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

by

C. Kirabo Jackson
Northwestern University
School of Education and Social Policy
2120 Campus Drive, Room 204
Annenberg Hall
Evanston, IL 60208
kirabo-jackson@northwestern.edu

and

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

PRELIMINARY – THIS VERSION: DECEMBER 7, 2010

ABSTRACT

We investigate the role of social networks in aligning the incentives of 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 due to a moral hazard associated with incomplete contracts. 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 via social sanctions; and (3) network benefits can help to explain the industry organization in terms of which drivers and owners form business relationships.

*We would like to thank Andrew Salkin at the NYC Taxi and Limousine Commission for helpful discussions and data, Ricard Gil, Justin Johnson, Jonathan Peterson, Lan Shi, Michael Waldman, and seminar participants at Cornell University and University of Washington for valuable comments, and James Cowan and Jonathan Peterson for excellent research assistance.

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 suboptimal 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]

Contracts that exploit social ties via social sanctions and lower monitoring costs have been proposed as a way to reduce moral hazard (Stiglitz 1990). Karlan et. al. (2009) argue that network connections generate "social collateral" because the possibility of losing friendships secures informal transactions the same way that physical collateral can secure formal lending. Banerjee, Besley, and Guinnane (1994), Besley and Coate (1995), and Wydick (1996) argue that it is the strength of group pressure on potential defaulters that matters rather than social ties *per se*. Portes and Sensenbrenner (1993) argue that altruism between members can align incentives.

Nearly all evidence on social ties and moral hazard relates to group-lending in developing countries (where markets are not well defined, credit is constrained, and property rights are not well enforced). [2] Moreover this evidence is mixed.[3] Given the importance of moral hazard, and theoretical work that social ties might mitigate moral hazard, we aim to provide empirical evidence on these issues in a developed nation context, in a well-defined market, and outside of

---

[1] For example, worker effort is higher under piece rate versus hourly wage (Foster and Rosenzweig 1994, Lazear 2000), and make-versus-buy patterns in trucking may reflect these concerns (Baker and Hubbard 2003, 2004).
[2] We are aware of only two empirical studies on social ties and moral hazard: Greif (1993), who finds that community enforcement limited moral hazard in 11th-century Mediterranean trade; and Bandiera, Barankay, and Rasul (2005, 2009), who find that social ties lead to lower productivity among college student fruit pickers.
[3] Social ties are argued to help explain the success of group-lending banks (Pitt and Khandker 1998, Banerjee, Duflo, Glennerster, and Kinnan 2009). Kandori (1992) and Ellison (1994) show theoretically how enforcement may operate, while Sharma and Zeller (1997), Karlan (2007), and Ahlin and Townsend (2007) show evidence for group-lending, though Giné and Karlan (2006, 2009) find no benefit of group lending. Social interactions are also found to be economically relevant elsewhere (Bertrand, Luttmer, and Mullainathan 2000, Glaeser, Liabson, Scheinkman, and Soutter 2000, Duflo and Saez 2003, Karlan 2005, Mas and Moretti 2009, Jackson and Bruegmann 2009).

the microcredit context. Specifically, we examine New York City (NYC) taxi drivers who lease their taxis, comparing the outcomes of drivers who lease from an owner from the same country of birth to the outcomes of drivers who lease from an owner from a different country of birth.

The taxi leasing market is attractive for studying moral hazard because lessee-drivers pay less or none of many of the variable costs they generate, including for vehicle maintenance, repair, replacement, and insurance, and hence have incentives to choose inefficient levels of vehicle care. The problem is that the driver's driving style cannot be monitored at reasonable cost, and hence leasing contracts cannot fully specify the driver's level of vehicle care.[4] Schneider (2010) finds that a significant fraction of accidents and driving violations for lessee-drivers is due to this moral hazard. This market is also attractive for studying social ties since most owners and drivers are immigrants to the United States, representing 146 countries, which allows us to compare outcomes both within and across many country-of-birth networks. We use country of birth to measure social ties since this has been found to be important elsewhere.[5]

Our approach is to compare the driving outcomes of drivers who lease from owners from the same country of birth (in-network) to those of drivers who lease from owners from a different country of birth (out-of-network). The empirical difficulty in isolating the causal (effort) effect of in-network driving is that the outcomes of in-network drivers versus out-of-network drivers may either be due to effort effects (i.e., in-network drivers may exert less/more effort in their driving style to limit the costs borne by the owner because of greater potential social sanctions, social collateral, social preferences, or a greater ease of monitoring) or because of differential selection into in-network leasing by driving ability (i.e., owners may be better positioned to identify and

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

hire high ability drivers when they are in-network, or alternately owners may be willing to hire lower ability drivers from in-network because they know that they will exert more effort). As such, our analysis will focus on isolating the casual (effort) effect of driving in-network.[6]

Because there is no experimental variation, our approach is not to provide one single estimate of the causal (effort) effect from a single research design, but to present a body of evidence based on several credible but imperfect research designs aimed at removing selection effects that rely on distinct sources of variation and different identifying assumptions.

Our first strategy is to compare the *change* in outcomes of drivers who switch between in-network and out-of network leasing over time to the outcomes of drivers who always drive in-network or never drive in-network over the same period ─ removing bias due to high ability drivers being more or less likely to lease in-network. Our second strategy exploits cross-sectional variation in the residential locations of drivers relative to owners *from the same country of birth*. We use the distance from a driver's residence to the nearest same-country owner's residence as an exogenous instrument for driving in-network while also conditioning on the place-of birth-network. We present empirical tests to support the validity of both these identification strategies. As supporting evidence, we then aggregate the data to the country level to compare driving outcomes for countries with high rates of in-network driving to outcomes for countries with low rates of in-network driving ─ removing selection bias *within* a county of birth network. While each of these three strategies individually may be subject to distinct sources of bias, it is unlikely that all three sources of variation are. Thus, while none of these designs alone are dispositive, consistent results across approaches would indicate a real causal effect.

We first show that in-network driving is much more common than would be expected by

[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.

random chance, and cannot be explained by residential clustering by country of birth— consistent with findings in Gil and Hartmann (2011) that markets may organize around common-origin communities to mitigate agency problems. Across all models, drivers have better outcomes when in-network versus out-of-network — indicating a positive effort effect. The effects are more negative in models that account for selection (relative to OLS), indicating *negative* selection into in-network driving — suggesting that owners do not screen for high ability in-network drivers but rather hire lower ability drivers in-network because these drivers exert more effort.

We then provide several robustness checks to provide evidence against owner selection, and to show that the largest effects are for summons types directly affecting owners' costs (as opposed to drivers' costs, which drivers would internalize). We also investigate the mechanisms. Network effects increase with the owner's network strength but not the driver's network strength — consistent with the presence of social sanctions in the form of community-enforced punishments or social collateral, which would likely emanate from the owner's network.

This paper adds breadth to the literature on social networks and moral hazard by being the first study to show directly the potential of social networks to reduce moral hazard in a labor market context and in a developed nation. That is, even in a setting in which credit is available to finance a taxi purchase, leasing is commonplace among members of the same country network, indicating that social networks significantly reduce the cost of using the market.[7]

The rest of the paper proceeds as follows. Section 2 describes the NYC taxi industry. Section 3 describes the data. Sections 4 describes our empirical frameworks. Section 5 describes the main results. Section 6 examines the underlying mechanisms. Section 7 concludes.

[7] An active credit market exists in which specialized lenders market loans to taxi drivers to buy their own taxis.

1. THE NEW YORK CITY TAXI INDUSTRY

To operate a taxi in NYC, a driver must possess a taxi license (often called a medallion).[8] These licenses come in two forms: An "owner-must-drive" license, which requires the license owner to drive full-time (i.e., 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. As such, drivers operate in one of three ways: Owner-driving, in which the driver owns the license and vehicle; driver-owned-vehicle-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 co-owner-drivers, or one owner-driver and one lessee-driver who leases from this owner-driver. For reasons explained below, we primarily examine lessee-drivers in this third type of driving arrangement, in which the lessee-driver leases the taxi (both the license and vehicle) for one twelve-hour shift per day from an owner-driver who operates the taxis for the second twelve-hour shift per day.

The costs and revenues of operation are split between lessee-drivers and owner-driver as follows. The owner is required by the NYC Taxi and Limousine Commission (TLC) to pay for no-fault and liability insurance and workers' compensation for the lessee-driver, licensing fees, vehicle inspections, and fines for improper taxi use, some of which may be due to lessee actions. Owners also pay all vehicle maintenance, repair, and replacement.[9] Lessees pay TLC fees, for

---

[8] 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).

[9] Note that the vehicle owner pays all repair costs. Since we are primarily examining drivers who lease from owner-

their own gas usage, tolls, parking tickets, and some of their DMV and TLC fines. 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. Lessee-drivers keep all fares and tips that they generate.

Long-term lessees often leave a modest deposit with the owner, which can be forfeited in the case of an at-fault accident. During the sample period, the deposit amount and lease fee were capped by the TLC at $500 and $650, respectively, which as mentioned industry participants indicate is respected. Note that these amounts are small compared to the cost of mechanical defects (approximately $7,000 per taxi/year on average), accident damage (up to $28,000 to replace the vehicle), and owners' insurance premiums (approximately $10,000 per taxi/year). Thus, while we cannot ensure that collateral and fee are uniform across drivers (we do not have these data), owners have limited leeway in varying contract terms relative to the operating costs.

Lessee-drivers generate three types of costs that are borne by owners. The first is repair costs from vehicle mechanical failures and maintenance. Given that NYC taxis are operated in congested urban conditions for 175 miles per day on average, driver care can quickly show up in vehicle condition.[10] The second is from accidents. Approximately one-third of taxis are involved in a serious accident per year (defined as having over $1000 in damage or involving injuries). Driver aggressiveness such as speeding or cutting across traffic to pick up a hailing passenger increase the chance of accidents. The third is fines to owners from driver behavior, ranging from $25 to $100 per instance, when drivers violate certain TLC regulations (discussed below). We can see that drivers pay only a fraction of the operating cost, but keep all of the fare revenue.

---

drivers, it is these owner-drivers who pay repair costs.

[10] For example, some taxi owners indicated to us that brakes must be replaced every few weeks. Thus, whether drivers aggressively accelerate and brake, take care to avoid curbs, and so on can determine an owner's repair costs.

Since revenues and costs are increasing in driver aggressiveness, drivers face a moral hazard.

2. THE DATA

A. Description of the TLC data

Data from the TLC contain information about all drivers of NYC yellow taxis during 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 accident and summons data originate from the New York City Police Department. 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 (the other two outcomes originate from the NYC police department, and were not available for 2007).[11] 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 greater variability, we focus our analysis on summonses (discussed below). However, we show that our results are robust for accidents and convictions.

The data also record each driver's country of birth and 2005 address, and for drivers

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

entering after 1997, an English-language test outcome and a written driving test score, both of which drivers must pass to obtain a taxi-driving license.[12] We calculate a driver's NYC taxi-driving experience from the number of months registered as an active NYC taxi driver, which we derive from his taxi driver's license numbers.[13]

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.[14] 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 these matches constitute the driver-of-record file.[15] 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 know the driver-taxi match on these two dates, we restrict analyses 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 additionally use Census data on 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),

---

[12] The driving test is on NYC geography and DMV and TLC regulations.
[13] Taxi-driving license numbers are issued sequentially, which allows us to identify when a license was issued.
[14] 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.
[15] 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.

which we match to drivers/owners using their addresses.[16] Table 1 reports summary statistics for the primary variables in our sample. Drivers have 11 years of NYC taxi-driving experience on average, 44 percent of drivers lease in-network, and approximately one in three drivers receives a summons, one in six a conviction, and one in twenty a serious accident per six month period.

B. Discussion of driving outcome measures

Our data do not contain measures of vehicle maintenance, repair, and replacement costs borne by the owner.[17] Instead 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.

Our third measure is summons, which are given for a range of TLC violations (as opposed to DMV driving violations). Some 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 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. While not all violations affect might owner costs directly, overall violations may still reflect general driving style and possibly a moral hazard via the level of vehicle usage

[16] The country of birth recorded in the TLC data is occasionally more precise than the country of birth in the Census data. Any additional countries are assigned to the broader Census categories, such as "Other western Europe."
[17] 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.

since the number of violations likely increases with time on the road, and drivers may operate for the taxi for more time within a lease period since they do not pay the full operating cost.

We believe some types of summonses 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

Our analysis relies on country-of-birth data being available for both the lessee-driver and also the owner from whom he leases, in order 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 NYC taxi driving experience of these drivers.[18] Country data for both the lessee and owner are available for 1,955 and 2,053 lessees in the 2005 and 2007 data respectively. For driver experience and all other observable characteristics (race, English language test, driving-test (not shown)), those with data are similar to those without, suggesting that the drivers we analyze are similar the drivers for which country data are missing.[19]

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).[20] 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 likely to be invested in the

[18] Table 1 reports slightly fewer observations than Table 2 because Table 1 describes only observations where country-level data are available and data are non-missing for all other variables in the regression analysis, while Table 2 only describes observations where country data are available regardless of the availability of other variables.
[19] 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.
[20] 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.

condition of their vehicle (since they drive most days) and interact with the lessee regularly, thereby facilitating a social connection, as opposed to non-driving owners, who often own hundreds of taxis, and hence may have more anonymous relationships with drivers. Table 2 shows that a small fraction of lessee-drivers that have country-level data available lease from non-driving owners. The results are similar and somewhat stronger when they are excluded.

D. Prevalence of In-Network Leasing

In our data, 44 percent of drivers are in-network.[21] This suggests there are benefits to in-network driving. However, it is possible that a "proximity" effect, based on residential clustering by country of birth, rather than an "in-network" effect, may be driving this result. We test this hypothesis directly. If the countries of birth of the owners and drivers do not determine the owner-lessee match, then conditional on having any owners in the same Census block, the owner's country should not affect the likelihood of leasing from an owner who lives in the same block. In fact, 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 (conditional on having any owner). Also, conditional on having an owner from the same country in the same block, having any owner (from any other country) in the block has a much smaller (though still significant) effect on leasing from that owner. This suggest that while proximity is certainly an important factor, the prevalence of in-network driving cannot be explained by proximity alone ─ indicative of real benefits to leasing in-network.

3. EMPIRICAL STRATEGY

We aim to: (1) remove any selection bias to uncover the causal (effort) effect of driving a

[21] See Table 9 in the online appendix for the distribution of driver countries for the largest owner countries.

taxi owned by someone from the same country; and (2) document the degree and direction of any selection on ability into in-network driving. To identify the casual effect of in-network driving, we use two distinct sources of arguably exogenous variation in in-network driving. Results from these distinct strategies offer robustness checks 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] $$Y_{it} = \beta \cdot Same_{it} + \gamma \cdot ln(Exp_{it}) + \alpha_i + \varepsilon_{it}$$

In [1], Y_{it} is the outcome of driver *i* at time *t*, which is the number of accidents, convictions, or summonses, $Same_{it}$ 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 (to the researcher) 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 β from [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. 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

models that isolate the casual effect of being a same-country driver will be informative about the selection process. We describe two strategies to remove the effects of selection below.

B. Within-driver identification strategy

The first 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,

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

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. This is one motivation for our use of another completely different source of variation to identify effort effects.[22]

C. Cross-sectional instrumental variables strategy

[22] Nevertheless, switchers and non-switchers are observably similar. 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 10 switchers are from Bangladesh, Pakistan, India, Haiti, and other countries, respectively, which is proportionally similar to non-switcher countries of birth.

In our second approach, 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. 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.[23] Figure 1 also shows that distance to a same-country owner is unrelated to driving experience; 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 4.[24]

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

---

[23] 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.

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

outcomes. Specifically we estimate the following system of equations by 2SLS,

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

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

In [3] and [4] all variables are defined as before (but there is no longer any time variation), θ_c is a driver country of birth fixed effect, and the error term ν includes both the idiosyncratic variation and driver ability. With driver country of birth fixed effects, this model exploits variation in residential locations of drivers *who come from the same country of birth* to remove selection to in-network driving. We also include fixed effects for the *owners* country of birth, θ_{co}, to account for any systematic differences across the places of birth of the owners.

4. RESULTS

A. Baseline analysis

Table 3 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 per six month period ($p < .01$), which is a modest-size effect relative to the standard deviation of the number of summonses per driver of 0.72. Both driver and owner country fixed effects are included in column (2) to address the possibility that drivers or owners 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.[25,26]

B. Within-driver analysis

Column (3) exploits the panel nature of the data to include driver fixed effects to remove

[25] Results are similar with Poisson and negative binomial regression models, and are available in an online appendix.
[26] In all specifications, including driver age and/or Census neighborhood characteristics, such as median income and education level, has at most a small effect on outcomes after controlling for experience, and no effect on the estimate of in-network effect. We therefore use the more parsimonious specification that includes only experience.

any bias due to (a) selection across countries into in-network driving, and (b) selection within countries but across drivers into in-network driving. The within-driver estimate indicates that drivers leasing in-network have 0.334 fewer summonses per six month period ($p < .01$). This is substantially larger than the OLS estimate.

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 the 25 drivers switching from in-network to out-of-network leasing between 2005 and 2007 versus the 33 drivers switching from out-of-network to in-network leasing between 2005 and 2007. If switching is endogenous (for example, if drivers tend to switch after an accident but then have good outcomes), then the effects of switching into versus out of network might be different. Column (4) 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 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 ($p=.18$).[27]

C. Instrumental-variables analysis

Column (5) and (6) 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 and owner country fixed effects are included.

[27] 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.

The instrumental variables estimate without country fixed effects indicates that drivers leasing in-network have 0.251 fewer summonses per six month period ($p < .05$). Again the estimate is substantially larger than the OLS estimates ($p=0.05$). When driver and owner country fixed effects are included, the estimate is very similar, though less precise given the weaker first stage.[28,29]

In Table 4, we provide evidence from falsification tests that the instrument is uncorrelated with driver ability by regressing our instrument on proxies for driver ability. Columns (1) and (2) show that the instrument has approximately zero effect on experience. Columns (3)-(6) show that the instrument also has essentially no effect on whether the driver passed an English language test or the driver's score on a driving test (these tests are described in Section 2). Finally, we examine a direct measure of unobserved driver ability. Specifically, because we have longitudinal data while our instrument relies on cross-sectional variation, we can obtain estimates of driver ability (α_i) from equation [1] using a fixed effects model and then see if these estimates are correlated with the instrument.[30] Columns (7) and (8) show that the relationship between ability and our instrument is approximately zero.

D. Country-level analysis (robustness check)

Our identification strategies above exploit variation in network status within country of birth networks. For additional evidence that our results are not driven by selection within a network, we also aggregate the data to the country level and use only cross-country variations.

We begin by showing in Figure 2 the relationship between the mean numbers of

---

[28] 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.

[29] 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.

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

summonses per driver by country versus the fraction of drivers from that country leasing from an owner from the same country.[31] The figure shows that driving outcomes are better for drivers from countries with more in-network leasing. We investigate the relationship more formally by estimating equation [1] aggregated at the country of birth level. To address to concern that in-network leasing may be correlated with average driver ability across countries, we control for the average number of summonses of *owners* from the same country of birth as the lessee.

Column (7) of Table 3 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). 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. Because the cross-country estimates could be subject to bias from selection *across* networks but not from selection *within* networks, the fact that the cross-country estimates are larger than the OLS and are similar to the within driver and 2SLS results suggests that (a) much of the selection occurs within country of birth networks, (b) there is negative selection into in network driving, and (c) both the proposed methods to remove selection are likely valid.

E. 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.

[31] Countries with at least five drivers are included (55 countries) and the circle area is proportional to the number of drivers. Country labels are not included due to a TLC request that country-level driving outcomes not be reported.

These results are more speculative since we cannot identify precisely the level of moral hazard associated with each summons type and because of the reduce test power due to splitting the data more finely. Nevertheless, the results in Table 5, which reports both OLS and instrumental variables estimates, are generally consistent with effort effects.[32] In-network driving has the most significant effect on the three summons types that directly result in an owner fine: 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.[33]

For the reasons discussed in Section 2, we have focused our analysis on summonses. However, we can also examine accidents and convictions for driving violations as alternative driving outcomes, albeit with less power 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. In columns (1)-(4) and (5)-(8) of Table 6, the dependent variables are number of accidents and the number of convictions for driving violations, respectively, by the driver during the lease period. 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. 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 have lower ability on average than out-of-network drivers.

---

[32] We do not include country or driver fixed effects due to insufficient variation at the summons level.
[33] t-statistics and not coefficients are reported since number of summonses varies by type and are not comparable.

F. 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, one concern might be that high-quality prospective drivers are more likely to enter the industry as in-network drivers, while low-quality prospective drivers do not enter the industry at all. Such sample selection could potentially bias the driver-level cross-sectional results (OLS and 2SLS). However, this selection would not drive the within-driver results (because all comparisons are within the same driver over time). Also 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). Given the robustness of our central findings across models that are, and are not, susceptible to sample selection bias, the results do not appear to be driven by sample selection. 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.

G. Discussion of owner selection

Even though our results suggest that selection on the part of drivers does not drive our results, one may wonder about selection on the part of owners. Specifically, if owners who lease to an in-network driver are more risk averse versus owners who lease to an out-of-network driver, then the owners leasing in-network may systematically use contracts with stronger incentives for better driving outcomes (for example, that require larger deposits). If so, our results might reflect owner selection rather than a mitigation of moral hazard.

This possibility is unlikely since, as discussed in section 2, there is only modest leeway for owners to vary contracts given the regulatory caps on the deposit amount ($500) and lease fee ($666), which industry participants indicate are respected in practice. That is, there is not much flexibility in contracting terms for explaining the large observed difference in outcomes between in-network and out-of-network drivers. However, we examine this issue more formally.

In Table 8 we present empirical evidence that driver selection does not drive our results. First we estimate our main OLS model with owner fixed effects instead of driver fixed effects to compare outcomes of in-network drivers and out-of-network drivers with the same owner. The results are almost identical to the basic OLS model ─ showing that behaviors specific to the owners (such a risk preferences or the types of contracts they offer drivers) do not drive our results. Second, we run instrumental variables models using as an instrument for whether the owner leases in-network an indicator for whether there is a same-country driver living in the same census block as the owner. The results again indicate that owner selection is not driving the results and that a causal network effect is present. Third, we present the country-level lessee driving outcomes versus the fraction of *owners* from that country who lease in-network. These results are similar to the original county-level model, and again are consistent with a network effect and not owner selection. We also run the basic OLS results when we restrict the sample to lessees who lease from owners who are not themselves drivers and show that the results are essentially the same.[34] As a final check on the possibility of owner selection, we show the instrumental-variables and within-owner models using the owners-driver driving outcomes and the instrument based on owner residential location, where "Lease in network" indicates that the

[34] For all of these results, the magnitudes of the estimates are smaller the estimates presented earlier that control for driver selection. This is as expected given that the models in Table 8 do not directly control for driver selection. (There is insufficient variation in the data to control for both driver and owner selection in the same model.)

owner leases to an in-network driver. For both of these models, the point estimates are small relative to those found for lessees, and the null hypothesis that in-network owners are the same as out-of-network owners cannot be rejected. In sum, the earlier results of an important in-network effect are robust to including various types of controls for owner selection explicitly.

7. EXPLORATION OF MECHANISMS

There are at least five mechanisms through which social network effects can operate, including screening, monitoring costs, social sanctions, social preferences, and social collateral. In this section we discuss which of these mechanisms is most consistent with our findings and we present suggestive evidence on the mechanisms.

First, we consider screening and monitoring costs. Since the OLS estimates show a significantly smaller benefit of in-network driving versus the within-driver, instrumental variables, and country-level estimates, we have evidence that selection is such that in-network drivers having lower ability on average than out-of-network drivers. Under screening, there should be selection of higher ability drivers into in-network leasing. Thus, screening is unlikely to be driving the driver-owner matches and outcomes.

Second, there are several reasons why monitoring costs are unlikely to explain the network effects. 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 lessees is infeasible at reasonable cost regardless of network status.[35] Second, most drivers in the sample lease from an owner who himself drives the

[35] One might argue that other drivers could observe a driver's driving style and report this information to the owner. However, this possibility seems remote due to the small chance any single driver observes another driver he knows exhibiting poor driving out of the approximately 40,000 drivers in the industry (short-term lessees, long-term

taxi, giving any owner some ability to monitor vehicle condition regardless of network status.

Third, 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 to measure network strength. Specifically, we measure driver/owner network strength as the fraction of residents in the driver/owner Census tract from the same country of birth. While this measure is imperfect, 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/driver networks generate larger improvements in outcomes, which may suggest which mechanisms are at play.[36]

Consider social sanctions, which would appear as community-enforced punishments against an offending group member. These sanctions would likely operate via the owner's network (and not a driver's network) and hence a relatively dense owner network may facilitate these 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 social collateral should be reflected in an owner network strength effect. In contrast, social preferences would likely operate independently of an owner's network (though perhaps it would operate via the driver's network if this represents the driver's level of connection to the community).

In Table 7 we report regression results where the dependent variable is the number of summonses and the key explanatory variables are an indicator for in-network leasing and the interaction of in-network leasing and network strength (an indicator variable denoting whether

---

lessees, and owner-drivers), also knows the owner, and wants to report the observed driving style to that owner.

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

more than fifteen percent of residents in the census tract are from the same place of birth).[37] Columns (5), (6), and (7)-(8) report results for driver network strength, owner network strength, and both, respectively. The results indicate that owner networks are more important than driver networks, and result can be seen most clearly in column (8), where driver network strength has no effect on summonses after conditioning on owner network strength. Also, the standalone in-network effect approaches zero, suggesting that the effects rely on a dense owner network.[38,39]

The estimates indicate that network effects operate via owner networks and not driver networks, and that network effects are weak when owner networks are weak. While these results are not dispositive, they suggest that social sanctions and/or social collateral are more important than social preferences for enabling the observed network effects.[40]

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 to mitigate moral hazard problems where contract are incomplete. While there is some evidence that social ties can mitigate moral hazard in group lending schemes in developing countries, little is known about the role of social ties in mitigating moral hazard outside of this particular context. In this study, we compare the driving

---

[37] Columns (1), (2), and (3)-(4) report similar results using the percentage of residents in the census tract instead of the non-linear specification described.
[38] The number of observations varies slightly across specifications due to differences in the availability of addresses.
[39] In columns (5)-(8), the continuous measure of network strength is replaced with an indicator 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, the results are more pronounced. We chose 15 percent since the marginal effect of network density appears to begin leveling off here.
[40] One way that social sanctions might operate is by making it easier for owners to collect damages from drivers since damages may often be identifiable but not legally recoverable.

outcomes of NYC taxi drivers who lease from owners from the same country to those of drivers who lease from owners from a different country. Across several models that exploit distinct and 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. We also find the marginal effects of driving in-network depend on the owner's network but not the driver's network, which suggests an important role for social sanctions.

An interesting policy implication from the group lending literature is the possibility of 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 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.

More generally, our results indicate that social ties can 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 improving economic outcomes.

REFERENCES

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

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.

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.

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, K. and E. 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.

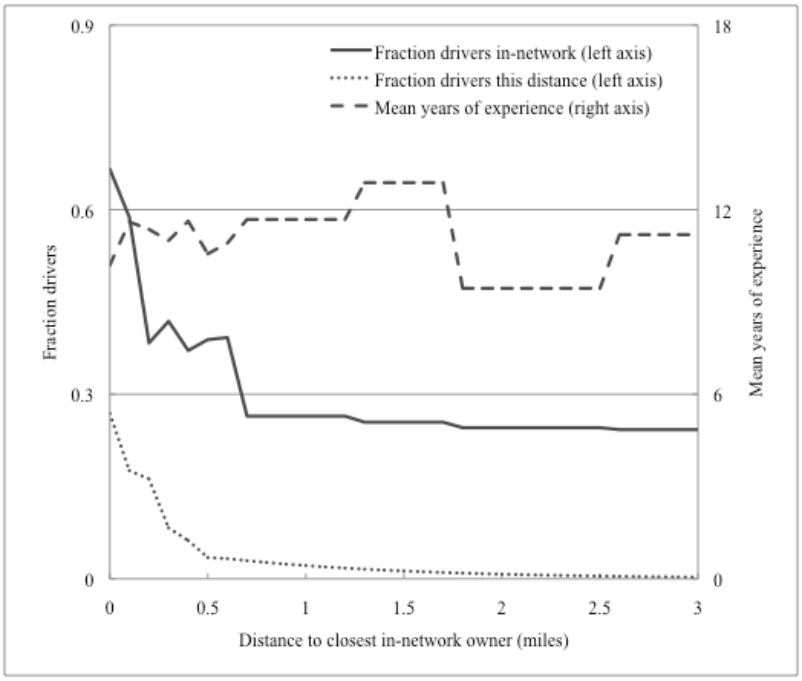
Sharma, M. and M. Zeller (1997) “Repayment performance in group-based credit programs in Bangladesh: An empirical analysis,” World Development.

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.

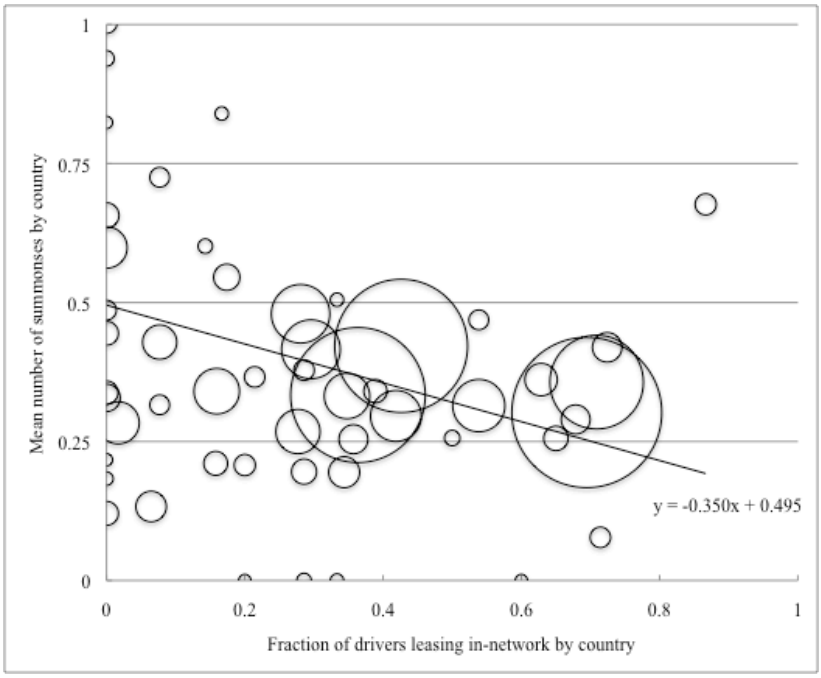
Wydick, B. (1996) “Credit access, human capital and class structure mobility,” *Journal of Development Studies*, 35(6) 131-152.

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



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 in network	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: Models of summonses

	(1) OLS	(2) OLS	(3) Within driver	(4) Within driver	(5) 2SLS	(6) 2SLS	(7) Country level
Lease in network	-0.085***	-0.070**	-0.334***	-0.489***	-0.251**	-0.333	-0.350**
	[0.024]	[0.033]	[0.093]	[0.151]	[0.128]	[0.389]	[0.138]
Lease in network x lease out of network 2005				0.271			
				[0.204]			
Log years of experience	-0.105***	-0.108***	-0.315*	-0.322*	-0.082***	-0.086***	-0.058
	[0.014]	[0.016]	[0.167]	[0.167]	[0.022]	[0.025]	[0.092]
2007 indicator	-0.064***	-0.054**	-0.015	-0.026			-0.225***
	[0.023]	[0.024]	[0.057]	[0.058]			[0.081]
Owner-driver summonses for lessee country							-0.143
							[0.209]
Constant	0.648***	1.162***	1.248***	0.665***	1.042	0.785***	0.636***
	[0.038]	[0.329]	[0.339]	[0.073]	[0.772]	[0.200]	[0.078]
First stage coefficient on lease in-network					0.312	0.115	
First stage F-statistic					128.8	29.9	
Fixed effects for driver country	No	Yes	No	No	No	Yes	No
Fixed effects for owner country	No	Yes	No	No	No	Yes	No
Fixed effects for driver	No	No	Yes	Yes	No	No	No
Observations	3646	3646	3646	3646	1629	1629	149
R-squared	0.025	0.091	0.039	0.041	0.007	0.110	0.086

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.

Table 4: 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	-0.088	-0.089	-0.012	-0.004	0.86	1.177	0.010	-0.018
	[0.056]	[0.060]	[0.027]	[0.034]	[0.987]	[1.143]	[0.039]	[0.045]
Constant	2.075***	1.414***	0.910***	0.941***	78.4***	87.6***	0.018	0.350
	[0.026]	[0.509]	[0.013]	[0.029]	[0.468]	[2.486]	[0.022]	[0.559]
Fixed effects for driver country	No	Yes	No	Yes	No	Yes	No	Yes
Fixed effects for owner country	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1629	1629	648	648	659	659	1629	1629
R-squared	0.002	0.191	0.000	0.161	0.001	0.168	0.000	0.131

Notes: The dependent variables are measures of driver ability (labeled at column head). The explanatory variable is the excluded instrument. 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 5: 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 number of drivers with a positive number of that summons type (N OLS, N 2SLS), and the t-statistic for regressions of in-network driving on the probability the driver incurred the summons type listed in the left-most column. Results for OLS (2005 and 2007) and 2SLS regressions (2005 only) are listed. Fixed effects for driver and owner country are not included (to preserve test power). The corresponding direct risk/cost to the owner is listed in the right-most column.

Table 6: 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**	-0.019	-0.095*	-0.027	-0.035*	-0.023	-0.068	-0.057
	[0.011]	[0.015]	[0.051]	[0.043]	[0.021]	[0.035]	[0.103]	[0.079]
Log years of experience	-0.016**	-0.012	-0.081**	-0.018**	-0.017	-0.055***	-0.104**	-0.017
	[0.006]	[0.007]	[0.036]	[0.008]	[0.012]	[0.014]	[0.049]	[0.014]
Owner-driver outcomes for lessee country			-0.148**				-0.125*	
			[0.070]				[0.073]	
Constant	0.067***	0.071**	0.225**	0.074***	0.221***	0.351**	0.411***	0.224***
	[0.014]	[0.034]	[0.086]	[0.018]	[0.030]	[0.156]	[0.124]	[0.040]
Fixed effects for driver country	No	Yes	No	No	No	Yes	No	No
Fixed effects for owner country	No	Yes	No	No	No	Yes	No	No
Observations	1813	1813	79	1647	1855	1855	81	1687
R-squared	0.006	0.094	0.145	0.009	0.014	0.114	0.078	0.005

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

Table 7: Models of the effect of network strength

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lease in network	-0.091**	-0.080**	-0.075*	0.055	-0.106***	-0.092**	-0.096**	0.025
	[0.041]	[0.040]	[0.043]	[0.061]	[0.037]	[0.038]	[0.039]	[0.055]
Lease in network x driver network strength	-0.239		0.077	-0.071	-0.014		0.074	0.057
	[0.230]		[0.326]	[0.371]	[0.091]		[0.109]	[0.129]
Lease in network x owner network strength		-0.386*	-0.603*	-0.573		-0.152**	-0.217***	-0.220**
		[0.200]	[0.323]	[0.376]		[0.068]	[0.077]	[0.094]
Log years of experience	-0.094***	-0.091***	-0.087***	-0.089***	-0.095***	-0.091***	-0.087***	-0.089***
	[0.022]	[0.022]	[0.023]	[0.026]	[0.022]	[0.022]	[0.023]	[0.026]
Constant	0.629***	0.622***	0.612***	0.867	0.631***	0.622***	0.612***	0.885
	[0.055]	[0.055]	[0.057]	[0.734]	[0.055]	[0.055]	[0.057]	[0.743]
Fixed effects for driver country	No	No	No	Yes	No	No	No	Yes
Fixed effects for owner country	No	No	No	Yes	No	No	No	Yes
Observations	1644	1645	1518	1518	1644	1645	1518	1518
R-squared	0.020	0.020	0.019	0.158	0.020	0.020	0.020	0.158

Notes: The dependent variable is driver-level number of summonses. Network strength in columns (1)-(4) is fraction of residents in the driver/owner 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 adjusted for clustering at the country level are reported in brackets. *, **, and *** indicate significance at the 10, 5, and 1 percent levels.

Table 8: Robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Lessee-driver					Owner-driver	
	Within owner	IV	Country level	Lessee-only taxis	Lessee-only taxis	Within owner	IV owner
Lease in network	-0.091	-0.140	-0.209*	-0.113*	-0.132	0.090	0.074
	[0.073]	[0.114]	[0.121]	[0.062]	[0.099]	[0.076]	[0.141]
Log years of experience	-0.066***	-0.086***	-0.073	-0.114***	-0.091***	-0.335	-0.107*
	[0.024]	[0.025]	[0.085]	[0.030]	[0.035]	[0.492]	[0.058]
2007 indicator	-0.092***		-0.206***	-0.077	-0.055	-0.030	
	[0.030]		[0.072]	[0.052]	[0.056]	[0.082]	
Owner-driver summons rate for lessee country			-0.203				
			[0.308]				
Constant	-0.059	0.651***	0.733***	0.659***	3.291***	0.358	0.447**
	[0.832]	[0.063]	[0.257]	[0.078]	[0.085]	[1.229]	[0.220]
Fixed effects for driver country	Yes	No	No	No	Yes	No	No
Fixed effects for owner country	No	No	No	No	Yes	No	No
Fixed effects for owners	Yes	No	No	No	No	Yes	No
Observations	3646	1615	156	769	769	1787	795
R-squared	0.069	0.009	0.071	0.028	0.214	0.021	0.015

Notes: The dependent variable in columns (1)-(5) is the number of summonses incurred by the lessee. In column (2), the instrument is an indicator for the owner living in the same Census block as any same-country driver. In column (3), "Lease in network" is the fraction of owners for the country leasing to same-country drivers. The sample in columns (4)-(5) are lessees of taxis operated exclusively by lessees (i.e., owned by a non-driving owner). In column (6)-(7), the dependent variable is the number of summonses incurred by the owner-driver, and "Lease in-network" indicates whether the lessee-driver to whom the owner-driver leases is in-network. Column (6) includes owner fixed effects. In column (7), the instrument is an indicator for the owner living in the same Census block as any same-country lessee.

APPENDIX: SUPPLEMENTARY RESULTS (FOR ONLINE REFERENCE ONLY)

Patterns of lessee-owner matches

Table 9 shows the country matches between owners and lessees. A high rate of in-network leasing is evident.

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.[41] 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.

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

[41] 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).

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, 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.

Table 9: 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 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.