

Research Group: Behavioral and Experimental Economics

July 10, 2009

Two is Company, N is a Crowd? Merchant Guilds and Social Capital

ROBERTA DESSI AND SALVATORE PICCOLO

Two is Company, N is a Crowd? Merchant Guilds and Social Capital*

Roberta Dessí[†] Salvatore Piccolo[‡]

July 10, 2009

Abstract

Local merchant guilds were ubiquitous in medieval Europe, and their development was inextricably linked with the development of towns and the rise of the merchant class. We develop a theory of the emergence of local merchant guilds as an efficient mechanism to implement collusion among merchants and rulers, building on the natural complementarity between merchants' market trading and mutual monitoring. Our model explains the main observed features of local merchant guilds' behavior, their rules and internal organization, including membership restrictions and exclusion, and their relationship with rulers. Moreover, it identifies the main channels through which the guilds' social capital influenced their ability to collude with rulers, and hence social welfare. As we show, the available historical evidence supports our theory, shedding new light on the role of the guilds' social capital. We then extend the model to analyze the key trade-offs faced by rulers in choosing whether to grant recognition to one or multiple guilds. This provides an additional rationale for the establishment of the *alien* merchant guilds first analyzed by Greif, Milgrom and Weingast (1994), helping us to understand the observed distribution of guilds and their characteristics.

Keywords: merchant guild, social capital, collusion, political economy, trade, taxation.

*For many helpful comments and discussions, we would like to thank Ran Abramitzky, Dirk Bergemann, Federico Boffa, Liam Brunt, Jeremy Edwards, Guido Friebel, Oscar Gelderblom, Avner Greif, Luigi Guiso, Denis Hilton, Dilip Mookherjee, Lyndon Moore, Sheilagh Ogilvie, Marco Pagano, Sven Rady, Giuseppe Russo, Klaus Schmidt, Joel Sobel, Giancarlo Spagnolo and Jean Tirole. We also thank all the participants in the CSEF-IGIER summer meeting (Anacapri, 2008). Corresponding author: Roberta Dessí, IDEI, Toulouse School of Economics, Manufacture des Tabacs, Aile Jean-Jacques Laffont, 21 Allée de Brienne, 31000 Toulouse, France (dessi@cict.fr).

[†]Toulouse School of Economics (GREMAQ and IDEI) and CEPR

[‡]Toulouse School of Economics, University of Naples and CSEF.

1. Introduction

Since the pioneering work by Greif, Milgrom and Weingast (1994) (henceforth GMW), merchant guilds have attracted considerable attention by economists, and for good reason: this celebrated historical institution dominated trade for several centuries, and its development was inextricably linked with the growth of medieval towns and the rise of the merchant class. The merchant guild has also been viewed by many as a shining example of *social capital*¹, bringing major economic and social benefits – suggesting a very valuable potential role for such social capital even in modern (e.g. development or transition) economies.²

Is this very positive view of merchant guilds' social capital justified, and if so, what made their social capital particularly valuable? We address this question theoretically and empirically. We begin by shedding new light on the reasons for the emergence of merchant guilds, and on the role played by these guilds. In a nutshell, the main idea here is that merchant guilds emerged as an efficient mechanism to implement *collusion among merchants and rulers*, building on the natural *complementarity* between merchants' market *trading* and mutual *monitoring*. Because of this complementarity, rulers seeking to extract surplus from trade found it advantageous to rely on merchant associations (guilds), able to monitor effectively the trading activities of members and non-members, rather than employing other parties to monitor and collect taxes from merchants. However, guilds could only generate sufficient surplus for rulers by sustaining profitable, collusive market outcomes: this in turn required that only guild members be allowed to trade. Moreover, it implied an upper bound on guild membership, which in some circumstances entailed membership restrictions and exclusion. As we demonstrate in Section 2, our theory is supported by a substantial body of historical evidence on merchant guilds' behavior, on the privileges they obtained from rulers and the transfers they made to rulers, and on their internal norms and organization, including membership restrictions.

What role did the guilds' social capital play, and what were its welfare implications? When applied to groups or networks, such as merchant guilds, the notion of social capital³ typically refers to cohesion and trust among members, and to their resulting ability to enforce group norms and engage in effective collective action. In this sense, as discussed in Section 2, merchant guilds did indeed possess substantial social capital. Our analysis identifies several channels through which this social capital affected the relationship

¹See for example Putnam et al. (1993).

²On this see Bardhan (1996), Dasgupta (2000), Raiser (2001) and Stiglitz (1999).

³For definitions see, among others, Bourdieu, 1986; Coleman, 1990; Spagnolo, 1999; Dasgupta and Serageldin, 2000; Glaeser, Laibson and Sacerdote, 2002; Lin, 2001; Putnam, 2000; as well as Sobel, 2002 for an excellent discussion.

between guilds and rulers, and social welfare. The main trade-off is readily apparent by considering the simplest and most widespread case, that of a single merchant guild in a given polity. In this case, a high level of social capital within the guild could generate important efficiency gains through its effect on the guild's *monitoring ability*. In particular, cohesion and trust among guild members facilitated coordination on efficient internal and external monitoring. *Internal* monitoring ensured that members who deviated from guild norms were detected, while *external* monitoring detected non-members attempting to undermine the guild's monopoly over local trade, granted by the ruler, as documented in Section 2. Both types of monitoring were crucial in sustaining the guild's ability to extract surplus from trade; indeed, alien merchants coming to English towns "were carefully watched, lest they should sell or buy under colour or cover of a faithless gild-brother's freedom, the latter being expelled from the fraternity, or otherwise severely punished, if found guilty of this offence" (Gross, 1890, p.48).

The reason is simple: medieval rulers could, and did, give *legal legitimacy* to the guilds' norms and privileges; however, their *enforcement* required monitoring to detect (and hence punish) deviations. It would have been very costly for rulers to hire third parties as monitors. Because of the complementarity between trading and monitoring trade, significant reductions in monitoring costs could be achieved if guild members undertook internal and external monitoring themselves, provided they possessed sufficient cohesion (social capital) to do so efficiently. On the other hand, cohesion and trust among guild members also increased their bargaining power in negotiating with rulers, and hence their ability to secure a share of the surplus from trade. Thus in some circumstances, high levels of social capital within the guild, combined with restrictions on membership, generated substantial inequality between guild members, who earned large rents, and all those who were excluded from membership. A trade-off therefore emerged between the *efficiency* and *equity* implications of the guilds' social capital.

Our model presents a very different, although potentially complementary, theory of the emergence and role of merchant guilds relative to GMW. The reason is that GMW developed a theory of *alien merchant guilds*; that is, associations of alien merchants supported by the rulers of the polities in which they traded. In their model, individual alien merchants trading in a medieval polity were potentially vulnerable to attacks and expropriation. This made it impossible to sustain efficient trade in the absence of a credible commitment by the polity's ruler to provide commercial security. GMW argued that merchant guilds emerged as a solution to this commitment problem: by organizing themselves in associations which could enforce trade embargoes in response to misbehavior by rulers, merchants were able to obtain commercial security. Moreover, rulers were willing

to support such organizations of alien merchants precisely because they made it possible to sustain efficient trade.

Historically though, as discussed in Section 2, most merchant guilds emerged as *local merchant guilds*, i.e., associations of local merchants that obtained recognition and privileges (including monopoly power over local trade) from their local rulers. Alien merchant guilds were typically formed by the members of local merchant guilds who were active in long-distance trade, and remained under the control and supervision of the guilds from the merchants' polities of origin. Moreover, only a subset of local merchant guilds went on to form such foreign 'branches', primarily in the main international trade centers. It is therefore of considerable interest to understand the economic rationale for the emergence of *local* merchant guilds, and the reasons why medieval rulers were willing to grant them recognition and privileges, as documented in Section 2.

This is all the more important as *local* merchants dominated medieval towns, many of which acquired considerable power and autonomy: thus understanding the roots of merchants' organizations and their relationships with rulers is crucial in understanding the wider political economy forces that shaped the development of towns and states. Indeed, it would be difficult to consider medieval merchant organizations and towns in isolation from each other: in England⁴, for example, charters were often granted by rulers to towns (boroughs), establishing the burgesses' right to a merchant guild with monopoly power over local trade, and at the same time exempting them from a variety of taxes, while requiring an annual fixed payment.⁵ Surviving records of proceedings following the grant of such charters confirm the very close link between merchant associations and towns.⁶

For this reason, and to keep our analysis as simple as possible, the model we develop in Section 3 abstracts from the distinction between guilds and towns. We focus instead on the key relationship, the dynamic game between the ruler of a medieval polity and a large number of potentially active merchants.⁷ In the absence of merchant organizations,

⁴The point applies with equal force to other countries such as France, Germany and Italy; on this see, among others, Dollinger (1970), Racine (2004) and Spruyt (1994).

⁵Thus many charters contained a clause similar to the following: "We grant a Gild Merchant ... so that... no one who is not of the Gild may merchandise in the said town, except with the consent of the burgesses" (Gross, p.8). Tax exemptions included "tall and stallage, lastage, passage, pontage, and all other customs"; in return, the burgesses would pay "annually at our Exchequer the right and customary ferm" (Gross, p.5, citing the charter given by King John to the town of Ipswich in 1200).

⁶For example, records of proceedings in Ipswich following the grant of the charter by King John detail the various elections that took place to appoint officers to govern the town and the Gild; they also document the requirement for "all who are of the freedom of the town" (i.e. enjoying the privileges, including tax exemptions, granted by the king) to "give their hanse to the said Gild" (i.e. make payments) (Gross, pp.23-25).

⁷For simplicity we also abstract from another important historical observation: in some cases, notably

the ruler hires a local agent to monitor trade and collect taxes on realized revenues from this trade. In each period, merchants individually decide whether to collude, deviate, or punish deviations (compete à la Bertrand), in the spirit of Rotemberg and Saloner (1986). Our first result is that *collusion among merchants is helped by the presence of the tax-collector*. The reason is that higher profits are taxed more, thereby reducing the potential gain from deviation.

However, the tax-collector regime is costly because the agent has to be induced to monitor and collect taxes. If collusion among merchants can be sustained in the absence of a tax-collector, the ruler can increase his revenues by granting a monopoly over local trade to an association ('guild') of merchants, in return for a transfer.⁸ Indeed, this is exactly what happened in very many polities. For this to work, *the number of merchants belonging to the guild cannot exceed an endogenous threshold level $\bar{n}(\delta)$* , since increasing membership beyond this threshold would reduce each member's share of current and future collusive profits too much relative to the static gain from a deviation, making it impossible to sustain collusion. For simplicity, we assume that deviations are perfectly detected, and punished by exclusion from the guild. Clearly the result will continue to hold if we allow for additional punishments (e.g. fines, jail, confiscation of property) and imperfect detection, as long as there is a positive expected static gain from deviation.⁹ This result provides a rationale for the restrictions on membership that were often imposed by guilds. At the same time, in our model guild membership cannot fall below a second threshold \underline{n} , because a very small number of guild members would not be able to monitor effectively and hence prevent the potential (unofficial and forbidden by the ruler) trading

many Italian city-states, the distinction between merchants and rulers was itself blurred, as the rulers were essentially merchant oligarchies. As we discuss in section 6, this can be viewed as an extreme case of our theory, where collusion between ruler and merchants becomes (almost) perfect. Many of the qualitative insights and predictions of our analysis continue to hold in this case, and are confirmed by the historical evidence.

⁸Dessi and Ogilvie (2003), in never published work, developed a simple model of merchant guilds in which local rulers grant recognition and privileges to associations of local merchants in order to maximize their fiscal revenue from trade. In their model, giving monopoly power over local trade to a guild, in return for an appropriate transfer, yields a higher revenue for the ruler than hiring a local agent to collect taxes on trade. The reason is that the agent, who has better information than the ruler concerning local demand shocks, will be able to earn informational rents ex post; moreover, he is assumed to be capital-constrained and hence unable to pay for these expected rents ex ante. In contrast, it is assumed that a sufficiently large group of merchants, by pooling their resources, will be able to pay the required ex-ante fee to the ruler, who can therefore extract all the surplus from trade. While there is considerable historical evidence that medieval tax-collectors were often capital-constrained, it might be argued that this problem could have been solved through associations of tax-collectors rather than merchant guilds. Our model, in contrast, provides a rationale for the emergence of merchant guilds that does not require tax-collectors to be capital-constrained: the key is the complementarity between trading and monitoring trade.

⁹On the other hand, rulers would not have allowed extreme punishments, reserved for greater offences.

activities of the many excluded merchants. This second threshold, as noted earlier, will depend on the degree of trust and cohesion among guild members, and the extent to which they are able to overcome free-riding incentives and cooperate to monitor effectively – in other words, on their *social capital*.

We therefore obtain two results concerning the ruler’s choice between the tax-collector regime and the guild regime. For sufficiently *high values of the discount factor* δ , *the ruler always prefers the guild regime*: Intuitively, it is easier to sustain collusion in the guild regime when merchants are more patient. For *lower values of the discount factor*, *the choice will depend on the relative efficiency of the tax-collector’s and the guild’s monitoring technologies* - which in turn depends on the quality of social capital. When the guild regime is chosen, the ruler and the guild bargain over how to share the resulting surplus. In particular, although the ruler will obviously obtain at least as much revenue as under the tax-collector regime, the guild may be able to secure a strictly positive share of the surplus. This will be the case when its members have sufficient cohesion, and there is no easy replacement for the guild. In the presence of membership restrictions, the result will be that a subset of potentially active merchants, the members of the guild, will be able to trade and earn strictly positive rents, while the remainder will be excluded from trade and obtain only their reservation utility. This illustrates the implications of the guild’s social capital for equity and efficiency: in terms of *efficiency*, more social capital among the guild members increases the guild’s monitoring ability, which may help to sustain efficient collusion, and avoid the need for costly monitoring by the tax collector. In terms of *equity*, more social capital reduces the guild’s minimum size, while increasing its bargaining power; this results in higher inequality between (possibly fewer) guild members and (possibly more numerous) excluded merchants.

Historical evidence on medieval merchant guilds shows that in some cases rulers granted recognition and privileges to more than one guild; for example, to a local merchant guild and to one or more alien merchant guilds. Existing theoretical models offer no explanation for this: GMW, for instance, predict that it is optimal for the ruler to recognize a single guild.¹⁰ In Section 5, we extend our analysis to study the trade-off faced by the ruler in choosing between recognizing one or two guilds. This is based on two main effects. The first is the effect on *monitoring efficiency*: by allowing each guild to target its monitoring so as to exploit its informational comparative advantage, recognizing two guilds instead of one may generate “economies of scale” in monitoring. This in turn relaxes the minimum size constraint for each guild, making it easier to provide the *individual* incentives required to sustain efficient collusion. For example, consider a city

¹⁰This is also true for Dessi and Ogilvie (2003).

such as Bruges, with its own local merchant guild and possibly an alien guild of merchants from, say, Genoa. Local guild members may have a comparative advantage in detecting "misbehavior" by other local merchants, while members of the Genoese guild may have a comparative advantage in detecting misbehavior by other Italian merchants. Overall, monitoring may then be more efficient if both guilds obtain recognition.

The second effect works in the opposite direction: in the presence of multiple guilds, there is an additional incentive constraint that needs to be satisfied. Specifically, members of each guild should not have a *collective* incentive to deviate from efficient collusion. A key variable here is the probability θ that the two guilds recartelize following a deviation: the higher this probability, the harder it is to sustain collusion. We therefore find that rulers are more likely to grant recognition and privileges to multiple guilds when θ is lower, and when each guild's social capital provides a monitoring advantage relative to a particular set of excluded merchants and potential unofficial trading activities. This result may help to explain why the establishment of a single guild of *local* merchants was the norm, while additional guilds of *alien* merchants were able to obtain recognition and privileges in a number of polities, and notably in important centers of international trade such as Constantinople or Bruges. Intuitively in the latter cases, national and cultural *homogeneity within* each of the guilds helped to generate the required internal cohesion and social capital, while *heterogeneity across* guilds made it difficult for trust to be restored after a deviation, reducing the probability of recartelization.

The paper is organized as follows. The remainder of this Section discusses further the relationship with the existing literature. Section 2 reviews the historical evidence on merchant guilds that will motivate our theoretical analysis. Our baseline model is introduced in Section 3. We then study the choice between the tax-collector regime and the guild regime in Section 4. The trade-off between establishing one or two guilds is examined in Section 5. We review the main implications of our theory and relate them to the historical evidence in Section 6. Our concluding remarks are contained in Section 7.

1.1. Relationship to the literature

Beyond the literature on merchant guilds in economics, discussed above, this paper is clearly related to two other important literatures.

First, our paper is obviously related to the literature on social capital. An important strand of this literature has studied the link between trust and economic outcomes: trust appears to be positively correlated with economic growth (Keefer and Knack, 1997, Knack and Zack, 2001), with judicial efficiency (La Porta, Lopez-de-Silanes, Shleifer and Vishny, 1997), with the size of a country's stock market (Guiso, Sapienza and Zingales,

2008), and with cross-country trading patterns (Guiso, Sapienza and Zingales (2009)). The determinants of trust (and trustworthiness) at the individual level have also been investigated (Alesina and La Ferrara, 2002, Glaeser, Laibson, Scheinkman and Soutter, 2000), as well as the determinants of participation in social activities in more or less heterogeneous communities (Alesina and La Ferrara, 2000). The relationship between trust and economic performance at the individual level has been explored by Butler, Giuliano and Guiso (2009). Yet another approach has been developed by Guiso, Sapienza and Zingales (2004), who study the link between social capital and financial development, using electoral participation and blood donation as measures of social capital. These studies have focused primarily on generalized trust and civic engagement, which correspond broadly to the notion of ‘bridging’ social capital used by sociologists. Our paper differs in this respect by focusing instead on an example of ‘group’ social capital, closer to the sociologists’ notion of ‘bonding’ social capital. A key issue of interest here, of course, is the relationship between these two forms of social capital. Indeed, some of the very positive views that have been expressed about merchant guilds’ social capital might be interpreted as positing a substantial degree of complementarity between the two (e.g. Putnam, Leonardi and Nanetti, 1993). Our theoretical analysis, on the other hand, suggests the opposite, and is consistent with the historical evidence reviewed in Section 2 and Section 6.

Second, our analysis shares also common features from the literature on collusion in oligopolies. Our paper is closest in spirit to Rotemberg and Saloner (1986): in modeling the dynamic interaction between merchants we allow, as they did in studying oligopolistic firms, for the possibility of fully efficient or inefficient collusion, depending on the realization of market demand. In our model though, monitoring abilities (and the social capital that generates them) play a crucial role, which was absent in their setting. Indeed, we find that the optimality of multiple guilds, in some circumstances, is driven by the possibility of achieving scale economies in monitoring, by exploiting the monitoring advantage due to different guilds’ social capital. This result is novel, to the best of our knowledge, and could also apply to other collusive organisations.

2. The historical evidence

This section reviews the historical evidence on merchant guilds, which will motivate our analysis. In particular, we review the evidence on the *origins* of merchant guilds, on the recognition and *privileges* they received from rulers (notably monopoly power over local trade, and a variety of tax exemptions), and on the *transfers* they made to rulers. We then

discuss the evidence on merchant guilds' *social capital*, and especially on the important role of internal and external *monitoring*, as well as guild *norms* intended to help members secure rents, and *sanctions* used to discourage members from breaching guild norms. finally, we review the evidence on guild membership restrictions and *exclusion*.

2.1. Origins of merchant guilds

Most merchant guilds emerged as associations of *local* merchants. These *local merchant guilds* were ubiquitous in medieval Europe, and were supported by their local rulers, who granted them official recognition and a variety of important privileges, including monopolies over local trade.¹¹ Some of these local merchant guilds then established foreign branches (“colonies” and “consulates”) in important trade centers¹², when a significant number of their members engaged in long-distance trade. However, while very many European towns had a local merchant guild (in England alone, for example, there were over one hundred towns with a local merchant guild¹³), only a small subset of these towns established colonies or consulates in the main international trade centers.¹⁴ These *alien merchant guilds* were closely linked to the local merchant guilds of their polities of origin, on whom they depended for their internal rules and governance, as well as for external recognition.¹⁵

2.2. Recognition and privileges

Local merchant guilds throughout medieval Europe obtained from their rulers a variety of privileges enabling them to restrict competition and secure rents. These privileges were sometimes granted as part of charters given to towns, which also gave the towns a degree of political, administrative and financial autonomy. This was the case in England, where many such town charters contained “a clause similar to the following: ‘We grant a Gild Merchant with a hanse and other customs belonging to the Gild, so that [or ‘and that’] no one who is not of the Gild may merchandise in the said town, except with the consent

¹¹See Bernard (1972), p.304; Dilcher (1984), pp. 72-76; Ehbrecht (1985), pp. 430, 449; Racine (1985), pp. 131-132, 134-138; Schulz (1985), p.411; Schütt (1980), p.79. The privileges granted to local merchant guilds are discussed below.

¹²See Planitz (1940), p.19; Racine (1985), pp.134-5; Schütt (1980), p.79; Woodward (2003), p.1.

¹³See Gross (1890), pp.9-20, for a list of all those for which there is explicit documentary evidence, many dating back at least to the eleventh and twelfth centuries. The actual number is likely to have been even greater, implying that by the thirteenth century, local merchant guilds were “one of the most prevalent and characteristic features of English municipalities.” (p.22).

¹⁴On this see, among others, Bernard (1972), pp.299, 304-5; Epstein (2000), pp.27-9; Johaneck (1999), pp.72, 76-7; Laiou (2000), pp.811-3; Prevenier (2000), pp.581-2.

¹⁵See Abulafia (1986), pp.530, 537-8; Abulafia (1997), pp. 54-5; Abulafia (2000), pp.660-1; Ashtor (1983), pp.13, 68-9, 78, 149, 411; Blom (1984), pp.20, 25; Choroskevic (1996), pp.71-2, 78, 86.

of the burghesses” (Gross, 1890, p.8). Thus English local merchant guilds were granted the right to *exclude any non-member from trade*. In continental Europe, the granting of monopoly privileges to local merchant guilds was not always linked to the granting of greater political autonomy to towns. A good example is the guild of the *mercatores aque* (“water merchants”) in Paris: in 1170, the French king, Philip Augustus, “granted them a virtual monopoly of the Seine traffic between the bridges of Paris and Mantes”.¹⁶ Yet under Philip Augustus, “Unlike those of most towns in the royal domain, the bourgeois of Paris were permitted no semblance of autonomy”.¹⁷

The privileges granted to local merchant guilds in many medieval European cities meant that *alien* merchants could be either excluded from trade¹⁸, or allowed to trade only subject to a number of restrictions clearly intended to favor local merchants. Among the most common of these restrictions were “staple” rights and brokerage rights. Local guilds’ “*rights of staple*” meant that alien merchants had to bring their merchandise to municipal warehouses where members of the local merchant guild could buy them at favorable prices.¹⁹ Local guilds’ *brokerage rights* meant that alien merchants could not trade directly with consumers or with other alien merchants: they had to use members of the local merchant guild as intermediaries (brokers).²⁰ Local merchant guilds often enjoyed several of these privileges: for example, the local merchant guilds in cities such as Bruges and Cologne enjoyed both staple rights and brokerage rights.²¹ At the same time, local merchant guilds could also exclude from trade *local* individuals who were not members of the guild.²²

¹⁶Baldwin (1986), p.348.

¹⁷Baldwin (1986), p.349. See also Luchaire (1902), p.239.

¹⁸See Hibbert (1963), pp.169-74; Irsigler (1985), p.59; Leguay (2000), p.121; Planitz (1940), p.25; Postan (1973), pp.189-91; Reyerson (2000), pp.59-60; Schultze (1908), pp.498-502, 506, 523, 526-7; Spufford (2000), p.177.

¹⁹Bernard (1972), p.302; Kuske (1939); Reyerson (2000), p.58; Schultze (1908), p.500; Volckart and Mangels (1999), p.444.

²⁰Bernard (1972), p.302; Choroskevic (1996), pp.84-6; Hibbert (1963), p.170; Schultze (1908), pp.498-502, 506, 523, 526-7; Spufford (2000), p.177.

²¹Bahr (1911), pp.21-2; Daenell (1905), p.14; Kuske (1939), pp.40-1; Prevenier (2000), p.593.

²²Postan (1973), pp.189-91; Schütt (1980), p.121. For some detailed examples, see Planitz (1940), pp.25-8, on the Flemish local merchant guild in Ghent and the French local merchant guild in Valenciennes; Schulze (1985), pp.379-81, on the German local merchant guild of Stendal; Schütt (1980), pp.398-9, on the German merchant guild of Halberstadt; Woodward (2003), p.3, on Catalan and Aragonian local merchant guilds.

2.3. Taxation and transfers

Local merchant guilds were granted exemptions from a variety of tolls and other taxes, and made regular direct transfers to their rulers.²³ Moreover, the transfers did not vary systematically with the profitability of trade, which is consistent with the model we develop in Section 3. In England, for example, the same town charters that granted legal recognition and monopoly privileges to the local merchant guild generally granted exemptions from all tolls and other taxes, in exchange for a fixed sum or farm (*firma burgi*) to be paid annually by the town to the ruler.²⁴ While local guild members enjoyed the right to trade freely and were generally exempt from all tolls, “unfranchised merchants, when allowed to practise their vocation, were hemmed in on every side by onerous restrictions. Of these the most irksome was probably the payment of toll on all wares that they were permitted to buy or sell.”²⁵ Membership of the local merchant guild also carried obligations of course, notably participation in the town’s assessments and payment of pecuniary charges - which ensured that the *firma burgi* was duly paid and the privileges granted in the charter maintained.²⁶

A similar pattern can be observed in France: one of the earliest examples is the town of St. Omer, which obtained freedom from all tolls and other taxes in 1128 in return for a fixed annual sum or farm.²⁷ The local merchant guild in St. Omer enjoyed a variety of monopolistic privileges and contributed to the provision of local public goods.²⁸ In Spain, local merchant guilds made regular payments to the ruler in return for their privileges, obtained exemptions from tolls, and collaborated in the collection and administration of taxes on trade.²⁹

2.4. Social capital

Cohesion and trust among members of the same merchant guild were fostered by repeated and close interaction in a range of different domains - economic, social and religious. Regular social gatherings, including assemblies and feasts, were very common, and often

²³For examples see, among others, Ehbrecht (1985), pp.425-6, on Germany; Freshfield (1938), p.17 and Racine (1985), p.139, on Constantinople; Hoffmann (1980), p.49, on Denmark; Racine (1985), pp.135-6, on Italy; Schütt (1980), pp.112-21, on Sweden.

²⁴Gross (1890), pp.6-7.

²⁵Gross (1890), p.43.

²⁶Gross (1890), pp.53-4, 57.

²⁷Lyon and Verhulst (1967), p.31.

²⁸See Dilcher (1984), p.70; Hilton (1992), p.90; Irsigler (1985), pp.57-8; Kemble (1876), p.533; Planitz (1940), p.21.

²⁹Smith (1940), pp.48, 61-5, 86; Woodward (2003), pp.3-4.

even compulsory, with members being fined for missing them.³⁰ When abroad, merchants typically lived and interacted closely with other merchants from the same polity of origin; for example, Italian merchants in the Byzantine Empire, the Levant and Africa obtained “special quarters where they could live according to their laws and beliefs” (Lopez (1971), p.64). Close and repeated interaction helped to generate social capital; this in turn enabled merchant guilds to enforce the privileges granted to them by rulers, and establish collective norms to restrict competition and secure rents.³¹

Enforcement of their monopoly over local trade required monitoring to detect possible deviations - for example by alien merchants, with the possible help of some local guild members. Thus as noted in the Introduction, alien merchants coming to trade in English towns “were carefully watched, lest they should sell or buy under colour or cover of a faithless gild-brother’s freedom, the latter being expelled from the fraternity or otherwise severely punished, if found guilty of this offence” (Gross, 1890, p.48). Guild members not only sanctioned ‘deviant’ members who breached guild rules; they also intervened directly against non-members who undermined their monopoly over local trade. Gross (1890), p.39, describes one such instance in detail: when Richard the Baker from Stafford bought some wool in Newcastle-under-Lyme, local guild members seized the wool, and then defended their action arguing that the purchase had been in breach of the guild’s privileges.

Internal and external monitoring, and the resulting ability of guild members to secure rents, were greatly facilitated by a variety of rules and restrictions. Here are just a few examples. In tenth-century Constantinople, local guild members were required to pool their resources and make purchases as a cartel.³² Alien merchants were not allowed to spend more than three months in Constantinople.³³ The guild of raw-silk dealers³⁴ required its members not to “sell unworked silk in their homes but in the market”, where mutual monitoring was easier. Members were also required not to sell to other merchants for resale outside the city. The widespread use of staple and brokerage rights throughout medieval Europe, described in subsection 2.2, also greatly facilitated monitoring of alien merchants, since the latter had to bring their merchandise to a specific place where it

³⁰See Dilcher (1984), p.70; Kohn (2003), pp. 42-43; Pertz (1925), pp.118-119; Schulze (1985), pp.379-380.

³¹On this see, for example, Bateson (1899), pp.205-7; Bernard (1972), p.320; Dilcher (1984), p.70; Freshfield (1938), pp.16-17, 19-22, 28-9; Irsigler (1985), pp.57-8; Planitz (1940), p.21; Racine (1985), p.139; Schulze (1985), pp.379-80.

³²Lopez and Raymond (2001), p.21-22.

³³Lopez and Raymond (2001), p.21-23.

³⁴Constantinople was unusual in having more than one local guild, but the local guilds specialized in trading different commodities and were not in competition with each other, which is consistent with the model we develop below.

could be easily inspected by local guild members, and could not trade directly with other alien merchants or consumers.

Monitoring played a key role in enforcing compliance with guild norms both at home and in other polities. For example, the statutes of the merchant guild of the Italian town of Piacenza describe the duties of the consuls of the colony of merchants from Piacenza in Genoa: these include monitoring and collecting fines imposed by the merchant guild in Piacenza.³⁵ The same statutes also specify that in any colony where there are at least three merchants from Piacenza, two consuls must be elected!³⁶

Merchant guilds throughout medieval Europe employed a variety of sanctions against members who violated guild norms, ranging from fines to exclusion from the guild, confiscation of property and imprisonment.³⁷ Exclusion typically represented a very important punishment; for example, “For very serious offences the guildsmen of Andover fulminated a decree of excommunication against the erring brother - commanding ‘that no one receive him, nor buy and sell with him, nor give him fire or water, nor hold communication with him, under penalty of the loss of one’s freedom.’” (Gross, 1890, p.32).

2.5. Exclusive membership

Did local merchant guilds restrict membership? The answer is yes: membership was often contingent on having “citizenship” or “burgess” or “free” status, from which many were excluded.³⁸ As towns grew, attracting large numbers of rural immigrants, this exclusion affected an increasing number of urban inhabitants. In England for instance, “big towns had populations most of whose members were not ‘free’ - two thirds, for example in late-thirteenth-century London, a half in Oxford and more than three quarters in Exeter” (Hilton, 1992, p.92).

A key requirement for membership of local merchant guilds was the payment of entry fees and a variety of dues³⁹, which is consistent with the model we develop in Section 3. This implied the exclusion of those who could not afford to pay the, often substantial,

³⁵Racine (1979), p.307.

³⁶Racine (1979), p. 307.

³⁷See Ashtor (1983), p.415; Bateson (1899), pp.205-7; Choroskevic (1996), pp.74, 84-6; Freshfield (1938), pp. 16-7, 19-22, 28-9; Moore (1985), p.298; Racine (1985), p.139; Schulze (1985), pp.379-80; Schütt (1980), pp.112-21; Volckart and Mangels (1999), p.440.

³⁸See Dilcher (1985), pp.88-9; Epstein (2000), pp.35-6; Leguay (2000), pp.110-1, 121-2; Schultze (1908), 475, 490-3; Schütt (1980), p.131.

³⁹For examples see Ehbrecht (1985), p.445, on entry fees for the German merchant guild of Goslar; Dilcher (1984), p.69, and Volckart and Mangels (1999), pp.437-8, on dues levied by the Flemish merchant guild of Tiel; Schütt (1980), pp.112-21, on the dues levied by the Swedish merchant guild of Flensburg; Störmer (1985), pp.366-7, on entry fees for the Austrian merchant guild of Laufen; Origo (1986), p.44, on entry fees for the Italian merchant guild of Prato.

entry fees, or who were unable to provide the required guarantees: “To become a guildsman...it was necessary to pay certain initiation-fees...The new comer was also required to produce sureties, who were responsible for the fulfilment of his obligations to the Guild - answering for his good conduct and for the payment of his dues”.⁴⁰ The historical evidence makes it clear that many of the towns’ inhabitants could not meet these requirements.⁴¹ Moreover, admission to local merchant guilds was sometimes controlled by requiring that the potential new member be approved by a majority of existing members, and this requirement appears to have been used to restrict membership.⁴²

In sum, local merchant guilds excluded an increasing proportion of the urban population, notably the least wealthy.

3. The baseline model

This section introduces our model. We consider a medieval polity with three types of player: a ruler, merchants, and a tax-collector. For simplicity, they are all assumed to be risk-neutral.⁴³

Merchants: There is a large number N of identical individual merchants who can sell a homogeneous good. They play an infinite horizon game so that in each trade period, throughout denoted by $\tau \in (1, \dots, +\infty)$, a collusive, deviation or punishment phase takes place. The static (market) game is a reduced form of Bertrand competition: in each trade period τ merchants set prices simultaneously and, given the market demand, they sell at constant marginal costs, which are normalized to zero for simplicity. The dynamic version of the game follows Rotemberg and Saloner (1986). In an ‘efficient’ collusive phase, that is when all merchants charge the monopoly price, each earns a per-period profit equal to π^τ/N , while each earns 0 if not colluding, i.e., if all price at marginal cost.⁴⁴ The profit that a merchant makes by deviating from an efficient collusive agreement is equal to π^τ , so that unilateral deviations from a cartel are profitable in a static sense and yield a revenue π^τ .⁴⁵ The variable π^τ can be interpreted as a measure of the total market size. In each period, it is determined by the realization of a random variable, $\tilde{\pi}^\tau$, which is identically

⁴⁰Gross (1890), p.28.

⁴¹See Hilton (1992), p.92; Schultze (1985), pp.379-81.

⁴²See Smith (1940), p.38.

⁴³Introducing risk aversion would not change the basic trade-offs characterized in the paper.

⁴⁴As in Rotemberg and Saloner (1986) we shall allow for inefficient collusion: given any realization of the market size π^τ , merchants may set prices above marginal cost but below the monopoly level so as to obtain any intermediate profit $\beta\pi^\tau$, with $\beta \in [0, 1]$.

⁴⁵This structure is standard in the literature analyzing collusion, see for instance Chang and Harrington (2007) and Chen and Rey (2007).

and independently distributed over time (iid) with continuously differentiable cdf $F(\pi)$ drawn on the compact support $\Pi \equiv [\underline{\pi}, \bar{\pi}]$.

The ruler: The ruler governs the polity: he provides certain public goods, such as law enforcement and defence, and finances these with various sources of revenue, including the taxation of trade. He also spends on activities that provide him with private benefits, such as military campaigns and court display.⁴⁶ For our purposes it is sufficient to treat his expenditures and his other sources of revenue as given exogenously, and to focus on the revenue he can raise from the taxation of trade. The ruler is assumed to maximize this revenue. This is a reasonable assumption for the historical period under consideration, when rulers typically attached a low weight to the well-being of ordinary consumers.⁴⁷

Bargaining power: The ruler has all the bargaining power relative to individual agents (tax-collector, individual merchants), whose reservation utilities are normalized to zero. On the other hand, an association of merchants could have some bargaining power. The idea is that merchants, once organized in a guild and used to acting together in their common interest, may acquire some bargaining power relative to the ruler, particularly in the absence of an easy replacement. This possibility will be allowed for by assuming that the merchant guild receives a share $1 - \alpha$ of any ‘surplus’ accruing from an agreement with the ruler ($\alpha \leq 1$). We can think of $1 - \alpha$ as a measure of the guild’s social capital (cohesion and trust among the members).

Commitment: We develop our analysis under the simplifying hypothesis that the ruler can make credible commitments. As we show, our theory provides a rationale for the emergence of merchant guilds which does not require imperfect commitment by the ruler. In particular, under full commitment, our objective is to disentangle in the clearest possible way the basic cost-benefit trade-offs associated with the establishment of merchant guilds even in the absence of any friction, such as those due to imperfect commitment, capital constraints, etc. As will become clear below, allowing for imperfect commitment by the ruler would entail less reliance on ex-ante fees and greater reliance on ex-post transfers to the ruler. While this would complicate the analysis, it would not in general make either the guild regime or the tax collector regime decisively more attractive to the ruler.⁴⁸ We therefore focus on the case of full commitment, which brings out clearly the

⁴⁶For historical evidence on the importance of these, see Brewer (1989).

⁴⁷The main exception would be that of essential commodities such as food, since excessively high food prices might well provoke urban riots. Thus the model should be thought of as applying to the taxation of other commodities.

⁴⁸This is because the need for ex-post transfers would make it harder to sustain collusion in the guild regime, by reducing each guild member’s expected future collusive profits; but it would also make it

basic trade-offs involved in choosing between the two regimes.

The guild regime: At time $\tau = 0$, the ruler may choose to grant recognition to a merchant guild, and endow it with privileges, notably the power to exclude individual merchants from trade (as discussed in Section 2, this power may be applied to individuals who are not members of the guild, and also to members who break guild rules). When the ruler grants recognition and privileges to a merchant guild, he requires an ex ante fee, call it R , which is shared equally among the guild’s members, that is, each merchant pays R/N .⁴⁹ This fee is set at the level which solves a standard Nash-bargaining game between the ruler and the guild with weights α and $1 - \alpha$, respectively. Once it has been granted recognition and privileges by the ruler, a guild can punish members who deviate from a collusive agreement by excluding them from future trade. The merchants remaining in the guild then recartelize with probability one.⁵⁰ We therefore assume that when a single merchant deviates from a collusive agreement, he is excluded from trade for the rest of the game⁵¹ while the remaining guild members keep colluding with probability 1, as long as this continues to be profitable.

The tax-collector regime: In the absence of merchant organizations, the ruler delegates trade taxation to an agent, call it the tax-collector. In order to make the case for the tax-collector regime as favorable as possible, we assume that the agent is not capital-constrained. As will become clear below, this means that it is optimal for the ruler to choose a very simple form of delegation: he endows the agent with the right to collect taxes on trade in the polity, in return for an ex ante royalty fee \hat{R} .⁵² The agent then imposes and collects, at the end of each period, a per-merchant lump-sum tax on trade $T_i^\tau(\hat{\pi}_i^\tau)$.⁵³ This tax is a function of each merchant i ’s realized profits $\hat{\pi}_i^\tau$, which in period τ depend on the merchants’ actual market strategy (pricing decisions) as well as the realization of the market size π^τ : that is, $\hat{\pi}_i^\tau = \pi^\tau/N$ in an efficient collusive phase, $\hat{\pi}_i^\tau = \pi^\tau$ if merchant i deviates in period τ given that his competitors are pricing at the monopoly level and $\hat{\pi}_i^\tau = 0$ in a punishment phase. Unlike the ruler, the agent in each period can observe

harder to induce efficient monitoring and surplus extraction in the tax collector regime.

⁴⁹As discussed in Section 2, guild members typically paid some entry fees, as well as a variety of other dues. The guild then made transfers to the ruler.

⁵⁰Obviously exclusion does not apply if *all* guild members deviate at the same time. In that case, it seems reasonable to assume that there is no exclusion, and that members subsequently recartelize.

⁵¹Note that it will never be in the remaining members’ interest to welcome an excluded member back into the guild.

⁵²This is essentially “tax farming”, a very widespread practice in medieval Europe (see for example Lyon and Verhulst, 1967, pp.33, 49; Webber and Wildavsky, 1986, p.202).

⁵³We focus on the revenue-maximizing taxes so as to make the case for the tax collector regime as strong as possible.

the aggregate state π^τ as well as each individual state $\hat{\pi}_i^\tau$ by paying a monitoring cost $c > 0$. The key assumption here is that the ruler did not have direct access to detailed information about these realizations. This was generally the case for medieval rulers, who could not rely on a civil service to provide them with such information.⁵⁴ Local tax-collectors, on the other hand, had greater access to local information. For simplicity, we capture this by assuming that they could observe the relevant information by incurring the per-period monitoring cost c . Essentially, the idea is that the tax-collector had to go to the market and observe trade in order to tax and collect his revenue. The cost c may be thought of as capturing both the cost of observing trade and the cost of actually collecting taxes.

Timing and strategies: At time $\tau = 0$ the ruler decides whether to grant recognition to a merchant guild or hire an agent as tax-collector. Accordingly, he collects the associated royalty fees, R or \hat{R} . In each generic period τ the sequence of events is as follows: (i) Merchants observe the ‘public’⁵⁵ history of the game up to period $\tau - 1$, which also includes the realization of the current state π^τ ; (ii) they post prices simultaneously; (iii) trade takes place and taxes are levied in case an agent has been hired by the ruler. We assume that if a deviation occurs in any period, other merchants observe the identity of the deviating merchant by the end of the period. This assumption seems realistic in the case of medieval polities. Indeed, most of the trading activities in that historical period were taking place in markets where merchants could easily monitor each other’s pricing strategy.⁵⁶

The (static) market game is repeated an infinite number of periods and all agents have a common discount factor $\delta \in [0, 1]$ which, as standard in the collusion literature, is assumed to be greater than $1/2$.⁵⁷

A strategy for the ruler is a choice at $\tau = 0$ between recognizing a guild or hiring a tax-collector, together with an associated royalty fee. A strategy for the tax-collector instead involves: (i) a choice as to whether to participate in the game; (ii) a monitoring decision profile $\{m^\tau(h^\tau)\}_{\tau=1}^{+\infty}$ with $m^\tau(h^\tau) \in \{1, 0\}$, where, at each trade period $\tau \geq 1$ and for any public history h^τ , $m^\tau(h^\tau) = 1$ if the tax-collector monitors the merchants in that period,

⁵⁴Indeed, medieval rulers had to rely on a variety of agents to collect taxes, and struggled to limit the extent to which these agents exploited their informational advantage for their own benefit. On this see Bisson (1984) and Spruyt (1994).

⁵⁵By public history of the game we shall mean the history observed by all the merchants, as well as by the tax collector, if any, provided he monitors.

⁵⁶Probably the most famous example is that of the Champagne fairs, but of course there were very many other markets, including numerous local ones accommodating primarily local trade.

⁵⁷As we shall explain in Section 3.2 this assumption will ensure that efficient collusion is possible in some states of the world when the ruler recognizes a merchant guild.

and $m^\tau(h^\tau) = 0$ otherwise; (iii) an intertemporal tax profile $\{T_i^\tau(\hat{\pi}_i^\tau|h^\tau)\}_{i \in N}^{\tau=1}^{+\infty}$, where in each period τ and for any public history h^τ , $T_i^\tau(\hat{\pi}_i^\tau|h^\tau)$ maps the individual state of each merchant i , $\hat{\pi}_i^\tau$, into the set of all feasible lump-sum taxes $T_i^\tau(\cdot)$. Essentially, a public history profile in each period τ includes past trades as well as past tax rates. A strategy for each merchant i specifies a per-period decision on whether to trade and a pricing behavior, both contingent on past history and the current state of nature. Moreover, we shall assume that whenever expected profits from collusion are at least equal to expected profits from deviating, merchants will prefer to collude.⁵⁸

We shall look for the (pure strategy) subgame perfect Nash equilibria (SPNE) of this game.

Technical assumptions: For expositional simplicity, in the rest of the analysis we will make the following assumptions:

(A1) *All agents in the polity are capital unconstrained.*

A1 guarantees that our results do not rely on binding capital constraints neither on the tax-collector nor on the guild's side.

(A2) *The monitoring cost c is lower than the expectation of the market value $\mathbb{E}[\pi]$, that is $c < \mathbb{E}[\pi]$.*

This assumption implies that if the tax-collector expects merchants to collude, he will always find it optimal to monitor. This renders the tax-collector regime non-trivial.

Let $n \leq N$ be the guild's size, i.e., the number of its active members,

(A3) *A guild requires a minimum number of active members to be profitable, that is $n \geq \underline{n}$, with $\underline{n} < N$ and $\inf \{\underline{n}\} = 2$.*

A3 rules out the unrealistic and uninteresting possibility of a 'single member' guild. This assumption is motivated by the historical evidence, and it captures, in the simplest possible way for our purposes, the idea that guilds with smaller density (fewer members) were more likely to be exposed to competition by excluded merchants since their ability to monitor trading activities taking place outside the 'regular' market, i.e., that officially permitted by the ruler, decreased with the number of active members. Indeed, preventing unofficial trade required a minimal amount of monitoring effort, whose cost was generally shared by the guild members. Thus we are assuming that the cost per member was

⁵⁸This is the standard assumption in the collusion literature and avoids cumbersome notation.

sufficiently small to be normalized to zero as long as the number of members did not fall below \underline{n} . In the following analysis we shall interpret the guild’s density \underline{n} as being a measure of the guild’s monitoring ability, that is, the larger is the lower-bound \underline{n} , the less efficient is the guild’s monitoring “technology”.

Social capital: To summarize, the guild’s social capital is captured in two ways in the baseline model. First, we assume that greater cohesion and trust among guild members makes it easier for them to overcome free-riding incentives and coordinate on efficient monitoring. This reduces \underline{n} , the minimum number of members needed to detect attempts by non-members to undermine the guild’s monopoly over local trade (by engaging in unofficial and forbidden trading activities). One way in which this reduction was achieved in practice, as discussed in Section 2, was through agreement on guild rules that simply made monitoring easier - e.g. by requiring alien merchants to bring their merchandise to a specific place, and trade only through intermediaries who were themselves members of the local merchant guild.

Second, we assume that greater cohesion and trust among guild members increases their bargaining power relative to the ruler; i.e. it reduces α .

4. Trade, taxation and collusion with a single guild

We begin by considering what the ruler can achieve when merchants are not organized in a guild, then proceed to examine the role of guilds.

4.1. Trade and taxation in the absence of merchant guilds

In the absence of merchant organizations, the ruler hires an agent who can observe in each period τ the realized market value π^τ as well as the individual profit $\hat{\pi}_i^\tau$ of each merchant upon paying the per-period monitoring cost c .⁵⁹ In this case, the agent is given the power to impose and collect a lump-sum tax $T_i^\tau(\hat{\pi}_i^\tau)$ contingent on every merchant i ’s realized profit. That is, at the end of each trade period τ , this agent collects $T_i^\tau(\hat{\pi}_i^\tau)$ from each merchant i , who is left with

$$u_i^t(\hat{\pi}_i^\tau) = \hat{\pi}_i^\tau - T_i^\tau(\hat{\pi}_i^\tau).$$

where, as described before, the variable $\hat{\pi}_i^\tau$ represents the actual profit of merchant i in period τ and it changes depending on whether the trade market is in a collusive,

⁵⁹As noted in Section 3, the cost c may be thought of as capturing both the cost of observing trade and the cost of collecting contingent taxes from merchants. Moreover, it seems reasonable to assume that if the ruler hired a merchant to collect taxes, this merchant would not be able to engage in trade at the same time; thus the cost c would still need to be incurred.

punishment or deviation phase.⁶⁰ We solve the model backward by analyzing the subgame between the merchants and the tax-collector assuming that this agent has been hired by the ruler.

Given our assumptions, there exists a class of simple and intuitive SPNE of this game where the tax-collector always monitors and sets a zero-profit tax rate which leaves merchants with no surplus in each trade period τ , that is $T_i^\tau(\hat{\pi}_i^\tau) = \hat{\pi}_i^\tau$ for each i and τ , while merchants collude at a price which guarantees an expected profit at least equal to c in each period.⁶¹ Indeed, since we have assumed that whenever indifferent between colluding and deviating merchants will collude, it immediately follows that this type of equilibria exist, and collusion is self-enforcing for every number of active merchants, N , and for any possible discount factor, δ . The reason is that taxation reduces the merchants' post-tax profits to zero when they deviate as well as when they collude, thereby eliminating any incentive to deviate. In order to make the case for the tax-collector regime as strong as possible, in what follows we focus on the equilibrium where collusion is fully efficient and the tax collector extracts the maximum surplus from trade.⁶²

Proposition 1: *When the ruler hires a tax-collector there exists a SPNE of the game between merchants and the tax-collector where: collusion is always efficient and, in each period, the equilibrium tax rate is $T_i^\tau(\hat{\pi}_i^\tau) = \pi^\tau/N$, which leaves merchants with no surplus. Moreover, the tax-collector's ex ante participation constraint binds,*

$$\hat{U}(\delta, c) = \frac{\mathbb{E}[\pi] - c}{1 - \delta} - \hat{R} = 0 \quad \forall (\delta, c).$$

and the ruler obtains a profit equal to

$$\hat{V}(\delta, c) = \frac{\mathbb{E}[\pi] - c}{1 - \delta}.$$

Clearly, the incentive for the ruler to hire the tax-collector increases the larger is the difference between the expected market size and the monitoring cost, and the larger is the discount factor.

⁶⁰Although merchants are assumed to be capital unconstrained, the tax-collector can never extract more than $\hat{\pi}_i^\tau$ at the market stage. Indeed, a merchant would have no incentive to bring his total wealth to the market, knowing that in this case the tax $T_i^\tau(\hat{\pi}_i^\tau)$ could exceed his market revenue $\hat{\pi}_i^\tau$.

⁶¹When N is small there may exist other SPNE of the game between merchants and the tax-collector, in which the tax-collector never monitors and merchants collude by setting a price such that total profits are equal to $c - \varepsilon$, with ε small enough. Clearly, in this case, there is no scope for hiring a tax-collector, thereby making the ruler's decision at $\tau = 0$ trivial. We therefore abstract from such equilibria in the remainder of the paper.

⁶²We therefore also abstract from any possible equilibria in which collusion is fully efficient but the tax collector only extracts part of the surplus from trade.

4.2. Merchant guilds: trade, taxation and privileges

A possible alternative for the ruler, enabling him to achieve a higher profit, is the following. A subset of merchants organize themselves as a group, able to act in the group members' collective interest: Call this group 'the guild'. The ruler grants privileges to the guild, and in particular monopoly power over trade - that is, only members of the guild are authorized to trade. Under what conditions can the guild implement a better outcome than the tax collector, from the ruler's point of view?

To answer this question, we begin by characterizing the incentives of individual guild members and their implications for collusion. Let $n \leq N$ be the subset of active merchants belonging to the guild. For any given realization of the market size $\pi \in \Pi$ in period τ , a merchant who charges the monopoly price, given that the other members of the guild do the same, obtains an expected discounted utility equal to:

$$u_i^{g,c}(\pi) = \frac{\pi}{n} + \frac{\delta \mathbb{E}[\pi]}{n(1-\delta)}, \quad (4.1)$$

while a deviation from a collusive agreement would secure him an intertemporal profit equal to

$$u_i^{g,d}(\pi) = \pi. \quad (4.2)$$

Where π is the spot gain from an individual deviation by merchant i in period τ given that the other $n - 1$ members of the guild are behaving collusively, while the continuation payoff following a deviation is equal to zero since the merchant will be excluded from the market starting from the next period $\tau + 1$.

Given (4.1) and (4.2), the condition for collusion to be sustainable in a generic state π is simply

$$\pi \leq \frac{\delta \mathbb{E}[\pi]}{(n-1)(1-\delta)}. \quad (4.3)$$

As might be expected, it is more difficult to sustain collusion when the guild's size (number of members) increases. Intuitively, the greater the number of merchants in the guild, the lower is the individual gain that each merchant can appropriate from collusion. Conversely, holding guild size constant, collusion is easier to implement the larger is the expectation of market value.

For sufficiently large realizations of π , it may not be possible to sustain efficient collusion, as the potential gain from deviation would be too high. In such cases, collusion will entail setting a price below the monopoly price, which makes the potential gain from deviation smaller. It seems reasonable, given that we are considering the incentives of guild members, to focus on equilibria in which the price is set at the highest level compatible

with sustaining collusion. Specifically, there may exist a threshold $\pi^* \in \Pi$, with $\pi^* < \bar{\pi}$, such that for all $\pi \leq \pi^*$ each merchant obtains a profit π/n from being in an efficient collusive phase, while for $\pi > \pi^*$ the maximal profit that a guild member can obtain is π^*/n ,⁶³ with π^* being defined by:

$$\pi^* = \frac{\delta \mathbb{E}[\pi]}{(n-1)(1-\delta)}, \quad (4.4)$$

Because of this cut-off rule, we have:

$$\mathbb{E}[\pi] = \int_{\underline{\pi}}^{\pi^*} \pi dF(\pi) + (1 - F(\pi^*))\pi^*. \quad (4.5)$$

Thus, when the market value is high, the guild will obtain a collective profit, π^* , which is lower than the monopoly one. Substituting (4.5) into (4.4) we have:

$$\pi^* = \frac{\delta}{(n-1)(1-\delta)} \left\{ \int_{\underline{\pi}}^{\pi^*} \pi dF(\pi) + (1 - F(\pi^*))\pi^* \right\}.$$

To pursue the analysis, we now specialize our model and impose the following simplifying assumption:

(A4) *The distribution of the market size is uniform and is drawn from the unit interval $\Pi \equiv [0, 1]$, that is $\tilde{\pi}^T \sim \text{U}[0, 1]$.*

A4 is made only for ease of computation and to work with tractable closed-form solutions. All our qualitative results can be generalized to the case of any generic cdf $F(\pi)$ which satisfies standard regularity conditions. Under this specification, it is easy to show that the cut-off value π^* depends only on δ and n , and is equal to:

$$\pi^*(\delta, n) = \frac{2(1 - n(1 - \delta))}{\delta}.$$

This cut-off value is obviously increasing in δ , since more patient merchants are less attracted by deviations. Moreover, the larger is the size of the guild, i.e., the larger is n , the lower is $\pi^*(.)$, since each merchant is tempted to deviate more often in an environment where the gain from collusion has to be shared among many. Thus to have $\pi^*(.)$ as large as possible, the guild cannot have too many members. In particular, it will never be

⁶³The case where $\pi^* < \bar{\pi}$ corresponds to instances where in order to support collusion in states higher than π^* merchants must charge prices lower than the monopoly one. Formally, this implies that the guild total profit in each of these states is $\beta\pi$ with $\beta = \pi^*/\pi \leq \bar{\pi}$.

profitable to have a guild when its size exceeds the maximum threshold level compatible with sustaining collusion in at least some state. This upper bound on the number of guild members is obtained by setting $\pi^*(\delta, n) = 0$, and is given by the following lemma.

Lemma 1: *The guild regime has the following key features:*

- (i) *For a guild to be profitable, its size (number of members) must be lower than an upper bound $\bar{n}(\delta)$, given by:*

$$n < \bar{n}(\delta) = \frac{1}{1 - \delta}.$$

- (ii) *A guild can be recognized if and only if $\delta > 1/2$ and $\underline{n} < \bar{n}(\delta)$.*

Thus if $N \geq \bar{n}(\delta)$ guild membership will be restricted: there will be some exclusion. Clearly the upper-bound on the number of guild members is higher when merchants are more patient, i.e., the higher is δ . The need for the upper-bound $\bar{n}(\delta)$, together with the minimum size requirement imposed in **A3**, leads to a necessary condition for a guild to represent a potentially valuable option for the ruler which is provided in part (ii). Intuitively, the ruler may have an interest in granting recognition to a merchant guild only when its members are sufficiently patient and able to protect their trading ‘territory’ from excluded local or alien merchants.

We can now analyze what happens when a guild is recognized. Letting

$$n^*(\delta) = \frac{2 - \delta}{2(1 - \delta)}$$

denote the value of n which solves $\pi^*(\delta, n) = 1$ (i.e., the highest number of guild members compatible with sustaining efficient collusion), we can establish the following result.

Proposition 2: *Whenever a guild is recognized the following properties hold:*

- (i) *if $\delta \geq 2/3$ and $\underline{n} \leq n^*(\delta)$, collusion is fully efficient in all states $\pi \in \Pi$. In this case, the optimal guild size is such that $\underline{n} \leq n \leq n^*(\delta)$, and each merchant obtains a profit π^τ/n in every period τ .*
- (ii) *if $\underline{n} > n^*(\delta)$, collusion cannot be fully efficient in all states $\pi \in \Pi$. In this case, the optimal guild size is equal to \underline{n} . In any period τ where $\pi^\tau \leq \pi^*$ each guild member obtains π^τ/\underline{n} , while in any period τ where $\pi^\tau > \pi^*$ each gets π^*/\underline{n} .*

Intuitively, collusion can be efficient in every state only if the discount factor is large enough and merchants' monitoring ability is sufficiently effective. Otherwise, efficient collusion cannot be enforced in all states even when the guild's size is reduced to its lower bound, $n = \underline{n}$. In this case the guild will not be able to extract all the potential surplus from trade. The ruler may then face a trade-off between hiring a tax-collector and granting recognition to a merchant guild. This trade-off will be explored in the next section.

4.3. Merchant guild versus tax-collector

We can now study the conditions under which recognizing a merchant guild is optimal for the ruler. Of course, when recognizing a guild and granting it a monopoly over local trade ensures efficient collusion in all possible states, the ruler will prefer not to hire the tax-collector, unless his monitoring cost is zero. When instead collusion cannot be fully efficient under the guild regime, that is if $\pi^*(\delta, \underline{n}) < 1$, the ruler faces an interesting trade-off. On the one hand, hiring a tax-collector ensures efficient collusion in all states, irrespective of the discount factor and the number of active merchants on the market. On the other hand, the tax-collector's monitoring activity is costly and thus there are limits to the surplus that the ruler can extract from this simple form of delegation. The trade-off between these two effects shapes the ruler's optimal choice and is analyzed in the next proposition:

Proposition 3: *If $\delta \leq 1/2$ or $\underline{n} \geq \bar{n}(\delta)$ the ruler's best strategy is to hire a tax-collector. When $\delta > 1/2$ and $\underline{n} < \bar{n}(\delta)$, the ruler's choice between granting recognition to a merchant guild or hiring a tax-collector is determined as follows:*

- (i) *If $\delta \geq 2/3$ and $\underline{n} \leq n^*(\delta)$ the ruler always prefers to recognize a guild rather than hiring a tax-collector.*
- (ii) *If $\underline{n} > n^*(\delta)$ there exists a positive function $c^*(\delta, \underline{n})$ such that the ruler prefers to recognize a guild rather than hire a tax-collector whenever $c \geq c^*(\delta, \underline{n})$, while the converse holds otherwise. Moreover, in the relevant range of parameters, $c^*(\delta, \underline{n})$ is increasing in \underline{n} and decreasing in δ .*

This result has a simple and immediate economic intuition. Of course, the ruler is unable to grant recognition to a guild if its ability to keep excluded merchants out of the market is sufficiently poor. Conversely, when the guild is endowed with a strong monitoring ability, the ruler always prefers to deal with an association of merchants instead

of hiring a tax-collector, as long as the merchants are sufficiently patient to sustain efficient collusion in every state of nature. Finally, when a guild cannot ensure efficient collusion in all states of nature, the ruler may still prefer the guild regime to hiring a tax-collector. This will be the case if the tax collector's monitoring cost is above a given threshold, which depends positively on the discount factor and negatively on the minimum guild size.

Clearly, the guild's monitoring ability plays a crucial role in the trade-off between the guild regime and the tax-collector regime. This monitoring ability will depend on several factors, including the geographical and population characteristics of the polity, which may make it easier or harder to detect alien or excluded local merchants who engage in unauthorized trade, as well as the polity's openness to foreigners, and its attractiveness to alien merchants. Monitoring ability will also depend, on the other hand, on the degree of trust and cohesion among guild members, and the extent to which they are able to overcome free-riding incentives and cooperate to monitor effectively - in other words, on their *social capital*. In this sense, the guild members' social capital is valuable not just for them, but also for the ruler.

5. Trade, taxation and collusion with multiple guilds

We now extend the previous analysis to the case where the ruler has the option of granting recognition and privileges to more than one guild. To be consistent with the earlier analysis, and thereby bring out clearly the implications of recognizing more than one guild, we continue to assume that the ruler has full commitment power and that merchants are capital unconstrained. For simplicity, we shall focus on the case where the ruler chooses between recognizing one or two guilds. As we shall see, this is sufficient to identify all the trade-offs we are interested in.⁶⁴

Consider a polity where $I = 2N$ merchants are potentially active and can sell a homogenous good. Each merchant belongs to one of two symmetric⁶⁵ groups ($j = 1, 2$), of size $N_j = N$.⁶⁶ A natural interpretation for the two groups, discussed below, would be local and alien merchants, or alien merchants from two different polities of origin. Merchants belonging to both groups compete in the same market by setting prices and

⁶⁴The analysis can be easily adapted to study the case of $M > 2$ guilds.

⁶⁵Symmetry is assumed only for expositional simplicity.

⁶⁶As will become clear in the remainder of the analysis, allowing for asymmetric groups (i.e., with different sizes) would not bring additional insights, since the individual incentive constraint that needs to be satisfied for collusion to be an equilibrium of the market game with multiple guilds depends only on the total number of merchants active in both guilds.

both the static and the dynamic games are modeled as in Section 4.⁶⁷ In each period τ the market size is captured once again by the realization of a random variable $\tilde{\pi}^\tau \sim U[0, 1]$ which is iid over time.

To study the ruler's choice between recognizing one or two guilds, we assume that in each of the two groups, a subset of merchants organize themselves as a subgroup, able to act in its members' common interest: i.e., a "guild". A first natural question one may want to ask is how the presence of two guilds modifies the bargaining game between ruler and merchants. The answer seems intuitive: if the ruler only needs to recognize *one guild* in order to maximize his revenue from trade, the presence of two potential candidates allows him to gain bargaining power by playing the two competing organizations one against the other. A simple way of modeling this is to assume that the ruler gains full bargaining power vis-à-vis each guild (whose reservation value is zero). This would be the case, for example, if recognition and privileges were assigned through an auction, forcing each guild to bid competitively. This will be our benchmark for the single-guild case.

Clearly, the ruler will only grant recognition to *both guilds* if this entails some (net) benefit relative to the single-guild case. Our analysis so far suggests a very likely potential advantage of granting recognition and privileges to both guilds: this may induce more efficient monitoring, and thereby relax the minimum size requirement for each guild. For example, a guild of local merchants is likely to have better access to information about possible unauthorized trading activities by excluded local merchants, while a guild of alien merchants from a particular polity of origin will be more easily informed about any trading activities by other citizens of that polity (who are not members of the guild). To capture this as simply as possible, we assume that when both guilds are recognized, the minimum size of each is reduced by a fraction ϕ , that is, $\underline{n}_j = \phi \underline{n}$. The parameter $\phi \leq 1/2$ captures the extent to which having the two guilds, rather than just one, increases monitoring efficiency: when $\phi < 1/2$, multiple guilds generate '*economies of scale*' in monitoring⁶⁸, and these scale economies are greater for lower values of ϕ .

Intuitively, while multiple guilds may increase monitoring efficiency, they may also make it harder to sustain collusion. To investigate the interaction between these two effects, and the resulting trade-off between recognizing one or two guilds, we start by analyzing the collective behavior of the two guilds, as well as the behavior of their individual members.

⁶⁷Obviously, it would be easy to provide a rationale for multiple guilds if we assumed that different groups sold different goods and were active in different markets. We abstract from this possibility to focus on the more interesting trade-off between one and two guilds when merchants sell the same good and are active in the same market.

⁶⁸These are "economies of scale" in the sense that more guilds make it possible to achieve the required level of monitoring with a smaller total number of merchants.

We retain the same assumptions as in the earlier analysis concerning behavior *within* each guild: in particular, an individual guild member who deviates from guild norms (e.g. by setting a lower price and thereby "stealing" trade from other guild members) will be punished by exclusion.⁶⁹ The main difference relative to the single-guild case concerns collusion *between* guilds. Here we assume that if the members of one guild collectively deviate from a collusive agreement with the other guild, collusion may be restored with probability $\theta \in [0, 1]$, after a period of punishment where each active merchant gets zero profits. This assumption differs from the assumption of infinite punishments in Rotemberg and Saloner (1986); however, as they argue, "such infinite-length punishments are unlikely to be carried out in practice". Indeed, two of the reasons they give for considering infinite punishments unrealistic in the case of oligopoly seem to apply with at least equal force to the case of merchant guilds: first, "once the punishment period has begun, the oligopoly would prefer to return to a more collusive arrangement". Second, "one can think the reason why firms succeed in punishing each other at all (even though punishments are costly) is because of the anger generated when a rival cheats on the implicit agreement. This anger, as any 'irritational' emotion, may be short-lived". Moreover, one important factor hindering collusion in an oligopoly setting is of course the possibility of being caught breaking the law, and the constraints this places on firms' ability to communicate effectively and trust each other. Merchant guilds, in contrast, would not have had to worry about any such sanctions and constraints imposed by rulers.⁷⁰ We therefore follow Chang and Harrington (2007) in assuming that 'recartelization' occurs with some probability. The parameter θ is meant to capture cultural, social and ethnic differences between merchants belonging to different groups. The idea is that greater heterogeneity between the two groups will entail a lower probability of recartelization θ . Essentially, sharper differences in cultural, ethnic and social characteristics make it harder for the two groups to communicate effectively and trust one another again once a collusive agreement has been broken by one group's collective deviation, hampering future recartelization.⁷¹

In line with the earlier analysis, we rule out the unrealistic and uninteresting case of 'single member' guilds by assuming:

⁶⁹Clearly, a merchant who undercuts his fellow guild members and is expelled from his guild will not be welcomed in the other guild.

⁷⁰In our model, the ruler will be simply indifferent about collusion among merchant guilds, once the ex-ante fees have been paid. If we introduced capital constraints and/or imperfect commitment, implying the need for ex-post transfers from the guilds to the ruler, recartelization after a deviation would be in the ruler's interest; this would make the case for allowing recartelization, with some probability, even stronger.

⁷¹There is substantial evidence showing that ethnic diversity can make it harder to achieve trust and cooperation; see Alesina and La Ferrara (2005) for an excellent survey.

(A5) $\phi \underline{n} \geq 2$ (implying that $\inf \{\underline{n}\} = 4$, to ensure consistency between the two conditions, $\phi \underline{n} \geq 2$ and $\phi \leq 1/2$).

Finally, we also assume that the ruler cannot create a single ‘mixed’ guild, i.e., an association of merchants consisting of members of both groups. There are two main reasons for this. First, substantial heterogeneity within a guild would make it harder to generate the required cohesion and trust among its members (social capital). Second, suppose the ruler of a given polity could establish a guild of local and alien merchants. Individual deviations from guild norms would then be punished by exclusion from this guild. However, this might be a rather weak punishment for alien merchants, who could always continue to trade in other polities, and back home. In contrast, if alien merchants from any given polity belonged to a foreign ‘branch’ of the local merchant guild in their polity of origin, exclusion would be a much more serious punishment, since it would apply both at home and abroad. Indeed, as discussed in Section 2, such foreign branches of local merchant guilds were common – mixed guilds were not.

5.1. Individual versus collective incentive constraints

We can now study individual and group behavior when two guilds are granted recognition and privileges. In this case, two types of incentive constraint must be satisfied for collusion to be feasible. First, collusion needs to be incentive compatible from each individual’s point of view: that is, given that the two guilds decide to cartelize, none of their members should find it profitable to deviate from such a ‘group’ strategy. It is straightforward to check that this individual incentive constraint is the same as in the case of a single guild analyzed above and is given by:

$$\pi \leq \frac{\delta \mathbb{E}[\pi]}{(n-1)(1-\delta)}, \quad (5.1)$$

where $n = n_1 + n_2$ now defines the total number of merchants active in the market and belonging to both groups, with $n_1 = n_2 = n/2$ because of symmetry. As before, this constraint implies that collusion can be enforced if the total number of active merchants is not too large, if merchants are patient enough, and if expected market profitability is sufficiently high.

However, we must now also make sure that both guilds find it profitable to collude: i.e., that collective deviations (at the guild level) are not profitable. Since we have assumed that following a collective deviation the two guilds recartelize with an exogenous

probability θ after a period of punishment, this incentive constraint can be written as:

$$\frac{\pi}{2} + \frac{\delta \mathbb{E}[\pi]}{2(1-\delta)} \geq \pi + \delta^2 \theta \frac{\mathbb{E}[\pi]}{2(1-\delta)}. \quad (5.2)$$

The left-hand-side of this inequality captures the (intertemporal) collective gain that each subgroup of merchants can obtain from a collusive agreement when two guilds are endowed with privileges. Its right-hand-side measures the gain from a collective deviation: the first term is the guild's spot gain from a deviation today, while the second term represents the expected gains from future recartelization, which materializes with probability θ one period after the punishment phase. Rewriting the condition as:

$$\pi \leq (1 - \delta\theta) \frac{\delta \mathbb{E}[\pi]}{1 - \delta}, \quad (5.3)$$

makes it easy to see that the collective incentive constraint becomes tighter the larger is the probability of recartelization θ : if future recartelization is more likely, collective deviations become more attractive. Note that since this condition does not depend on the number of active merchants, only the discount factor δ affects simultaneously (5.1) and (5.3).

The analysis proceeds along similar lines to the one-guild case of Section 4. If neither incentive constraint binds, efficient collusion is feasible. When one of the two constraints binds, collusion can only be sustained if merchants price below the monopoly level. In particular, the incentive constraints (5.1) and (5.2) yield two thresholds, call them π^* and $\hat{\pi}$ respectively, such that for all states of nature satisfying $\pi > \min\{\pi^*, \hat{\pi}\}$, merchants price below the monopoly level, and each guild gains a total profit $\pi_j = \min\{\pi^*, \hat{\pi}\}/2$. On the other hand, collusion is fully efficient for all $\pi < \min\{\pi^*, \hat{\pi}\}$. In these states the two guilds obtain a profit equal to $\pi_j = \pi/2$.

Two cases must then be distinguished. First, if $\min\{\pi^*, \hat{\pi}\} = \pi^*$, the analysis follows the same lines as that developed in the previous section. Hence, for any pair (δ, n) one has

$$\pi^*(\delta, n) = \frac{2(1 - n(1 - \delta))}{\delta}.$$

If, instead, $\min\{\pi^*, \hat{\pi}\} = \hat{\pi}$ one must have:

$$\mathbb{E}[\pi] = \int_0^{\hat{\pi}} \pi dF(\pi) + (1 - F(\hat{\pi}))\hat{\pi},$$

where, for any pair $(\delta, \theta) \in [0, 1]^2$, the cut-off $\hat{\pi}$ solves:

$$\hat{\pi}(\delta, \theta) = \frac{\delta(1 - \delta\theta)}{1 - \delta} \left\{ \int_0^{\hat{\pi}} \pi dF(\pi) + (1 - F(\hat{\pi}))\hat{\pi} \right\}.$$

Using the uniform specification over the unit support, simple integration yields:

$$\hat{\pi}(\delta, \theta) = \frac{2(2\delta(1 - \theta) - 1)}{\delta(1 - \delta\theta)}.$$

Which of these cut-off values is lower depends on the discount factor, the total number of guild members, and the probability of recartelization. Note first that since $\pi^*(\delta, n)$ is decreasing in n , an exclusion result similar to the one illustrated in Lemma 1 obtains:

Lemma 2: *If privileges are recognized to both subgroups of merchants, there is an upper bound on the number of members for each guild; that is,*

$$n_j < \frac{\bar{n}(\delta)}{2} = \frac{1}{2(1 - \delta)} \quad \forall j = 1, 2.$$

As before, in order to sustain collusion, the size of each guild cannot be too large. At the same time, it cannot be too small, because of the need to monitor effectively and prevent unauthorized trade (and in particular, prevent excluded merchants from undercutting guild members). In addition, we now have a collective incentive constraint that can only be satisfied if the probability of recartelization is not too large. Let

$$\bar{\theta}(\delta) = \frac{2\delta - 1}{2\delta}$$

be the value of θ which solves $\hat{\pi}(\delta, \theta) = 0$, and

$$\bar{n}(\delta, \phi) = \frac{1}{2\phi(1 - \delta)}.$$

be the value of \underline{n} which solves $2\phi\underline{n} = \bar{n}(\delta)$, respectively. Moreover, denote by $\delta^*(\phi)$ the discount factor solving $\bar{n}(\delta, \phi) = \inf\{\underline{n}\}$. The following result provides the necessary conditions for the ruler to be able to grant recognition to both subgroups of merchants.

Lemma 3: *The ruler will consider recognizing two guilds if, and only if, merchants are sufficiently patient, $\delta \geq \max\{1/2, \delta^*(\phi)\}$, the probability of recartelization is not too large, $\theta < \bar{\theta}(\delta)$, and the merchants' monitoring technology is not too inefficient, $\underline{n} < \bar{n}(\delta, \phi)$.*

This result provides necessary conditions for two guilds to be a viable option for the

ruler. It captures the simple economic intuitions discussed above. First, for a ruler to find it worthwhile to grant recognition to both subgroups of merchants these must be sufficiently patient; moreover, their ability to prevent excluded merchants from trading and undercutting the cartel must be sufficiently effective, that is $\underline{n} < \bar{n}(\delta, \phi)$. Finally, the probability of recartelization must be sufficiently small, so that collective deviations are not too attractive, $\theta < \bar{\theta}(\delta)$. The idea here is that if guilds anticipate that the likelihood of recartelization is sufficiently large in any subgame following a collective deviation, the temptation to break the cartel will be so strong as to make it impossible to construct equilibria where collusion can be sustained.

Note that for $\phi \leq 1/2$ one has $\bar{n}(\delta, \phi) \geq \bar{n}(\delta)$, with equality only at $\phi = 1/2$. Therefore, since $\bar{n}(\delta, \phi)$ is decreasing in ϕ , when $\phi < 1/2$ there will be cases in which recognizing a single guild is not feasible, while it is possible to recognize both.

We now study the conditions under which collusion is fully efficient with two guilds. Let

$$n^{**}(\delta, \phi) = \frac{2 - \delta}{4\phi(1 - \delta)}$$

be the value of \underline{n} which solves $2\phi\underline{n} = n^*(\delta)$, and

$$\theta^*(\delta) = \frac{3\delta - 2}{\delta(4 - \delta)}$$

denote the value of θ such that $\hat{\pi}(\delta, \theta) = 1$. Moreover, let $\delta^{**}(\phi)$ be the value of δ such that $n^{**}(\delta, \phi) = \inf\{\underline{n}\}$. We have:

Proposition 4: *Granting recognition and privileges to both subgroups of merchants makes it possible to sustain efficient collusion in all states of nature if, and only if, $\delta \geq \max\{\delta^{**}(\phi), 2/3\}$, $\underline{n} \leq n^{**}(\delta, \phi)$ and $\theta \leq \theta^*(\delta)$.*

The intuition underlying this result is similar to that provided for Proposition 2, with the additional requirement that in the case of multiple guilds the collective incentive constraint must also be satisfied in all states of nature (for collusion to be fully efficient). In particular, granting recognition to both subgroups of merchants allows the two guilds to jointly implement monopoly profits in the region of parameters where both the individual and collective incentive constraints are satisfied in all states of nature. This is true when: (i) the discount factor is large enough so as to satisfy both constraints (5.1) and (5.2); (ii) the probability of recartelization is sufficiently small so as make collective deviations unattractive; and (iii) the guilds' ability to prevent excluded merchants from undercutting the cartel is sufficiently high.

This clearly illustrates the key potential *cost and benefit* of recognizing two guilds instead of one, when it comes to implementing efficient collusion: the cost is that fully efficient collusion may be harder to sustain - in particular, we have an additional condition that has to be satisfied, $\theta \leq \theta^*(\delta)$. When the probability of recartelization is too high, efficient collusion is not feasible with two guilds, whereas it may be feasible with a single guild. When the probability of recartelization is not too large, on the other hand, efficient collusion may be feasible with two guilds but not with one. The reason is that, in the presence of economies of scale in monitoring, it may be possible to satisfy the minimum size constraints when two guilds are recognized but not when only one guild is recognized: for $\phi < 1/2$, $n^{**}(\delta, \phi) > n^*(\delta)$. This trade-off will be important in determining the ruler's optimal strategy; in addition, we will need to take into account what happens when fully efficient collusion is not feasible - but some collusion can nevertheless be sustained. We do this in the following section.

5.2. The ruler's optimal strategy with multiple guilds

The previous section described the individual and collective incentive constraints that need to be satisfied for collusion to be enforceable when two guilds are recognized. We can now turn to the ruler's optimization program for this case. Let

$$W_j(\delta, n, \theta) = \frac{\int_0^{\min\{\hat{\pi}(\cdot), \pi^*(\cdot)\}} \pi dF(\pi) + (1 - F(\min\{\hat{\pi}(\cdot), \pi^*(\cdot)\})) \min\{\hat{\pi}(\cdot), \pi^*(\cdot)\}}{2(1 - \delta)}, \quad (5.4)$$

be guild j 's expected intertemporal profit, and denote by $U_j(\delta, n, \theta, R_j) = W_j(\delta, n, \theta) - R_j$ its ex ante utility net of the royalty fee. Our objective in this section is to study the trade-off between recognizing one or two guilds due to factors other than bargaining power. In particular, it may be that if there are only two guilds and the ruler wants to recognize both, the guilds have some bargaining power: the ruler cannot simply force them to bid competitively in an auction (as in the single-guild case). However, we are only considering the case with two (as opposed to multiple) guilds for simplicity, and if, for example, we had three guilds and the ruler could take full advantage of possible economies of scale in monitoring by recognizing just two guilds, an auction with competitive bidding would again be feasible. Thus it seems reasonable, and more interesting, to abstract from *ad hoc* differences in bargaining power and assume that the ruler has full bargaining power in both cases. In the case where he recognizes two guilds, the ruler will then choose R_j and n so as to solve the following optimization program:

$$\hat{\mathcal{P}} : \begin{cases} \max_{(R_j, n)_{j \in \{1,2\}} \in \mathbb{R}_+^2 \times [2\phi \underline{n}, \bar{n}(\delta)]} \sum_{j \in \{1,2\}} R_j \\ \text{s.t.} \\ U_j(\delta, n, \theta, R_j) \geq 0 \quad \forall j \in \{1, 2\}. \end{cases}$$

In a symmetric equilibrium the ruler's optimal (total) transfer, which defines his profit, is given by:

$$V^{**}(\delta, n, \theta) = 2R^{**}(\delta, n, \theta) = 2W(\delta, n, \theta).$$

Since the ruler extracts all the surplus from both guilds, his profit is equal to the total market surplus.

We can now study the ruler's optimal strategy. To make the exposition easier to follow, it is useful to distinguish and study in turn three parameter regions.

5.2.1. Tax collector

First, it is immediate to see that when δ is sufficiently small or \underline{n} is very large the only feasible action for the ruler is to hire a tax-collector. In this parameter region, it is not possible to satisfy the minimum size requirements *and* sustain collusion, whether one guild is recognized (Lemma 1) or two guilds (Proposition 3):

Proposition 5: *If $\delta < \max\{1/2, \delta^*(\phi)\}$ or $\underline{n} \geq \bar{n}(\delta, \phi)$ the ruler can only hire a tax-collector.*

5.2.2. Tax collector or two guilds?

The second case obtains when $\underline{n} \in \Gamma(\delta, \phi) = [\bar{n}(\delta), \bar{n}(\delta, \phi)]$. In this parameter region, the ruler's action space includes two options: hiring a tax-collector or granting recognition to both guilds. In particular, as observed before, when $\phi < 1/2$ the upper-bound on a single guild size, $\bar{n}(\delta)$, is lower than that required when two guilds are recognized, $\bar{n}(\delta, \phi)$, so that the interval $\Gamma(\delta, \phi)$ is non-empty. As a consequence, for all $\underline{n} \in \Gamma(\delta, \phi)$ it is not possible to sustain collusion with one guild, whereas collusion can be sustained if both guilds are granted recognition and privileges. The ruler's best action is to recognize two guilds if

$$2W(\delta, n, \theta) \geq \hat{V}(\delta, c),$$

while a tax-collector will be hired otherwise. The choice will therefore depend on the discount factor δ , the tax-collector's monitoring cost c , the probability of recartelization θ

and the guild's minimum size requirement ϕn . Of course, if the probability of recartelization is too large, i.e., if $\theta > \bar{\theta}(\delta)$, collusion can never be efficient with two guilds and hence the ruler's optimal choice is to hire a tax-collector. If $\theta < \bar{\theta}(\delta)$, on the other hand, the ruler will recognize the two guilds when the tax-collector's monitoring technology is not sufficiently effective, that is, c is relatively large. Otherwise, he will hire a tax collector. The result can be summarized as follows.

Proposition 6: *Assume $\phi < 1/2$, $\delta \geq \max\{1/2, \delta^*(\phi)\}$ and $\theta < \bar{\theta}(\delta)$, then for all $\underline{n} \in \Gamma(\delta, \phi)$ there exists a positive function $c^{**}(\delta, \underline{n}, \theta, \phi)$ such that if $c \geq c^{**}(\delta, \underline{n}, \theta, \phi)$ granting recognition to both guilds is optimal for the ruler. A tax-collector is hired otherwise.*

This result is similar in spirit to Proposition 3. The difference is that now the ruler may face a trade-off between hiring a tax-collector and granting recognition to two guilds (rather than one as in the earlier analysis). As already pointed out, hiring a tax-collector delivers efficient collusion, but at a price, because of the agent's cost of monitoring trade. When two guilds are granted recognition, on the other hand, there are no monitoring costs but collusion is not necessarily efficient because of the interplay between the individual and the collective incentive constraints (5.1)-(5.3). As a consequence, the ruler will prefer to grant recognition to both guilds when the tax-collector's monitoring cost is higher than the loss of profits due to inefficient collusion.

5.2.3. One guild, two guilds or tax collector?

Finally, perhaps the most interesting case occurs when $\underline{n} < \bar{n}(\delta)$. In this parameter region, the ruler's action set is the largest possible and it includes the option of granting privileges to a single guild or to both, as well as the possibility of delegating taxation to an agent. This is a natural case to study if one is interested in describing the basic trade-offs faced by a ruler who can decide whether to grant recognition only to a local guild or to accommodate also foreign merchants organized as an independent association. Leaving aside the option of hiring a tax-collector, the relevant trade-off is the following: granting recognition to a single guild is better for the purpose of enhancing collusion, as in this case there is no collective incentive constraint to satisfy, which is instead a key requirement when recognition is extended to multiple guilds. However, in the presence of scale economies in monitoring (i.e., when $\phi < 1/2$), granting recognition to two guilds may enable the ruler to relax the individual incentive constraint, relative to the case in which the single-guild regime entails minimum size. The ruler will then choose to recognize two guilds if:

$$2W(\delta, n, \theta) \geq \max_{j \in \{1,2\}} \left\{ \hat{V}(\delta, c), W(\delta, n_j) \right\},$$

where $W(\delta, n_j)$ is the surplus he would obtain when dealing only with guild j . The next proposition summarizes the result:

Proposition 7: *Assume $\delta > \max\{1/2, \delta^*(\phi)\}$. Then, the following properties are satisfied:*

- (i) *The ruler weakly prefers to recognize one guild instead of two if one of the following conditions hold: $\phi = 1/2$, $\underline{n} < n^*(\delta)$, or $\theta > \bar{\theta}(\delta)$. In this parameter region his choice will be as described in Proposition 3.*
- (ii) *Assume $\phi < 1/2$ and $\underline{n} \in [n^*(\delta), \bar{n}(\delta)]$. Then, if $c \geq c^*(\delta, \underline{n})$ there exists a function $\hat{\theta}(\delta, \underline{n}, \phi) < \bar{\theta}(\delta)$ such that for all $\theta \leq \hat{\theta}(\delta, \underline{n}, \phi)$ the ruler prefers to recognize both guilds. Otherwise he prefers to have a single guild.*
- (iii) *Assume $\phi < 1/2$ and $\underline{n} \in [n^*(\delta), \bar{n}(\delta)]$. Then, if $c < c^*(\delta, \underline{n})$ the same qualitative results as in Proposition 6 obtain; that is, the ruler prefers to recognize two guilds only if c is sufficiently large. Otherwise he hires a tax-collector.*

The intuition for this result is simple. Clearly, the ruler prefers to recognize a single guild, rather than multiple guilds, in the following cases: (a) if having two guilds does not bring any efficiency gain in monitoring; (b) if merchants' monitoring ability is sufficient to deliver efficient collusion with a single guild; (c) if the probability of recartelization is so large that efficient collusion cannot be sustained with multiple guilds. The trade-off between a single guild and a tax collector is then the same as in the earlier analysis.

On the other hand, when the single-guild regime generates inefficient collusion because the minimum size constraint is binding, recognizing two guilds may increase efficiency (of collusion), as long as there are economies of scale in monitoring, and the probability of recartelization is not too high. The trade-off between two guilds and a tax collector is then the same as in Proposition 6.

6. Merchant guilds: theory and evidence

In this section we review the main implications of our theory and relate them to the available historical evidence. To begin with, the theoretical analysis developed in Section 4 is able to account for the emergence of merchant guilds as a mechanism to sustain collusion among merchants, and between merchants and rulers, bypassing the need for costly monitoring by other parties hired as tax-collectors. In doing so, the model also provides a rationale for the observed privileges granted by rulers to merchant guilds (notably the

right to exclude non-members from trade, as well as members who have deviated from guild rules), the transfers made by guilds to rulers, the tax exemptions obtained by guilds, and the membership restrictions imposed by guilds. All these implications of the model are amply borne out by the historical evidence reviewed in Section 2.

In addition, our theory sheds light on a number of other historical observations. First, it clearly follows from our analysis that rulers had an interest in supporting *merchant* guilds, but not also separate guilds of producers/suppliers, such as producers of agricultural commodities supplied to merchants and then sold by the latter to consumers. The emergence of such agricultural guilds would have given rise to inefficiencies associated with double marginalization - and indeed, medieval rulers did not support the establishment of such guilds. A similar argument applies to craft guilds, except of course where craftsmen were also merchants, trading the commodities they produced. Again, the historical evidence is consistent with this implication of our theory: craftsmen often belonged to guilds that combined production and trade⁷², thereby obtaining recognition and privileges from rulers; on the other hand, they struggled to obtain support from rulers when their interests were in conflict with those of merchant guilds⁷³.

Second, another key feature of our analysis is to clearly identify the trade-offs faced by a ruler choosing between tax-farming, a single merchant guild or multiple merchant guilds. As discussed in Section 2, merchant guilds were ubiquitous in medieval Europe, suggesting that the trade-off between tax-farming and merchant guilds was typically favorable to guilds. In terms of our model, this would be the predicted outcome when merchants are sufficiently patient, and when merchant guilds' monitoring ability (which depends on cohesion and trust among members, and their resulting ability to overcome free-riding incentives and cooperate to monitor effectively) is sufficiently high. The historical evidence on merchant guilds' social capital and monitoring ability suggests that this was very much the case: merchant guilds were social groups or networks whose members participated in a variety of social and religious activities together; they held regular assemblies and feasts; when abroad, they lived in their own quarters of foreign cities and interacted closely. Trust was clearly perceived to be very important: symbolically, English merchant guilds required "an oath of fealty to the fraternity" from new members⁷⁴. In practice, close and repeated interaction facilitated monitoring and the exchange of information; this, together with the establishment of a variety of norms and sanctions for members who breached

⁷²See Luchaire (1902), p.241, for some French examples, and Gross (1890), p.107, for some English ones. In some cases, craftsmen were, to begin with, highly dependent on merchants, who supplied them with necessary raw materials from other polities, and sold the finished products in international markets. The relationship between the two was then dominated by merchants (see Racine (1979), pp. 297, 714-715).

⁷³See for example Dilcher (1984), p.71 and Frölich (1934), pp. 36-37.

⁷⁴Gross (1890), p.29 (quotation).

them or helped others to breach them, helped to sustain trust. The social capital thereby generated by merchant guilds was, according to our model, a key factor influencing rulers' support for the guilds over tax-farming.

Just as importantly, the analysis developed in Section 5 sheds light on the choice between establishing one or multiple guilds. In particular, we have shown that recognition of multiple guilds should have been more likely when (a) each guild could target its monitoring so as to exploit an informational comparative advantage, generating efficiency gains in monitoring, and (b) the probability of different guilds recartelizing after a deviation was sufficiently low. In small towns with low levels of international trade, the scope for efficiency gains in monitoring by multiple guilds must have been quite limited. In these cases, as our model would predict, a single merchant guild was typically recognized and granted privileges by the ruler. In contrast, rulers of important international trade centers (e.g. Bruges, Constantinople) generally recognized a number of guilds. Each guild's membership was normally quite homogeneous in terms of nationality, ethnicity and culture (e.g. Catalan, Genoese, Pisan, Venetian...) and very cohesive, with correspondingly high abilities to monitor potential trading activities by non-members from their polities of origin. Thus they satisfied condition (a) above. Cultural and ethnic *homogeneity within* guilds also implied significant *heterogeneity between* guilds, suggesting that condition (b) was satisfied as well; indeed, there is plenty of evidence of conflict, sometimes violent, between guilds.⁷⁵

Finally, it is interesting to consider the implications of our analysis for city-states ruled by merchant oligarchies. These represent, in a sense, an extreme form of collusion between ruler and local merchants. Does this make local merchant guilds unnecessary? Our model suggests that the answer is no: there is still a need for norms to sustain collusive profits and extract maximal surplus from trade, and monitoring to ensure enforcement of those norms. Indeed, most city-states whose governments were dominated by mercantile interests possessed local merchant guilds. A notable exception was Venice, a special case in many respects. Venice gained its autonomy (from the Byzantine emperor) much earlier than other Italian city-states. Long before the tenth century, its ruling class owned some land but was also involved in maritime trade: "As early as 829 the will of Venetian Doge Justinian Partecipazio mentioned among his assets a substantial sum (1,200 silver pounds) invested in oversea commercial ventures" (Lopez (1971), p.63). Thus mercantile interests played a key role in Venice from very early on. Indeed, it is perhaps the historical example that comes closest to 'perfect' collusion between ruler and local merchants. Yet it did not

⁷⁵On this see Abulafia (1978, 1986); Bahr (1911); Daenell (1905); De Roover (1963); Dollinger (1970); Greif et al. (1994); Lloyd (1991); Postan (1973); Pryor (2000); Reyerson (2000); Schütt (1980) and Smith (1940).

possess a local merchant guild. The puzzle is easily explained by observing that the city government itself took on the tasks that local merchant guilds performed elsewhere: it established maritime regulations and withdrew the right to trade from any Venetian merchant who breached them, or who breached financial contracts; it imposed fines for smaller breaches of mercantile rules; it monitored closely Venetian merchant colonies abroad⁷⁶. It also excluded alien merchants from many types of commercial activity in Venice and required them to trade, when allowed to, only through Venetian brokers⁷⁷. Thus the norms that allowed extraction of surplus from trade by merchant guilds elsewhere were also adopted in Venice.

7. Concluding remarks

Understanding why merchant guilds emerged, and the role they played, matters not only for historical interest, but also for current debates over institutions and social capital. Merchant guilds have been widely regarded as an example of how social capital can benefit whole societies and economies. We have revisited the rationale for the emergence of merchant guilds, examining the implications of their comparative advantage in monitoring and hence their role as an efficient mechanism to sustain collusion. Our theory suggests that merchant guilds did indeed generate some efficiency gains, but these gains benefited only part of the population (in particular, rulers and gilded merchants). By modeling explicitly the dynamic incentives of individual merchants, and analysing the role of the guilds' social capital, we have been able to shed light on a number of important historical observations. Moreover, our analysis provides a theoretical framework capable of accounting for the basic trade-offs involved when a polity's ruler had to choose between granting recognition to a single or multiple guilds. This helps us to understand the observed distribution of guilds, and provides a rationale for the establishment of both *local* and *alien* merchant guilds, consistent with the historical evidence.

References

- [1] Abulafia, D., "The rise of Aragon-Catalonia", in Abulafia, David (ed.), *The new Cambridge medieval history*, vol. V: *c. 1198 - c. 1300*, (Cambridge, 2000), 644-67.

⁷⁶Ashtor (1983), pp.68-9, 78, 411; Choroskevic (1996), pp.84-6; Gonzales de Lara (2001), pp.25-6.

⁷⁷Ashtor (1983), p.71; Mueller (1997), pp.89-90, 266-7.

- [2] Abulafia, D., “East and west: comments on the commerce of the city of Ancona in the Middle Ages”, in Ghezzi, M. P. (ed.) *Città e sistema adriatico alla fine del medioevo*, (Padua, 1997), 49-66.
- [3] Abulafia, D., “The Anconitan privileges in the Kingdom of Jerusalem and the Levant trade of Ancona”, in Airaldi, G., and Kedar, B. Z. (eds.), *I comuni italiani nel regno crociato di Gerusalemme*, (Genova, 1986), 525-70.
- [4] Abulafia, D., “Pisan commercial colonies and consulates in twelfth-century Sicily”, *English Historical Review*, 93, (1978), 68-81.
- [5] Alesina, A., and La Ferrara, E., “Ethnic diversity and economic performance”, *Journal of Economic Literature*, 43(3), (2005), 762-800.
- [6] Alesina, A., and La Ferrara, E., “Who trusts others?”, *Journal of Public Economics*, 85(2), (2002), 207-234.
- [7] Alesina, A., and La Ferrara, E., “Participation in heterogeneous communities”, *Quarterly Journal of Economics*, 115(3), (2000), 847-904.
- [8] Ashtor, E., *Levant trade in the later Middle Ages*, (Princeton, NJ, 1983).
- [9] Bahr, K., *Handel und Verkehr der Deutschen Hanse in Flandern während des vierzehnten Jahrhunderts* (Leipzig, 1911).
- [10] Baldwin, J., *The Government of Philip Augustus*, (Berkeley, 1986).
- [11] Bardhan, P., “The nature of institutional impediments to economic development”, (March 3, 1996), Center for International and Development Economics Research Paper C96-066. <http://repositories.cdlib.org/iber/cider/>.
- [12] Bateson, M. (ed.), *Records of the Borough of Leicester*, (London, 1899), vol. I.
- [13] Bernard, J., “Trade and finance in the Middle Ages: 900 to 1500”, in Cipolla, C. M., and Borchardt, Knut (eds.), *The Fontana economic history of Europe*, vol. I: *The Middle Ages*, (London, 1972), 274-329.
- [14] Bisson, T. N., *Fiscal accounts of Catalonia under the early Count-Kings (1151-1213)*, (Berkeley, 1984).
- [15] Blom, G. A., “Der Ursprung der Gilden in Norwegen und ihre Entwicklung in den Städten während des Mittelalters”, in Friedland, Klaus (ed.), *Gilde und Korporation in den nordeuropäischen Städten des späten Mittelalters*, (Köln/Wien, 1984), 5-28.

- [16] Bourdieu, P., “The forms of capital”, in Richardson, J. G. (ed.) *Handbook of Theory and Research for the Sociology of Education*, (Westport, CT, 1986).
- [17] Brewer, J., *The sinews of power: war, money and the English state, 1688-1783* (London, 1989).
- [18] Butler, J., P. Giuliano and L. Guiso, "The Right amount of Trust", (2009), mimeo.
- [19] Chang, M., and Harrington J., “Modelling the Birth and Death of Cartels with an Application to Evaluating Antitrust Policy,” (2007), mimeo.
- [20] Chen, Z., and Rey, P., “On the Design of Leniency Programs,” (2007), IDEI Working Paper, #452.
- [21] Choroskevic, A. L., “Der deutsche Hof in Novgorod und die deutsche Herberge in Venedig im 13./14. Jahrhundert. Eine vergleichende Vorstudie”, in Pelc, Ortwin, and Pickhan, Gertrud (eds.), *Zwischen Lübeck und Novgorod*, (Lüneburg, 1996), 67-87.
- [22] Coleman, J. S., *Foundations of social theory*, (Cambridge, MA, 1990).
- [23] Daenell, E. R., *Die Blütezeit der deutschen Hanse*, 2 vols. (Berlin, 1905).
- [24] Dasgupta, P., and Serageldin, I. (eds.), *Social capital: a multifaceted perspective*, (Washington, 2000).
- [25] De Roover, R., “The organization of trade”, in Postan, M. M., Rich, E. E., and Miller, Edward (eds.), *The Cambridge economic history of Europe*, vol. III: *Economic organization and policies in the Middle Ages*, (Cambridge, 1963), 42-118.
- [26] De Roover, R., *Money, banking and credit in medieval Bruges*, (Cambridge, MA, 1948).
- [27] Dessí, R., and Ogilvie, S., “Social capital and collusion: the case of merchant guilds”, CES-Ifo Working Paper (2003), University of Munich.
- [28] Dilcher, G., “Die genossenschaftliche Struktur von Gilden und Zünften”, in Schweineköper, Berent (ed.), *Gilden und Zünfte* (Sigmaringen, 1985), 71-112.
- [29] Dilcher, G., “Personale und lokale Strukturen kaufmännischen Rechts als Vorformen genossenschaftlichen Stadtrechts”, in Friedlan, Klaus (ed.), *Gilde und Korporation in den nordeuropäischen Städten des späten Mittelalters*, (Köln/Wien, 1984), 65-78.
- [30] Dollinger, P., *The German Hansa*, (London, 1970).

- [31] Ehbrecht, W., “Beiträge und Überlegungen zu Gilden im nordwestlichen Deutschland, (vornehmlich im 13. Jahrhundert)”, in Schwineköper, Berent (ed.), *Gilden und Zünfte*, (Sigmaringen, 1985), 413-50.
- [32] Epstein, S. A., “Urban society”, in Abulafia, David (ed.), *The new Cambridge medieval history*, vol. V: *c. 1198 - c. 1300*, (Cambridge, 2000), 26-37.
- [33] Freshfield, E. H., *Roman Law in the later Roman Empire: Byzantine guilds, professional and commercial. Ordinances of Leo VI c. 895 from the Book of the Eparch*, (Cambridge, 1938).
- [34] Glaeser, E. L., Laibson, D., and Sacerdote, B., “An economic approach to social capital”, *Economic Journal*, 112, (2002), 437-58.
- [35] Glaeser, E. L., Laibson, D., Scheinkman, J. A., and Soutter, C. L., “Measuring trust”, *Quarterly Journal of Economics*, (2000), 811-846.
- [36] Greif, A., Milgrom, P., and Weingast, B. R., “Coordination, commitment, and enforcement: the case of the merchant guild”, *Journal of Political Economy*, 102 (1994), 745-76.
- [37] Gross, C., *The Gild Merchant: A Contribution to British Municipal History*, Volume I (Oxford, 1890).
- [38] Guiso, L., Sapienza, P., and Zingales, L., “The role of social capital in financial development”, *American Economic Review*, 94, (2004), 526-556.
- [39] Guiso, L., Sapienza, P., and Zingales, L., “Trusting the stock market”, *Journal of Finance*, 63(6), (2008), 2557-2600.
- [40] Guiso, L., Sapienza, P., and Zingales, L., “Cultural biases in economic exchange?”, *Quarterly Journal of Economics*, forthcoming.
- [41] Hibbert, A. B., “The economic policies of towns”, in Postan, M. Michael, Rich, E. E., and Miller, Edward (eds.), *The Cambridge economic history of Europe*, vol. III, *Economic organization and policies in the Middle Ages*, (Cambridge, 1963), 157-229.
- [42] Hilton, R. H., *English and French Towns in Feudal Society*, (Cambridge, 1992).
- [43] Hoffman, E., “Beiträge zur Geschichte der Stadt Schleswig und des westlichen Ostseeraums im 12. und 13. Jahrhundert”, *Zeitschrift der Gesellschaft für Schleswig-Holsteinische Geschichte*, 105 (1980), 27-76.

- [44] Irsigler, F., “Zur Problematik der Gilde- und Zunftterminologie”, in Schweineköper, Berent (ed.), *Gilden und Zünfte*, (Sigmaringen, 1985), 53-70.
- [45] Johanek, P., “Merchants, markets and towns”, in Reuter, Timothy (ed.), *The new Cambridge medieval history*, Vol. 3: *c. 900-c. 1024*, (Cambridge, 1999), 64-94.
- [46] Keefer, P., and Knack, S., “Does social capital have an economy payoff? A cross-country investigation”, *Quarterly Journal of Economics*, 112(4), (1997), 1251-1288.
- [47] Kemble, J. M., *The Saxons in England: a history of the English commonwealth till the period of the Norman conquest*, (London, 1876).
- [48] Knack, S., and Zack, P., “Trust and growth”, *Economic Journal*, 111(470), (2001), 295-321.
- [49] Kohn, M., “Merchant associations in pre-industrial Europe”, Working Paper 03-11, Dartmouth College (2003).
- [50] Kuske, B., “Der Kölner Stapel und seine Zusammenhänge als wirtschaftspolitisches Beispiel”, *Jahrbuch des Kölnischen Geschichtsvereins*, 21, (1939), 1-46.
- [51] Laiou, A. E., “The Byzantine empire in the fourteenth century”, in Jones, Michael (ed.), *The new Cambridge medieval history*, vol. VI: *c. 1300 - c. 1415*, (Cambridge, 2000), 795-824.
- [52] La Porta, R., Lopez de Silanes, F., Shleifer, A., and Vishny, R., “Trust in large organizations”, *American Economic Review Papers and Proceedings*, 87(2), (1997), 333-338.
- [53] Leguay, J. P., “Urban life”, in Jones, Michael (ed.), *The new Cambridge medieval history*, vol. VI: *c. 1300 - c. 1415*, (Cambridge, 2000), 102-25.
- [54] Lin, N., *Social capital: a theory of social structure and action*, (Cambridge, 2001).
- [55] Lloyd, T. H., *England and the German Hanse, 1157-1611: a study of their trade and commercial diplomacy*, (Cambridge, 1991).
- [56] Lopez, R. S., and Raymond, I. W., *Medieval Trade in the Mediterranean World: Illustrative Documents*, New York, Columbia University Press, (2001).
- [57] Luchaire, A., “Philippe Auguste et son temps. 1137-1226”, in Lavissee, Ernest (ed.), *Histoire de France des Origines à la Révolution*, (1902).

- [58] Lyon, B., and Verhulst, A., *Medieval Finance*, (Providence, 1967).
- [59] Moore, E. W., *The fairs of medieval England: an introductory study*, (Toronto, 1985).
- [60] Origo, I., *The merchant of Prato, Francesco di Marco Datini, 1335-1410*, (Boston, 1986).
- [61] Pertz, G. H. (ed.), *Monumenta Germaniae Historica, Scriptorum Tomus IV*, (Hanover, 1925).
- [62] Planitz, H., “Kaufmannsgilde und städtische Eidgenossenschaft in niederfränkischen Städten im 11. und 12. Jahrhundert”, *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte, germanistische Abteilung*, 60 (1940), 1-116.
- [63] Postan, M. M., “The economic and political relations of England and the Hanse from 1400 to 1475”, *Medieval trade and finance*, (Cambridge, 1973), 232-304.
- [64] Preuenier, W., “The Low Countries, 1290-1415”, in Jones, Michael (ed.), *The new Cambridge medieval history*, vol. VI: *c. 1300 - c. 1415*, (Cambridge, 2000), 570-94.
- [65] Pryor, J. H., “The Maritime Republics”, in Abulafia, David (ed.), *The new Cambridge medieval history*, vol. V: *c. 1198 - c. 1300*, (Cambridge, 2000), 419-46.
- [66] Putnam, R. D., *Bowling alone: the collapse and revival of American community*, (New York, 2000).
- [67] Putnam, R. D., Leonardi, R., and Nanetti, R. Y., *Making democracy work: civic traditions in modern Italy*, (Princeton, NJ, 1993).
- [68] Racine, P., *Les villes d'Italie: du milieu du XIIIe siècle au milieu du XIVe siècle*, (Paris, 2004).
- [69] Racine, P., “Associations de marchands et associations de métiers en Italie de 600 à 1200”, in Schweineköper, Berent (ed.), *Gilden und Zünfte*, (Sigmaringen, 1985), 127-50.
- [70] Racine, P., *Plaisance du Xe à la fin du XIIIe siècle: essai d'histoire urbaine*, (Paris, 1979).
- [71] Raiser, M., “Informal institutions, social capital and economic transition”, in Cornia, Giovanni Andrea, and

- [72] Reyerson, K., “Commerce and communications”, in Abulafia, David (ed.), *The new Cambridge medieval history*, vol. V: *c. 1198 - c. 1300*, (Cambridge, 2000), 50-70.
- [73] Rotemberg, J. J., and Saloner, G., “A Supergame-Theoretic Model of Price Wars during Booms”, *The American Economic Review*, 76, (1986), 390-407.
- [74] Schultze, A., “Über Gästerecht und Gastgerichte in den deutschen Städten des Mittelalters”, *Historische Zeitschrift*, 101, (1908), 473-528.
- [75] Schulz, K., “Patriziergesellschaften und Zünfte in den mittel- und oberrheinischen Bischofsstädten”, in Schwineköper, Berent (ed.), *Gilden und Zünfte*, (Sigmaringen, 1985), 311-36.
- [76] Schulze, H. K., “Kaufmannsgilde und Stadtentstehung im mitteldeutschen Raum”, in Schwineköper, Berent (ed.), *Gilden und Zünfte*, (Sigmaringen, 1985), 377-412.
- [77] Schütt, H. F., “Gilde und Stadt”, *Zeitschrift der Gesellschaft für Schleswig-Holsteinische Geschichte*, 105 (1980), 77-136.
- [78] Smith, R. S., *The Spanish guild merchant: a history of the consulado, 1250-1700*, (Durham, NC, 1940).
- [79] Sobel, J., “Can we trust social capital?”, *Journal of Economic Literature*, 40, (2002), 139-54.
- [80] Spagnolo, G., “Social Relations and Cooperation in Organizations,” *Journal of Economic Behavior and Organization*, 38, (1999), 1-26.
- [81] Spufford, P., “Trade in fourteenth-century Europe”, in Jones, Michael (ed.), *The new Cambridge medieval history*, vol. VI: *c. 1300 - c. 1415*, (Cambridge, 2000), 155-208.
- [82] Spruyt, H. *The sovereign state and its competitors*, (Princeton, New Jersey, 1994).
- [83] Stiglitz, J., “New bridges across the chasm: institutional strategies for the transition economies” (World Bank, 8 Dec. 1999). <http://wbln0018.worldbank.org/eca/eca.nsf/>.
- [84] Störmer, W., “Vergesellschaftungsformen des Meliorats und des Handwerks in den Städten des bayerisch-österreichischen Raums”, in Schwineköper, Berent (ed.), *Gilden und Zünfte*, (Sigmaringen, 1985), 337-376.
- [85] Volckart, O., and Mangels, A., “Are the roots of the modern *lex mercatoria* really medieval?”, *Southern Economic Journal*, 65, (1999), 427-50.

[86] Webber, C., and Aaron, W., *A History of Taxation and Expenditure in the Western World*, (New York, 1986).

[87] Woodward, R. L. Jr., “Merchant guilds (*Consulados de Comercio*) in the Spanish World”, in Northrup, Cynthia (ed.), *World trade: a historical encyclopedia of economics, politics, society, and culture*, (Westport, CT, 2003).

8. Appendix

Proof of Proposition 1: The proof of this result is immediate. Indeed, given **A2**, one can easily show that it is strictly dominant for the tax-collector to pay the monitoring cost c in each period and extract all the surplus from the merchants if he expects them to collude efficiently. Moreover, if merchants expect the tax collector to monitor in each period, they will charge the monopoly price. In fact, in this case, they would be indifferent between colluding and price at the marginal cost anyway as all the surplus they get from the market is taken away by the zero-profit tax imposed by the agent. Hence, there exists a SPNE of the game where the tax collector monitors in each period and merchants collude. Notice that if $N > 1/(1 - \delta)$ this equilibrium is also unique. Indeed, in this case, a SPNE where the tax-collector does not monitor cannot exist because individual deviations will always be profitable for any equilibrium candidate where merchants get a profit lower than c . ■

Proof of Lemma 1: Part (i) can be immediately established by the definition of $\pi^*(\delta, n)$. Indeed, for this cut-off value to be positive one must have $n < \bar{n}(\delta)$. The proof of part (ii) rests, instead, on the idea that in order for the ruler to grant recognition to a merchant guild, the minimum size condition $n \geq \underline{n}$ must be compatible with the necessary condition for collusion to be enforceable at least in some state π , that is $n < \bar{n}(\delta)$. Given the definition of $\bar{n}(\delta)$, and the fact that $\inf \{\underline{n}\} = 2$, it is immediate to show that the necessary conditions for a guild to be recognized are $\delta > 1/2$ and $\underline{n} < \bar{n}(\delta)$. ■

Proof of Proposition 2: First, observe that $\pi^*(\delta, n) \geq 1$ is a necessary condition for collusion to be fully efficient in all states $\pi \in \Pi$. Hence, solving $\pi^*(\delta, n) = 1$ yields the maximum n , in the text denoted by $n^*(\delta)$, which ensures that collusion is fully efficient in the whole support Π . Then, observe that if $\underline{n} > n^*(\delta)$, granting recognition to a merchant guild is incompatible with efficient collusion, while the converse holds otherwise. Therefore, for collusion to be efficient in all states of nature one needs $\underline{n} < n^*(\delta)$. But, as we have assumed $\inf \{\underline{n}\} = 2$, a necessary condition for this to be true is that $n^*(\delta) > 2$. This inequality yields immediately $\delta > 2/3$, which provides a second necessary condition

for collusion to be efficient in all states. Moreover, in this case, the optimal guild size is such that $n \in (\underline{n}, n^*(\delta)]$ provided that this interval is non-empty. When $\underline{n} > n^*(\delta)$ the minimum size constraint binds as $\pi^*(\delta, \underline{n}) < 1$. In this case, the guild reaches its minimum size \underline{n} and collusion cannot be efficient, i.e., for all states $\pi > \pi^*(\delta, \underline{n})$ merchants need to price below the monopoly level. ■

Proof of Proposition 3: Of course, when $\delta \leq 1/2$ or $\underline{n} \geq \bar{n}(\delta)$ the ruler cannot implement the guild as in this case the minimum size requirement is incompatible with the maximum size condition implied by Lemma 1. Hence, in this region of parameters the ruler is forced to hire a tax-collector. If $\delta > 1/2$ and $\underline{n} < \bar{n}(\delta)$, instead, the ruler will always grant recognition to a merchant guild whenever this ensures efficient collusion in all states of nature, that is, if $\delta > 2/3$ and $\underline{n} \leq n^*(\delta) < \bar{n}(\delta)$ as shown in Proposition 2. Differently, when $\delta < 2/3$ or $n > n^*(\delta)$ collusion cannot be fully efficient under the guild regime. In this case, if the ruler decides to grant recognition to a guild he obtains a intertemporal expected utility equal to

$$V^*(\delta, \underline{n}) = \hat{V}(\delta, c) + \alpha \left[W(\delta, \underline{n}) - \hat{V}(\delta, c) \right],$$

where

$$\begin{aligned} W(\delta, \underline{n}) &= \frac{\int_0^{\pi^*(\delta, \underline{n})} \pi dF(\pi) + (1 - F(\pi^*(\delta, \underline{n}))) \pi^*(\delta, \underline{n})}{1 - \delta} \\ &= \frac{2(1 - \underline{n}(1 - \delta))(\underline{n} - 1)}{\delta^2}. \end{aligned}$$

Hence, granting recognition to a merchant guild is optimal if and only if $W(\delta, \underline{n}) \geq \hat{V}(\delta, c)$, that is

$$\frac{2(1 - \underline{n}(1 - \delta))(\underline{n} - 1)}{\delta^2} - \frac{\mathbb{E}[\pi] - c}{1 - \delta} \geq 0,$$

since $\mathbb{E}[\pi] = 1/2$ under the uniform specification $\tilde{\pi}^\tau \sim U[0, 1]$ the above inequality implies:

$$c \geq c^*(\underline{n}, \delta) \equiv \left[\frac{2\underline{n}(1 - \delta) - (2 - \delta)}{\delta\sqrt{2}} \right]^2. \quad (8.1)$$

Clearly, for $c < c^*(\underline{n}, \delta)$ hiring a tax-collector will be the ruler's best choice. Finally, showing that the threshold $c^*(\cdot)$ is decreasing in δ and increasing in \underline{n} within the relevant range of parameters is immediate from (8.1). ■

Proof of Lemma 2: The proof of this result follows the same logic as the proof of Lemma 1. Since the individual incentive constraint (5.1) does not change when the ruler considers granting recognition to both subgroups of merchants, in order for each

merchant not to find it optimal to deviate from a ‘grand’ collusive agreement, the total number of active merchants must be limited. Using symmetry, this implies immediately $n_j < \bar{n}(\delta)/2 = 1/2(1 - \delta)$ for all $j = 1, 2$, thus $n < \bar{n}(\delta)$. ■

Proof of Lemma 3: The proof of this result rests on the idea that recognizing privileges to both subgroups of merchants can be profitable only if this guarantees collusion to be efficient at least in some state π . For this to be true one needs to have $\hat{\pi}(\delta, \theta) \geq 0$ and $\pi^*(\delta, n) \geq 0$ for all $n \geq \underline{n}$. It is then straightforward to show that for $\theta < \bar{\theta}(\delta)$ and $\underline{n} < \bar{n}(\delta, \phi)$ these inequalities are satisfied altogether. However, $\bar{\theta}(\delta) > 0$ only if $\delta > 1/2$ and $\bar{n}(\delta, \phi) > \inf\{\underline{n}\} = 4$ only if $\delta > \delta^*(\phi)$, where

$$\delta^*(\phi) = 1 - \frac{1}{8\phi}.$$

Hence, the ruler will consider granting recognition to both guilds only if $\underline{n} < n(\delta, \phi)$, $\theta < \bar{\theta}(\delta)$ and

$$\delta > \max\{1/2, \delta^*(\phi)\}.$$

This concludes the proof. ■

Proof of Proposition 4: The proof of this result follows closely the logic of the proof of Proposition 2. For collusion to be efficient in the whole space of parameters it must be $\hat{\pi}(\delta, \theta) \geq 1$ and $\pi^*(\delta, n) \geq 1$. It can be immediately verified that these inequalities can be jointly satisfied if and only if $\delta \geq \max\{\delta^{**}(\phi), 2/3\}$, $\underline{n} \leq \bar{n}^*(\delta, \phi)$ and $\theta \leq \theta^*(\delta)$. Conversely, if one of these conditions does not hold, collusion cannot be efficient. ■

Proof of Proposition 5: The proof of this proposition is an immediate consequence of Lemma 3 and the facts that: (i) $\bar{\theta}(\delta) \leq 0$ for all $\delta \leq 1/2$, and (ii) $\bar{n}(\delta, \phi) \geq \bar{n}(\delta)$ for $\phi \leq 1/2$. ■

Proof of Proposition 6: The proof of this result follows the same lines as the proof of Proposition 3. First, it is immediate to verify that $\Gamma(\delta, \phi)$ is non-empty for $\phi < 1/2$ and that only in the region of parameters where $\delta > \max\{1/2, \delta^*(\phi)\}$, $\theta < \bar{\theta}(\delta)$ and $\underline{n} < \bar{n}(\delta, \phi)$, the ruler will consider to grant recognition to both guilds. Then, it is straightforward to show that if $c = 0$ the ruler will always prefer to hire the tax-collector as this will provide efficient collusion at no monitoring costs; while, if $c = \mathbb{E}[\pi] = 1/2$, he will strictly prefer to grant recognition to both subgroups of merchants. Therefore, a simple continuity argument implies that there must be a function $c^{**}(\delta, \underline{n}, \theta, \phi)$ such that for all $c < c^{**}(\delta, \underline{n}, \theta, \phi)$ a tax-collector will be hired, differently privileges will be granted to both subgroups of merchants. ■

Proof of Proposition 7: First, observe that restricting attention to cases where $\delta > \max\{1/2, \delta^*(\phi)\}$ implies that the regime with two guilds is feasible. In the opposite case, following the logic of Proposition 3, the result would be to recognize one guild or to hire the tax collector.

In order to prove part (i) it is useful to remember that in the region of parameters where $\theta \geq \bar{\theta}(\delta)$ it must be true that $\hat{\pi}(\delta, \theta) \leq 0$, so that the ruler will always prefer to deal with one guild. Indeed, in this case, granting recognition to both subgroups renders collusion impossible via the collective incentive constraint. Moreover, the same result holds if $\phi = 1/2$. In this case, there is no difference between the individual incentive constraint with one or two guilds, i.e., having two guilds does not bring any efficiency gain; thus, dealing with one guild must be weakly preferred to deal with two. Finally, for $\underline{n} < n^*(\delta)$ it must be true that one guild is efficient, hence the ruler must (weakly) prefer to deal with one guild.

The proof of part (ii) is simple. Indeed, when $\phi < 1/2$, $c \geq c^*(\delta, \underline{n})$ and $\underline{n} \in [n^*(\delta), \bar{n}(\delta))$ the ruler has two options available: either he recognizes one or two guilds. However, for $\underline{n} > n^*(\delta)$ having one guild does not guarantee efficient collusion. In this region of parameters, the ruler will clearly prefer to deal with two guilds if θ is small enough since this would imply $\pi^*(\delta, \underline{n}) < \pi^*(\delta, \phi \underline{n})$. Differently, if θ is large enough, say for instance close to $\bar{\theta}(\delta)$, having two guilds will never be optimal since, in such a case, fully efficient collusion will be enforced in a smaller subset of the support Π relative to a single guild regime. A simple continuity argument then implies that there must exist a function $\hat{\theta}(\underline{n}, \phi, \delta)$ such that for all $\theta < \hat{\theta}(\delta, \underline{n}, \phi)$ the ruler prefers to grant recognition to both guilds, and the converse holds true otherwise.

Finally, to prove part (iii) notice that for $c < c^*(\delta, \underline{n})$ the ruler prefers the tax-collector to a single guild regime. Hence, the same qualitative results as in Proposition 6 obtain.

■

The bargaining game between the ruler and a single guild: Here we provide the details of the bargaining game between the ruler and a single guild, which were omitted from the main text for expositional convenience. Assuming that $\pi^* \in [0, 1]$ and using the fact that π is iid, one can verify that the guild's expected intertemporal profit, $W(\delta, n)$, is:

$$W(\delta, n) = \frac{\int_0^{\pi^*(\delta, n)} \pi dF(\pi) + (1 - F(\pi^*(\delta, n))) \pi^*(\delta, n)}{1 - \delta}, \quad (8.2)$$

Once again, this equation captures the fact that the guild is successful in enforcing efficient collusion only if $\pi \leq \pi^*(\cdot)$. When instead $\pi > \pi^*(\cdot)$, the profit of every guild member has to fall below the fully efficient collusion level, π/n , because the incentive

constraint (4.3) would not be satisfied otherwise. Clearly, if n can be chosen so as to imply efficient collusion in all states, i.e., $\pi^*(\delta, n) \geq 1$, one has

$$W(\delta, n) = \frac{\mathbb{E}[\pi]}{1 - \delta},$$

in the region of parameters where $\pi^*(\delta, n) < 1$, instead, simple integration yields

$$W(\delta, n) = \frac{2(1 - n(1 - \delta))(n - 1)}{\delta^2}.$$

Now, let

$$U(\delta, n, R) = W(\delta, n) - R,$$

be the guild's expected intertemporal payoff net of the royalty fee. The ruler chooses n and R so as to solve the following Nash bargaining program:

$$\mathcal{P} : \max_{(R, n) \in \mathbb{R}_+ \times [\underline{n}, \bar{n}(\delta)]} (R - \hat{V}(\delta, c))^\alpha U(\delta, n, R)^{1-\alpha}.$$

Where $\hat{V}(\delta, c)$ measures the ruler's outside option, which is the expected payoff he would obtain by hiring a tax-collector. On the other hand, the outside option of the guild is zero since single merchants would make zero profits if a tax-collector is hired. It is then immediate to check that the solution of \mathcal{P} entails:

$$R^*(\delta, \alpha, c) = \hat{V}(\delta, c) + \alpha \left[W(\delta, n) - \hat{V}(\delta, c) \right], \quad (8.3)$$

which also defines the ruler's profit from granting recognition and privileges to a guild, i.e., $V^*(\delta, \alpha, c) = R^*(\delta, \alpha, c)$. The guild obtains an intertemporal payoff equal to

$$U^*(\delta, \alpha, c) = (1 - \alpha) \left[W(\delta, n) - \hat{V}(\delta, c) \right], \quad (8.4)$$

which is shared equally among its n members. As standard, equations (8.3) and (8.4) are the solution of the Nash bargaining problem with weights α and $1 - \alpha$, according to which the ruler and the guild share the total surplus defined by the difference between the intertemporal surplus from trade and the value of the ruler's outside option, i.e., $W(\delta, n) - \hat{V}(\delta, c)$.