial decision.

- The no–participation case $((L_1(x), L_q(x)) = (0, 0))$ is stable on $]x_a, x_a^*]$. To show this, observe that this case is characterized by the participation constraint $R_a > (\pi + (1-\pi)q)(y - xt - u)$, where the RHS is the expected plot price in case of a land sale (given the proportions of 1 and q-owners). If a mass ϵ of owners starts selling, buyers will expect these deviant owners to include $\epsilon\pi$ 1-owners and $\epsilon(1-\pi)$ q-owners, leading to $\pi(x) = \pi$. As in the previous case, the participation constraint is unchanged, so that deviant owners come back to their initial decision.

- The partial participation cases $((L_1(x), L_q(x)) \in]0, \pi[\times]0, 1-\pi[)$ are not stable on $]x_a^q, x_a^*[$. To show this, observe that these cases are characterized by equality $\left[\frac{L_1(x)(1-q)}{L_1(x)+L_q(x)} + q\right](y - xt - u) = R_a$, which reflects owners' indifference between participating and not participating in the market. We need to consider two sub-cases here, depending on whether $x \in]x_a^q, x_a[$ or $x \in]x_a, x_a^*[$.

  - First, consider $x < x_a$. This is equivalent to $[\pi(1-q) + q]\,(y - xt - u) > R_a$. And, as we know that $[\pi(x)(1-q) + q]\,(y - xt - u) = R_a$, we clearly have: $\pi(x) < \pi$. Consider now that a mass ϵ of owners start selling. Buyers expect that $\frac{\pi - L_1(x)}{1 - L_1(x) - L_q(x)}\epsilon$ of them are 1-sellers and that $\frac{1-\pi-L_q(x)}{1-L_1(x)-L_q(x)}\epsilon$ are q-sellers.[47] Denoting $\eta = \frac{\epsilon}{1-L_1(x)-L_q(x)}$, the price of a land plot becomes $[\Pi(x,\eta)(1-q) + q]\,(y - xt - u)$ where $\Pi(x,\eta) = \frac{L_1(x)+\eta(\pi-L_1(x))}{L_1(x)+\eta(\pi-L_1(x))+L_q(x)+\eta(1-\pi-L_q(x))}$. It can easily be shown that $\frac{\partial \Pi}{\partial \eta} = \frac{\pi(L_1(x)+L_q(x))-L_1(x)}{[(L_1(x)+L_q(x))(1-\eta)+\eta]^2} > 0$ because $\pi(x) < \pi$. The new price of land is therefore increased by the deviation and becomes strictly greater than R_a. Consequently, the deviation triggers a cumulative process whereby all owners enter the market until the stable configuration with full participation $(L_1(x), L_q(x)) = (\pi, 1-\pi)$ is reached.

  - Second, consider $x > x_a$. This is equivalent to $[\pi(1-q) + q]\,(y - xt - u) < R_a$. And, as we know that $[\pi(x)(1-q) + q]\,(y - xt - u) = R_a$, we clearly have $\pi(x) > \pi$. Consider that a

[47] This is because all owners that were initially not selling are equally likely to start selling. The formula is obtained by recognizing that a mass $\pi - L_1(x)$ of 1-owners and $1 - \pi - L_q(x)$ of q-owners was initially not participating in the market.

mass ϵ of owners start selling. Here again, buyers expect that $\frac{\pi - L_1(x)}{1-L_1(x)-L_q(x)}\epsilon$ of them are 1-sellers and that $\frac{1-\pi-L_q(x)}{1-L_1(x)-L_q(x)}\epsilon$ are q-sellers. The price of a land plot becomes $\Pi(x,\eta)$, as defined in the previous case. However, we now have $\frac{\partial \Pi}{\partial \eta} < 0$ because $\pi(x) > \pi$. The new price of land is therefore decreased by the deviation and becomes strictly smaller than R_a. This triggers a cumulative process whereby all owners stop participating in the market until we reach the stable equilibrium with no participation $(L_1(x), L_q(x)) = (0,0)$.

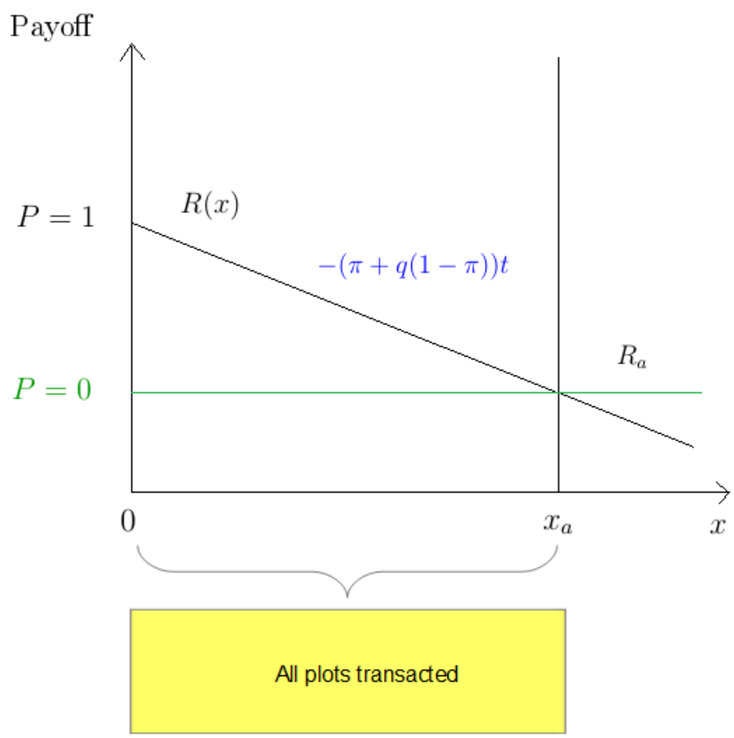
On each interval for x, we select the only stable cases. This leads to a unique possible configuration with full sale of all 1-plots and q-plots on $[0, x_a]$ and no sale on $]x_a, +\infty[$.

Compatibility with equilibrium conditions It is easy to verify that the selected stable Pareto-optimal configuration satisfies the 4 equilibrium conditions for each $x \in [0, x_a]$:

- $L_1(x) + L_q(x) = 1$ so that (3) is verified.
- $B(P = 1|x, Q \in \{q,1\}, R) = (\pi(1-q)+q)(y - tx - u) > R_a$ so that (4) is verified.
- $R(x) = (\pi(1-q)+q)(y - tx - u)$ so that (5) is verified.
- $\pi(1-q)+q)(y_u - tx_a - u) = R_a$ so that (6) is verified.

B.2. Payoffs of land owners

Figure B1: Payoffs of land owners



Note: This figure represents the equilibrium payoffs of sellers as a function of distance to the city center and their market participation decisions. The slope of the payoff curve for market participants is indicated in blue.

B.3. Suboptimality of the equilibrium

The market equilibrium involves an externality insofar as agents do not internalize the effect of their market participation decision on the composition of transacted plots, which in turn affects other agents' decisions. Following Fujita (1989), we define the surplus as the city production (sum of wages) minus the costs to organize the city (transport costs, composite good consumption, and foregone agricultural production). In the competitive equilibrium, recognizing that the composite good consumption is u, the surplus can be written as:

$$\Gamma(q, \pi, u) = \int_0^{x_a} \pi(y - xt - u - R_a) + (1-\pi)(q(y - xt - u) - R_a)dx$$

The optimal city configuration corresponds to a situation where 1-plots and q-plots are allocated to migrant workers until boundary zones denoted by $x_a^{1,opt}$ and $x_a^{q,opt}$ respectively. Indeed, observe that, if a plot is not allocated, its contribution to the surplus is zero. If a 1-plot is allocated, it contributes to the city surplus by an amount $y - xt - u - R_a$. If q-plot is allocated, it contributes to the city surplus by an amount $q(y-xt-u)-R_a$, which are decreasing functions of x. $x_a^{1,opt}$ and $x_a^{q,opt}$ are then defined as the locations where these contributions become zero. We have $x_a^{1,opt} = \frac{1}{t}(y - R_a - u) = x_a^*$ and $x_a^{q,opt} = \frac{1}{t}(y - \frac{R_a}{q} - u) = x_a^q$. The optimal city configuration is thus the same as the city equilibrium configuration in the model without information asymmetry so that:

$$\Gamma^{opt}(q,\pi,u) = \Gamma^{sym}(q,\pi,u) = \pi \int_0^{x_a^*} y - xt - u - R_a dx + (1-\pi) \int_0^{x_a^q} q(y - xt - u) - R_a dx$$

Appendix C - Adding ethnic matching to the model

Before solving for the equilibrium, we derive in Section C.1 two lemmas regarding sorting among transaction types. In Section C.2, we make use of these lemmas to characterize the spatial equilibrium (Proof of Proposition 2) and identify the stable and Pareto-dominant equilibria for all potential values of the registration fee and the social penalty. Section C.3 represents the equilibrium payoffs of landowners. Section C.4 compares the surplus in the extended and in the benchmark model.

C.1. Lemmas

Lemma C1: *In a stable equilibrium,* 1*-owners selling informally only sell to ethnic cousin buyers.*

Proof:

We reason by contradiction. Let's assume that, at a stable equilibrium, 1-sellers do not only sell to ethnic cousins, then they either (i) sell to both ethnic cousins and non-cousins or (ii) they sell only to non-cousins. To show that 1-sellers only sell to ethnic cousins, we sequentially show that (i) and (ii) cannot be true:

- If there is a stable equilibrium in which 1-sellers sell both to ethnic cousins and non-cousins, then we have $\pi^c(x) > 0$ and $\pi^{nc}(x) > 0$ and the payoff of 1-sellers is the same when selling to an ethnic cousin or a non-cousin, which means that:

$$\psi(x, u|C = c) = \psi(x, u|C = nc)$$

Given that we have, by definition:

$$\begin{cases} \psi(x, u|C = c) = \{\pi^c(x) + q(1 - \pi^c(x))\}(y - tx - u) \\ \psi(x, u|C = nc) = \{\pi^{nc}(x) + q(1 - \pi^{nc}(x))\}(y - tx - u) \end{cases}$$

the equality of payoffs implies $\pi^c(x) = \pi^{nc}(x)$. Then, the payoff of q-sellers selling to ethnic cousins is $\psi(x, u|C = c) - J$ and the payoff of q-sellers selling to non-cousins is $\psi(x, u|C = nc) = \psi(x, u|C = c)$. Thus, q-sellers all prefer to sell to non-cousins and do so. Therefore, $\pi^c(x) = 1$ (all sellers selling to ethnic cousins are 1-sellers) while $\pi^{nc}(x) < 1$ (because 1-sellers only make up a limited proportion of the sellers selling to non-cousins). This contradicts $\pi^c(x) = \pi^{nc}(x)$ and our initial assumption.

- If there is a stable equilibrium in which 1-sellers sell only to non-cousins, then we have $\pi^{nc}(x) > \pi^c(x) = 0$ and the payoff of 1-sellers is strictly larger when selling to a non-cousin than when selling to an ethnic cousin, which means that:

$$\psi(x, u|C = nc) > \psi(x, u|C = c)$$

Then, the payoff of q-sellers selling to ethnic cousins is $\psi(x, u|C = c) - J$ and the payoff of q-sellers selling to non-cousins is $\psi(x, u|C = nc)$, which is strictly larger than $\psi(x, u|C = c) - J$. Thus, q-sellers all prefer to sell to non-cousins and do so. Therefore, all sellers sell to non-cousins and $\pi^{nc}(x) = \pi$. A deviating 1-seller selling to an ethnic cousin would get a payoff of $\{\pi + q(1 - \pi)\}(y - tx - u)$, as the buyer would assume that a probability π for the seller to be a 1-seller. This payoff is exactly equal to $\psi(x, u|C = nc)$, so that the deviating seller would not come back to selling to a non-cousin and would increase $\pi^c(x)$ to 1, thereby triggering a transition away from the equilibrium where 1-sellers sell only to ethnic cousins. Thus, this equilibrium would be unstable, which contradicts our initial assumption.

Both cases lead to a contradiction. Consequently, in a stable equilibrium, 1-sellers only sell to ethnic cousins.

Lemma C2: *In a stable equilibrium, if there are transactions of plots between non-cousins, they must always involve insecure plots ($Q = q$).*

Proof: This is a direct consequence of Lemma C1. Because 1-owners never sell to non-cousins, any transaction of land between non-cousins must therefore involve insecure plots ($Q = q$).

In the equilibrium, it is thus only possible to acquire a secure plot if transacting with an ethnic cousin. Yet, transactions between ethnic cousins involve information asymmetry, as insecure plots may also be sold to ethnic cousin buyers. On the contrary, there is no information asymmetry in transactions between non-cousins who always exchange insecure plots. With Lemmas C1 and C2, we see that transactions between ethnic cousins pool risky and non-risky plots, whereas transactions between non-cousins clearly separate a subset of risky plots. With these lemmas in mind, let us now solve for the equilibrium.

C.2. Proof of Proposition 2 - Competitive equilibrium in the model with ethnic matching

We provide here a more detailed version of Proposition 2, where we now indicate the equilibrium quantities of transacted land in the different zones of the city:

Proposition C1: *In equilibrium, the city is organized in three zones. Denoting the boundary zone thresholds $\underline{x}(J) = \frac{1}{t}\left(y - \frac{R_a+J}{\pi(1-q)+q} - u\right)$ and $\bar{x}(J) = \frac{1}{t}(y - (R_a + J) - u)$, we have:*

- ***Zone 1 (informal residential zone, full market participation)**: On $]0, \underline{x}(J)]$, all owners (q- and 1-owners) sell their plot exclusively to ethnic cousins. In each location x, the price for these informal sales is $R^c(x) = (\pi(1-q)+q)(y - xt - u)$.*
- ***Zone 2 (mixed informal residential and agricultural zone, partial market participation of q-owners)**: On $]\underline{x}(J), \bar{x}(J)]$, all 1-owners and some q-owners sell their plots exclusively to ethnic cousins. The rest of q-owners drop out of the market. The mass of q-owners selling their plot in x to ethnic cousins is $L^c_q(x, J) = \frac{\pi(1-q)(y-xt-u)}{R_a+J-q(y-xt-u)} - \pi$. The price in each location x is $R^c(x) = R_a + J$.*
- ***Zone 3 (mixed informal residential and agricultural zone, all q-owners dropping out of the market)**: On $]\bar{x}(J), x^*_a]$, all 1-owners sell their plot exclusively to ethnic cousins and all*

q-owners drop out of the market. The price in each location x is $R^c(x) = y - xt - u$.

The city boundary is at $x_b = x_a^* = \frac{1}{t}[y - R_a - u]$.

Proof:

To derive the competitive equilibrium, we study, in each x, all possible combinations of participation and ethnic matching decisions that 1- and q-owners may take to satisfy (8) subject to (9) and (10). We then study their stability. In some locations, we will see that more than one combination is possible and stable. In that case, we select the Pareto-dominant combination that unambiguously benefits owners the most (we will see that 1-owners and q-owners prefer the same combinations).

We then verify that the selected configuration satisfies the equilibrium conditions (7)-(11).

Determination of sellers' possible participation and ethnic matching decisions Given Lemmas C1 and C2, it is clear that the decisions taken by all sellers in location x can be uniquely characterized by the triple of variables $(L_1^c(x), L_q^c(x), L^{nc}(x))$, where $L_1^c(x)$ (resp. $L_q^c(x)$) denotes the quantity of land plots sold by 1-owners (resp. q-owners) to an ethnic cousin buyer in location x and $L^{nc}(x)$ denotes the quantity of land plots sold by either a 1-owner or a q-owner to a non-cousin buyer. We therefore can have the following combinations:

- If $(L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, 1-\pi, 0)$, then, denoting $\underset{\sim}{x}(J) \equiv \frac{1}{t}\left(y - \frac{J}{\pi(1-q)} - u\right)$, $\underline{J} = \pi R_a \frac{1-q}{q}$ and $\bar{J} = R_a \frac{1-q}{q}$,, the payoff maximization constraint has different implications depending on the values of J:

$$(L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, 1-\pi, 0)$$
$$\Rightarrow (J > \underline{J} \text{ and } \underset{\sim}{x}(J) > x)$$
$$\text{or } (\underline{J} > J \text{ and } \underline{x}(J) > x)$$

- If $(L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, 0, 1-\pi)$, then the implications of the payoff maximization constraint also depend on J. Denoting $\widetilde{x}(J) \equiv \frac{1}{t}\left(y - \frac{J}{1-q} - u\right)$ and $x_a^q = \frac{1}{t}\left(y - \frac{R_a}{q} - u\right)$, these implications are:

$$(L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, 0, 1-\pi)$$
$$\Rightarrow (J > \bar{J} \text{ and } x_a^q > x > \widetilde{x}(J))$$

- If $(L_1^c(x), L_q^c(x), L^{nc}(x)) = (0,0,0)$, then the payoff maximization constraint implies:

$$(L_1^c(x), L_q^c(x), L^{nc}(x)) = (0,0,0)$$
$$\Rightarrow x > x_a$$

- If $(L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, \beta, 1-\pi-\beta)$ where $\beta \in]0, 1-\pi[$, then the payoff maximization constraint implies:

$$(L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, 1-\pi-\beta, \beta) \text{ where } \beta \in]0, 1-\pi[$$
$$\Rightarrow (J > \bar{J} \text{ and } \widetilde{x}(J) > x > \underset{\sim}{x}(J))$$
$$\text{or } (\bar{J} > J > \underline{J} \text{ and } x_a^q > x > \underset{\sim}{x}(J))$$

- If $(L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, 0, 0)$, then the payoff maximization constraint implies:

$$(L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, 0, 0)$$
$$\Rightarrow (J > \bar{J} \text{ and } x_a^* > x > x_a^q) \text{or } (\bar{J} > J \text{ and } x_a^* > x > \bar{x}(J))$$

- If $(L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, \beta, 0)$ where $\beta \in]0,1\text{-}\pi[$, denoting $J_{aux} = (k(\pi(1-q)+q) - R_a(1-\pi)(1-q)$, then the payoff maximization constraint implies:

$$(L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, \beta, 0) \text{ where } \beta \in]0,1\text{-}\pi[$$
$$\Rightarrow (\bar{J} > J > \underline{J} \text{ and } \bar{x}(J) > x > x_a^q)$$
$$\text{or } (\underline{J} > J \text{ and } \bar{x}(J) > x > \underline{x}(J))$$

In this case, q-sellers are indifferent between selling their plot to an ethnic cousin or keeping it under agricultural land use. This implies that the payoff of a q-seller in location x is the same for these two decisions:

$$\frac{\pi + L_q^c(x,J)q}{\pi + L_q^c(x,J)}(y - xt - u) - J = R_a$$

where $L_q^c(x,J)$ corresponds to the number of q-sellers in location x who sell their land plot to an ethnic cousin and where the first fraction corresponds to the probability that a plot transacted among cousins in location x will be kept by its buyer in the future. From this, we can derive the number of q-sellers in location x who sell their land plot to an ethnic cousin:

$$L_q^c(x, J) = \frac{\pi(1-q)(y-xt-u)}{R_a + J - q(y-xt-u)} - \pi$$

- In all other combinations of decisions where some q-owners participate in the market (proportion $L_q^c(x) + L^{nc}(x) \in]0, 1-\pi[$) and all 1-owners participate in the market ($L_1^c(x) = \pi$), the payoff maximization constraint implies that the corresponding interval for x is reduced to a singleton or the empty set.

- The payoff maximization constraint implies that the six following cases are reduced to a singleton or the empty set:
  - 1-sellers do not participate in the market but q-sellers do.
  - q-owners do not participate in the market and 1-owners participate, at least partially, with ethnic matching decisions different from the cases previously analyzed.

Stability of the equilibria Using the same approach as before, it can easily be shown that:

- All "corner" combinations (where all 1-owners in a given location adopt the same decisions, and all q-owners adopt the same decisions) are stable. As before, this is done by showing that a small enough deviation in participation and ethnic matching decisions does not change the strict ranking of owners' decisions so that they return to their initially optimal decisions.

- The case $(L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, \beta, 1-\pi-\beta)$ where $\beta \in]0, 1-\pi[$ is stable if and only if $J > 0$:
  - If a mass ϵ of owners who used to sell to ethnic cousins stop selling (or start selling to non-cousins), buyers expect these deviant owners to include $\epsilon\frac{\pi}{\beta+\pi}$ 1-owners and $\epsilon\frac{\beta}{\beta+\pi}$ q-owners. Therefore, the land price for transactions between ethnic cousins and the land price for transactions between non-cousins remain unaltered. All payoffs and the payoff maximization constraint are preserved, so that deviant owners come back to their initial decisions.
  - If a mass ϵ of owners who used to sell to non-cousins stop selling, land prices and thus payoffs remain unaltered. Therefore, deviant owners come back to their initial decisions.

– If a mass ϵ of owners who used to sell to non-cousins start selling to ethnic cousins (those can only be q-owners), the land price for transactions between ethnic cousins is reduced from $\left(\frac{\pi(1-q)}{\beta+\pi}+q\right)(y-xt-u)-J$ (which is equal to $q(y-xt-u)$ since q-owners are indifferent between selling to ethnic cousins and non-cousins) to $\left(\frac{\pi(1-q)}{\beta+\epsilon+\pi}+q\right)(y-xt-u)-J$. It is easy to see that the new price is lower than the former price, so that q-owners now strictly prefer selling to non-cousins, while 1-owners still prefer selling to ethnic cousins. Therefore, 1-owners do not change their decisions and some q-owners shift from selling to ethnic cousins to selling to non-cousins. This cumulative shift lasts until the benefit obtained by q-owners when selling to ethnic cousins becomes equal to their benefit when selling to non-cousins. At this point, we are back to the initial combination of owners' decisions. Note that if $J=0$, on the opposite, this case is unstable.

- The case $(L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, \beta, 0)$ where $\beta \in]0, 1-\pi[$ is stable if and only if $J > 0$:

  – If a mass ϵ of owners stop selling to ethnic cousins, then buyers assume that deviant owners include $\epsilon\frac{\pi}{\beta+\pi}$ 1-owners and $\epsilon\frac{\beta}{\beta+\pi}$ q-owners. Therefore, the price of land transacted between ethnic cousins is unaltered, payoffs are preserved and deviant owners come back to their initial decisions.

  – If a mass ϵ of owners starts selling to non-cousins, payoffs are unaltered and they come back to their initial decisions.

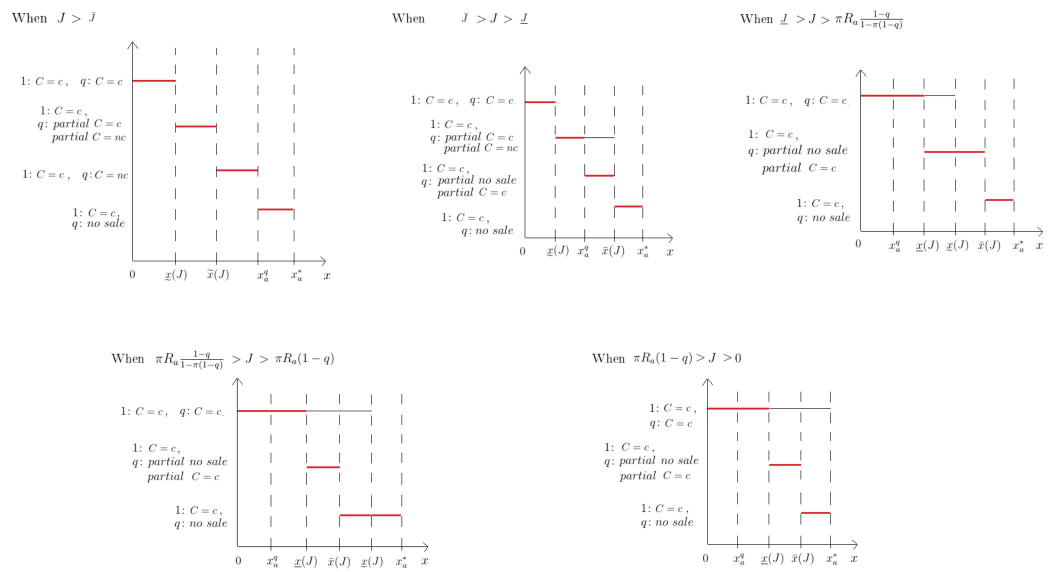
  – If a mass ϵ of owners starts selling to ethnic cousins, then they must be q-owners as all 1-owners are already selling. The price of land transacted between ethnic cousins is reduced from $\left(\frac{\pi(1-q)}{\beta+\pi}+q\right)(y-xt-u)-J$ (which is equal to R_a) to $\left(\frac{\pi(1-q)}{\beta+\epsilon+\pi}+q\right)(y-xt-u)-J$, which is strictly below the agricultural rent R_a. Therefore, although 1-owners still prefer selling to ethnic cousins, q-owners now strictly prefer keeping their land under agricultural use. Consequently, 1-owners do not change their decisions and some q-owners shift from selling to ethnic cousins to keeping their land under agricultural use. This cumulative shift lasts until the benefit obtained by q-owners when selling to ethnic cousins becomes equal to the agricultural rent. At this point, we are back to the initial combination of owners' decisions. Note that if $J=0$, on the opposite, this case is unstable.

The stable configurations are presented in Figure C1 for all possible J values. In those figures, we indicate with "1 :" and "q :" the decision of 1-owners and q-owners respectively, where "$C=c$" and "$C=nc$" refer to their decision to sell to ethnic cousins and non-cousins respectively, "*no sale*" refers to staying out of the market, and "*partial*" qualifies any of the above decisions to indicate that only a fraction of 1-owners or q-owners take that decision. For instance, on the first graph of Figure C1, "$1 : C = c, q : C = nc$" indicates that all 1-owners participate in the market and sell to ethnic cousins and all q-owners participate in the market and sell to non-cousins.

Selection of the Pareto-dominant configurations When several stable configurations are possible for given values of J, we select the one that benefits owners the most (as 1-owners and q-owners prefer the same combinations). The Pareto-dominant configuration is highlighted in red for each J value. For example, when $\pi R_a(1 - q) > J > 0$ (i.e. bottom right graph in Figure C1), there are two possible configurations on $x \in [\underline{x}(J), \bar{x}(J)]$, one in which all q-owners sell informally to cousins and one in which only a portion of them do so while the others do not sell. q-sellers prefer the second configuration, because it allows to increase informal land prices.

Compatibility with equilibrium conditions It is easy to verify that the stable and Pareto-dominant configuration satisfies the five equilibrium conditions (7)-(11).

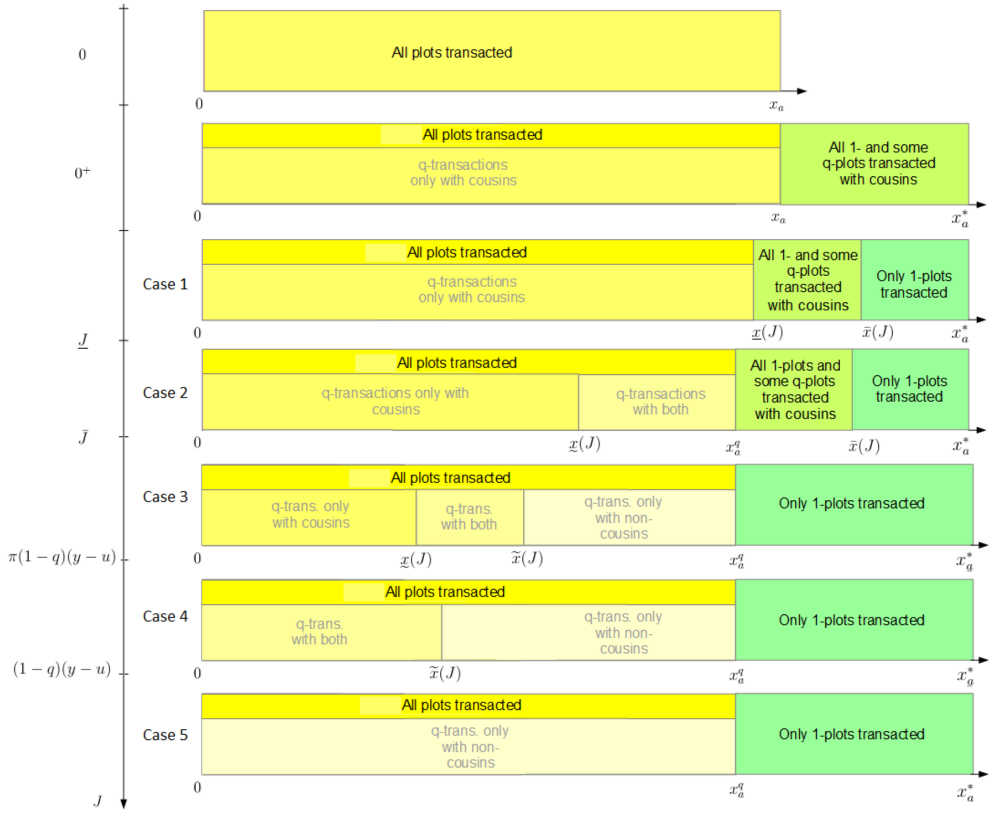
Figure C1: Stable and Pareto-dominant equilibria



Note: This figure represents the stable equilibria for varying values of the social penalty J. In case of multiple equilibria, the dominant equilibria are highlighted in red. Non participation in the market is not represented.

We now present the spatial structure of the city for all values of J on Figure C2.

Figure C2: Equilibrium city structure depending on the value of the social penalty

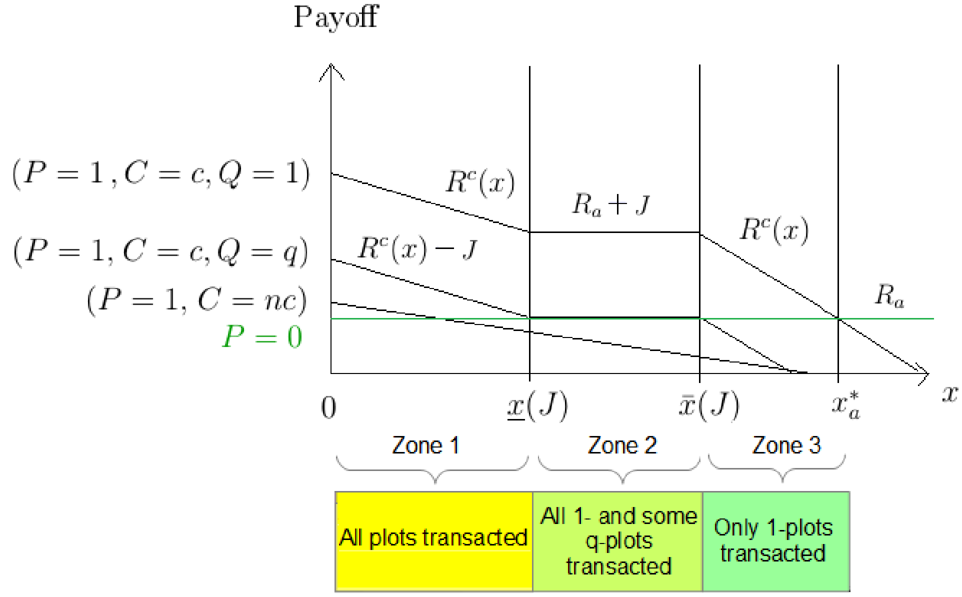


Note: This figure represents the city structure as a function of distance to the city center for varying levels of the social penalty (J). As 1-plots are always sold to cousin buyers, the cousinage link is generally indicated only for q-transactions.

C.3. Payoffs of land owners

Figure C3 shows the equilibrium payoffs of transacting land owners depending on their decisions to sell to a cousin and on the plot risk in the case where $J < \underline{J}$. Selling to a cousin always dominates selling to a non-cousin. On the zone $[\underline{x}(J), \bar{x}(J)]$, the fraction of land owners selling to a cousin is such that land owners are indifferent between selling to a cousin and not participating in the market.

Figure C3: Payoffs of land owners, depending on their participation and ethnic matching decisions



Note: This figure represents the equilibrium payoffs of landowners as a function of distance to the city center, their market participation and ethnic matching decisions when $J < \underline{J}$.

C4. Proof of Proposition 3

It is clear that the surplus increases with J for all $J > 0$, because, in each location beyond x_a^q, q-plot sales (which are, in this part of the city, surplus-reducing as we have seen when determining the optimal city structure in the benchmark model) decrease : $L_q^c(x, J) = \frac{\pi(1-q)(y-xt-u)}{R_a+J-q(y-xt-u)} - \pi$, $\underline{x}(J) = \frac{1}{t}\left(y - \frac{R_a+J}{\pi(1-q)+q} - u\right)$ and $\bar{x}(J) = \frac{1}{t}(y - (R_a + J) - u)$ are all decreasing in J.

The only ambiguity resides in the comparison between the benchmark model and the case with infinitely low cousinage (i.e. $J = 0^+$). Indeed, the introduction of very low cousinage increases both the zone over which 1-plots and the zone over which q-plots are sold. The first effect increases the overall surplus, while the second effect decreases it. The net impact of these two effects on the surplus in a given location $x \in]x_a; x_a^*[$ is: $\pi(y - xt - u - R_a) + L_q^c(x, J)(q(y - xt - u) - R_a)$, which is positive if and only if: $(y - xt - u - R_a)(R_a + J - q(y - xt - u)) + (y - xt - u - R_a - J)\,(q(y - xt - u) - R_a) > 0$, which is always true. Thus, even for the lowest value of the cousinage penalty, the introduction of cousinage increases the surplus.

Consequently, the introduction of cousinage always increases the surplus.

Appendix D - Adding registration to the model

In Section D.1, we make use of the two lemmas of Section C.1 to characterize the spatial equilibrium (Proof of Proposition 4) and identify the stable and Pareto-dominant equilibria for all potential values of the registration fee and the social penalty. Section D.2 represents the equilibrium payoffs of landowners. Section D.3 proves Proposition 5 on the substitutability between registration and ethnic matching. Sections D.4 and D.5 detail the proofs of Propositions 6 and 7 regarding surplus properties.

D.1. Proof of Proposition 4 - Competitive equilibrium in the model with registration and ethnic matching

We provide here a more detailed version of Proposition 4, where we now indicate the equilibrium quantities of transacted land in the different zones of the city:

Proposition D1: *In equilibrium, the city is organized in four zones. Denoting the boundary zone thresholds* $\mathring{x}(k,J) = \frac{1}{t}\left(y - \frac{k-J}{(1-q)(1-\pi)} - u\right)$, $\underline{x}(J) = \frac{1}{t}\left(y - \frac{R_a+J}{\pi(1-q)+q} - u\right)$ *and* $\bar{x}(J) = \frac{1}{t}(y - (R_a + J) - u)$*, we have:*

- ***Zone 1 (mixed formal and informal residential zone, full market participation):*** *On* $[0, \mathring{x}(k,J)]$*, all owners (irrespective of the initial tenure security level of their plot) participate in the market. Some q-owners register their plot before the sale (in quantity* $L_{fq}(x,J,k) = 1 - \frac{\pi(1-q)(y-xt-u)}{(1-q)(y-xt-u)+J-k}$*), although some do not and sell exclusively to their ethnic cousins. 1-owners do not register their secure plots and sell them exclusively to their ethnic cousins. The informal price in each location x is* $R^c(x) = y - xt - u - k + J$ *and the formal price for registered plots is* $R_f(x) = y - xt - u$*.*

- ***Zone 2 (informal residential zone, full market participation):*** *On* $]\mathring{x}(k,J), \underline{x}(J)]$*, all owners (q- and 1-owners) sell their plot informally and exclusively to ethnic cousins. In each location x, the price for these informal sales is* $R^c(x) = (\pi(1-q)+q)(y - xt - u)$*.*

- ***Zone 3 (mixed informal residential and agricultural zone, partial market participation of q-owners):*** *On* $]\underline{x}(J), \bar{x}(J)]$*, all 1-owners and some q-owners sell their plots exclusively to ethnic cousins. The rest of q-owners drop out of the market. The mass of q-owners selling their plot in x to ethnic cousins is* $L^c_q(x,J) = \frac{\pi(1-q)(y-xt-u)}{R_a+J-q(y-xt-u)} - \pi$*. The price in each location x is* $R^c(x) = R_a + J$*.*

- ***Zone 4 (mixed informal residential and agricultural zone, all q-owners dropping out of the market):*** *On $]\bar{x}(J), x_a^*]$, all 1-owners sell their plot exclusively to ethnic cousins and all q-owners drop out of the market. The price in each location x is $R^c(x) = y - xt - u$.*

The city boundary is at $x_b = x_a^* = \frac{1}{t}[y - R_a - u]$.

Proof:

To derive the competitive equilibrium, we study, in each x, all possible combinations of participation, ethnic matching, and registration decisions that 1- and q-owners may take to satisfy (13) subject to (14), (15) and (16), relying on the two Lemmas proved in Section C.1. We then study their stability. In some locations, we will see that more than one combination is possible and stable. In that case, we select the Pareto-dominant combination that unambiguously benefits owners the most (we will see that 1-owners and q-owners prefer the same combinations).[48] We then verify that the selected configuration satisfies the equilibrium conditions (12)-(17).

Determination of sellers' possible participation and ethnic matching decisions

Given Lemmas C1 and C2, it is clear that the decisions taken by all sellers in location x can be uniquely characterized by the 5-uple of variables $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x))$, where $L_{f1}(x)$ (resp. $L_{fq}(x)$) denotes the quantity of formalized land plots sold by 1-owners (resp. q-owners) in location x, $L_1^c(x)$ (resp. $L_q^c(x)$) denotes the quantity of informal land plots sold by 1-owners (resp. q-owners) to an ethnic cousin buyer in location x and $L^{nc}(x)$ denotes the quantity of informal land plots sold by either a 1-owner or a q-owner to a non-cousin buyer. We therefore can have the following combinations:

- If $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, 1-\pi, 0)$, then, denoting $\underset{\sim}{x}(J) \equiv \frac{1}{t}\left(y - \frac{J}{\pi(1-q)} - u\right)$, $\underline{J} = \pi R_a \frac{1-q}{q}$ and $\bar{J} = R_a \frac{1-q}{q}$,, the payoff maximization constraint has different implications depending on the values of k and J:

  - If $k > \bar{J}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, 1-\pi, 0)$$
$$\Rightarrow (\pi k > J > \underline{J} \text{ and } \underset{\sim}{x}(J) > x > \mathring{x}(k, J))$$
$$\text{or } (\underline{J} > J \text{ and } \underline{x}(J) > x > \mathring{x}(k, J))$$

[48] We assume, without changing the model's main results, that $\bar{k} > R_a \frac{1-q}{\pi q}$,which allows to reduce the number of possible cases to be studied.

– If $\bar{J} > k > \underline{k}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, 1-\pi, 0)$$
$$\Rightarrow (k(\pi(1-q)+q) - R_a(1-q)(1-\pi) > J \text{ and } \underline{x}(J) > x > \mathring{x}(k, J))$$

– If $\underline{k} > k$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, 1-\pi, 0) \text{ is impossible.}$$

- If $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, 0, 1-\pi)$, then the implications of the payoff maximization constraint also depend on k and J. Denoting $\ddot{x}(k) \equiv \frac{1}{t}\left(y - \frac{k}{1-q} - u\right)$, $\widetilde{x}(J) \equiv \frac{1}{t}\left(y - \frac{J}{1-q} - u\right)$ and $x_a^q = \frac{1}{t}\left(y - \frac{R_a}{q} - u\right)$, these implications are:

  – If $k > \bar{J}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, 0, 1-\pi)$$
$$\Rightarrow (J > k \text{ and } x_a^q > x > \ddot{x}(k)) \text{ or } (k > J > \bar{J} \text{ and } x_a^q > x > \widetilde{x}(J))$$

  – If $k < \bar{J}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, 0, 1-\pi) \text{ is impossible.}$$

- If $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, 0, 0, 0, 1-\pi)$, then the payoff maximization constraint implies that this case is impossible.

- If $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, 1-\pi, 0, 0, 0)$, then, denoting $\check{x}(k) \equiv \frac{1}{t}\Big(y - \frac{k}{(1-q)(1-\pi)} - u\Big)$ and $\hat{x}(k) = \frac{1}{t}\left[y - R_a - k - u\right]$, the payoff maximization constraint implies:

  – If $k > \underline{k}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, 1-\pi, 0, 0, 0) \Rightarrow \check{x}(k) > x$$

  – If $k < \underline{k}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, 1-\pi, 0, 0, 0) \Rightarrow \hat{x}(k) > x.$$

- If $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 1-\pi, \pi, 0, 0)$, the payoff maximization constraint implies:

– If $k > \bar{J}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) \\ = (0, 1-\pi, \pi, 0, 0) \Rightarrow (J > k \text{ and } \ddot{x}(k) > x)$$

– If $k < \bar{J}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) \\ = (0, 1-\pi, \pi, 0, 0) \Rightarrow (J > k \text{ and } \hat{x}(k) > x).$$

- If $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, 0, 0, 0)$, then the payoff maximization constraint implies:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, 0, 0, 0) \\ \Rightarrow (k > \underline{k} \text{ and } x > x_a) \text{ or } (k < \underline{k} \text{ and } x > \hat{x}(k)).$$

- If $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, \beta, \pi, 1-\pi-\beta, 0)$ where $\beta \in]0, 1-\pi[$, then the payoff maximization constraint implies:

– If $k > \bar{J}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, \beta, \pi, 1-\pi-\beta, 0) \quad \text{where } \beta \in]0, 1-\pi[\\ \Rightarrow (\pi k > J \text{ and } \mathring{x}(k, J) > x) \text{ or } (k > J > \pi k \text{ and } \ddot{x}(k) > x)$$

– If $\underline{k} < k < \bar{J}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, \beta, \pi, 1-\pi-\beta, 0) \text{ where } \beta \in]0, 1-\pi[\\ \Rightarrow ((\pi(1-q)+q)k - R_a(1-\pi)(1-q) > J \text{ and } \mathring{x}(k, J) > x) \\ \text{or } (k > J > (\pi(1-q)+q)k - R_a(1-\pi)(1-q) \text{ and } \hat{x}(k) > x)$$

– If $k < \underline{k}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, \beta, \pi, 1-\pi-\beta, 0) \text{ where } \beta \in]0, 1-\pi[\\ \Rightarrow (J < k \text{ and } \hat{x}(k) > x).$$

In these three cases, q-sellers are indifferent between selling their plot after registration and selling their plot informally to an ethnic cousin. This implies that the payoff of a q-seller in location x is the same whether he takes one decision or the other:

$$\frac{\pi + (1 - \pi - L_{fq}(x, J, k))q}{1 - L_{fq}(x, J, k)}(y - xt - u) - J = y - xt - u - k$$

where $L_{fq}(x, J, k)$ corresponds to the number of q-sellers in location x who sell their land plot after registration and where the first fraction corresponds to the probability that an informally transacted plot in location x will be kept by its buyer in the future. From this, we can derive the number of q-sellers in location x who sell their land plot after registration:

$$L_{fq}(x, J, k) = 1 - \frac{\pi(1 - q)(y - xt - u)}{(1 - q)(y - xt - u) + J - k}$$

- If $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, \beta, \pi, 0, 1 - \pi - \beta)$ where $\beta \in]0, 1 - \pi[$, then the payoff maximization constraint implies that it is only possible on the singleton $x = \ddot{x}(k)$. As this case has measure zero, we disregard it.

- If $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, \beta, 1 - \pi - \beta)$ where $\beta \in]0, 1 - \pi[$, then the payoff maximization constraint implies:

  – If $\pi k > \bar{J}$:

$$\begin{aligned}(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) &= (0, 0, \pi, 1 - \pi - \beta, \beta) \text{ where } \beta \in]0, 1 - \pi[\\ &\Rightarrow (k > J > \pi k \text{ and } \widetilde{x}(J) > x > \ddot{x}(k)) \\ &\quad \text{or } (\pi k > J > \bar{J} \text{ and } \widetilde{x}(J) > x > \underset{\sim}{x}(J)) \\ &\quad \text{or } (\bar{J} > J > \underline{J} \text{ and } x_a^q > x > \underset{\sim}{x}(J))\end{aligned}$$

  – If $k > \bar{J} > \pi k$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, 1-\pi-\beta, \beta) \text{ where } \beta \in]0, 1-\pi[$$
$$\Rightarrow (k > J > \bar{J} \text{ and } \widetilde{x}(J) > x > \ddot{x}(k))$$
$$\text{or } (\bar{J} > J > \pi k \text{ and } x_a^q > x > \ddot{x}(k))$$
$$\text{or } (\pi k > J > \underline{J} \text{ and } x_a^q > x > \underset{\sim}{x}(J))$$

– If $\bar{J} > k$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, 1-\pi-\beta, \beta)$$
$$\text{where } \beta \in]0, 1-\pi[\text{ is impossible.}$$

- If $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (\alpha, 0, \pi-\alpha, 1-\pi-\beta, \beta)$ where $\alpha \in]0, \pi[$ and $\beta \in]0, 1-\pi[$, then the payoff maximization constraint implies that this case is possible on at most one singleton. We disregard this case because it is of measure zero.
- If $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, 0, 0)$, then the payoff maximization constraint implies:

  – If $k > \bar{J}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, 0, 0)$$
$$\Rightarrow (J > \bar{J} \text{ and } x_a^* > x > x_a^q) \text{or } (\bar{J} > J \text{ and } x_a^* > x > \bar{x}(J))$$

  – If $k < \bar{J}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, 0, 0)$$
$$\Rightarrow (J > k \text{ and } x_a^* > x > \hat{x}(k)) \text{ or } (k > J \text{ and } x_a^* > x > \bar{x}(J)).$$

- If $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, \beta, 0)$ where $\beta \in]0,1\text{-}\pi[$, denoting $J_{aux} = (k(\pi(1-q)+q) - R_a(1-\pi)(1-q)$, then the payoff maximization constraint implies:

  – If $k > \bar{J}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, \beta, 0) \text{ where } \beta \in]0,1\text{-}\pi[$$
$$\Rightarrow (\bar{J} > J > \underline{J} \text{ and } \bar{x}(J) > x > x_a^q)$$
$$\text{or } (\underline{J} > J \text{ and } \bar{x}(J) > x > \underline{x}(J))$$

– If $\underline{k} < k < \bar{J}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, \beta, 0) \text{ where } \beta \in]0,1\text{-}\pi[$$
$$\Rightarrow (k > J > J_{aux} \text{ and } \bar{x}(J) > x > \hat{x}(k))$$
$$\text{or } J_{aux} > J \text{ and } \bar{x}(J) > x > \underline{x}(J))$$

– If $k < \underline{k}$:

$$(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, \beta, 0) \text{ where } \beta \in]0,1\text{-}\pi[$$
$$\Rightarrow (k > J > 0 \text{ and } \bar{x}(J) > x > \hat{x}(k)).$$

In these three cases, q-sellers are indifferent between selling their plot informally to an ethnic cousin or keeping it under agricultural land use. This implies that the payoff of a q-seller in location x is the same for these two decisions:

$$\frac{\pi + L_q^c(x, J)q}{\pi + L_q^c(x, J)}(y - xt - u) - J = R_a$$

where $L_q^c(x, J)$ corresponds to the number of q-sellers in location x who sell their land plot informally to an ethnic cousin and where the first fraction corresponds to the probability that a plot transacted informally among cousins in location x will be kept by its buyer in the future. From this, we can derive the number of q-sellers in location x who sell their land plot informally to an ethnic cousin:

$$L_q^c(x, J) = \frac{\pi(1 - q)(y - xt - u)}{R_a + J - q(y - xt - u)} - \pi$$

– In all other combinations of decisions where some q-owners participate in the market (proportion $L_q^c(x) + L^{nc}(x) + L_{fq}(x) \in]0, 1 - \pi[$) and all 1-owners participate in the informal market ($L_1^c(x) + L_{f1}(x) = \pi$), the payoff maximization constraint implies that the corresponding interval for x is reduced to a singleton or the empty set.

- The payoff maximization constraint implies that the six following cases are reduced to a singleton or the empty set:

- $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (\alpha, 0, \pi - \alpha, 1 - \pi, 0)$ where $\alpha \in]0, \pi[$
- $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (\alpha, 0, \pi - \alpha, 0, 1 - \pi)$ where $\alpha \in]0, \pi[$
- $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (\alpha, \beta, \pi - \alpha, 1 - \pi - \beta, 0)$ where $\alpha \in]0, \pi[$ and $\beta \in]0, 1 - \pi[$
- $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (\alpha, \beta, \pi - \alpha, 0, 1 - \pi - \beta)$ where $\alpha \in]0, \pi[$ and $\beta \in]0, 1 - \pi[$
- $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (\pi, \beta, 0, 0, 1 - \pi - \beta)$ where $\beta \in]0, 1 - \pi[$
- 1-sellers do not participate in the market but q-sellers do.
- q-owners do not participate in the market and 1-owners participate, at least partially, with ethnic matching decisions different from the cases previously analyzed.

Stability of the equilibria Using the same approach as before, it can easily be shown that:

- All "corner" combinations (where all 1-owners in a given location adopt the same decisions, and all q-owners adopt the same decisions) are stable. As before, this is done by showing that a small enough deviation in participation and ethnic matching decisions does not change the strict ranking of owners' decisions so that they return to their initially optimal decisions.
- The case $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, \beta, \pi, 1 - \pi - \beta, 0)$ where $\beta \in]0, 1 - \pi[$ is stable, except if $J = 0$:
  - If a mass ϵ of owners stop selling to ethnic cousins (i.e. they start formalizing, or they start selling to non-cousins, or they stop selling altogether), then buyers assume that deviant owners include $\epsilon\frac{\pi}{1-\beta}$ 1-owners and $\epsilon\frac{1-\pi-\beta}{1-\beta}$ q-owners. Therefore, the price of land transacted between ethnic cousins is unaltered, payoffs are preserved and deviant owners come back to their initial decisions.
  - If a mass ϵ of owners start selling to ethnic cousins (i.e. if a mass ϵ of q-owners stops formalizing), then buyers assume that deviant owners include only q-owners, as 1-owners were already all selling to ethnic cousins. Therefore, the price of land transacted between ethnic cousins is reduced. Then, there are two cases:
    * Either $J > 0$, in which case it becomes strictly preferable for q-owners to sell after registration than to sell informally to ethnic cousins (while it remains preferable for 1-owners to sell informally to ethnic cousins, because they do not face the social penalty).

Therefore, the mass of q-owners selling informally to ethnic cousins decreases in favor of the mass of q-owners selling after registration until q-owners become indifferent between the two options. We are back to the initial configuration.

  * Either $J = 0$, in which case it becomes strictly preferable for both q-owners and 1-owners to sell after registration than to sell informally to ethnic cousins. Therefore, both q-owners and 1-owners gradually shift to selling after registration until full registration is reached. This case is thus unstable.

- Other deviations to owners' behaviors (e.g. when a mass ϵ of owners stop formalizing and sell to non-cousins) do not affect land prices (in our example, it is q-owners who start to sell to non-cousins, which does not affect the price of informal land traded between non-cousins). Because land prices are not affected, these deviations do not affect the ranking of payoffs and the deviations are reversed back to the initial configuration.

- The case $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, \beta, 1 - \pi - \beta)$ where $\beta \in]0, 1 - \pi[$ is stable if and only if $J > 0$ for exactly the same reason than in the Proof of Proposition 2.

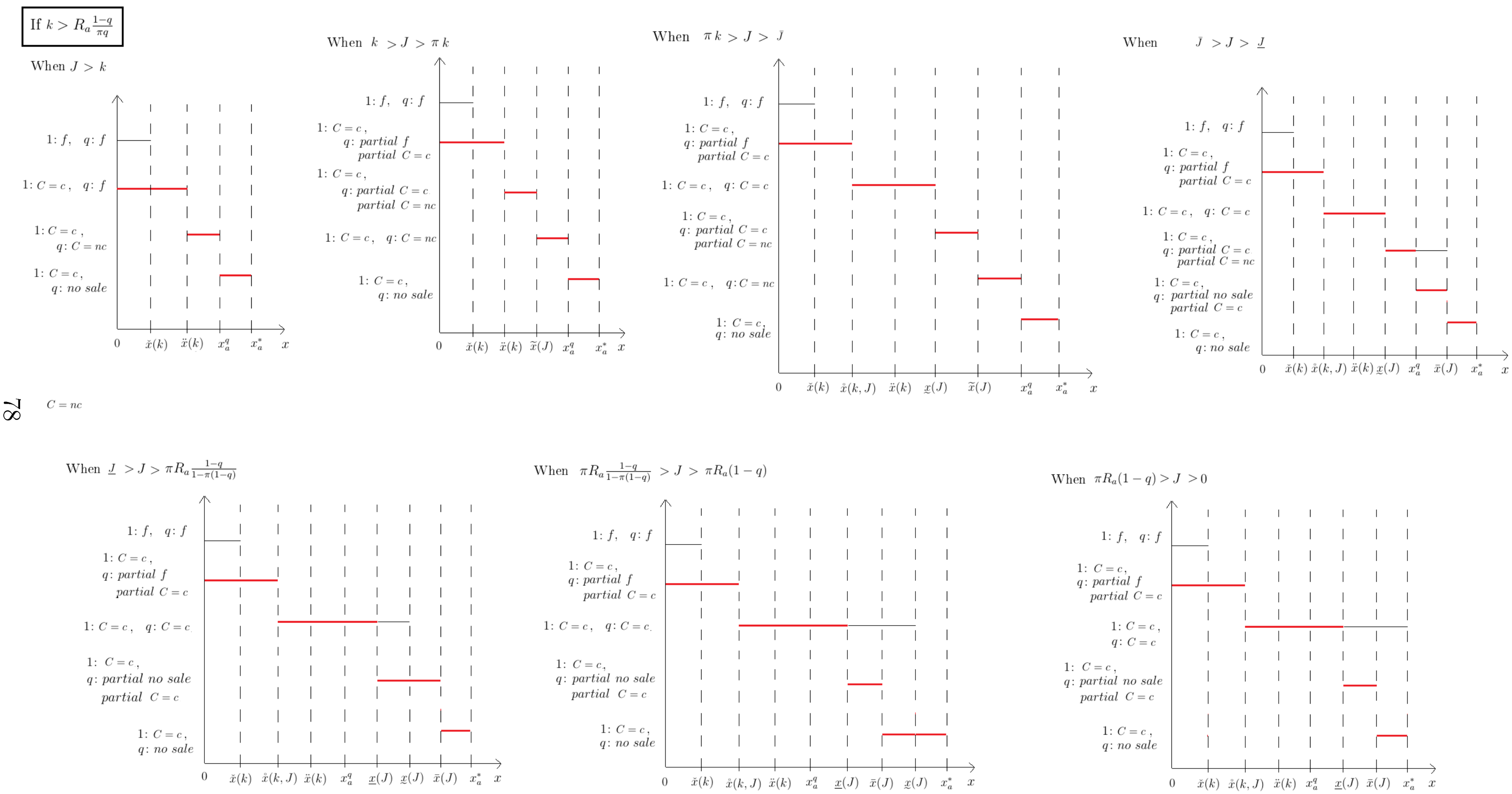
- The case $(L_{f1}(x), L_{fq}(x), L_1^c(x), L_q^c(x), L^{nc}(x)) = (0, 0, \pi, \beta, 0)$ where $\beta \in]0, 1 - \pi[$ is stable if and only if $J > 0$ for exactly the same reason as in the Proof of Proposition 2.

The stable configurations are presented in Figures D1, D2 and D3 for all possible combinations of k and J values. In those figures, we use the same notations as in Appendix C figures with, in addition, "f" referring to registering and participating in the market.

Selection of the Pareto-dominant configurations When several stable configurations are possible for given values of J and k, we select the one that benefits owners the most (as 1-owners and q-owners prefer the same combinations). The Pareto-dominant configuration is highlighted in red for each k and J value.

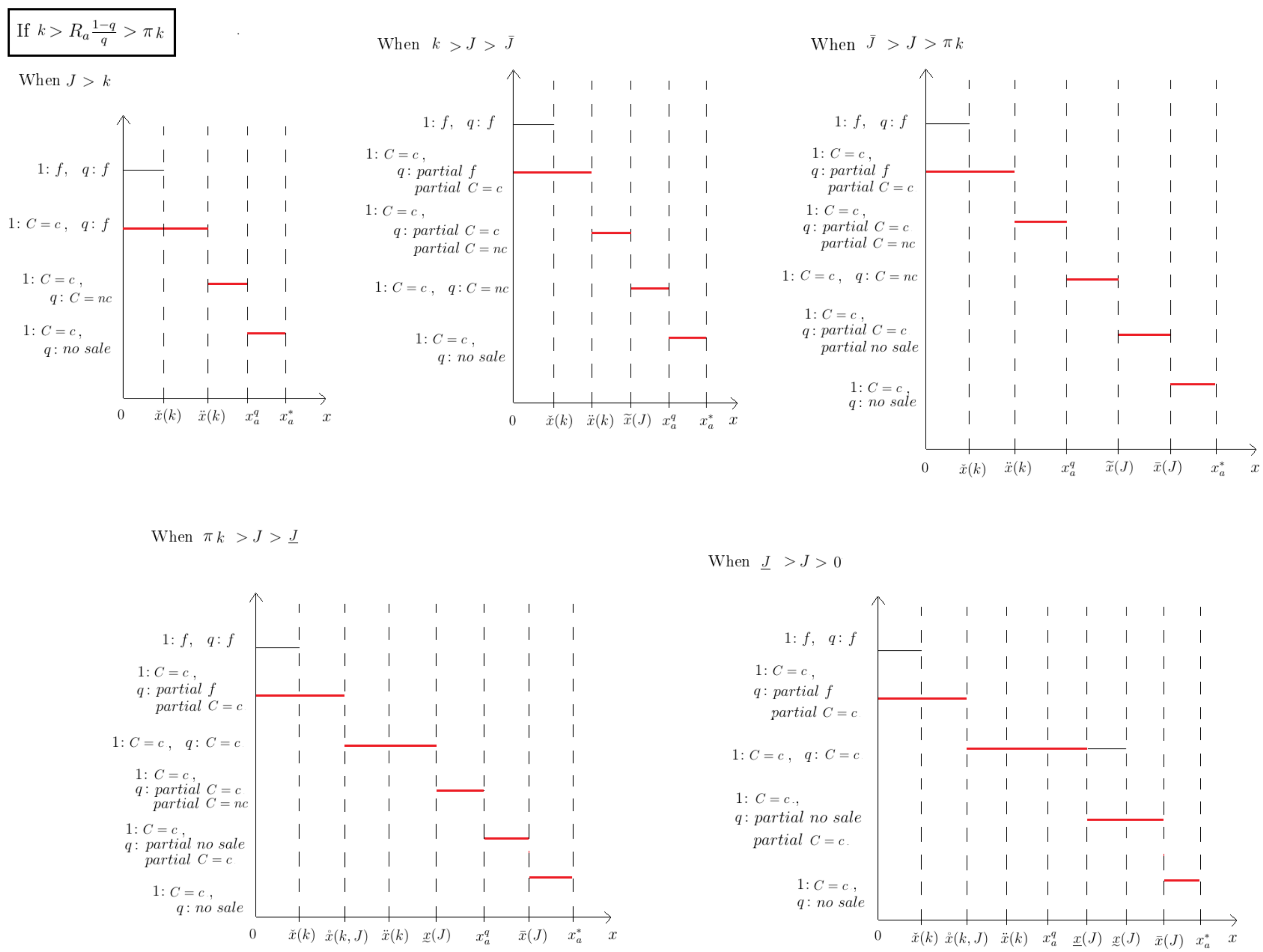
Compatibility with equilibrium conditions It is easy to verify that the stable and Pareto-dominant configuration satisfies the six equilibrium conditions (12)-(17).

Figure D1: Stable and Pareto-dominant equilibria (high registration cost)



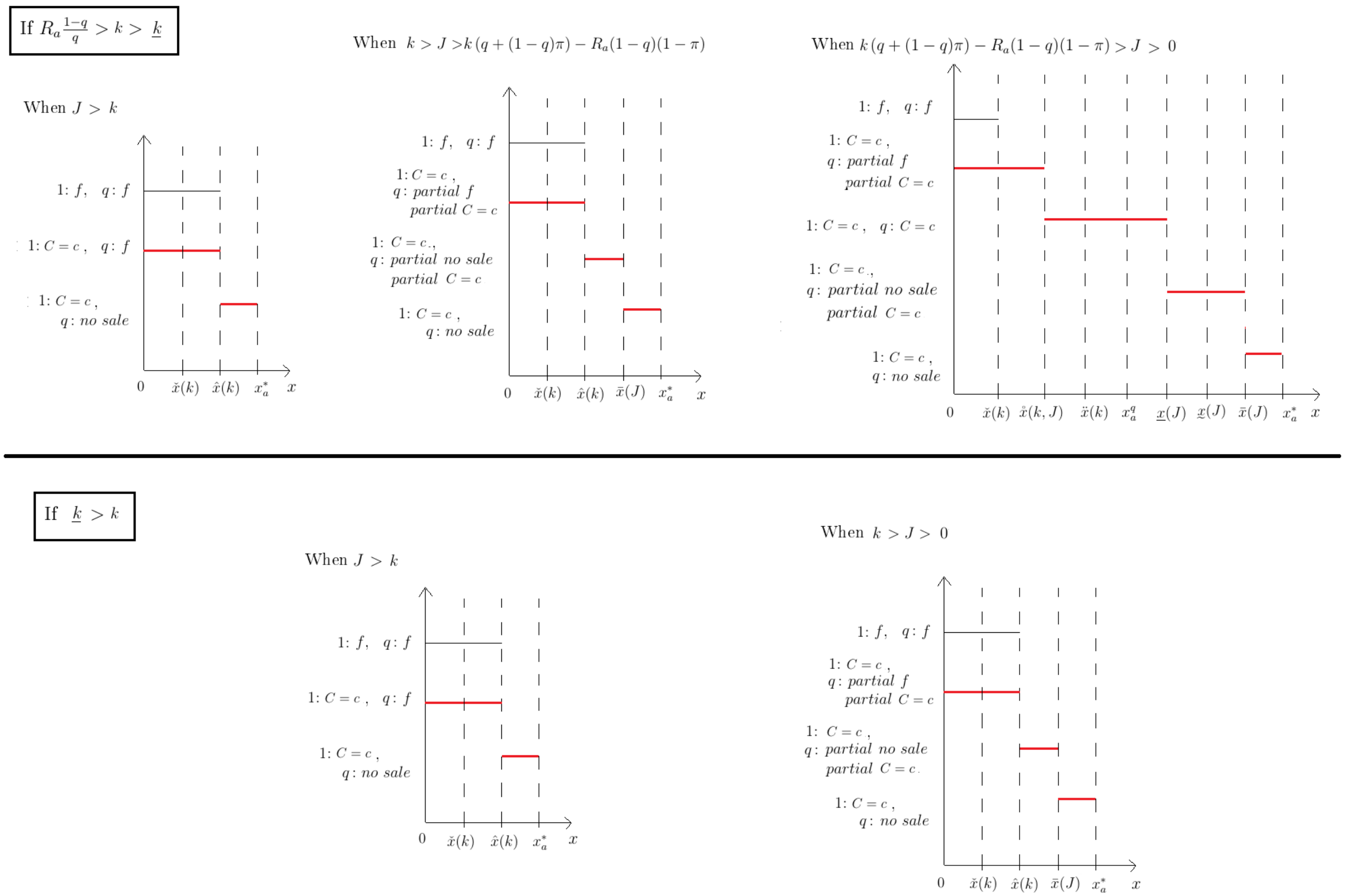
Note: This figure represents the stable equilibria for varying values of the registration cost k and the social penalty J. In case of multiple equilibria, the dominant equilibria are highlighted in red. Non participation in the market is not represented.

Figure D2: Stable and Pareto-dominant equilibria (intermediate registration cost)



Note: This figure represents the stable equilibria for varying values of the registration cost k and the social penalty J. In case of multiple equilibria, the dominant equilibria are highlighted in red. Non-participation in the market is not represented.

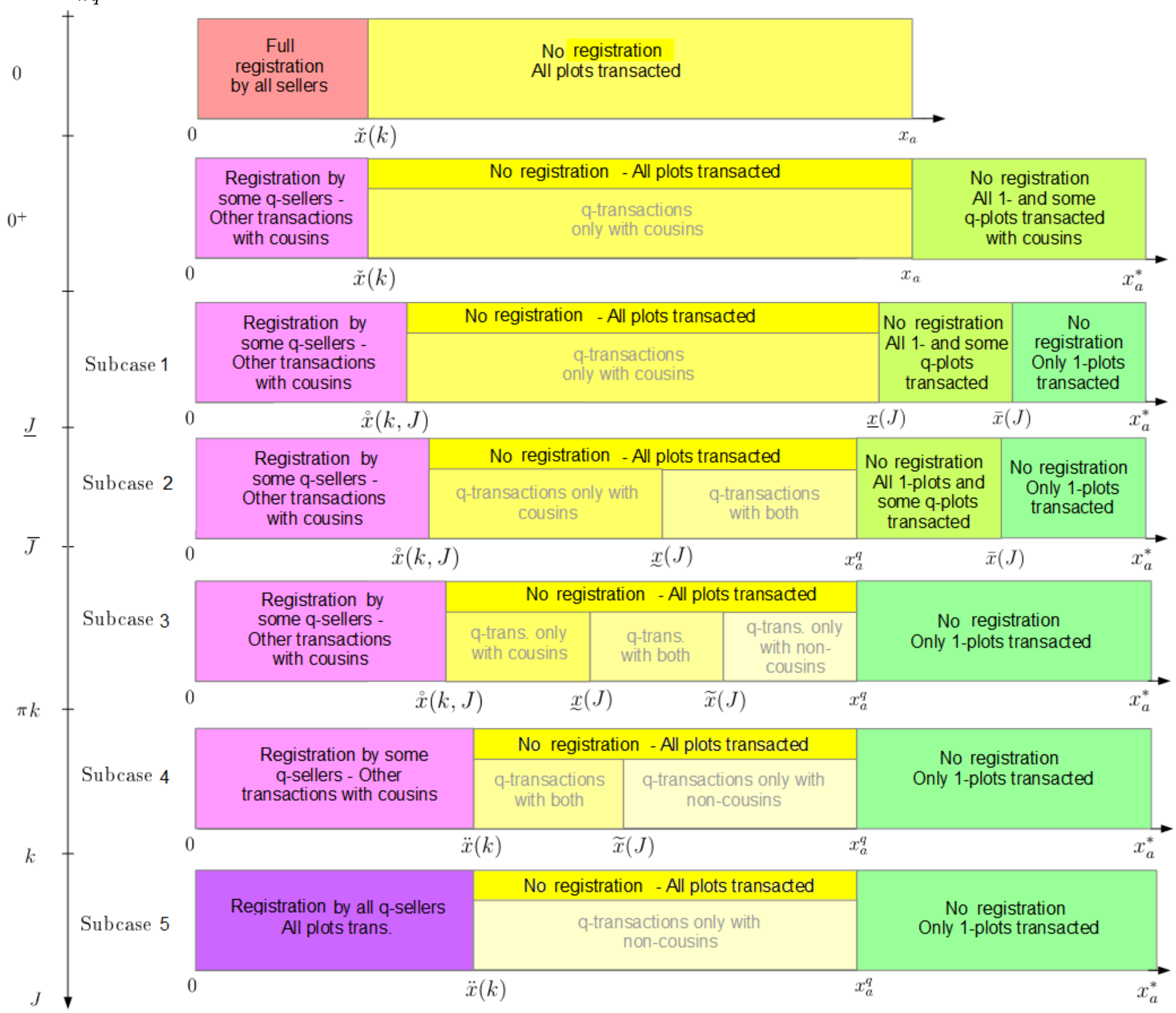
Figure D3: Stable and Pareto-dominant equilibria (low and very low registration cost)



Note: This figure represents the stable equilibria for varying values of the registration cost k and the social penalty J. In case of multiple equilibria, the dominant equilibria are highlighted in red. Non-participation in the market is not represented.

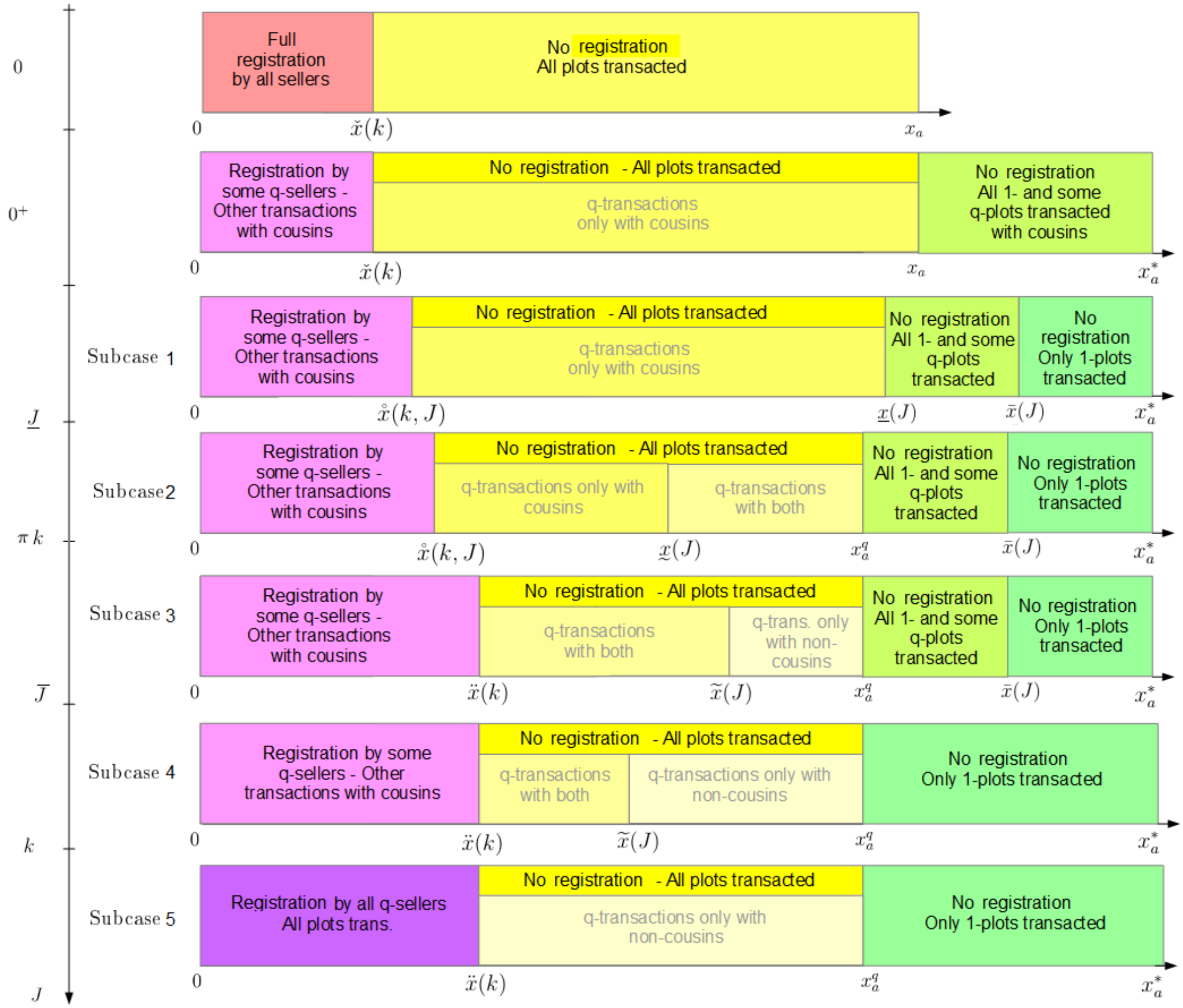
We now present the spatial structure of the city for all values of $k < \bar{k}$ and J on Figures D4., D4, D5, D6 and D7. When k further increases above $\bar{k}$, the zones in Figure D4 remain unchanged, except that the registration zone shrinks: It first disappears for low values of J and then for larger values as k increases (demonstration available upon request). When k reaches $\ddot{k} \equiv (y-u)(1-q)$, registration is totally abandoned, whatever the value of J and we are back to the case with only ethnic matching depicted in Appendix Section C.

Figure D4: Equilibrium city structure depending on the value of the social penalty (high registration cost: $\bar{k} > k > R_a \frac{1-q}{\pi q}$)



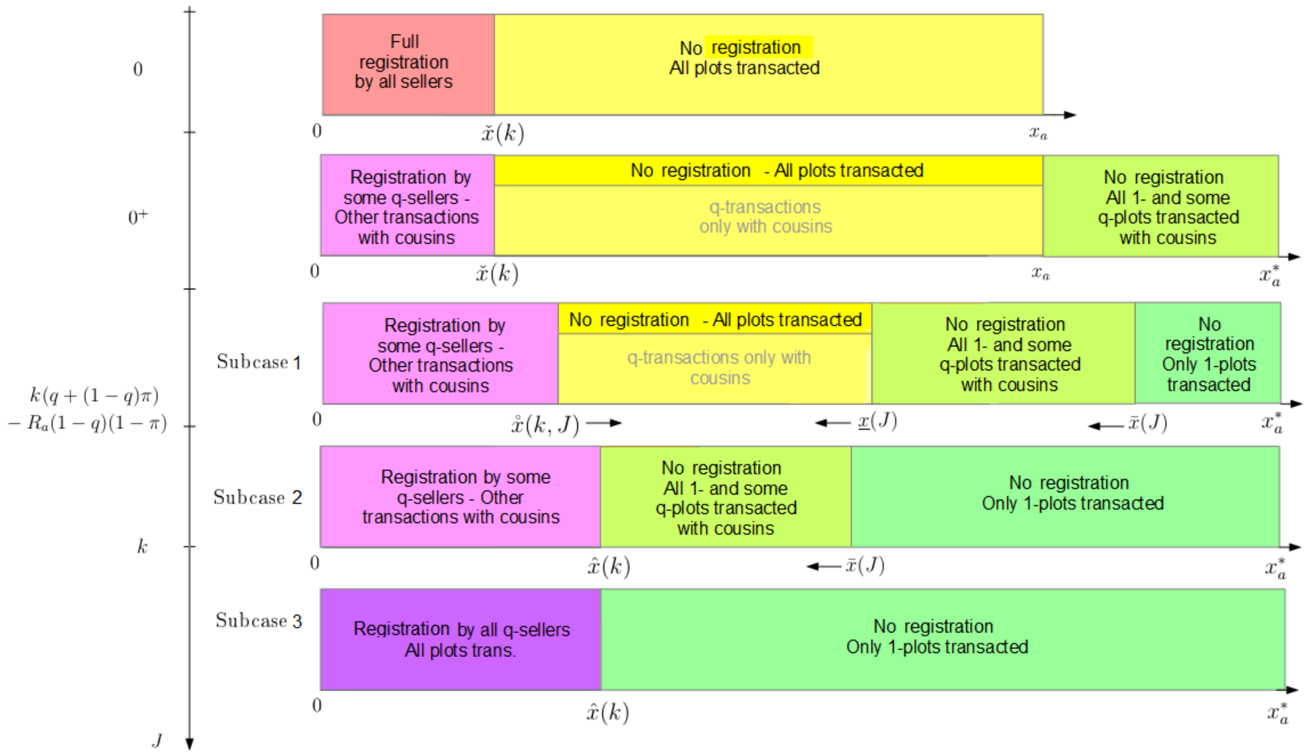
Note: This figure represents the city structure as a function of distance to the city center for high registration costs (k) and for varying levels of the social penalty (J). As 1-plots are always sold to cousin buyers, the cousinage link is generally indicated only for q-transactions.

Figure D5: Equilibrium city structure depending on the value of the social penalty (intermediate registration cost: $R_a\frac{1-q}{\pi q} > k > R_a\frac{1-q}{q}$)



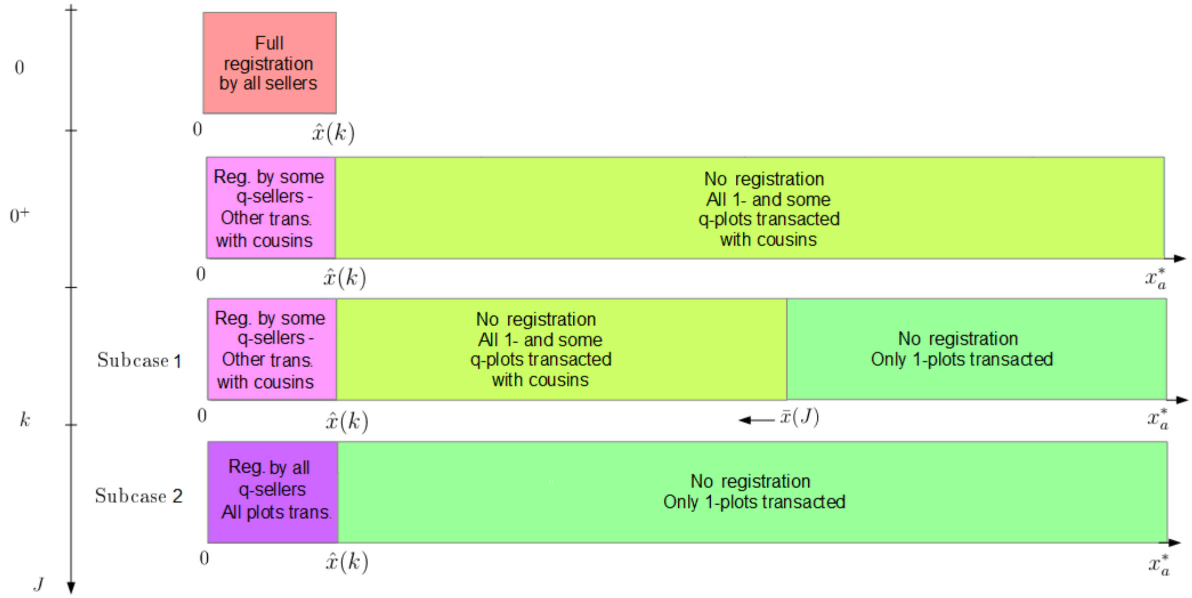
Note: This figure represents the city structure as a function of distance to the city center for intermediate registration costs (k) and for varying levels of the social penalty (J). As 1-plots are always sold to cousin buyers, the cousinage link is generally indicated only for q-transactions.

Figure D6: Equilibrium city structure depending on the value of the social penalty (low registration cost: $R_a \frac{1-q}{q} > k > \underline{k}$)



Note: This figure represents the city structure as a function of distance to the city center for low registration costs (k) and for different levels of the social penalty (J). As 1-plots are always sold to cousin buyers, the cousinage link is generally indicated only for q-transactions.

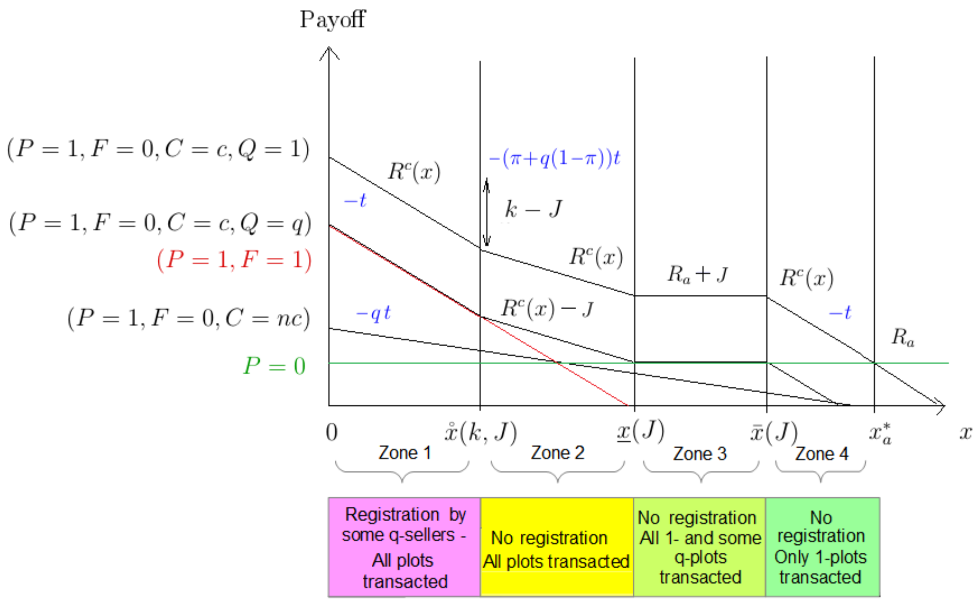
Figure D7: Equilibrium city structure depending on the value of the social penalty (very low registration cost:$\underline{k} > k$)



Note: This figure represents the city structure as a function of distance to the city center for very low registration costs (k) and for different levels of the social penalty (J). As 1-plots are always sold to cousin buyers, the cousinage link is generally indicated only for q-transactions.

D.2. Payoffs of land owners

Figure D8: Payoffs of land owners, depending on their participation, registration and ethnic matching decisions



Note: This figure represents the equilibrium payoffs of landowners as a function of distance to the city center, their market participation, registration and ethnic matching decisions when $J < \underline{J}$ and $k > \underline{k}$. The slopes of the payoff curves are indicated in blue.

D.3. Proof of Proposition 5

We start by focusing on the case where $0 < J \leq \underline{J}$ and $R_a \frac{1-q}{q} < k < \bar{k}$ or $0 < J \leq k(q + (1 - q)\pi) - R_a(1 - q)(1 - \pi)$ and $\underline{k} < k < R_a \frac{1-q}{q}$ (i.e. when the city corresponds to that described in Proposition 4). Let us first look at an increase in k. Inspection of $\mathring{x}(k, J)$ and $L_{fq}(x)$ shows that they are decreasing functions of k and increasing functions of J. It follows that an increase in k reduces both the zone over which plots are registered (Zone 1) and the proportion of landowners registering their plot in each location of this zone, resulting in an unambiguous reduction in the overall number of registered plots. Landowners who do not register their plots anymore all resort to ethnic matching. The resulting effect is an unambiguous increase in the overall number of transactions under trusted ethnic relationships.

Let us now focus on an increase in J. It shifts $\mathring{x}(k, J)$ to the right, $\underline{x}(J)$ and $\bar{x}(J)$ to the left, it reduces $L^c_q(x, J)$ and increases $L_{fq}(x, k, J)$. Thus, it is easy to see that fewer landowners resort to trusted ethnic relationships and more landowners decide to register their plot.

Using the same kind of reasoning, the substitution between k and J can be easily shown for the other values of k and J.

D.4. Proof of Proposition 6

First, let us derive the optimal allocation of plots within the city. It is easy to see that there are two cases. If $k < R_a \frac{1-q}{q}$, then the optimal city allocation corresponds to a situation where, between 0 and $\ddot{x}(k) = \frac{1}{t}\left(y - \frac{k}{1-q} - u\right)$, all plots are allocated to a migrant with all q-plots being registered (and all 1-plot not registered), then, between $\ddot{x}(k)$ and x_a^q, all plots are allocated to a migrant and remain informal and, between x_a^q and x_a^*, all 1-plots are allocated to a migrant while all q-plots remain agricultural. If, on the contrary, $\bar{k} > k > R_a \frac{1-q}{q}$, then the optimal city allocation corresponds to a situation where, between 0 and $\hat{x}(k) = \frac{1}{t}[y - R_a - k - u]$, all plots are allocated to a migrant with all q-plots being registered (and all 1-plot not registered), and, between $\hat{x}(k)$ and x_a^*, all 1-plots are allocated to a migrant while all q-plots remain agricultural. Indeed, the surplus contributions of secure and insecure land plots depending on their allocation (to agricultural use or urban use) and their registration status are as follows: If a plot is not allocated to a migrant, its net contribution to the surplus is zero. If a 1-plot is allocated to a migrant, it contributes to the city surplus by an amount $y - xt - u - R_a$. If a q-plot is allocated to a migrant informally, it contributes to the city surplus by an amount $q(y - xt - u) - R_a$. If an allocated plot is registered, it contributes to the surplus by an amount $y - xt - u - k - R_a$. Comparing these surpluses, it is easy to see that the previously described allocation is indeed surplus maximizing.

Now, we can compare the boundary of the registration zone found in the various city configurations depicted in Figures D4 to D7 with the optimal allocation. It is clear that, for all $J > 0$, the boundary of the registration zone (i.e. $\mathring{x}(k, J)$, $\ddot{x}(k)$ or $\hat{x}(k)$ depending on the case) is lower than or equal to the optimal boundary of the registration zone. Additionally, only q-plot owners register their plot within the registration zone when $J > 0$. Thus, wherever registration takes place, it is surplus-increasing. However, in most cases, there is not enough registration and the registration zone is too small. The only cases where optimal registration decisions by all sellers are obtained are when $J > k$ (as can be seen in Figures D4 to D7). We can also see that, for all $J \in]0, k[$, either the boundary of the registration zone (i.e. $\mathring{x}(k, J)$, $\ddot{x}(k)$ or $\hat{x}(k)$) or the fraction of q-owners formalizing their plot ($L_{fq}(x, j, k)$) or both are increasing in J. Consequently, for all $J \in]0, k[$, surplus is strictly increasing in J.

Looking now more specifically at the case where there is information asymmetry and registration but no ethnic matching (i.e. the first city structure corresponding to the case $J = 0$ in each Figure D4 to D7), it appears that the absence of ethnic matching leads all plot owners to register within the registration zone. It is not a priori evident to determine whether surplus is larger in the absence of ethnic matching ($J = 0$) or in the presence of a very small ethnic matching penalty ($J \to 0^+$) because, in the absence of ethnic matching, while the fact that all q-plot owners formalize within the registration zone tends to increase the surplus, the fact that all 1-plot owners formalize too reduces the surplus. To find out whether the surplus is increased or reduced in the overall, we can see that the combination of these two changes in a given location x increases the surplus if and only if $\pi k - (1 - \pi - L_{qf}(x))((1 - q)(y - xt - u) - k) > 0$, which is equivalent to $x < \frac{1}{t}\left(y - u - \frac{k}{1-q}\right)$, which is always true within the registration zone. Thus, for a very low cousinage penalty J, the introduction of both registration and cousinage increases more the surplus than only introducing registration and, as the surplus increases with J, it is also true for all $J \geq 0$.

Eventually, it is clear that the overall surplus is increased when the registration cost k decreases, as the registration decision brings a higher surplus in each location (i.e. $y - xt - u - k - R_a$ increases when k decreases).

D.5. Proof of Proposition 7

To study the impact of the subsidy, let us first write the overall surplus obtained when the city structure is that described in Proposition 4 (taking as a reference the overall surplus in the model with ethnic matching only, $\Xi(J)$):

$$\begin{aligned}\Sigma^{sub}(J,k,s) &= \Xi(J) + \int_0^{\mathring{x}(k-s,J)} L_{fq}(k-s)((1-q)(y-xt-u)-(k-s))dx - s\int_0^{\mathring{x}(k-s,J)} L_{fq}(k-s)dx \\ &= \Xi(J) + \int_0^{\mathring{x}(k-s,J)} L_{fq}(k-s)((1-q)(y-xt-u)-k)dx\end{aligned}$$

In the first expression of the surplus, we can observe that it can be decomposed into the sum of the surplus in the absence of registration, $\Xi(J)$, and the additional surplus contribution of the registration zone when the registration cost is $k - s$, minus the total cost of the subsidy (i.e. the last term of the line). After simplification, we can see that the subsidy only impacts the size of the registration zone and the number of q-plot owners registering their plot in each location of the registration zone (but not the contribution to the surplus of each registration decision, because the monetary effect of the subsidy cancels out).

Then, it is clear that, when $J > k$, as the optimal city structure already prevails (with an optimally sized registration zone), introducing a registration subsidy will lead to an undesirable extension of the registration zone beyond its optimal boundary and thereby reduce the surplus.

When $0 < J < min(\pi k, (k + R_a)(\pi(1 - q) + q) - R_a) < k$ and $k > \underline{k}$, on the contrary, it is clear that both the size of the registration zone and the number of q-plot owners in each of its location are too small. Thus, introducing a registration subsidy of the right magnitude will unambiguously increase the surplus. This subsidy, however, must not be too large. Indeed, when s increases, the registration zone reaches and then goes beyond its optimal boundary (this happens when $s = min(\pi k, (k + R_a)(\pi(1 - q) + q) - R_a) - J$). At this point, it may still be surplus-enhancing to further increase the subsidy because not all q-owners register their plot in each location of the registration zone. But, when the subsidy reaches $k - J$, all q-owners register in each location of the registration zone and the registration is (still) too large so that it becomes unambiguously surplus-reducing to further increase the subsidy.