DO SOCIAL CONNECTIONS REDUCE MORAL HAZARD? EVIDENCE FROM
THE NEW YORK CITY TAXI INDUSTRY

C. Kirabo Jackson
Henry S. Schneider

Do Social Connections Reduce Moral Hazard? Evidence from the New York City Taxi Industry
C. Kirabo Jackson and Henry S. Schneider
NBER Working Paper No. 16279
August 2010
JEL No. D52,D62,D64,D82,J01,J41

ABSTRACT

This study investigates the role of social networks in aligning the incentives of economic agents in settings with incomplete contracts. We study the New York City taxi industry where taxis are often leased and lessee-drivers have worse driving outcomes than owner-drivers as a result of a moral hazard associated with incomplete leasing contracts. Using instrumental variables and fixed-effects analyses, we find that: (1) drivers leasing from members of their country-of-birth community exhibit significantly reduced effects of moral hazard; (2) network effects appear to operate primarily via social sanctions; and (3) network benefits can help to explain the organization of the industry in terms of which drivers and owners form business relationships.

C. Kirabo Jackson
Cornell University, ILR School
Department of Labor Economics
Ives Hall East
Ithaca, NY 14853-3901
and NBER
ckj5@cornell.edu

Henry S. Schneider
Johnson Graduate School of Management
Cornell University
326 Sage Hall
Ithaca, NY 14853
henry.schneider@cornell.edu

1. INTRODUCTION

When an economic agent does not bear the full cost or full fruit of her actions, a moral hazard may arise in which the agent, through doing what is personally optimal, behaves in ways that are sub-optimal from a social standpoint (Holmstrom 1979). Moral hazard plays a central role in our understanding of numerous contractual relationships and researchers have found evidence consistent with moral hazard in many contexts.[1] The standard approach to alleviating moral hazard is to use contracts that align the incentives of the agent and principal; for example, paying workers in stock options to induce them to care about the welfare of the company or requiring collateral to reduce borrowers' incentives to default. However, in many settings, such as those with poor access to collateral, such options are not available.

The use of contracts that exploit social ties, within which monitoring costs are low and members can apply social sanctions on each other, has been proposed as a way to enhance the willingness of individuals to act in the interest of the broader society (Stiglitz 1990). However, Banerjee, Besley, and Guinnane (1994), Besley and Coate (1995), and Wydick (1996) argue that a sufficient strong and credible threat of social sanctions against a defaulting group member may be necessary to deter moral hazard in a credit contract, suggesting that it is the strength of group pressure against potential defaulters that matters rather than social ties *per se*. Karlan, Mobius, Rosenblat, and Szeidl (2009) model network connections between individuals as generating "social collateral" that facilitates secure informal borrowing and they document that network *strength* predicts informal borrowing in Peru. In their model, the possibility of losing valuable friendships secures informal transactions the same way that the possibility of losing physical

---

[1] For example, worker effort and productivity are higher when workers are paid by a piece rate versus an hourly wage (Foster and Rosenzweig 1994, Lazear 2000), farm field work may be mechanized when supervisory monitoring costs of workers are high (Smith 1986), and make-versus-buy patterns in the trucking industry may reflect an inability of firms to ensure the efficient use of their resources (Baker and Hubbard 2003, 2004).

collateral can secure formal lending. Altruism between group members has also been identified

as a nonmarket channel through which incentives are aligned (Portes and Sensenbrenner 1993).

Virtually all the empirical evidence that social ties might mitigate moral hazard is in the

group-lending context in developing countries (where markets are not well defined, credit is

constrained, and property rights are not well enforced). Moreover this evidence is mixed.[2,3]

Given the central role that moral hazard plays in explaining the structure of contracts in various

domains, and a theoretical literature highlighting that contractual arrangements that exploit social

ties might mitigate moral hazard, it is important to provide empirical evidence on these issues in

a developed nation context, in a well defined market, and outside of the microcredit context. As

such, we aim to extend the scope of the empirical literature on the importance of social ties in

mitigating moral hazard by presenting the first clean evidence that social ties reduce moral

hazard in a labor market setting in the United States. Specifically, we analyze the outcomes of

New York City taxi drivers who lease (as opposed to own) their taxis and compare the outcomes

of those who lease from an owner from the same country of birth to the outcomes of those who

lease from an owner from a different country of birth.

The taxi leasing market is a prime candidate for studying moral hazard because lessee-

drivers pay less or none of many of the variable costs they generate, including vehicle

---

[2] To our knowledge, the only empirical studies on the ability of social networks to reduce moral hazard outside of group lending are Greif (1993), who finds that community enforcement was effective at limiting moral hazard in 11th-century Mediterranean trade; and Bandiera, Barankay, and Rasul (2005, 2009), who find that social connections lead to a misallocation of effort and hence lower productivity among college student fruit pickers. Gil and Hartmann (2011) provide evidence on market organization through social networks in response to moral hazard, which we discuss later in the section.

[3] This ability of social ties to reduce moral hazard is argued to be largely responsible for the success of group-lending banks such as the Grameen Bank of Bangladesh (Pitt and Khandker 1998 and Banerjee, Duflo, Glennerster, and Kinnan 2009). Kandori (1992) and Ellison (1994) study theoretically how community enforcement may operate, while Sharma and Zeller (1997), Karlan (2007), and Ahlin and Townsend (2007) provide additional empirical evidence about how social ties operate in group-lending, though Giné and Karlan (2006, 2009) find no benefit to group versus individual lending. Social interactions have been identified as determinants of economic actions in other contexts as well (for example, Bertrand, Luttmer, and Mullainathan 2000, Glaeser, Liabson, Scheinkman, and Soutter 2000, Duflo and Saez 2003, Karlan 2005, Mas and Moretti 2009, and Jackson and Bruegmann 2009).

maintenance, repair, replacement, and insurance, and hence have incentives to choose inefficient levels of vehicle usage, care, and risk. The heart of the problem is that a lessee-driver's driving style cannot be monitored at reasonable cost, and hence the ability to specify in the leasing contract the driver's level of vehicle care is limited.[4]

As a result of these contracting limitations, Schneider (2010) finds significantly higher rates of accidents and driving violations for lessee-drivers versus drivers who own their taxis due to moral hazard. The NYC taxi driver labor market is also uniquely well suited for the study of social ties because most NYC taxi owners and drivers are immigrants to the United States, with owners representing 112 countries and drivers representing 146 countries, which allows us to compare outcomes both within and across numerous country-of-birth social networks.

To present some of the first evidence that social ties can mitigate the effects of moral hazard in a labor market and developed country setting we compare the driving outcomes of drivers who lease from owners from the same country of birth (in-network) to the driving outcomes of drivers who lease from owners from a different country of birth (out-of-network). We use country of birth as our measure of social connectedness because this has been found to be an important social connection in other contexts. For example, country-of-birth networks have been found to be important for immigrant assimilation (Borjas 2000) and job seeking (Munshi 2003, Edin, Fredriksson, and Aslund 2003).[5]

To guide the interpretation of the empirical results, we lay out the theoretical reasons why the outcomes of in-network versus out-of-network drivers may vary: first, owners may be better positioned to identify and hire high ability drivers when they are in-network; second, in-network

---

[4] Drivers however keep all fare revenue so effort to generate fares for the owner is not relevant in this context.
[5] Granoveter (1974), Montgomery (1991), Fernandez and Weinberg (1997), and Pellizzari (2004) are other studies on the role of social networks in employment search.

drivers may exert more effort in their driving style to limit the costs borne by the owner. These effort effects could be due to several factors, including social sanctions, social collateral, social preferences, or a better ability of owners to monitor drivers. To formalize these notions, we present a search model connecting owners to drivers that gives the following empirical predictions: (a) in the presence of either better screening or higher effort for in-network drivers, in-network driving may be prevalent; (b) there is positive/negative selection of drivers into in-network driving based on their ability if the results are driven by screening/effort effects; and (c) in the presence/absence of effort effects there exists a positive/no relationship between in-network driving and driving outcomes (after taking selection on ability into account).[6]

The empirical difficulty in isolating the causal (effort) effect of in-network driving is that in-network drivers may differ from out-of-network drivers in important unobserved ways. Because this paper is not based on experimental variation our approach is not to provide one single estimate of the causal effect from a single research design, but to present a compelling body of evidence based on three credible but imperfect research designs that rely on distinct sources of variation and rely on different identifying assumptions.

Our first approach is to aggregate the data to the country-of-birth level to compare the outcomes of drivers from countries with high rates of in-network driving to drivers from countries with low rates of in-network driving ─ removing any bias due to selection within a network; that is, selection that leads one driver to lease in-network and another to lease out-of-network from the same country of birth. Our second strategy exploits longitudinal data to compare the *change* in outcomes of drivers who switch from driving in-network to out-of network (and vice versa) over time to the outcomes of drivers who always drive in-network or

---

[6] Note that we use the term ability broadly to indicate intrinsic driver characteristics that affect the driving outcomes of interest. These characteristics include motor skills and also factors such as conscientiousness.

never drive in-network over the same time period — removing any bias due to high ability drivers being more or less likely to lease in-network. Our third strategy exploits cross-sectional variation in the residential locations of drivers relative to owners *from the same country of birth*. Under the assumption that the distance from a driver's residence to the nearest same-country owner's residence is unrelated to underlying driver ability (an assumption for which we present empirical support), we use this distance measure as an exogenous instrument for driving in-network *while also conditioning on the place-of birth-network*. That is, we compare the outcomes of drivers who lease in-network because they live close to an in-network owner to those of drivers who lease out-of-network because they do not live close to an in-network owner.

Although none of these research designs is dispositive, the combination of evidence is compelling because these estimation strategies control for and are subject to different sources of bias. Specifically, if our effects are driven by selection *within* a place-of-birth network, they would disappear once we aggregate to the place-of-birth level to rely exclusively on variation across place-of-birth networks. However, if the results are driven by selection *across* place-of-birth networks (which would bias our aggregate results), then they would disappear in models that condition on the place-of-birth network of the driver. Also, if the results are driven by sample selection or survivorship bias (which would bias the aggregate results) then they would disappear in models that use within-driver variation. Finally, if the results are driven by *changes* in driver ability that co-vary with *changes* in in-network driving status (which would bias the within-driver estimates), then the results would disappear in the instrumental variables and the country-level models. As such, while each strategy individually may be subject to some form of selection bias, it is highly implausible that all three sources of variation are. Therefore, consistent

results across these methodologies would indicate a real causal effect. We additionally provide supporting robustness checks related to all of these approaches.

We first find that, consistent with there being an important benefit to driving in-network, in-network driving is much more common than would be expected by random chance, and cannot be explained by residential clustering by country of birth. That is, the patterns are consistent with the taxi industry organizing around social networks to mitigate agency problems. We then find that across all three strategies that account for selection, drivers have better outcomes when driving in-network versus out-of-network — indicative of a real causal effort effect. Our naive OLS estimates are smaller than in all three models that account for selection, indicating negative selection into in-network driving as predicted by the theoretical model under effort effects and inconsistent with the model under screening. To provide further evidence that our effects operate through the mitigation of moral hazard, we split out in-network effects by summons type and find the largest effects for types directly affecting owners' costs (as opposed to drivers' costs which would be internalized by the driver).This provides further support that selection is not driving our results.

Finally, given the large apparent effort effects, we investigate the likely mechanisms at play. We use as a measure of owner/driver network strength the density of residents in the owner's/driver's immediate neighborhood from the same country of birth, and find that the marginal effects of driving in-network increase with the strength of the owner's network but not the driver's network, and that in-network effects are weak when the owner's network strength is weak. These patterns are consistent with the presence of social sanctions in the form of community-enforced punishments, which would likely emanate from the owner's and not the driver's network, but inconsistent with altruism, which should operate independently of the

owner's network.

Our results support the findings in Gil and Hartmann (2011) who also show that markets may organize around common-origin communities to mitigate agency problems. Our study improves upon Gil and Hartmann by examining richer data for identifying market organization, and also by providing evidence directly that social networks reduce moral hazard. That is, we actually observe important benefits to organizing around social networks.

More generally, this paper adds breadth to the literature on social networks and moral hazard by being the first study (to our knowledge) to show directly the potential of social networks to reduce moral hazard in a labor market context and in a developed nation. Also given that the micro-credit literature is mixed, this study contributes to this literature by providing compelling evidence that social networks can reduce moral hazard. The current results are particularly interesting because the New York City taxi industry offers a practical alternative to taxi leasing that alleviates most of the moral hazard problem: An active credit market exists in which specialized lenders specifically market loans to taxi drivers to buy their own vehicles. Thus, even in a setting in which credit is available, transactions are commonplace among members of the same country network, indicating that social networks significantly reduce the cost of using the market. We can infer that social networks would be even more valuable if contracting alternatives were not available.

The rest of the paper proceeds as follows. Section 2 gives background on the New York City taxi industry. Section 3 describes the data. Sections 4 and 5 provide the theoretical and empirical frameworks that underlie our analysis. Section 6 describes the main results. Section 7 examines how social ties may operate in the current context. Section 8 concludes.

2. THE NEW YORK CITY TAXI INDUSTRY

Licenses and driving arrangements

To operate a taxi in New York City, a driver must possess a taxi license (often called a medallion).[7] There are two types of taxi licenses: An "owner-must-drive" license, which requires the license owner to be a full-time driver (specifically, to drive at least 210 shifts of at least nine hours per year); and a corporate license, which is free of owner-driving requirements and typically is owned by corporations that lease the licenses (often with vehicles) to drivers. The two license type are immutable. Except in rare instances, new licenses have not been issued since the 1930s, leading to an active trade in the existing licenses between private parties for hundreds of thousands of dollars each.

Under this licensing structure, drivers can operate in three ways: Owner-driving, in which the driver owns the license and vehicle; driver-owned-vehicle-driving (DOV-driving), in which the driver leases the license but owns the vehicle; and lessee-driving, in which the driver leases both the license and vehicle. Lessee-drivers can operate under either long-term leases, which typically extend for six months at a time, or short-term leases, which are twelve hours long. Taxis are typically operated by two drivers for two shifts per day (a day and night shift). Thus, a taxi is typically operated by two lessee-drivers, two owner-drivers who co-own the taxi, or one owner-driver and one lessee-driver who leases from this owner-driver.

For reasons explained below, the drivers we examine are primarily lessee-drivers operating under this third type of driving arrangement, where the owner-driver operates the taxi for one shift twelve-hour shift per day and a lessee-driver who leases the taxi (both the license and vehicle) from the owner-driver operates the taxis for the second twelve-hour shift per day.

---

[7] Much of the information in this section comes from personal discussions with industry participants, TLC Rules documents made available by the TLC, and Urbanomics (2004).

Discussions with industry participants indicate that owner-drivers often find lessees through family-members, neighbors, and friends, though anonymous matches are also common.

The costs and revenues of operation are split between the lessee-driver and the owner-driver as follows. The owner is required by the TLC to pay for no-fault and liability insurance ($7,000 to $13,000 per year) and workers' compensation for the lessee-driver ($1,400 to $3,000 per year), licensing fees (approximately $1,000 per year), vehicle inspections ($50), and fines for improper usage of the taxi, some of which may be due to the lessee's actions. Owners also pay for all vehicle maintenance and repair ($4,000 to $10,000 per year) and for vehicle replacement (approximately $28,000).[8] Repairs are typically conducted at garages chosen by the owner-driver that specialize in servicing taxis but are also sometimes conducted at other garages.

Lessees pay for their own gas usage ($5,000 to $10,000 per year), tolls, parking tickets, fines for DMV and TLC violations, and TLC fees (approximately $100 per year) that they themselves generated. Lessees also pay a flat lease fee to the owner. The lease fee is capped by the TLC at a maximum amount that owners can charge lessees per lease period, which industry participants indicate is respected in practice and nearly always binds (TLC (2006)). Lessees often provide a modest deposit, which is sometimes forfeited to the owner if the lessee has an at-fault accident. During the sample period, the deposit amount was capped by the TLC at $500 for long-term lessees.[9] Lessee-drivers keep all fares and tips that they generate.

Measures of owners' costs that are generated by drivers

As described above, lessee-drivers generate three types of costs that are borne by taxi

---

[8] Note that is the vehicle owner who pays all repair costs. Since we are primarily examining drivers who lease from owner-drivers, it is these owner-drivers who pay repair costs.

[9] There exist industry programs that could provide incentives for better driving outcomes (they stipulate license suspension and revocation for repeated violations), but they have typically not been enforced. Another program reduces insurance premium if all lessees of the medallion have a good driving record.

owners, which we discuss in detail now. The first is repair costs associated with vehicle mechanical failures and other maintenance work. Given that New York City taxis are operated in heavily congested urban conditions for an average of 175 miles per day, the driver's level of vehicle care can quickly show up in vehicle condition.[10] The second type of cost is from accidents. Approximately one-third of taxis are involved in a serious accident per year (defined as having over $1000 in damage or involves injuries). Again, driver aggressiveness such as speeding or cutting across traffic to pick up a hailing passenger (a common site in Manhattan) increases the likelihood of accidents. The third type of cost is for fines to owners stemming from driver behavior. Owners receive fines ranging from $25 to $100 per instance when drivers do not adequately record information on passenger trips they provide, when the vehicle is not in satisfactory condition, either regarding cleanliness or safety, and when items such as the driver's taxi license are not properly displayed.[11]

Drivers typically pay little or none of the repair and maintenance costs, a fraction or none of the costs of an accident, and only a fraction of the costs of fines from violations given they do not pay the owner's portion (drivers may receive a fine themselves for failing to adequately record trip information, keep the taxi in good condition, or properly display all required information). However, drivers keep all the fare revenues they generate. Given that revenues and costs are increasing in driver aggressiveness, we can see that drivers face a moral hazard.

3. THE DATA

---

[10] For example, in conversations with taxi owners, some indicated to us that the brakes must be replaced every few weeks. Whether the driver is aggressively accelerating and braking, taking care to avoid potholes and curbs, and so on can have an important effect on an owner's repair costs.

[11] New York City police has jurisdiction over enforcing many New York City taxi regulations, and the violations are typically assigned during police stops of the taxi during regular taxi operation.

A.   Description of the TLC data

Data were provided by the TLC under a proprietary data-sharing agreement and contain information about all drivers of New York City yellow taxis during two time periods: the spring of 2005 and fall of 2007. The 2005 data contain driver-level records on three types of driving outcomes: Accidents, convictions, and summonses. The accident records describe all accidents involving taxis with injuries or property damage exceeding $1000 to any vehicle involved. The violations records describe all drivers' convictions for driving violations during taxi operation for which points are issued against a DMV driver's license. The summons records describe all summonses, which are issued for TLC violations (as opposed to DMV driving violations) such as passenger service refusals, using the taxi for unlawful purpose, and missing required items from the taxi such as the display of the driver's taxi-driving license.

The 2007 data only contain records for the summons outcomes and not the accident and violations outcomes (the latter outcomes originate from the New York City police department, and were not available for 2007).[12] In addition to being available only for 2005, the accidents and convictions records also have more limited variation than the summons records. Approximately 24 percent of drivers receive at least one summons during a six-month period versus 14 percent having a conviction and 4 percent having an accident. Because of the added power for tests, and its panel nature, we focus our analysis on summonses. However, we are careful to show that our results are robust for accidents and convictions.

Summonses are given for a range of TLC violations. Some of these driver violations generate costs for owners directly: owners receive fines for missing items, missing trip records, and vehicle condition violations stemming from the actions of drivers to whom they lease. Some

---

[12] We top-code outcomes to three to limit the impact of outliers, though less than one percent of drivers are affected.

of these driver violations generate costs for owners indirectly: using the taxi for an unlawful purpose may risk the taxi (for example, temporary impounding), hazardous moving violations, and possibly shift violations, in particular smoking in the taxi. Some driver violation may not affect owner costs directly, but may reflect general driving style and also level of vehicle usage. Since violations frequency likely increases monotonically with usage, and lessee-drivers may operate their taxis for longer than owner-drivers because they do not pay the full cost of operation, higher usage may also reflect moral hazard.

The driver-level data also record each driver's country of birth and 2005 address, and for drivers entering after 1997, whether the driver passed an English-language test and his score on a written driving test, both of which the driver must pass to obtain a taxi-driving license.[13] We calculate a driver's New York City taxi-driving experience from the number of months registered as an active New York City taxi driver, which we derive from his taxi driver's license numbers.[14] Driver-level records are available for both lessee-drivers and owner-drivers.

Two driver-of-record files identify the taxi each long-term lessee-driver and owner-driver was registered to operate on April 2, 2005 and October 21, 2007, which we use to merge the driver and owner data.[15] Owners receive a significant discount on insurance premiums when they register specific "named-drivers" with the TLC and the insurance provider. These named-drivers are then the exclusive operators of that taxi (including the owner himself when he drives), and

---

[13] The English-language test is aimed to ensure that drivers can communicate with passengers, while the written driving test is on New York City geography and DMV and TLC regulations.

[14] Taxi-driving license numbers are issued sequentially, which allows us to identify when a license was issued. We thank Bruce Schaller for providing this algorithm.

[15] Taxis not in the driver-of-record file have "unspecified" drivers, and TLC officials and industry participants indicate they are nearly always short-term lessees. Since short-term lessees operate different taxis every day, they cannot be matched in the data to specific taxis and hence owner countries, and are excluded from the analysis.

these matches constitute the driver-of-record file.[16] Taxis with such named-drivers represent 9,535 of the 12,779 taxis in 2005 and 9,025 of the 12,953 taxis in 2007. Since we only know the driver-taxi match on these two dates, we restrict the majority of the analysis to the periods surrounding these dates.

For the long-term lessees, the driver-of-record files record the current lease period, which typically are six months intervals (50 percent are between 177 and 183 days long, and 95 percent are between 147 and 189 days long). We match the dates of offense for the summons, convictions, and accidents to these lease periods to identify the drivers' outcomes associated with particular taxis. To make the driving outcomes comparable across drivers with different lease period lengths, we use the recorded lengths to normalize the number of incidences to 182 days.

We supplement the TLC data with Census data on driver/owner neighborhood characteristics, most importantly the fraction of residents in the driver/owner Census tract from the same country of birth (resident country of birth is only reported at the tract level), which we match to drivers/owners using their addresses.[17]

Table 1 reports summary statistics for the primary variables of interest for our sample. It shows that drivers have 11 years of NYC taxi-driving experience on average, that almost half of drivers lease in-network, and that approximately one in three drivers receives a summons, one in six has a conviction, and one in twenty has a serious accident per six month period.

B.   Discussion of driving outcome measures

---

[16] Because of the large financial incentives to accurately report named-drivers, we believe these records are accurate. Furthermore, summonses for inaccurate named-driver records are very infrequent.

[17] The country of birth recorded in the TLC data is occasionally more precise than the country of birth in the Census data. For example, the Census and TLC data both record the European countries, England, Ireland, Sweden, Austria, France, Germany, Netherlands, Greece, Italy, Portugal, Spain, Czech Republic or Slovakia, Hungary, Poland, Romania, Belarus, Russia, Ukraine, and Yugoslavia. However, the TLC data record the additional European countries, Belgium, Switzerland, Monaco, Bulgaria, Latvia, Lithuania, and Moldova. These additional countries are assigned to the Census categories, "Other western Europe," "Other southern Europe," and "Other eastern Europe."

Our data do not contain measures of vehicle maintenance, repair, and replacement costs borne by the owner.[18] Instead as mentioned above we examine three measures that either directly or indirectly reflect costs generated by the driver but that are borne by the owner. One measure is accidents, which, unlike maintenance and repairs, can be attributed directly to individual drivers. The second measure is convictions for driving violations, such as speeding or disobeying other traffic laws, which are enforced and processed in the same way as ordinary motorists' driving violations. Both accidents and convictions directly reflect driver ability and aggressiveness, and hence are likely to reflect on average risk for owners from unsafe driving. In fact, Gebers (2003) finds that prior violations more accurately predict future accidents than prior accidents themselves (which is also a strong predictor of future accidents).

Our third measure is summons, which are issued for TLC violations (as opposed to DMV driving violations) based on a driver's action for which owners receive fines directly, such as when a required item is missing from the taxi like the driver's taxi-driving license (as mentioned above), and also for actions for which the owner does not directly receive a fine, for example for failing to satisfy a reasonable passenger request or smoking in the taxi, which reflects a general tendency regarding conscientious versus careless driving. We believe some types of summons better reflect a driver's driving style and conscientiousness than others. However, we have no way to identify precisely which summons types are most reflective, and hence we aggregate all summonses for the primary analysis. However we also provide results broken out by summons type to show that individual types indeed conform to our expectations of which are more likely to be susceptible to moral hazard.

C.   Availability of country-of-birth data

---

[18] This data would not reveal directly the costs generated by the lessee-driver anyway since attributing repair costs to the owner-driver versus the lessee-driver would be challenging given both operate the taxi nearly every day.

Our primary analysis relies on country-of-birth data being available for both the lessee-driver and also the owner from whom he leases. Country data for both parties must be present to identify whether the driver is leasing from a same-country owner. Table 2 reports the numbers of drivers with available country data for both the driver and owner, along with mean New York City taxi driving experience of these drivers.[19]

The table shows that country data for both the lessee and owner are available for 1,955 and 2,053 lessees in the 2005 and 2007 data respectively, and that lessees with country data have only slightly more experience than lessees without data. Drivers with country data also have a similar distribution over race than drivers without country data (not shown in the table), suggesting that the distribution over countries is similar for drivers with and without country data.[20] Drivers with and without country data also have similar English-language and driving-test results (also not shown). Therefore, based on the observable driver characteristics that are available to us, we have no reason to believe that the drivers that we analyze are systematically different than the drivers for which country data are missing.

Most lessee-drivers with available country data for both the driver and the owner share a taxi with an owner-driver (that is, they operate owner-driver taxis, as opposed to non-owner-driver taxis).[21] Thus, the sample we analyze consists primarily of lessee-drivers who lease from owner-drivers. This sample is attractive for studying social networks because owner-drivers are most likely to be invested in the condition of their vehicle (since they drive most days) and

---

[19] Note that Table 1 reports slightly fewer observations than Table 2 because Table 1 reports summary statistics only for observations for which country-level data are available and data are non-missing for all other variables used in the regression analysis, while Table 2 only examines the number of drivers for which country data are available regardless of the availability of other variables.

[20] Fifty-four, 23, and 13 percent of drivers with country data were Asian (including South Asian), Black, and White, respectively, while 58, 22, and 8 percent of drivers without country data were Asian, Black and White.

[21] The reason country data are available primarily only for lessee-drivers on owner-driver taxis is because country data are available for most drivers, including owner-drivers, but not for non-driving owners.

interact with the lessee regularly, thereby facilitating a social connection, as opposed to non-driving owners, who often own many hundreds of taxis, and hence likely have more anonymous relationships with drivers regardless of whether they share a common country of birth. Table 2 shows that a small fraction of lessee-drivers that have country-level data available lease from non-driving owners and hence are included in our sample. The results throughout are similar and modestly stronger when they are excluded.[22]

4. THEORETICAL FRAMEWORK

There are two important potential benefits of in-network driving. The first is better screening such that owners can better assess the ability of drivers from the same country and as such are more likely to hire drivers from the same country. The second is that drivers from the same country exert higher levels of effort than drivers who lease from owners from a different country (we explain reasons for this in greater detail below).

A.   Potential benefits of in-network leasing

Even in the case where in-network drivers do not exert more effort than out-of-network drivers, one might observe high rates of in-network leasing and different driving outcomes on average for in-network drivers versus out-of-network drivers. This would be due to screening.

i.   The screening technology may be better in network. Specifically, due to information flows about prospective drivers via the country-of-birth network, owners may be better able to determine the ability of same-country drivers versus different-country drivers, and therefore select different ability drivers for in-network versus out-of-network.

---

[22] Note that we do not examine owner-driver outcomes in the main text of this study because we are interested in comparing in-network versus out-of-network lessee-drivers. However, owner-driver results are available in Schneider (2010) as well as the online appendix.

17

There are a few reasons why *all else equal* drivers who lease from an owner from the same country of birth might exert higher levels of effort than drivers who do not.

ii. Social preferences on the part of drivers may lead drivers to exhibit more altruistic behavior toward owners from the same country versus owners from a different country. It is well documented that people tend to be willing to give up some of their own material self-interest in order to improve the welfare of others with whom they have close social ties.[23]

iii. Social ties may create social collateral (Karlan, Mobius, Rosenblat, and Szeidl 2009) such that individuals derive positive utility from ties and have a disincentive to behave in ways that harm them. Benefits may be non-pecuniary, such as enjoying a friendship or gaining spousal prospects, or arrive in the future, such as having a resource for job information or a loan in times of need. Persons from the same country are much more likely to have a strong social tie and hence have significant social collateral than persons from different countries.

iv. Individuals are better able to impose social sanctions on others from the same group versus a different group. For the same reasons that social collateral creates a disincentive to behave in ways that harm a relationship, group membership creates disincentives to behave in ways that lead others to exclude one from the benefits of membership. In the current context, owners have more leverage over drivers from the same country versus a different country, leading drivers to exert more effort toward the objectives of same-country owners, *all else equal*.

v. Owners' costs of monitoring the behavior of drivers might be lower for drivers from the same country versus a different country given the more frequent interactions between driver

---

[23] For example, Fong and Luttmer (2009) find in a charitable-giving experiment that respondents who report feeling close to their racial or ethnic group give substantially more when victims are of the same race and Hungerman (forthcoming) finds that the charitable activity of all-white religious congregations decreases in the fraction of blacks in the community. It has even been documented that food giving is much more common in the animal kingdom among kin (Stevens and Gilby (2003)), suggesting that affinity for group members is a basic instinct.

and owner through network ties. Specifically, drivers may have a higher chance of being caught driving with insufficient care when they lease in network versus out-of-network, giving in-network drivers greater disincentive to exert insufficient care, *all else equal*.[24]

B.   Theoretical model

We now present a simple theoretical model of how owners choose which drivers to hire. The model provides testable implications that will help us disentangle the effort effects versus the screening effects.

Using a modified version of a standard job search model, we consider the owner's decision of whether to lease to a potential driver to whom he has come in contact. This is a discrete time, infinite-horizon model, where, for simplicity, the owners are infinitely lived and have discount factor β. As long as the owner has not hired a driver, the owner comes in contact with one potential driver per period. The potential driver's ability, α, is drawn i.i.d. from a continuous distribution F with mean $\alpha^+>0$ and support A, and represents the innate attributes of the driver that may limit the owner's costs.[25]

If the owner hires a driver with ability α, he gets a payoff of $\alpha+e$ each period forever, where $e \geq 0$ represents the level of effort exerted by the driver to limit the owner's costs. If the owner rejects the driver, the owner receives a payoff of 0 this period and draws again next period. We also assume that the probability of drawing an ability $\alpha < \beta \alpha^+$ is greater than zero. This condition guarantees that some drivers will be rejected in equilibrium.

---

[24] We list this possibility because it has been used in the micro-credit loan market to argue that individuals may be able to monitor borrowers from the same group better than a large bank can monitor borrowers (for example, Stiglitz (1990)). However, later we will argue that this channel is unlikely to be important in the taxi market.

[25] To keep the model simple, we assume a new pool of potential drivers enters each period such that we have a stationary distribution over the abilities of potential drivers. This implies that if a driver is rejected, he does not come in contact with another owner-driver from whom he could lease (although of course he may remain in the industry and lease from a non-driving owner).

Owners and drivers belong to one of at least two networks. Each period, the probability that the potential driver being considered by an owner from network j is from the same network is δ_j (again, the distribution over ability within network is the common distribution F).

Effort effects

We now solve the model in the case that owners fully observe the abilities of both in-network and out-of-network drivers. The importance of the network in this case comes from the additional effort that driving in-network induces. We normalize in-network effort to be e=e'>0 and out-of-network effort to be e=0. To solve for an equilibrium, we set up the standard Bellman equation,

$$V_j(\alpha+e) = \max\{((\alpha+e)/(1-\beta)), \beta\delta_j\int_A V_j(\alpha'+e')f(\alpha')d\alpha' + \beta(1-\delta_j)\int_A V_j(\alpha')f(\alpha')d\alpha'\}$$

The value function, $V(\cdot)$, which is bounded and weakly increasing, has a unique solution.[26]

The decision rule for an owner is simple. The owner will hire the potential driver if his ability plus effort is greater than the expected payoff of waiting for another potential driver. Thus, the owner uses a minimum cutoff level of his payoff $\alpha+e$ in deciding whether to hire a potential driver. The expected payoff of waiting will be the same regardless of whether the potential driver is in-network or not. However, conditional on ability level, in-network drivers generate a higher payoff for the owner due to the positive in-network effort effects. Therefore, the owner will use a lower cutoff for ability for in-network drivers relative to out-of-network drivers in the hiring decision. The opportunity cost of searching for another driver, and this lower ability cutoff, generates a lower average ability of in-network relative to out-of-network drivers.

---

[26] The transformation $TV_j(\alpha+e) = \max\{((\alpha+e)/(1-\beta)), \beta\delta_j\int_A V_j(\alpha'+e')f(\alpha')d\alpha' + \beta(1-\delta_j)\int_A V_j(\alpha')f(\alpha')d\alpha'\}$ satisfies Blackwell's (1965) sufficient conditions to be a contraction mapping. Therefore the value function has a unique solution.

Proposition 1: Under effort effects, the average ability of in-network drivers is lower than the average ability of out-of-network drivers.

Proof: The lowest ability driver that an owner will hire will be $K(1-\beta)-e$, where $K=\beta\delta_j\int_A V(\alpha+e')f(\alpha)d\alpha+\beta(1-\delta_j)\int_A V(\alpha)f(\alpha)d\alpha$. This cutoff value is smaller for in-network drivers since in-network drivers exert higher effort. Therefore the average ability of in-network drivers will be lower that the average ability out-of-network drivers.

Screening effects

We now consider the effect of screening. For simplicity, we assume that effort is the same for both in-network and out-of-network drivers and normalized to $e=0$ (however, it will be straightforward to see that the prediction also holds under in-network effort effects). Under screening, owners can still observe the ability of potential in-network drivers (which he learns about through contacts within the network), but now cannot observe the ability of potential out-of-network drivers. Thus, if the owner comes in contact with an out-of-network driver in some period, the owner treats the driver as if he has the average ability for the population, α^+. The Bellman equation representing the owner's problem is now,

$$V_j(\alpha) = \max\{(\alpha/(1-\beta)), \beta\delta_j\int_A V_j(\alpha')f(\alpha')d\alpha' + \beta(1-\delta_j)V_j(\alpha^+)\}$$

which again satisfies the conditions for a unique solution.

As before, an owner will hire a potential driver if the driver's ability is above some threshold. This threshold may lie above or below the average ability of out-of-network drivers. This implies that, depending on the parameter values, owners hire either all out-of-networks or no out-of-network drivers that they encounter. The owner's knowledge about the ability of

drivers in their network allows the owner to be more selective regarding drivers in their network.

Proposition 2: If out-of-network drivers are hired, they will have the same average ability as the population of potential drivers. The average ability of in-network drivers who are hired, however, will be higher than the average ability of the population of potential drivers.

Proof: To show this we need only to show that some in-network drivers will be rejected. Let us show this by contradiction. Suppose that all applicants are hired. This implies that $V_j(\alpha)=(\alpha/(1-\beta))$ for all α. Therefore it is better to hire the lowest ability type α_L than to reject and draw again. Accepting every applicant implies $((\alpha_L/(1-\beta))\geq \beta(\alpha^+/(1-\beta)))$. But this contradicts the earlier assumption that there are some possible ability types that lie below $\beta\alpha^+$. There must be some ability types low enough that they are rejected in equilibrium. The average ability of in-network drivers accepted must be above the average ability of population of potential drivers.

As before, information about a driver's ability is useful to the owner. Some in-network drivers have sufficiently low ability that in-network owners reject them. Thus, the average ability of in-network drivers that are accepted is higher than the average ability of potential drivers. Since out-of-network drivers appear to the owner to have average ability, owners hire all out-of-network drivers they encounter if his hiring cutoff level is below the average ability in the population, and rejects all out-of-network if his cutoff level is above the average ability in the population. Thus, if there is an owner who accepts out-of-network drivers, the average ability of such drivers will be lower than in-network drivers.

Finally, note that we have not specified the arrival rate of in-network versus out-of-

22

network drivers. However, it is apparent that proportionally more in-network owners arrive versus out-of-network owners compared to their ratio in the population, as we might expect, then many or most hired drivers will be in-network, in particular under effort effects. We also expect that networks with higher ratios of drivers to owners have a higher arrival rate of in-network drivers. As Table 3 shows, these ratios vary considerably across countries, and the fraction of drivers who are in-network appears to be closely related to these ratios.

C.   Disentangling screening versus effort effects

From the model it is clear that the high rate of in-network leasing could be due to either effort effects or to screening, depending on the arrival rate of in-network versus out-of-network drivers and the parameter values. However, we also know that better in-network outcomes conditional on ability can only be due to effort effects. From Propositions 1 and 2, we now also know that in-network drivers will have higher average ability than out-of-network drivers under screening (assuming out-of-network driving occurs at all, and regardless of effort effects) but lower average ability than out-of-network drivers under effort effects but no screening. Thus, the outcomes of in-network drivers conditional on ability should be better than the outcomes of in-network drivers unconditional on ability under no screening but worse than the outcomes of in-network drivers unconditional on ability under screening. In summary, the effect of in-network driving on outcomes conditional on ability provides a test of effort effects, while the effect of conditioning the outcomes on ability provides a test of screening.

5. EMPIRICAL STRATEGY

Among the aims of this paper are (1) to document the degree and direction of any selection on driver ability into driving a taxi owned by someone from the same country, and (2)

to remove any selection bias to uncover the causal effect (the effort effect) of driving a taxi

owned by someone from the same country. To identify the casual effect of in-network driving,

we use three distinct sources of arguably exogenous variation in in-network driving. Results

from these separate sources of variation provide a robustness check on each other.

A.   Baseline model

Our basic empirical strategy is to compare the driving outcomes of drivers who lease

from an owner from the same country of birth to the driving outcomes of drivers who lease from

an owner from a different country of birth. Specifically, we estimate the following model by

Ordinary Least Squares (OLS),

[1] $Y_{it} = \beta \cdot Same_{it} + ln(Exp_{it}) \cdot \gamma + \alpha_i + \varepsilon_{it}$

In [1], Y_{it} is the outcome of driver i at time t, which is the number of accidents, convictions, or

summons, $Same_{it}$ is an indicator variable denoting whether driver i leases from an owner from

the same country of birth at time t, Exp_{it} is the experience level of driver i at time t, α_i is the

unobserved time invariant ability level of driver i, and ε_{it} is the idiosyncratic error term. Because

driver ability is not observed, the total error term is $\alpha_i + \varepsilon_{it}$.

OLS estimates of [1] are unlikely to yield the causal effect of in-network driving because

owners may select the drivers to which they lease based on unobserved driver ability. As such,

OLS estimates of β may be biased depending on the nature of the selection. In the presence of

screening, where owners can better observe ability for same-country drivers such that same-

country drivers are higher ability, there would be positive selection and the OLS estimates of β

would be biased downward (more negative). Alternatively, if owners know that same-country

drivers will exert more effort, making owners more willing to hire same-country drivers of lower

ability, there would be negative selection and the OLS estimates of β would be biased upward

(less negative). As such, comparing the OLS estimates to estimates obtained from a model that isolates the casual effect of being a same-country driver will be informative about the selection process. We propose three strategies to remove the effects of selection, which we describe now.

B.   Cross-country identification strategy

The first strategy to isolate the effort effect is to aggregate driving outcomes to the country level and to use cross-country variation in leasing rates to test whether countries with relatively higher in-network lease rates have relatively better driving outcomes. Aggregating driving outcomes to the country level controls for any selection on ability into in-network leasing that occurs within groups of individuals from the same country of birth. This strategy will uncover a causal effect so long as differences in in-network driving across countries are not systematically related to differences in unobserved driver ability across countries. To implement this strategy we estimate the following model by OLS,

[2] $Y_{ct} = \beta \cdot Same_{ct} + (ln(Exp_t))_c \cdot \gamma + \varepsilon_{ct}$

where the c subscript indicates a country-level mean. Since the model assumes that mean driver ability by country (α_c) is uncorrelated with in-network lease rate by country ($Same_{ct}$) after conditioning on mean experience by country ($(ln(Exp_t))_c$), the estimate of the network effect isolates the effort effect.

C.   Within-driver identification strategy

The second strategy to isolate the effort effect is to use within-driver variation in same versus different-country leasing. Specifically, with two observations per driver over time, in 2005 and 2007, we can compare the outcomes of drivers who leased from a same-country owner in one period and a different-country owner in another period. By comparing the outcomes of the same drivers over time we effectively remove the contribution of time-invariant ability on

driving outcomes (the contribution due to selection), thus isolating the effort effect. Specifically, we estimate the following differenced (within-driver) model by OLS,

[3] $\qquad \Delta Y_i = \beta \cdot \Delta Same_i + \Delta ln(Exp_i) \cdot \gamma + \Delta \varepsilon_i$

All variables are defined as before, but now the time-invariant ability component is differenced away, so that the error term $\Delta \varepsilon_i$ does not reflect the effects of selection. Identification of the effort effect β comes from drivers who switch from leasing from a same-country owner to a different-country owner (and vice versa). The identifying assumption is that any changes in unobserved characteristics that affect driving outcomes are orthogonal to changes in same-country driving status over time. It is worth noting that this model estimates a local treatment effect for switchers who make up only 3.2 percent of the sample (58 drivers), so while the effect may be well identified it does not necessarily generalize to all drivers (note that this limitation applies to all within-entity models). This is one motivation for our use of two other completely different sources of variation to identify effort effects.[27]

D.   Cross-sectional instrumental variables strategy

Finally, we exploit the variation in leasing from a same-country owner that is due to the residential locations of drivers versus same-country owners. The idea is that drivers may be more likely to lease from an owner who lives nearby versus far away all else equal. Thus a driver who happens to live near a same-country owner should be more likely to lease from a same-country owner simply because he lives close to that owner. Using the 2005 addresses of drivers and owners, we compute the minimum distance of each driver to a same-country owner in the data.

---

[27] Nevertheless, the observed characteristics of switchers are unremarkable. On average, switchers have 10.9 years of experience and are 44.4 years old versus 11.1 years of experience and 44.5 years old for non-switchers. Sixteen, 12, 11, 9, and 9 switchers are from Bangladesh, Pakistan, India, Haiti, and other countries, respectively, which is proportionally similar to non-switcher countries of birth. Twenty-five drivers switched to in-network leasing between 2005 and 2007 while 33 drivers switched away from in-network leasing between 2005 and 2007.

We then use this distance as an exogenous instrument for in-network leasing to isolate the causal effect of in-network driving.

To be a valid instrument, this distance measure must (a) be correlated with in-network driving, (b) not be correlated with unobserved ability, (c) have a monotonic relationship with being a same-country driver. Figure 1 shows clearly that distance to the closest same-country owner predicts leasing from a same-country owner. The effect however is most pronounced for distances of less than two-tenths of a mile, and for this reason we specify our instrument as an indicator for whether the driver lives in the same Census block as a same-country owner.[28] Figure 1 also shows that distance to a same-country owner is unrelated to driving experience, which is the best measure of driver ability in our data. It is also clear from the figure that the relationship between distance to a same-country owner and being a same-country driver is monotonic. We present formal econometric tests of the validity of the instrument in Section 6.[29]

In summary, we exploit this variation in residential locations of drivers vis-à-vis owners to predict network status, and then use these predicted values in a regression on driving outcomes. Specifically we estimate the following system of equations by 2SLS,

[4] $\quad\quad\quad Same_i = Dist_i \cdot \delta_1 + ln(Exp_i) \cdot \delta_2 + \theta_{1c} + \nu_{1i}$

[5] $\quad\quad\quad Y_i = \beta \cdot Same_i + ln(Exp_i) \cdot \gamma + \theta_{2c} + \nu_{2i}$

In [4] and [5] all variables are defined as before (but there is no longer any time variation), θ_c is a country of birth fixed effect, and the error term ν includes both the idiosyncratic variation and

---

[28] Using same Census tract or log of distance to the nearest same-country owner gives similar though modestly weaker results due to a weaker first stage.

[29] A slightly different interpretation of the instrumental variable estimate is that leasing from a same-country owner who also lives nearby facilitates stronger network effects, for example, because an owner can better enforce social sanctions. However, the OLS estimates of the network effect are similar for drivers living nearby the owner versus far away from the owner, indicating that the existence of network effects is not predicated on geographic proximity. Regardless, the alternate interpretation would still indicate the presence of network effects.

driver ability. This model exploits variation in residential locations of drivers *who come from the same country of birth* to remove selection to in-network driving.

To present evidence that our instrumental variables model is valid, we tests whether our distance measure is correlated with both our observed measures of driver ability and an estimate of the unobserved component of driver ability. In most situations where instrumental variables are used, authors cannot directly test the identifying assumption of their model (that is, that the instrument is uncorrelated with the unobserved error term). In our context, because we have longitudinal data but our instrument relies on cross-sectional variation, we can identify driver ability and then test whether our estimates of unobserved ability are correlated with our instrument. Specifically, we can obtain estimates of driver ability (α_i) from our driver fixed effects model and then see if these estimates are correlated with the instrument.[30]

6. RESULTS

A.   Determinants of in-network driving

One of the principal predictions of the model is that where there are screening effects or effort effects there will be much more in-network driving observed in the data than would be predicted by random chance or by residential sorting alone. In this sub-section, before presenting evidence on the effects of in-network driving on outcomes, we provide compelling evidence that this central prediction of our model holds in the data.

---

[30] We also use a mixed effects estimator to obtain the Best Linear Unbiased Predictors (BLUPs) of driver ability and find that these are also not correlated with our instrument. This mixed effects method for identifying BLUPs is used in Jackson (2010) and Woodcock (2009). See Woodcock (2009) for detailed discussion of their properties.

Table 3 reports the number of lessees by lessee country broken out by the corresponding owner country.[31] The patterns suggest that owners lease to drivers within their network when they can. For countries with more drivers than owners, owners mostly lease to drivers from the same country, but many drivers do not lease from same-country owners. For example, there are 291 Bangladeshi drivers but Bangladeshi owners lease to only 148 total drivers total, and consequently only 133 of 291 Bangladeshi drivers lease from Bangladeshi owners, while 133 of 148 drivers leasing from Bangladeshi owners were Bangladeshi. For countries with fewer drivers than owners, owners often lease to different-country drivers, while most drivers lease from same-country owners. For example, there are 12 Greek lessees but Greek owners lease to 52 drivers, and 10 of 12 Greek drivers lease from Greek owners, while only 10 of the 52 drivers to whom Greek owners lease are Greek.

These patterns are consistent with drivers being more likely to lease in-network than would be predicted by random chance, and with there being a benefit to in-network driving. However, this is not a conclusive test. Given that many drivers may live in ethnic enclaves, a preference for proximity would also generate more in-network leasing than random chance. Given that 24 percent of drivers in our data lease from an owner who lives in the same Census block, it is plausible that the patterns are picking up a "proximity" effect rather than an "in-network" effect. As such, showing that the pattern of in-network driving is not determined by proximity alone is key to showing that all else equal the likelihood of leasing is higher in-network than out-of-network.

To show this we test the hypothesis that the patterns observed in the data are driven by proximity alone and that the leasing decisions are not affected by the network status of the

---

[31] The table reports 2005 data, but 2007 patterns are similar.

drivers and owners. To do this we create an indicator variable equal to one if a driver leases from an owner who lives in the same Census block as themselves and zero otherwise. Under the null hypothesis that the countries of birth of the owners and drivers have no effect on the decision of from whom to lease, then conditional on having any owners in the same Census block, the country of birth of the owners should have no effect on the likelihood of leasing from an owner who lives in the same block.

Table 4 shows results from a regression where the dependent variable is an indicator for whether the driver leases from an owner in the same block and the explanatory variables are indicators for having any owners in the same block and any in-network owners in the same block. The results show that the likelihood of leasing from a neighbor is 32 percentage points higher (p < .01) when there is an owner from the same country in the block. Also, conditional on whether there is an owner from the same country in the same block, having any owner in the block has a relatively small effect on leasing from that owner. The results provide a clear indication that the prevalence of in-network driving cannot be explained by proximity alone ─ indicative of there being benefits to leasing in-network. Having established the high prevalence of in-network driving (a central prediction of our model), we now turn to our estimates of driving in-network on outcomes.

B.   Baseline analysis

Table 5 provides regression evidence regarding the effect of in-network driving on driving outcomes. Column (1) shows OLS results for a driver's number of summonses. Drivers leasing in-network have 0.09 fewer summonses (p < .01), which is a moderate-size effect relative to the standard deviation of the number of summonses per driver of 0.72. Driver country-of-birth fixed effects are included in column (2) to address the possibility that drivers from some

countries are systematically better or worse in ways that are correlated with the probability of in-network leasing and the results are similar.[32] In the OLS and all other specifications, when driver age is included, the estimate of the age effect is small in magnitude and not statistically significant after controlling for experience, nor does it affect the estimate of in-network driving effect. Including Census neighborhood characteristics, such as median income and education level, also does not affect the estimate of the in-network driving effect (although Census data are only available for 2005 outcomes due to address data being available only for this year). We therefore use the more parsimonious specification that includes only experience.

C.   Country-level analysis

We begin by showing in Figure 2 the relationship between the mean numbers of summonses per driver by country versus the fraction of drivers from that country leasing from an owner from the same country.[33] The figure shows clearly that driving outcomes are better for drivers from countries with more in-network leasing, with a switch of all drivers from out-of-network to in-network leasing corresponding to 0.35 fewer summonses per driver.

We can investigate the relationship more formally however by estimating equation [2] in order to control for the effect of any differences in average driver experience across countries. Column (3) of Table 5 provides these results, showing that the in-network leasing effect is substantially larger than in columns (1) and (2) and is statistically significant (p=.01). As discussed in Section 5, the country-level estimates identify the true causal effect of social ties on driver effort (separate from selection on ability) by eliminating any within-country correlation

---

[32] Results are similar when the models are estimate with 2005 and 2007 data separately, and with Poisson and the more flexible negative binomial regression models, and are available in an online appendix.

[33] Data for countries with at least five drivers are included (55 countries) with the area of the bubble being proportional the number of drivers. Country labels for the points in the figure are not included because the TLC requested that country-level driving outcomes not be reported.

between unobserved driver ability and in-network leasing. The identifying assumption is that any country-level differences in ability are uncorrelated with country-level in-network lease rates after controlling for country-level differences in driving experience. The estimate indicates that a country with all in-network driving will have 0.35 fewer summonses per driver than a country with no in-network driving.

D.   Within-driver analysis

Column (4) exploits the panel nature of the data to include driver fixed effects. The effect is substantially larger than the OLS estimate and is statistically significant ($p < .01$). As discussed in Section 5, this estimate is likely to isolate the true causal effect of social ties on driver effort as it controls for time-invariant driver ability, and hence removes any bias due to (a) selection across countries into in-network driving, and (b) selection within countries but across drivers into in-network driving.

As a check of the identifying assumption that changes in in-network status between 2005 and 2007 are not associated with unobserved changes in factors that may affect driving outcomes, we break out the effect separately for drivers switching from in-network to out-of-network leasing between 2005 and 2007 versus from out-of-network to in-network leasing between 2005 and 2007. If the switching process is endogenous (for example, drivers tend to switch taxis after an accident but subsequently have safe driving outcomes), then the effect of switching into versus away from in-network driving might be different. Column (5) breaks out the effect for drivers switching away from in-network driving between 2005 and 2007 ("Lease in network") versus switching into in-network leasing between 2005 and 2007 ("Lease in network x lease out of network in 2005," which is additive to the "Lease in network" estimate). While drivers switching away from in-network driving show a larger effect, the estimates for both types

of switchers are much larger than the OLS estimates, and the two types are not statistically different (p=.18).[34]

E.   Instrumental-variables analysis

Column (6) and (7) contains the instrumental variables results, where the instrument for in-network driving is an indicator for whether the driver lives in the same Census block as an owner from the same country. The first-stage F-statistic for the excluded instrument is 131 and 41 depending on whether driver country fixed effects are included, indicating a strong instrument. Again the estimate of the effect of in-network leasing is statistically significant and substantially larger than the OLS estimates (p=0.05). When driver country fixed effects are included, the estimate is very similar, though not surprisingly is less precise given the weaker first stage F-statistic.[35,36]

In Table 6, we provide evidence from falsification tests that the instrument is uncorrelated with driver ability. In columns (1)-(6) we find no evidence of a systematic relationship between the instrument and observed measures we expect to reflect driver ability. Columns (1) and (2) show in a regression of experience, the strongest predictor of driving outcomes in our data, versus the instrument, that the instrument has approximately zero effect on experience. Columns (3)-(6) show that the instrument also has essentially no effect on driver test outcomes we might expect to predict driver ability, which is whether the driver passed an English language test and the driver's score on a driving test (these tests are described in Section 3). Finally, we examine a direct measure of unobserved driver ability, the estimate of α_i from the

_____

[34] Note that even though all drivers in the panel have two additional years of experience in 2007 versus 2005, the effect of log of experience is identified due to the nonlinearity of the measure. The nonlinear measure of experience is suitable given that the marginal benefit of experience almost certainly decreases with experience.

[35] Columns (6) and (7) also report the estimated coefficients on the excluded instrument in the first stage regression, showing that the instrument to be a very strong predictor of network status.

[36] Note that the sample size for these specifications is smaller than previous specifications since the instrument is based on driver and owner addresses, which are only available in the 2005 data.

within-driver model. This puts us in the unusual position of being able to test directly the identifying assumption of the instrumental variables model, that the instrument is uncorrelated with the unobserved error term. Columns (7) and (8) show the effect to be approximately zero.

F.   Analysis by summons type and alternative driving outcomes

If effort effects are indeed driving the observed network effects, we would expect network effect to be largest for the types of summonses that directly impact owners' costs. For example, we expect to find a relatively large effect for hazardous driving summonses since they may reflect more reckless driving and hence higher associated costs from accidents and mechanical defects for the owner; and for trip sheet, vehicle condition, and missing items summonses since owners receive a fine when drivers to whom they lease incur these summonses.

These results are more speculative since we cannot identify precisely the level of moral hazard associated with each summons type and because test power is significantly lower when analyzing individual summons types. Nevertheless, the results in Table 7, which reports both OLS and instrumental variables estimates, are generally consistent with effort effects.[37] In-network driving has the most significant effect on the three summons types that directly result in an owner violation: trip record violations, vehicle condition violations, and missing item violations. While the t-statistics for hazardous moving violations and for using the taxi unlawful purposes, which may also directly generate costs for the owner, are smaller, they have the correct sign, and the instrumental variables estimates are among the most negative.[38]

For the reasons discussed in Section 3, we have focused our analysis on summonses. However, we can also examine accidents and convictions for driving violations as alternative

---

[37] We do not include country fixed effects, nor country-level or within-driver estimates, since there is even less power to identify these models when summons are broken out by type. The general patterns however appear similar.
[38] Results from t-tests and not coefficient estimates are reported since the baseline numbers of summonses across types varies and hence are not directly comparable.

driving outcomes, albeit with less power for tests, and without a panel data component since only

2005 outcomes are available. Accident rate in particular is an attractive measure because it

captures an outcome that affects owners directly.

Columns (1)-(4) of Table 8 report results where the dependent variable is number of

accidents by the driver during the lease period. Columns (5)-(8) report results where the

dependent variable is number of convictions for driving violations. For each measure, we report

baseline OLS estimates, OLS estimates with fixed effects for driver country, country-level

estimates, and instrumental variables estimates. Reassuringly, all of the estimates have the

correct sign and a reasonable magnitude (although the country-level accident effect is somewhat

larger). Furthermore, the magnitudes of the effects of same-country leasing on accidents,

convictions, and summonses are all approximately proportional to their relative frequencies in

the data. The country-level and instrumental variables estimates, which are likely to isolate the

causal effect of in-network driving on driver effort, while imprecise, are again larger than the

OLS estimates, indicating an important network effect and also that in-network drivers are lower

ability on average than out-of-network drivers.

G.   Discussion of sample selection

Given that we only have data on individuals who became drivers, but not on prospective

drivers not already in the industry, readers may wonder whether our findings are driven by some

sample selection bias (as distinct from selection into in-network driving conditional on being in

the sample). Specifically, given that our sample consists of drivers in the industry and not

prospective drivers, readers may wonder if the better outcomes among in-network drivers are

driven by high-quality prospective drivers being more likely to enter the industry as in-network

drivers, while low-quality prospective drivers not entering the industry at all. Such sample

selection could potentially bias the driver-level cross-sectional results (OLS and 2SLS).

However, it is important to note that it would not drive the aggregate country-level results as

long as the sample selection process was the same across countries (because this selection within

country-of-birth will be averaged out at the country level), and it would not drive the within-

driver results (because all comparisons are within the same driver over time). Given the

robustness of our central findings across models that are, and are not, susceptible to sample

selection bias, we are confident that our results uncover a real causal effect. Moreover, all of our

strategies that account for selection (conditional on being in sample) indicate that the selection

into in-network driving is negative rather than positive, which is generally inconsistent with

sample selection driving our results.

7. EXPLORATION OF MECHANISMS

The preceding section provides a range of evidence indicating that social ties between

owners and drivers lead to significantly better driving outcomes in the NYC taxi-leasing market.

However, as discussed in Section 4, there are at least five mechanisms through which social

networks can operate, including screening, monitoring costs, social sanctions, social preferences,

and social collateral, and we would like to know more about which mechanisms are important.

First, we consider screening and monitoring costs. As discussed in Section 5, the

estimates of the effect of in-network driving from the country-level, within-driver, and

instrumental variables models are likely to reflect the true causal effect of in-network driving on

outcomes, while the estimates of the effect of in-network driving from the OLS model represent

both the causal effect of in-network driving and the effect of any selection of drivers into in-

network driving based on ability. Since the OLS estimates show a significantly smaller benefit of

in-network driving versus the other estimates, we have evidence that selection leads to in-network drivers having lower ability on average than out-of-network drivers. As the predictions from the theoretical model in Section 4 indicates, under screening, there should be selection of higher ability drivers into in-network leasing; while under no screening, there should be selection of lower ability drivers into in-network leasing. Since in-network drivers appear to have lower average ability than out-of-network drivers, screening is unlikely to be an important determinant of driver-owner matches and outcomes.

There are several reasons why monitoring costs are unlikely to explain the better in-network driving outcomes. First, taxis are operated primarily in Manhattan and typically far from the location of owners and members of the community (which are primarily in Brooklyn and Queens), and hence direct monitoring of lessee driving is infeasible at reasonable cost regardless of network status.[39] Second, most drivers in the sample lease from an owner-driver who himself regularly drives the taxi, giving the owner-driver some ability to regularly monitor vehicle condition again regardless of network status.

Second, we examine the effect of network density on driving outcomes to learn about social sanctions, social collateral, and social preferences. Following Bertrand, Luttmer, and Mullainathan (2000) and Gil and Hartman (2011), we use Census data on the density of residents from a particular country in a neighborhood as a measure of the network strength among residents from that neighborhood and country. Specifically, we measure driver/owner network strength as the fraction of residents in the driver/owner Census tract from the same country of

---

[39] One might argue that other drivers from the same country could observe a driver's driving style while on the road and report this information to the owner. However, we believe this scenario is very unlikely because of the small chance that any single driver while operating his taxi would observe another driver he knows exhibiting poor driving style out of the approximately 40,000 drivers that are actively driving in the industry (short-term lessees, long-term lessees, and owner-drivers), also happen to know the owner of that taxi, and be interested in reporting the observed driving style to that owner.

birth. While this measure may not capture network strength completely, it nonetheless should reflect the level of communication and ties between residents, for example, through the number of community organizations and events. The idea is to test whether denser owner or driver social networks generate larger improvements in outcomes, which may tell us about which mechanisms are important.[40]

Consider social sanctions, which would take the form of community-enforced punishments against an offending group member. These social sanctions would most likely operate through the owner's network (and not a driver's network) and hence a relatively dense owner network is likely central to facilitating social sanctions. Social collateral could operate similarly to social sanctions in that owners may provide or withhold favors to drivers based on driving outcomes and these favors could operate through the owner's network. Hence the effect of social sanctions and possibly social collateral would be reflected in an owner network strength effect. In contrast, social preferences would likely operate independently of an owner's or driver's network.

In Table 9 we report regression results where the dependent variable is a driver's number of summonses and the explanatory variables of interest are an indicator for in-network leasing and the interaction of in-network leasing and network strength. Columns (1), (2), and (3) report results driver network strength, owner network strength, and both, respectively. Comparing the estimates of the interaction terms indicates that owner networks are far more important than driver networks for generating good outcomes. This result can be seen most clearly in column (4), where a stronger owner network leads to significantly better outcomes for in-network drivers. Furthermore, driver network strength has no effect on summonses after conditioning on

---

[40] We examine 2005 outcomes only since address data, required to identify tracts, are available for 2005 only.

owner network strength, and when fixed effects for driver country are included, the standalone in-network effect approaches zero, suggesting that network effects rely in large part on the density of the owner's social network.[41]

In columns (5)-(8), the continuous measure of network strength is replaced with an indicator variable for whether at least 15 percent of residents in the owner/driver Census tract have the same country of birth. The idea is that a minimum density of residents might be necessary for a mechanism that operates through a local community to be effective, but that increasingly dense networks have marginally smaller benefits. Indeed, results from this measure are similar, but are now significantly more pronounced.[42]

The result that network effects appear to operate in large part through owner networks and not driver networks, and that the in-network effect is weak when the owner network is weak, suggests that social sanctions may be driving the observed effort effects. Social collateral could also play a role since there are ways in which social collateral could depend on owner network strength.[43] The limited stand-alone and driver network effects, however, are inconsistent with an important role for social preferences.

8. CONCLUDING REMARKS

Moral hazard plays a central role in our understanding of numerous contractual relationships, and researchers have found evidence consistent with moral hazard in many contexts. However, much less is known about how one might mitigate moral hazard problems

---

[41] The number of observations varies slightly across specifications due to differences in the availability of addresses.
[42] Fifteen percent was chosen because it corresponds to the level at which the marginal effect of network density appears to begin leveling off.
[43] One way that social sanctions and social collateral might operate is by making it easier for owners to collect damages from drivers with bad outcomes since damages may often be identifiable but not legally recoverable.

where contract are necessarily incomplete. While there is some mixed evidence that social ties can mitigate moral hazard problems in group lending schemes in developing countries, little is known on the potential role of social ties in mitigating moral hazard outside of this particular developing-country context.

In this study we investigate the role of social networks in aligning the incentives of economic agents in a United States labor market setting with incomplete contracts. Specifically, we compare the driving outcomes of drivers who lease from owners from the same country of birth to the driving outcomes of drivers who lease from owners from a different country of birth.

We find that across three models that exploit three distinct plausibly exogenous sources of variation to account for selection, drivers have better outcomes when driving in-network versus out-of-network ─ indicative of a real causal effort effect. Consistent with social ties reducing moral hazard, we find that these effects are largest for outcomes that impose a cost to the owner. Also the marginal effects of driving in-network increase with the strength of the owner network but not the driver network, and appear to be weak when the owner network is weak. This is consistent with social sanctions, in the form of community-enforced punishments, playing a primary role in facilitating the network effects.

Given the prevalence of in-network driving and the strength of the observed network effects in the taxi industry, our results indicate that social ties have incentive benefits in addition to any positive selection effects. In many labor markets, jobs are frequently obtained through referrals via social ties. It is commonly thought that the main benefit of these referrals is to overcome asymmetric information over candidate quality. However, our results suggest these referrals may also help to overcome the moral hazard problem. More generally, our findings provide direct evidence that even in developed nations with strong institutions, social

connections can play a valuable role in achieving better economic outcomes.

Finally, an interesting policy implication of the literature on micro-credit group lending is the possibility of introducing a joint liability program among groups of taxi drivers from the same network when there is an insufficient number of in-network owners from whom to lease. Given that responsibility for mechanical defects typically cannot be contractually assigned to any single driver, and that the costs of bad outcomes like accidents often cannot be recovered from individual drivers, group liability may induce a higher level of effort among these drivers and allow more flexibility in which owners and drivers can optimally form business relationships.

REFERENCES

Ahlin, C. and Townsend, R. (2007) "Using repayment data to test across models of joint liability lending," Economic Journal.

Alesina, A., R. Baqir, and W. Easterly (1999) "Public goods and ethnic divisions," Quarterly Journal of Economics, 114(4), 1243-84.

Baker, G. and T. Hubbard (2003) "Make versus buy in trucking: Asset ownership, job design, and information," American Economic Review.

Baker, G. and T. Hubbard (2004) "Contractibility and asset ownership: On-board computers and governance in the U.S. Trucking Industry," Quarterly Journal of Economics.

Bandiera, O., I. Barankay, and I. Rasul (2005) "Social preferences and the response to incentives: Evidence from personnel data," Quarterly Journal of Economics.

Bandiera, O. I. Barankay, and I. Rasul (2009) "Social connections and incentives in the workplace: Evidence from personnel data," Econometrica, 44(4), 1047-1094.

Banerjee, A. T. Besley, and T. Guinnane (1994) "The neighbor's keeper: The design of credit cooperatives with design and test," Quarterly Journal of Economics.

Banerjee, A., E. Duflo, Glennerster, R., and C. Kinnan (2009) "The miracle of microfinance? Evidence from a randomized trial," working paper.

Bertrand, M., Luttmer, E., and S. Mullainathan (2000) "Network effects and welfare culture," Quarterly Journal of Economics.

Besley, T. and S. Coate (1995) "Group lending, repayment incentives and social collateral," Journal of Development Economics, 46 (1), 1-18.

Blackwell, D. (1965) "Discounted dynamic programming," Annals of Mathematical Statistics, 36(1), 226-235.

Borjas, G. (2000) "Ethnic enclaves and assimilation," Swedish Economic Policy Review, 7, 89-122.

Duflo, E., A. Banerjee, R. Glennerster, and C. Kinnan (2009) "The miracle of microfinance? Evidence from a randomized evaluation," working paper.

Duflo, E. and E. Saez (2003) "The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment," Quarterly Journal of Economics, 118(3): 815-842.

Edin, P.-A., P. Fredriksson, and O. Aslund (2003) "Ethnic enclaves and the economic success of immigrants: Evidence from a natural experiment," The Quarterly Journal of Economics, 118(1), 329-357.

Ellison, G. (1994) "Cooperation in a prisoner's dilemma with anonymous random matching," Review of Economic Studies.

Fernandez, R. and N. Weinberg (1997) "Sifting and sorting: Personal contacts and hiring in a retail bank," American Sociological Review.

Fong, C. and E. Luttmer (2009) "What determines giving to Hurricane Katrina victims? Experimental evidence on racial group loyalty," American Economic Journal: Applied Economics.

Foster, A. and M. Rosenzweig (1994) "A test for moral hazard in the labor market: contractual arrangements, effort, and health," The Review of Economics and Statistics, 76(2), 213-227.

Gebers, M. (2003) "An inventory of California driver accident risk factors," California Department of Motor Vehicles, Research and Development Branch, RSS-03-204.

Gil, R. and W. Hartmann (2011) "Airing your dirty laundry: Vertical integration, reputational capital, and social networks," Journal of Law, Economics, and Organizations.

Giné, X. and D. Karlan (2006) "Group versus individual liability: Evidence from a field experiment in the Philippines," Yale University Economic Growth Center working paper 940.

Giné, X. and D. Karlan (2009) "Group versus individual liability: Long-term evidence from Philippine microcredit lending group," working paper.

Glaeser, E., D. Liabson, J. Schneinkman, and C. Soutter (2000) "An economic approach to social capital," Economic Journal.

Granovetter, M. (1974) *Getting a job: A study of contracts and careers*. The University of Chicago Press.

Greif, A. (1993) "Contract enforceability and economic institutions in early trade: The Maghribi Traders' coalition," American Economic Review.

Holmstrom, B. (1979) "Moral hazard and observability," Bell Journal of Economics, 10(1), 74-91.

Hungerman, D. (forthcoming) "Race and charitable church activity," *Economic Inquiry*.

Jackson, C. Kirabo and Elias Bruegmann, (2009). "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers," *American Economic Journal: Applied Economics*, 1(4), 85-108.

Jackson, K. (2010) "Match quality, worker productivity, and worker mobility: Direct evidence from teachers," NBER Working Paper. w15990.

Kandori, M. (1992) "Social norms and community enforcement," Review of Economic Studies.

Karlan, D. (2005) "Using experimental economics to measure social capital and predict financial decisions," American Economic Review.

Karlan, D. (2007) "Social connections and group banking," Economic Journal.

Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl (2009) "Trust and social collateral," The Quarterly Journal of Economics, 124(3), 1307-1361.

Lazear, E. (2000) "Performance pay and productivity," American Economic Review, 90(5), 1346-1361.

Luttmer, E. (2001) "Group loyalty and the taste for redistribution," *Journal of Political Economy*, 109(3), 500-528.

Mas, A. and E. Moretti (2009) "Peers at work," American Economic Review.

Montgomery, J. (1991) "Social networks and labor-market outcomes: Towards an economic analysis," American Economic Review.

Munshi, K. (2000) "Networks in the modern economy: Mexican migrants in the U.S. labor market," The Quarterly Journal of Economics, 118(2), 549-597.

Pellizzari, M. (2010) "Do friends and relatives really help in getting a good job?," The Industrial and Labor Relations Review.

Pitt, M. and S. Khandker (1998) "The impact of group-based credit programs on poor households in Bangladesh: Does the gender of participants matter?," Journal of Political Economy, 106.

Portes, A. and J. Sensenbrenner (1993) "Embeddedness and immigration: Notes on the social determinants of economic actions," American Journal of Sociology.

Rai, A. and T. Sjostrom (2004) "Is Grameen lending efficient? Repayment incentives and insurance in village economies," Review of Economic Studies.

Schneider, H. (2010) "Moral hazard in leasing contacts: Evidence from the New York City taxi industry," Journal of Law and Economics.

Sharma, M. and M. Zeller (1997) "Repayment performance in group-based credit programs in Bangladesh: An empirical analysis," World Development.

Smith, J. (1986) "Moral hazard, costly supervision, and agricultural mechanization" *The Journal of Developing Areas*, 21(1), 75-86.

Stevens, J. and Gilby, I. (2003) "A conceptual framework for non-kin food sharing: timing and currency of benefits," *Animal Behavior*, 67(4), 603-614.

Stiglitz, J. (1990) "Peer monitoring and credit markets," *World Bank Economic Review* 4, 351-366.

Urbanonomics (2004) "Taxi & Limousine Commission environmental impact statement socioeconomic conditions report," internal document prepared for the TLC.

Woodcock, S. (2009) "Match effects," working paper.

Wydick, B. (1996) "Credit access, human capital and class structure mobility," *Journal of Development Studies*, 35(6) 131-152.

Figure 1: Graphical evidence in support of the validity of the instrument
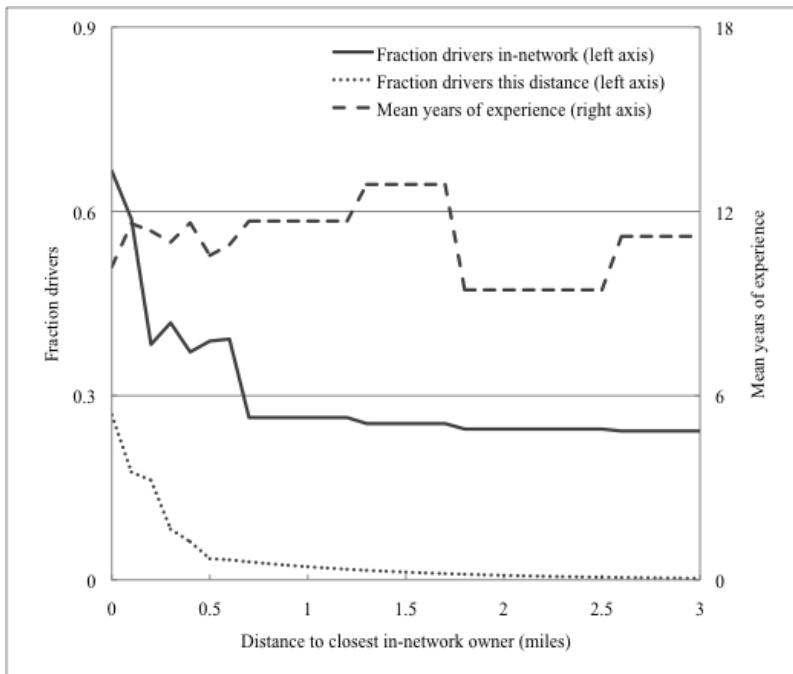
Notes: From top to bottom, the curves represent mean experience, fraction of drivers leasing in-network, and fraction of drivers at that distance, all versus distance to an in-network owner.

Figure 2: Country-level relationship between in-network leasing and summons

$$y = -0.350x + 0.495$$
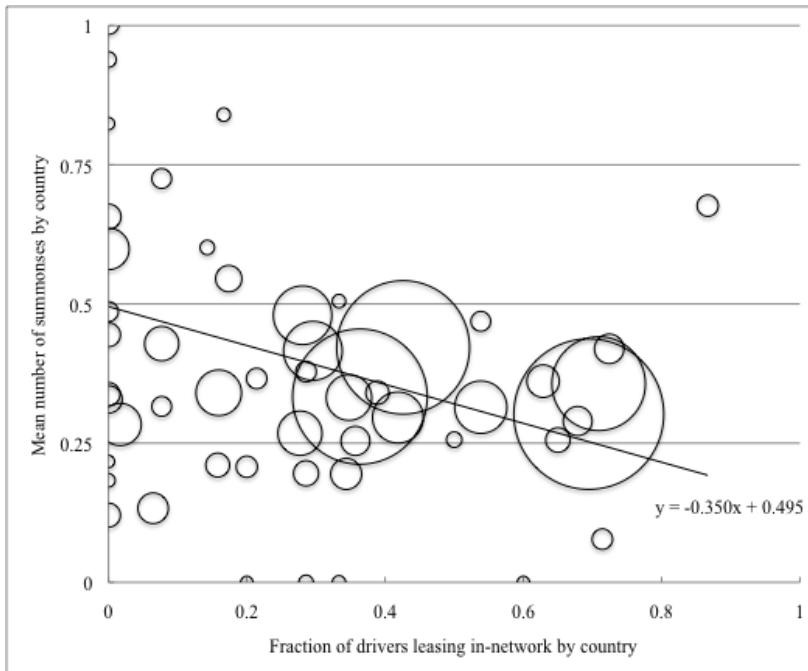
Notes: Circle size represents the number of drivers from that country. The line is the fitted relationships between fraction of drivers leasing in-network by country versus mean number of summonses by country, unweighted by number of drivers by country (the weighted fit is similar).

Table 1: Summary statistics

| | 2005 | | | | | | 2007 | | | |
	N	Mean	SD	Min	Max		N	Mean	SD	Min	Max
Years of NYC taxi-driving experience	1824	11.0	7.5	0.5	33.0		1877	11.2	8.0	0.8	35.4
Lease from same-country driver	1824	0.45	0.50	0	1		1877	0.42	0.49	0	1
Number of summons	1824	0.39	0.76	0	3		1877	0.33	0.68	0	3
Number of accidents	1824	0.04	0.21	0	2						
Number of convictions for driving violations	1824	0.17	0.46	0	3						
Passed English language test on first try	731	0.90	0.30	0	1		957	0.92	0.28	0	1
Driving test score on first try	743	78.6	10.5	0	100		967	78.2	10.5	12	100
Fraction residents in driver's Census tract from same country	1677	0.06	0.09	0.00	0.85						
Fraction residents in owner's Census tract from same country	1678	0.08	0.15	0.00	0.94						

Notes: The table shows summary statistics of the sample of drivers used in the regression analyses. Fewer observations are available for the English-language and driving test results because these data are only available for drivers entering the industry after 1997, and for fraction residents in driver/owner Census tract from the same country because of missing addresses. Accident, conviction, and address (and hence Census) data are only available for 2005.

Table 2: Availability of data on driver and owner country of birth

| | 2005 | | | | 2007 | | | |
| | | Country data available | | | | Country data available | | |
			In network	Out of network			In network	Out of network
All long-term lessees (N)	13,523	1,955	874	1,081	12,683	2,053	850	1,203
Mean experience (years)	9.3	10.6	11.1	10.2	9.6	10.7	11.8	9.9
Drive owner-driver taxis (N)	2,540	1,504	782	722	2,548	1,604	770	834
Mean experience (years)	10.3	10.9	11.1	10.6	10.7	10.9	11.8	10.0
Drive non-owner-driver taxis (N)	10,983	451	92	359	10,135	449	80	369
Mean experience (years)	9.1	9.9	11.3	9.5	9.3	10.2	11.9	9.8

Notes: The table shows number of drivers for whom country-of-birth data are available for both the driver and the corresponding owner, by ownership arrangement, along with mean experience levels.

Table 3: Patterns of matches between owners and drivers

	Owner country												
Driver country	Bangladesh	China	Colombia	Egypt	Greece	Haiti	India	Pakistan	Russia	U.S.	Ukraine	Other	Total
Afghanistan	0	0	0	0	0	1	0	0	0	1	0	13	15
Bangladesh	124	5	2	2	7	16	18	4	22	17	2	53	272
China	0	11	0	0	0	1	0	0	0	1	0	4	17
Colombia	1	3	26	0	1	2	2	0	0	2	0	13	50
Dom. Rep.	0	3	3	0	2	5	1	1	1	9	0	13	38
Ecuador	0	1	5	1	1	0	2	0	1	3	0	19	33
Egypt	1	1	0	18	3	9	3	2	11	2	0	14	64
Ghana	0	1	1	2	1	6	4	0	3	7	0	28	53
Greece	0	0	0	0	9	1	0	0	0	0	0	0	10
Haiti	0	3	1	0	3	101	1	0	8	12	4	18	151
Hong Kong	1	4	0	0	0	1	0	0	0	0	0	4	10
India	6	2	5	3	5	15	278	11	19	8	6	34	392
Morocco	1	0	1	1	1	6	1	1	1	1	0	10	24
Nigeria	0	0	1	0	0	4	0	0	2	0	1	6	14
Pakistan	0	0	1	6	5	29	36	106	22	21	4	55	285
Peru	0	0	2	2	0	1	0	0	0	0	0	5	10
Puerto Rico	1	0	2	0	2	1	0	1	0	4	1	6	18
Romania	0	0	0	0	1	0	0	0	2	0	0	13	16
Russia	0	0	0	1	0	2	0	1	21	2	10	14	51
Senegal	0	1	0	0	0	2	0	1	2	5	2	3	16
Turkey	0	0	0	0	0	0	2	1	1	0	0	3	7
U.S.	0	1	0	0	2	7	0	1	0	12	0	12	35
Ukraine	0	0	0	0	0	1	0	0	6	0	6	5	18
Vietnam	0	5	0	0	0	1	0	0	0	1	0	8	15
Other	3	3	12	7	7	23	11	6	27	19	4	88	210
Total	138	44	62	43	50	235	359	136	149	127	40	441	1,824

Notes: The table shows the country of birth for drivers and corresponding owners for the 2005 sample. Patterns for the 2007 sample are similar.

Table 4: Predictors of leasing from owner in same block

Driver resides in same block as in-network owner	0.321***
	[0.020]
Driver resides in same block as any owner	0.061***
	[0.017]
Constant	0.009
	[0.009]
Observations	1663
R-squared	0.249

Notes: The dependent variable is an indicator equal to one if the driver leases in-network and zero otherwise, and is limited to the 2005 sample since addresses are only available for this year. Heteroskedasticity robust standard errors are reported in brackets. *** indicates significance at the 1 percent level.

Table 5: Models of summonses

	(1) OLS	(2) OLS	(3) Country level	(4) Within driver	(5) Within driver	(6) 2SLS	(7) 2SLS
Lease in network	-0.085***	-0.078***	-0.308**	-0.323***	-0.472***	-0.248**	-0.233
	[0.023]	[0.026]	[0.125]	[0.092]	[0.146]	[0.126]	[0.227]
Lease in network x lease out of network in 2005					0.268		
					[0.201]		
Log years of experience	-0.103***	-0.108***	-0.095	-0.288*	-0.298*	-0.082***	-0.086***
	[0.014]	[0.015]	[0.079]	[0.167]	[0.167]	[0.022]	[0.025]
2007 indicator	-0.058**	-0.055**	-0.238***	-0.015	-0.013		
	[0.023]	[0.024]	[0.082]	[0.056]	[0.028]		
Constant	0.640***	0.647***	0.775***	1.102***	27.013	0.658***	0.686*
	[0.037]	[0.039]	[0.200]	[0.331]	[56.905]	[0.071]	[0.368]
First stage coefficient on lease in-network						0.312	0.189
First stage F-statistic						131.4	40.9
Fixed effects for driver country	No	Yes	No	No	No	No	Yes
Fixed effects for driver	No	No	No	Yes	Yes	No	No
Observations	3701	3701	174	3701	3701	1663	1663
R-squared	0.024	0.063	0.084	0.033	0.035	0.007	0.063

Notes: The dependent variable is number of summonses per driver. The 2SLS sample is limited to 2005 since addresses are only available for this year. Heteroskedasticity robust standard errors that are adjusted for clustering at the country of birth level are reported in brackets. *, **, *** indicate significance at the 10, 5, and 1 percent levels.

48

Table 6: Falsification tests for the excluded instrument

	(1) Log years of experience	(2) Log years of experience	(3) Passed English test	(4) Passed English test	(5) Driving test score	(6) Driving test score	(7) Driver effect from panel	(8) Driver effect from panel
In-network owner in block	-0.086	-0.075	-0.003	-0.018	0.880	1.520	0.005	-0.001
	[0.055]	[0.058]	[0.027]	[0.031]	[0.974]	[1.085]	[0.038]	[0.042]
Constant	2.074***	2.072***	0.903***	0.906***	78.363***	78.210***	0.015	0.016
	[0.026]	[0.025]	[0.013]	[0.013]	[0.460]	[0.470]	[0.021]	[0.021]
Fixed effects for driver country	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1663	1663	663	663	674	674	1663	1663
R-squared	0.001	0.144	0.000	0.108	0.001	0.086	0.000	0.073

Notes: The dependent variables are measures of driver ability (reported at the top of each column) and the explanatory variable of interest is the excluded instrument (in-network owner in block). The sample is limited to 2005 since addresses are only available for this year. Heteroskedasticity robust standard errors that are adjusted for clustering at the country of birth level are reported in brackets. *** indicates significance at the 1 percent level.

Table 7: Results by individual summons type

Summons for TLC driver violation	N OLS (2005, 2007)	t-stat OLS (2005, 2007)	N 2SLS (2005)	t-stat 2SLS (2005)	Corresponding TLC owner violation or risk
Non-hazardous moving violation	169	0.95	84	0.28	
Stationary vehicle violation	104	0.00	30	-0.06	
Shift violation (primarily cell phone and smoking)	79	-0.37	35	-2.39	
Abuse (primarily verbal abuse)	64	-0.41	39	-0.10	
Passenger refusal	118	-0.43	43	-0.26	
Hazardous moving violation	95	-0.44	60	-1.99	Risk to taxi
Passenger request violation	61	-0.69	20	1.10	
Noncooperation with TLC	104	-0.72	74	-0.12	
Using taxi for unlawful purpose	211	-0.98	96	-1.32	Risk to taxi
Off-duty procedures violation	56	-1.02	36	-0.81	
Discourteous to passenger	85	-1.32	44	0.03	
Trip records violation	153	-1.70	90	-2.12	Missing driver records violation
Vehicle condition violation (primarily safety related)	84	-2.06	34	0.36	Vehicle condition violation
Item missing from taxi (primarily trip records)	184	-3.42	97	-1.82	Item missing from taxi (primarily trip records)

Notes: The table reports sample size and the t-statistic for regressions of in-network driving on the probability of the driver incurring the summonses type listed in the left-most column. T-statistics for OLS regressions and for 2SLS regressions (available for 2005 only) are listed. The corresponding risk to the owner from the particular summons type is listed in the right-most column.

Table 8: Models of alternative driving outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Accidents				Convictions		
	OLS	OLS	Country level	2SLS	OLS	OLS	Country level	2SLS
Lease in network	-0.022**	-0.020*	-0.093**	-0.031	-0.035*	-0.030	-0.068	-0.057
	[0.010]	[0.012]	[0.047]	[0.038]	[0.021]	[0.025]	[0.103]	[0.079]
Log years of experience	-0.005	-0.002	-0.067**	-0.007	-0.017	-0.017	-0.104**	-0.017
	[0.006]	[0.007]	[0.030]	[0.007]	[0.012]	[0.013]	[0.049]	[0.014]
Constant	0.067***	0.059***	0.225**	0.074***	0.221***	0.220***	0.411***	0.224***
	[0.014]	[0.016]	[0.086]	[0.018]	[0.030]	[0.032]	[0.124]	[0.040]
Fixed effects for driver country	No	Yes	No	No	No	Yes	No	No
Observations	1839	1839	96	1674	1870	1870	97	1704
R-squared	0.003	0.065	0.086	0.005	0.003	0.072	0.057	0.002

Notes: The dependent variable in columns (1)-(4) is number of accidents and in columns (5)-(8) is number of convictions for driving violations per driver. The sample is limited to 2005 since accidents and convictions data are only available for this year. Heteroskedasticity robust standard errors that are adjusted for clustering at the country of birth level are reported in brackets. *, **, *** indicate significance at the 10, 5, and 1 percent levels.

Table 9: Models of the effect of network strength

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lease in network	-0.093**	-0.079**	-0.074*	-0.026	-0.107***	-0.092**	-0.096**	-0.048
	[0.040]	[0.040]	[0.042]	[0.050]	[0.037]	[0.037]	[0.038]	[0.044]
Lease in network x driver network strength	-0.211		0.114	0.076	-0.004		0.078	0.086
	[0.226]		[0.327]	[0.353]	[0.092]		[0.108]	[0.119]
Lease in network x owner network strength		-0.397*	-0.665*	-0.634*		-0.152**	-0.218***	-0.221***
		[0.214]	[0.355]	[0.384]		[0.067]	[0.075]	[0.085]
Log years of experience	-0.094***	-0.091***	-0.087***	-0.087***	-0.095***	-0.091***	-0.087***	-0.087***
	[0.021]	[0.022]	[0.022]	[0.025]	[0.021]	[0.022]	[0.022]	[0.025]
Constant	0.625***	0.616***	0.608***	0.587***	0.626***	0.616***	0.608***	0.586***
	[0.054]	[0.054]	[0.056]	[0.060]	[0.054]	[0.054]	[0.056]	[0.060]
Fixed effects for driver country	No	No	No	Yes	No	No	No	Yes
Observations	1677	1678	1550	1550	1677	1678	1550	1550
R-squared	0.020	0.020	0.020	0.109	0.020	0.020	0.020	0.110

Notes: The dependent variable is number of summonses per driver. Network strength in columns (1)-(4) is fraction of residents in the driver's/owner's Census tract from the same country, and in columns (5)-(8) is an indicator for at least 15 percent of residents from the same country. Heteroskedasticity robust standard errors that are adjusted for clustering at the country of birth level are reported in brackets. *, **, and *** indicate significance at the 10, 5, and 1 percent levels.

APPENDIX: SUPPLEMENTARY RESULTS (FOR ONLINE REFERENCE ONLY)

Drivers leasing from a family member

Previous work in the context of micro-credit group lending finds evidence that network effects are weaker in groups where members have sufficiently close relationships, such as family members, suggesting a breakdown of the social sanctions mechanism (Sharma and Zeller (1997), Ahlin and Townsend (2007), and Karlan (2007) provide empirical evidence and Rai and Sjostrom (2004) have a theoretical analysis). We examine this issue in the taxi industry by comparing the outcomes of drivers who lease from a family member versus the outcomes of drivers who lease in-network but not from a family member and drivers who lease out-of-network. Family member are identified as drivers who are registered as leasing from an owner with the same last name.[44] The sample consists of 326 drivers who lease from a family member (9 percent), 1,290 lessees who lease in-network but not from a family member (35 percent), and 2,085 lessees who lease out-of-network (56 percent). The categories are defined to be mutually exclusive such that in-network and family member do not overlap.

The regression results are provided in Table 10 and show that drivers leasing from a family member have 0.13 and 0.10 fewer summons than out-of-network drivers versus non-family in-network drivers having 0.08 and 0.08 fewer summons than out-of-network drivers, depending on whether fixed effects for the driver's country of birth are included. Thus, the results suggest that family members have even better driving outcomes than non-family in-network drivers, and that the effect of social ties is monotonic in the strength of that relationship,

---

[44] This method of identifying family members is imperfect: Drivers identified as family members are very likely to actually be family members since sharing a last name with drivers operating the same taxi is unlikely to be coincidental. However, some drivers who are family members may not be identified as such since drivers related to owners by marriage, for example, would not have the same last name. Nevertheless, family members classified in the data as non-family members are likely to represent only a small fraction of total lessees, and hence this misclassification is unlikely to affect the results meaningfully (and if anything will bias the estimate of the family effect towards zero).

although of course it is possible that the selection process by which drivers come to lease from a family member versus a non-family member but same-country owner may differ in systematic ways.

In-network drivers versus owner-drivers

Table 11 shows how in-network lessees fare versus the owner-drivers from whom they lease, both relative to out-of-network lessees. As expected, the owner-drivers show somewhat better outcomes in terms of fewer summonses for driving violations. Note however that the selection process based on ability by which drivers come to lease in-network versus become an owner-driver may be very different, and hence a comparison of the magnitudes of these estimates is only suggestive.

Alternative models and samples

Table 12 reports results from the basic model of number of summons versus in-network leasing, but uses alternative modeling functions, including a Poisson regression model and a more flexible negative binomial regression model. Estimates for the later two models are reported as marginal effects. The results are nearly identical. The estimates in Table 13 provide results regarding the effects of in-network leasing for alternative samples. Columns (1)-(4) provide results for the 2005 and 2007 samples individually, showing that the in-network leasing effect appears robust to the sample period.

Table 10: Outcomes of drivers leasing from a family member

	(1)	(2)
Lease in network from non-family	-0.077***	-0.076***
	[0.025]	[0.027]
Lease from family	-0.132***	-0.101**
	[0.039]	[0.042]
Log years of experience	-0.104***	-0.109***
	[0.014]	[0.015]
2007 indicator	-0.058**	-0.055**
	[0.024]	[0.024]
Constant	0.643***	0.649***
	[0.037]	[0.039]
Fixed effects for driver country	No	Yes
Observations	3701	3701
R-squared	0.025	0.063

Notes: The dependent variable is number of summonses per driver. Robust standard errors adjusted for clustering at the country of birth level are reported in brackets. ** and *** indicate significance at the 5 and 1 percent levels.

Table 11: Outcomes of owner-drivers

	(1)	(2)
Lease in-network	-0.084***	-0.068***
	[0.023]	[0.025]
Owner-driver	-0.122***	-0.105***
	[0.021]	[0.023]
Log years of experience	-0.106***	-0.104***
	[0.012]	[0.014]
2007 indicator	-0.042**	-0.043**
	[0.017]	[0.017]
Constant	0.637***	0.624***
	[0.033]	[0.036]
Fixed effects for driver country	No	Yes
Observations	5989	5989
R-squared	0.034	0.060

Notes: The dependent variable is number of summonses per driver. Robust standard errors adjusted for clustering at the country of birth level are reported in brackets. ** and *** indicate significance at the 5 and 1 percent levels.

Table 12: Models of summonses with alternative regression functions

	(1)	(2)	(3)
	OLS	Poisson	Negative binomial
Lease in-network	-0.085***	-0.084***	-0.084***
	[0.023]	[0.023]	[0.023]
Log years of experience	-0.103***	-0.092***	-0.091***
	[0.014]	[0.011]	[0.011]
2007 indicator	-0.058**	-0.055**	-0.057**
	[0.023]	[0.022]	[0.022]
Constant	0.640***		
	[0.037]		
Fixed effects for driver country	No	No	No
Observations	3701	3701	3701

Notes: The dependent variable is number of summonses per driver. Estimates in columns (2) and (3) are reported as marginal effects evaluated at the means of other regressors. Robust standard errors adjusted for clustering at the country of birth level are reported in brackets. ** and *** indicate significance at the 5 and 1 percent levels.

Table 13: Models of summonses with alternative samples

	(1)	(2)	(3)	(4)
	2005	2005	2007	2007
Lease from same-country owner	-0.108***	-0.055	-0.063**	-0.101***
	[0.035]	[0.040]	[0.031]	[0.034]
Log years of experience	-0.096***	-0.095***	-0.111***	-0.117***
	[0.021]	[0.023]	[0.018]	[0.020]
Constant	0.636***	0.610***	0.589***	0.617***
	[0.052]	[0.055]	[0.046]	[0.051]
Fixed effects for driver country	No	Yes	No	Yes
Fixed effects for driver	No	No	No	No
Observations	1824	1824	1877	1877
R-squared	0.021	0.097	0.025	0.064

Notes: The dependent variable is number of summonses per driver. Robust standard errors adjusted for clustering at the country of birth level are reported in brackets. ** and *** indicate significance at the 5 and 1 percent levels.