



The Business School
for the World®

Working Paper

2023/45/EFE

(Revised version of 2022/24/EFE)

Social Diversity in Corporate Boards and Firm Outcomes

Suresh Bhagavatula
IIM Bangalore, sureshbh@iimb.ernet.in

Manaswini Bhalla
IIM Bangalore, manaswinib@iimb.ernet.in
Corresponding author

Manisha Goel
Pomona College, manisha.goel@pomona.edu

Balagopal Vissa
INSEAD, balagopal.vissa@insead.edu

We examine how firm performance is associated with social diversity among corporate directors, proxied by diversity along religion and caste, a deep-rooted institution dividing India's Hindu society into hundreds of communities. To identify directors' social identities, we build one of the first data-driven mappings of last names to caste and religion. We find that Indian corporate boards are strikingly homogeneous (i.e. lack diversity) during 1999-2015. Using four instrumental variable strategies, we find that board homogeneity is negatively related to firm performance. These perverse outcomes appear to be due to overlapping views and networks of caste-proximate directors and cronyism impairing their monitoring and advising roles.

JEL Classifications: M14, L25, G30, Z13

Keywords: Diversity; Firm Performance; Corporate Governance; Caste; Corporate Culture

Electronic copy available at: <https://ssrn.com/abstract=4021311>

Working Paper is the author's intellectual property. It is intended as a means to promote research to interested readers. Its content should not be copied or hosted on any server without written permission from publications.fb@insead.edu

Find more INSEAD papers at <https://www.insead.edu/faculty-research/research>

Copyright © 2023 INSEAD

1 Introduction

In recent years, the lack of diversity in corporate boards has garnered considerable attention. Several measures have been proposed to diversify corporate boards, with a key principle being that diversity of identities and backgrounds brings diversity of perspectives and attitudes.¹ Our paper measures social diversity using a relatively unexplored instrument which relates deeply to people's sense of identity. In particular, we focus on diversity of religion and caste – an informal social institution that divides India's Hindu society into hundreds of communities. Our findings reveal that India's corporate boards are strikingly homogeneous (i.e. lack diversity) in their religious and caste composition. We then ask how this lack of diversity is associated with firm outcomes and show that firm performance is negatively related to social homogeneity of boards. In addition, we delve into plausible mechanisms underlying this effect. Our evidence is consistent with the notion that socially homogeneous boards are less likely to offer novel perspectives (thus impairing their advisory role) or challenge management (thus impairing their monitoring role) thereby adversely affecting firm performance outcomes.

Social differences among corporate directors are proxied well by diversity of their castes, since the caste system is a deeply rooted and temporally stable structural aspect of Indian society. Although there is a multiplicity of views on the precise historical origins of the caste system, there is scholarly consensus that it is an important societal feature of contemporary India. Caste has been shown to affect education (Hnatkowska et al. (2012), Hnatkowska et al. (2013)), inter-generational occupational mobility (Munshi and Rosenzweig (2013)), lending (Fisman et al. (2017)), and more.² As Munshi (2019) puts it, "Caste plays a role at every stage of an Indian's economic life."³

¹There are several notable examples of such measures. Since 2008, Norway has required public and state-owned companies to have 40% of their boards be constituted by females. Since 2013, India has required large public companies to have at least one female director.

²Marriages, residence, occupations, voting patterns, public good provision, etc. are all influenced by the fine-grained measure. See Joshi et al. (2018), Kumar and Somanathan (2017), Beteille (1996), and Srinivas (1995).

³The strong influence of caste carries beyond India. California's Department of Fair Employment and Housing has on-going litigation alleging caste-based discrimination at Cisco Systems, Inc.

Caste can be measured coarsely as five categories (termed *varnas*) or finely as hundreds of categories (termed *jatis*). However, most previous studies on caste use a binary measure of “upper” and “lower” caste, in part, because there is no systematic documentation enabling researchers to identify individuals’ coarse- or fine-grained caste (*varna* and *jati*, respectively). We use a novel computational methodology to develop a data driven mapping of last names to social identity at three levels of granularity – religion, coarse- and fine-grained caste. In doing so, we exploit the facts that individuals’ last names are indicative of their caste, and that marriages are predominantly intra-religion and, among Hindus, intra-caste. We obtain data from three prominent matrimonial websites on the names of nearly six million registered users and their self-reported castes. Since the mapping between last name and religion/caste is not always one-to-one, we use these data to assign probabilities with which a last name belongs to each caste (at coarse- and fine-grained levels). This method helps us map 16,637 unique last names into eight religions, five coarse-grained castes, and 471 distinct fine-grained castes. We apply this mapping to directors of public and private firms, whose names are taken from an annual firm-level database.

We use these data to develop a rich set of stylized facts about caste and religion diversity in Indian corporate boards during 1999-2015. We find systematically low levels of diversity on boards along all three measures – fine-grained caste, coarse-grained caste, and religion diversity. Importantly, the low diversity in boards is not simply coincidental or driven by low diversity in the supply pool of directors. Also, this low diversity is pervasive across states and industries throughout the sample period. We further show that it varies systematically across firm types such that lower performance and poorer corporate governance are associated with lower diversity in boards.

Next, we investigate the relationship between diversity of boards and firm performance. Ex ante, this relationship is unclear. On one hand, homogeneous board members may not bring a wide range of perspectives to bear upon the decisions they make for the firm, worsening their advisory roles. They may also be more prone to cronyism (which refers to their inability to challenge the firm’s management and have tough conversations needed to improve firm performance), hurting

their monitoring role and, hence, the firms they serve. On the other hand, socially homogeneous directors may have greater trust, which reduces conflicts during strategy execution, thus improving firm performance.

Regression analysis, therefore, provides us with estimates of the net effect of these mechanisms. We use four instrumental variable strategies to examine how caste and religion diversity of boards affects key accounting measures of firm performance and market related variables. To instrument for diversity, we use the diversity of a firm's director supply pools, measured as the sets of directors in the firm's state and industry. In a second approach, we additionally use as instruments the distance of a board's caste/religion composition from that of the supply pool composition. In a third strategy, we exploit a change in corporate governance requirements that induced changes in board memberships during our sample period. In a fourth strategy, we use as an instrument the fraction of a board's directors who serve on at least one other board whose dominant caste is different from the directors' own caste. This instrument, inspired by Adams and Ferreira (2009), captures the extent to which a board's directors are exposed to castes other than their own. Results from all four analyses show that diversity on corporate boards is negatively associated with firm value and performance.

Finally, our moderation tests (implemented using interaction terms in a regression framework) provide suggestive evidence for mechanisms through which high caste homogeneity on boards adversely affects firm performance. In essence, we provide illustrative evidence on the channels through which caste homogeneity might hamper boards monitoring and advisory roles and thereby cause poor firm performance. First, we show that firms with homogeneous boards fare even worse than others during slow GDP growth years. This finding is consistent with the possibility that caste homogeneous boards' overlapping external networks and viewpoints inhibits them from providing novel and useful advisory input to the firm precisely when such advice is needed for improving firm performance. Next, we show that firms with homogeneous boards perform even more badly when key board committees (e.g., audit, shareholder grievance, and compensation and nomination committees) that impact the board's monitoring role have an over-representation of

the dominant caste. In addition, there is suggestive evidence that firms with homogeneous boards perform even more badly when the CEO shares the same caste as the dominant caste on the board. Taken together, these findings are consistent with the possibility that caste homogeneity makes it less likely that boards effectively discharge their monitoring role of challenging management's actions, thereby leading to poor firm performance. Overall, we interpret these patterns of findings as consistent with the notion that caste homogeneity on the board causes poor firm performance in part because homogeneity inhibits the efficacy of boards' monitoring and advisory roles.

Several studies have focused on how socio-cultural identity shapes networks (Currarini et al., 2009), hiring (Åslund et al., 2014; Giuliano et al., 2009; Giuliano and Ransom, 2013; Petersen et al., 2000), and economic exchange in dyads such as lender-borrower, manager-employee, venture capitalist VC-entrepreneur, VC partners, research collaborators, and teacher-student (Gompers et al., 2016; Glover et al., 2017; Shayo and Zussman, 2011; Fisman et al., 2017; Bengtsson and Hsu, 2015; Hegde and Tumlinson, 2013; Claes and Vissa, 2020; Freeman and Huang, 2015; Dee, 2005; Fairlie et al., 2014). Our paper is different from all of these in that we analyze the effect of social diversity in teams of multiple agents and their joint decisions in high stakes economic settings.

The literature on corporate governance has also analyzed the effects of board diversity, but the dimension examined is almost exclusively gender (Adams and Ferreira, 2009; Ahern and Dittmar, 2012; Kim and Starks, 2016; Sila et al., 2016; Bertrand et al., 2018). Terjesen et al. (2009) provide an excellent review. A few exceptions include Giannetti and Zhao (2019) who use ancestral origins, Arnaboldi et al. (2018) and Gennaro et al. (2018) who develop a multidimensional diversity index, and Kramarz and Thesmar (2013) who use shared alma maters of directors. We add to this body of work by considering the social diversity of directors, as measured by the traditional institutions of caste and religion. The impact of such deep rooted traditional institutions on board composition and firm performance has not been previously explored.

This paper also contributes to the literature examining the economic effects of caste. Previous studies mainly compare socio-economic outcomes of disadvantaged "lower" castes to those of

advantaged “upper” castes (Hnatkowska et al., 2012, 2013; Iyer et al., 2013; Ghani et al., 2014; Damodaran, 2008; Thorat and Neuman, 2012; Jodhka, 2010; Varshney et al., 2012). However, we approach the economic effects of caste through a different lens – does shared caste identity influence economic outcomes, regardless of whether the caste itself is underprivileged or not? Only a few studies have taken a similar approach. See Fisman et al. (2017), Munshi and Rosenzweig (2013), Munshi and Rosenzweig (2016), and Vissa (2011). Besides examining a different economic outcome, we differ from these studies in a few important respects. First, while they focus on rural areas, specific cases, start-up ventures or traditional businesses, we show that caste influences economic outcomes nationally, even in urban, elite, and high-stakes corporate environment typical of large, established corporates. Our focus on board composition and firm performance also distinguishes us from Chen et al. (2015) who consider caste proximity between equity analysts and CEOs and Damaraju and Makhija (2018) who consider caste proximity between CEOs and firm owners or chairpersons.

The rest of the paper is organized as follows. In Section 2, we describe our data. Section 3 shows that diversity in corporate boards is systematically low. In Section 4, we demonstrate that social diversity is positively associated with firm performance. In Section 5, we present instrumental variable analyses to examine the association between social diversity and firm performance. Section 6 explores the mechanisms underlying our results. Section 7 concludes.

2 Building the Database

We combine data on names and caste and religion identity from matrimonial websites with data on Indian firms and their boards, as described below.

2.1 Data from Matrimonial Websites

We develop a novel dataset mapping Indian last names to religions and castes.⁴ Our approach exploits two features of Indian society – last names are indicative of caste and religion (Dumont, 1980), and marriages are overwhelmingly intra-religion, with Hindu marriages being intra-caste. Both aspects are reflected in individual profiles on matrimonial websites, where prospective suitors self-report their names, religions, and castes (specifically, fine-grained castes).⁵ We obtain data on over six million such profiles from three popular matrimonial sites – *Shaadi.com*, *Jeevansathi.com* and *Bharatmatrimony.com*.⁶ Specifically, we have information on users’ first and last names, native language, religion, and fine-grained caste.⁷ In our final cleaned data, we have 5,447,129 profiles, spanning 16,637 unique last names, eight religions, five coarse-grained castes, and 471 fine-grained castes.

The same last name may be associated with more than one religion or caste, often depending on the geographical region. We exploit the fact that different regions of India have different dominant languages. Therefore, our mapping from last name to caste is conditioned on users’ self-reported first language.

We also account for the possibility that the same last name may have several spelling variations. The vast majority of names in our dataset are words from Indian languages like *Hindi*, *Tamil*, *Marathi*, *etc.*, whereas the websites from which our data are culled use the English language. Therefore, we have an English equivalent (not translation) of last names in Indian languages. This

⁴Our work is one of the first academic exercise that links individual Indian last names to a likelihood distribution over religion, coarse- (and fine-grained) caste. Online Appendix Section A lists other data sources that collect caste information.

⁵Since people want to marry within caste, users who do report their caste have an incentive to report it truthfully.

⁶In India’s fast evolving socio-economic setting, where the role of traditional kinship networks and local matchmakers is rapidly diminishing, matrimonial websites act as an alternative to traditional marriage brokers by nationalizing the pool of prospective spouses. According to a study conducted by KPMG in 2016, the Indian online matrimonial market was expected to rise from USD 0.11 bn to USD 0.26 bn by the end of 2022, growing at an annual growth rate of 19%.

⁷To build a robust mapping, we drop all last names that appear only once in the database.

may result in multiple English spelling variations of the same last name. To address this, we use two different word matching algorithms (Levenshtein distance and Ratcliff/Obershelp pattern matching) to predict the similarity of differently spelled last names. If the similarity predictions from both algorithms are above a certain threshold, then the two last names are considered to be the same and their caste mappings are combined.⁸ This name matching results in groups of similar sounding last names that have different spellings.

Following this approach, we assign all spelling variants of the same last name (conditional on same language) a frequency equal to the number of records with that last name in our data.⁹ We take all religions and fine-grained castes reported for a last name across all corresponding records and count their respective occurrences. Dividing the number of times a last name is associated with a particular caste and religion by the total number of times the last name appears for a given language in the database gives us the probability with which the last name belongs to a particular

⁸Our first algorithm is a modified Levenshtein distance algorithm. In this method, a distance measure between two strings is calculated using a dynamic programming approach, with each replacement letter adding one to the distance measure. The standard Levenshtein distance measure is appropriate for English words, whereby each difference in letters between two words contributes equally to the distance measure. However, since we want to match the phonetic translation of Indian languages, we develop a modified Levenshtein distance algorithm. In this method, differences in letters that constitute the same sound in Indian languages are assigned a zero distance measure. The final distance calculated using this algorithm provides a measure of how similar two words are in an Indian language. In our approach, two words that have a distance of less than three are deemed to have the same spelling in the Indian language and, hence, be the same name. Our second method is a modified version of the Ratcliff/Obershelp pattern matching algorithm. This algorithm looks for matches in the longest contiguous matching subsequence of letters in two words and assigns a matching score. The original algorithm was developed to find sequence matches between two sentences, and was found to be appropriate for matching words that are phonetic translations from other languages. For our purposes, if the algorithm provides a match score of greater than 85%, the two last names are deemed to be matched. As a final step, we consider two last names to be fully matched if both algorithms predict that the corresponding strings are matched. This approach is borrowed from the concept of bagging used in the machine learning based classification literature where votes from different classifiers are used together to increase the robustness of the final prediction (Friedman et al., 2001).

⁹The resulting average total record count for each last name is 362.68.

caste and religion . Doing so over all religions/castes associated with a last name gives us the probabilistic mapping of each last name to religions and fine-grained castes.¹⁰ For names identified as non-Hindu (3070 in number), we assign the corresponding religion instead of caste.

We also map last names to coarse-grained caste. For this, we rely on Government of India reports, Wikipedia entries, and several other sources to build a mapping from fine- to coarse-grained caste.¹¹ In our final mapping, last names are associated with an average of 1.57 and a maximum of 6 religions, with 56.59% of last names being associated with a single religion. The probabilistic mass is concentrated in the top two most likely religions, with them jointly accounting for about 99% of the total likelihood on average. As for caste, last names are associated with an average of 3.6 (10.4) and a maximum of 11 (138) coarse (fine)-grained castes. Table 1 of the Online Appendix shows the most likely religion, coarse- and fine-grained caste composition in the final matrimonial sample. Hindus, at 80.69%, form the overwhelming majority of names. Muslims and Christians, respectively, account for the next highest proportions. Other religions together constitute about 5% of the sample. This distribution is not far from the religion composition of the aggregate population of India.¹²

¹⁰In one region of southern India (the state of Tamil Nadu), the last name of an individual is often their father's first name. Hence, the last name changes across generational cohorts of a family. For these cases, we map the first name of the individual to their caste instead of the last name. Though Tamil first names do vary systematically by caste, we expect the caste mapping of Tamil names to be noisier than the mapping based on last names for the rest of India, since the latter remain invariant across generations.

¹¹The caste system divides the Hindu society into four hierarchical categories – *Brahmin*, *Kshatriya*, *Vaishya*, *Shudra*. There is an additional de facto fifth category – *Dalit*. Our mapping assigns every last name a probability mapping over these five categories. For Non-Hindu last names, we assign the corresponding religion as their coarse- and fine-grained caste. Additionally, we assign “unknown” to names for which we are unable to find a fine- to coarse-grained caste mapping, or if a fine-grained caste maps to multiple coarse-grained castes.

¹²According to the 2011 census, Hindus constitute 80% of the population, Muslims are the next largest group (14.2%), followed by 2.3% Christians. Other religions together account for just over 5% of the total population. The census does not provide population composition by coarse- or fine-grained caste. In our data, however, we see the maximum representation of the coarse caste category of *Shudra* followed by *Brahmin*. For space considerations, Table 1 of the Online Appendix only reports the percentages of last names for the top eleven most frequently occurring

2.2 Firm Level Data

We obtain firm-level data from Prowess, a database provided by the Centre for Monitoring the Indian Economy (CMIE). The data provide information on financials and corporate governance of large firms, as reported in their annual reports, financial statements, and profit and loss accounts. The database includes large publicly listed and unlisted firms.¹³ While the database does not cover the universe of all firms, those included account for a substantial proportion of economic activity; in 2009, they contributed 84% of GDP, 55% of exports, 70% of imports, 47% of the total output of non-agricultural and non-government services sector, and 58% of all corporate and excise taxes collected by the government. Although the database follows firms longitudinally, most firms appear in the data only for a few years.

We combine firms' board members' data with our last name to caste/religion mapping to measure boards' social diversity. To identify the religions and castes of firms' directors, we match their last names to those in the matrimonial data, assigning each matched director last name the same probabilistic distribution over religions and castes as that constructed using the matrimonial data. We retain only those firm-year observations for which we can (probabilistically) identify the caste/religion of all board members.¹⁴ We also consider only those firm-years that have at least two directors serving on their boards. In our matched sample, we have 23,819 unique firms with a total of 576,579 directorships.¹⁵

Table 2 in the Online Appendix profiles the caste and religion mapping for directors. For religion, we see that 69% of directors' last names are associated with a single religion, and

fine-grained castes in the data.

¹³As of 31 March, 2009, 7,86,774 companies were registered with the Registrar of Companies, an administrative arm of the Ministry of Corporate Affairs. Of these, 82,058 were public limited companies and 704,716 were private limited companies. Prowess contains information on about 24,000 public companies.

¹⁴There are a few directors whose names suggest that they may not be of Indian origin. We are unable to match these names with those in the matrimonial database. These directors are assigned a religion, coarse-, and fine-grained caste category of "NA."

¹⁵Over 1999-2015, we are able to fully match 63.52% of firm-year boards and 57.11% of all directorships in Prowess.

92% names are associated with up to two religions. As expected, the mapping for coarse- and fine-grained caste is noisier, with 45% (40%) last names associated with a single coarse- (fine-) grained caste and 85% (65.5%) last names associated with up to five coarse- (fine-) grained castes. Although the mapping is probabilistic, the probabilities are front loaded. The top two most likely religions account for 99.5% of the total likelihood, on average. Similarly, the top five most likely coarse- (fine-) grained castes account for 99.3% and 89% of the total likelihood, on average.

Homophily Index: To measure the degree of boards' social diversity we calculate the religion, coarse- and fine-grained caste HHI for every board. The caste/religion HHI for a board is the sum of squared shares of directors belonging to the various identities represented on the board. A higher HHI with respect to an identity represents lower diversity, or greater homophily, in a board. For example, consider a board with five members – three Hindus and one each with the most likely religion as Muslim and Christian. The religion HHI of this board is $0.44(= (0.6)^2 + 2 * (0.2)^2)$. Continuing with this example, suppose that of the three Hindu board members, the most likely coarse-grained caste of two of them is *Brahmin* and the other is *Kshatriya*. The coarse-grained caste HHI (replacing coarse-grained caste by the religions of non-Hindu directors) for this board is $0.28(= (0.4)^2 + (0.2)^2 + 2(0.2)^2)$. Finally, suppose that the two *Brahmin* directors have the fine-grained castes *Brahmin Iyer* and *Pandey*, and the *Kshatriya* director has the fine-grained caste *Khatri*. In this case, the fine-grained caste HHI (replacing fine-grained caste by the religions of non-Hindu directors) for this board is $0.2(= 5 * (0.2)^2)$. This example illustrates that for the same board composition, the measure of concentration reduces as the lens of identity becomes finer from religion to fine-grained caste.

We also use firms' financial data from Prowess. Refer to Table 1 for definitions of financial and other variables used in the paper. We deflate all nominal variables by all-India CPI (2001=100). A few key summary statistics are provided in Table 2. We note that the highest proportions of firms at both the start and end of our sample period belong to manufacturing, trade, and finance, insurance, and real estate sectors. Over the entire sample period, firms with fine-grained caste homophily above the median have poorer average characteristics than the ones with homophily below median.

For example, the mean real asset value is Rs. 100.82 million for below-median homophily firms and Rs. 5.3 million for above-median homophily firms. The average % of independent directors and average board size of firms with below-median caste homophily is larger than that of firms with above-median homophily.

2.3 Indian Boards Database

Prowess does not identify unique individuals serving as directors. Since more than one director may have the same name, Prowess cannot be used to identify unique directors. To do this, we use the Indian Boards Database, maintained by the Prime database group, which provides a unique identification code for each individual serving as a director, along with the firms on which he/she serves, for the period 2012-2015. This helps us measure the degree of interlocks across boards as well as construct an instrument that closely follows Adams and Ferreira (1999). Using these data, we have information on 17,608 unique directors across 1,501 firms.

3 Diversity in India's Corporate Boards

In this section, we describe the state of caste and religion diversity in Indian corporate boards.

India's corporate boards are not diverse – the average homophily index is high, at 0.45 for fine-grained caste. To assess if this is simply a result of the caste or religion composition of the entire pool of directors, we compare the observed diversity levels to those in several random simulated samples. In the first method, we consider all directors across all firms in a year as the potential pool of directors available to each firm in that year. From this “supply pool” of directors, we randomly assign directors to each firm, equal in number to its observed board size. We create a hundred such simulated samples of boards for each year, calculate the mean board homophily for all boards across the hundred iterations, and compare it to the corresponding mean in the observed data. In the second (third) method, we define the director supply pool for a firm in a year as the set of all

directors in that year across all firms in the same state (two-digit industry) as that firm.¹⁶

Figure 1 shows the yearly means of firms' fine-grained caste homophily for the observed and simulated samples. For the simulated means, we also present the 5% confidence intervals. The figure presents these means for all three approaches described above: unconditional, conditional on firms' state, and conditional on firms' industry. In all cases, we see that the mean observed fine-grained caste homophily of boards is significantly higher than the corresponding simulated mean in every year. Similar results are presented for coarse-grained caste and religion in Figures 1 and 2 of the Online Appendix. Table 3 of the Online Appendix also presents hypothesis tests for comparisons of observed and simulated homophily means. The t-statistics are large, indicating that the observed mean homophily is significantly different from the simulated means.

Further results presented in the Online Appendix show that diversity in boards has been persistently low over time across states and sectors.¹⁷ We observe that larger firms (with firm size measured by assets, sales and profits) have more diverse boards (Figure 7 of the Online Appendix). Figure 8 of the Online Appendix demonstrates that older firms have board members from more diverse caste backgrounds. Figure 9 of the Online Appendix shows that exporting firms have significantly more diverse boards, on average, than non-exporting firms. In Figure 10 of the Online Appendix, we observe lower fine-grained caste diversity among firms that do not belong to business groups relative to those that do. Figure 11 of the Online Appendix shows that firms with larger corporate boards have higher caste diversity among their directors.

¹⁶As mentioned earlier, we cannot identify individual directors since we do not have unique identification codes for them. So, we do the simulations by defining the supply pool in two ways. In one approach, we consider every name as a distinct director, i.e., we consider directorships rather than directors. Alternatively, we consider all occurrences of the same name as the same individual director. We present results from the first approach in the paper. Results from the second approach are extremely close and available upon request.

¹⁷Figures 3, 4, and 5 of the Online Appendix respectively present state-wise mean fine-grained caste, coarse-grained caste, and religion HHIs for the first and last years of our sample. While the average homophily stayed high in most states over the sample period, states did change their relative quartile positions in the overall distribution. Sectoral average board diversity also stayed low over the sample period (Figure 6 of the Online Appendix). These figures indicate that no one state or sector drives the lack of diversity on boards observed in the aggregate.

4 How does social homophily relate to firm performance?

Next, we examine whether and to what extent low caste diversity in firm boards relates to key measures of firm performance.

Consider the following regression equation:

$$P_{it} = \beta_0 + \beta_1 H_{it} + \beta_2 X_{it} + \beta_3 B_{it} + \beta_4 I_{jt} + \delta_1 I_j + \delta_2 T_t + \varepsilon_{ijt} \quad (4.1)$$

where P_{it} denotes the value or performance of firm i in year t , H_{it} , the key variable of interest, is the fine-, coarse-grained caste or religion homophily of firm i 's board in year t , X_{it} is a vector of time varying firm characteristics, B_{it} is a vector of time varying board characteristics, I_{jt} denotes a vector of time varying industry characteristics, I_j denotes a vector of two-digit industry (National Industrial Classification (NIC) 2008) fixed effects, and T_t is a vector of year fixed effects. We cluster the standard errors by year and industry (two-digit) and correct them for arbitrary heteroskedasticity.

Our dependent variables (P_{it}) are operating income relative to sales in natural logs (accounting measure), and Tobin's Q and firm volatility (market based indicators). Operating income is defined as the difference between sales and operating expenses. Tobin's Q is calculated as the market value of a firm divided by the replacement value of the firm's assets (where we proxy replacement value with the book value of assets). Volatility is measured as the standard deviation of returns on a firm's security over a year.^{18 19}

Firm level control variables (X_{it}) include firm age, firm size, tangibility, book leverage, and indicators for whether the firm is listed on the stock market, whether it belongs to a business group,

¹⁸We measure Tobin's Q and firm volatility in two ways. One measure uses stock prices of the entire year between two annual reports and the other uses stock prices for a month around the reporting date. We present results for the former but results remain close with the latter method. Some firms' stocks are traded on the National Stock Exchange (NSE), and some are traded on the Bombay Stock Exchange (BSE), and some on both. Throughout the paper, we use firm returns at NSE when their stocks are traded on only NSE or both, and BSE when they are only traded on BSE. Results using BSE returns for firms whose stocks are traded on both markets are close.

¹⁹We also use two additional accounting measures of firm performance – operating cash flow and profits (both relative to sales and in natural logs) and find qualitatively similar results for how they associate with board diversity.

and whether it is an exporter. We define a firm's age as the number of years since incorporation of the firm. Firm size is measured by real assets (in natural logs). Tangibility is defined as the fraction of tangible assets in the total assets of a firm. Book leverage is calculated as the ratio of the total debt of a firm to its total assets.²⁰

Board characteristics (B_{it}) include size of the board and its square, where size of the board is defined as number of directors on the board of a firm in a year. Coles et al. (2008) find that big firms, which have greater advising requirements, have larger boards than small firms. They also find that the relation between Tobin's Q and board size is U-shaped. Therefore, we include a quadratic in board-size across all regressions in our analysis. Different social groups have different levels of cohesion or sense of affinity. To account for this we include a vector of indicator variables to control for the dominant coarse-grained caste (for Hindus) and religion (for non-Hindus) on the firm's board. The dominant caste/religion of the board is the most likely caste/religion for the largest number of board members.

Time varying industry characteristics (I_{jt}) include market-concentration of the industry (measured as the three-digit NIC HHI at the national level), proportion of business group firms in the industry (three-digit NIC), and the shares of firms in the industry (three-digit NIC) with a particular dominant religion/coarse-grained caste.²¹ Including the industry's market-concentration helps us control for the competitiveness of the industry which can influence a firm's performance. Controlling for the proportion of business group firms in the industry allows us to account for the possibility that there may be firms with close ties that adopt similar practices, which, in turn, affect a firm's performance. Controlling for the shares of firms in an industry with various dominant religions or coarse-grained castes allows us to account for the caste/religion composition of firms in the industry.

In the above regression, β_1 captures the association between a board's caste or religion homophily and firm performance. We present estimates of this coefficient as obtained from fixed effects regressions in Table 3. Column (1) shows results for the association between fine-grained caste homophily of boards and the accounting measure of firm performance – log (operating

²⁰All financial variables are winsorized at 1% and 99% for the entire sample period.

²¹Alternatively, we also estimate industry HHI at two-digit NIC-state level. Results remain close.

income/sales). Columns (2)-(3) present analogous results for homophily and market based measures – Tobin’s Q and volatility, respectively. We observe that firm performance measures are worse for firms with less diverse boards, i.e., those with higher homophily. A one unit increase in fine-grained caste homophily is associated with 0.26 log points decline in operating income relative to sales. Tobin’s Q is also negatively associated with homophily. Higher board homophily is also significantly correlated with greater stock market volatility for the firm. A one unit increase in fine-grained caste homophily is associated with a 0.008 increase in the standard deviation of the firm’s stock market returns. The corresponding associations between coarse-grained caste and religion homophily and firm outcomes are similar, although the estimated coefficients are smaller in magnitude (See Tables 4 and 5 in the Online Appendix).

5 Instrumental Variable Analysis

The fixed effects estimates of β_1 presented in Table 3 (and Tables 4 and 5 in the Online Appendix) are not causal since homophily is an endogenous regressor. The endogeneity can result from both omitted variable bias and reverse causality. An unobservable time varying firm characteristic (for example, adoption of new management practices) can drive both homophily and firm outcomes. Firm performance can also influence homophily. For instance, as a firm’s value grows, it may become increasingly prestigious for directors to serve on its board. This can influence board composition.

To overcome this endogeneity, we employ four instrumental variable strategies. In the first approach, we use two instruments: (1) the caste/religion HHI for all directors in the two-digit industry that the firm belongs to and (2) the caste/religion HHI for all directors in the state where

the firm is located.^{22 23} These two variables provide us a measure of the caste/religion composition of the set of directors that constitute the firm’s “supply pool”, as described in Section 3.1. In the second approach, we augment our list of excluded instrumental variables with: (3) the Euclidean distance of the vector representing the board’s caste/religion composition from that of the full set of directors in the corresponding industry and (4) the Euclidean distance of a board’s caste/religion composition vector from that of the full set of directors in the corresponding state. In the third approach, we construct an instrument that exploits changes in board memberships necessitated by a set of requirements announced by the Securities and Exchange Board of India (SEBI), commonly referred to as “Clause 49.” In the fourth approach, similar to Adams and Ferreira (2009), we instrument for board homophily with the fraction of the firm’s directors with board membership in at least one other firm whose dominant caste is different from that of the directors’ own caste.

In the following sub-sections we describe each of the instrumental variable strategy in greater detail and present associated results.

5.1 Approach 1: Homophily of Industry and State Director Supply Pools

Industry and state director caste HHIs are relevant excluded instruments since a firm’s board composition may be similar to that of other firms in the same industry or state. Previous studies have shown that both industry and geography influence the supply of directors that firms can choose from (Knyazeva et al., 2013; Dass et al., 2013). We show that this holds in our setting too by

²²A more dis-aggregated classification is unsuitable for two reasons. First, directors may not serve on closely competing firms’ boards due to conflict of interest. Second, the narrower the classification level, the fewer the number of firms in each industry so that the influence of each firm in determining the overall pool of directors in the full industry may be high, invalidating the instrument. A less dis-aggregated classification level, on the other hand, is undesirable as it will not yield enough variation in the industry level homophily index.

²³We measure homophily of state and industry level directors in two ways. In the first approach, each name is considered to represent a distinct director. In doing so, we effectively measure the homophily of directorships rather than unique directors. In the second method, we assume that all occurrences of the same name represent the same unique director and measure homophily using unique names in a state/industry. In the paper, we present results using the first approach. Results using the second approach are extremely close.

documenting that (a) a non-negligible proportion of a firm's directors are also directors of other firm(s) in the same industry and state, and (b) the caste/religion composition of directors on firm boards is very similar to that in the industry or state. Table 6 of the Online Appendix documents within-industry board interlocks for one-digit industries for the year 2015.²⁴ To identify these interlocks, we use the Indian Boards Database which, unlike Prowess, allows us to identify unique directors, albeit for a smaller sample of firms. Using these data, we identify a within-industry interlock if a director on a firm is currently, or has been in the past, a director on at least one other firm that belongs to the same industry. We then calculate the percentage of all directors in a firm that are interlocked within-industry. We observe that the average interlocks range from 0% to 31% across these broad industries, while the maximum degree of interlocks can be as high as 100%. Looking at two-digit and three-digit industries, we see that interlocks are present even at these narrower levels, albeit to a smaller degree. The mean interlock in two-digit industries in 2015 is 5.2% (3.4% in three-digit), although the maximum interlock is over 80% in many industries.

Nonetheless, there are several firms with no directors that serve (or have served in the past) on other firm(s) in the same broad industry. However, even across these firms, the caste/religion composition of directors is similar to that of directors in the industry. We show this by comparing the distribution of directors in firms that have below (and above) median interlocks to that of the industry using two Kolmogorov-Smirnov tests (K-S tests).²⁵ In the first test, we compare the distribution of the top caste/religion of all the unique directors in the pool of firms with interlocks below (and above) median to that of the entire industry. In the second set of K-S tests we compare the distribution of the dominant caste/religion of firms with below (and above) median interlocks to that of all firms in the entire industry. Results for fine-grained caste composition from these tests are presented in Table 7 of the Online Appendix. The table shows that for one-digit industry, we are unable to reject the null hypothesis that the samples of directors in firms below (and above) median

²⁴For space considerations, we present this evidence only for 2015. Similar results hold in other years of our sample period.

²⁵The Kolmogorov-Smirnov test (K-S test) examines the null hypothesis that two samples are drawn from the same continuous, one dimensional probability distribution.

and the aggregate industry are drawn from the same distribution. The same conclusion is reached when we alternatively look at the samples of firms according to their dominant coarse-grained caste and religion.

The validity of using industry and state director caste HHIs as the excluded instruments is also plausible for several reasons. First, to the extent that industry and state level homophily indices are associated with some unobservable characteristics of the industry or state that can have an independent effect on firm performance, that possibility is controlled for by including industry and state fixed effects. Note, however, that we are unable to include both sets of fixed effects simultaneously, in addition to year fixed effects and other time-invariant firm characteristics including listing and export status. This is because, the number of firms within the resulting cells is often small so that we do not have enough variation left in a large proportion of cells in the samples. Second, as explained above, we define the industry broadly at the two-digit level. The number of firms in a two-digit industry tends to be large, so that any single firm is unlikely to strongly influence homophily among the set of directors in the entire industry. Analogous intuition applies to the state-level homophily index.

Third, we control for several other firm and industry characteristics in our regression to ensure that the excluded instruments only affect firm performance through board diversity or through their correlation with included controls. Directors and firms could match in an assortative manner based on factors such as reputation, causing an omitted variable bias invalidating the instruments. We measure firm performance variables relative to size (as proxied by firm sales) and additionally control for firm assets. We also include other firm characteristics – it's listing and exporting status, leverage, tangibility, age, whether it belongs to a business group and the dominant coarse-grained caste of the firm. This last firm characteristic also helps account for the possibility that some caste/religion groups are more cohesive than others such that diversity may have different implications for firm performance. In a robustness check, we split our sample into sub-groups of firms dominated by various coarse-grained castes (for Hindus) and religions (for non-Hindus).²⁶

²⁶To control for firm ownership beyond whether it belongs to a business group, in one specification we additionally

We also control for industry characteristics to account for the possibility that the industry is dominated by similar castes or closely related firms or has low competition. In particular, we control for industry's market competition, proportion of business group firms, and the shares of firms dominated by various castes or religions. All of these could be related to both supply pools and individual firm performance. In another robustness check, we adjust for industry by taking differences of firm performance variables from the corresponding industry median and regressing them on the above-mentioned control variables.

Results: Table 4 Panel (B) (columns (1), (2) and (3)) presents the first stage results, showing coefficients only for the excluded instruments (fine-grained caste homophily levels of the industry and state director supply pools for a given firm). Since the samples differ somewhat due to missing observations for the second-stage dependent variables, there is a different first stage regression equation estimated for each of the three dependent variables that we consider. The table shows that both instruments are strongly and significantly positively associated with an average firm's board homophily. The corresponding first stage F-statistics are above 10, indicating that the instruments explain a significant proportion of the variation in the endogenous regressors. Online Appendix Table 8 Panel (B), presents similar first stage results when the first stage dependent variables are coarse-grained caste and religion homophily of boards, respectively.

Table 4 Panel (A) (Columns (1), (2) and (3)) presents second stage results. The estimated coefficients on fine-grained caste homophily show that all performance measures are worse when boards are less diverse. For example, Tobin's Q falls by 2.32 points when fine-grained caste

include whether the firm is a family firm, percentage of major equity investments coming from institutional investors, and the proportion of promoters on the board. Due to data limitations these variables are measurable only for a considerably smaller number of firms. Second stage results for the second IV approach show that lower diversity in boards relates to significantly worse firm performance. We find that CEO duality is negatively associated with firm performance and that a firm that has greater institutional investors is associated with better firm performance. A one unit increase in fine-grained caste homophily leads to almost 0.28 log points drop in operating income and 0.22 point drop in Tobin's Q.

homophily increases by one unit. Volatility also increases 0.007 points, although the estimate is not statistically significant. Similar results for coarse-grained caste and religion homophily are presented in Table 8 Panel (A) of the Online Appendix.

Industry Adjusted Regressions: In our main regressions, to ensure validity of our instruments we control for several industry characteristics that can influence both our instruments and firm performance. We conduct an alternative robustness check wherein we use industry-adjusted firm performance measures as our dependent variables. For this purpose, we take the difference of all firm performance variables in each year from the corresponding industry (two-digit NIC) medians in that year and regress these on the same regressors as in equation (4.1). Note that we drop industry-fixed effects in this specification since fixed effects would de-mean all variables by the industry mean, invariant across years.

Columns (4), (5) and (6)) of Table 4 present the industry-adjusted first and second stage results for the first IV approach for fine-grained caste homophily.²⁷ In Panel (B) we see that homophily of industry and state supply pools of directors are strongly and significantly positively associated with an average firm's fine-grained caste homophily. First stage F-statistics are above 10 in all cases. Table 4 Panel (A) presents the second stage results. The estimated coefficients on fine-grained caste homophily show that all performance measures are worse when boards are less diverse. For example, Tobin's Q falls by 3 points when fine-grained caste homophily increases by one unit.

5.2 Approach 2: Euclidean Distance between Caste Composition of Firm's Board and Industry and State Director Supply Pools

In our second instrumental variable approach, we additionally use the distance between a firm's board and its industry/state with regard to their caste/religion composition. Note that several different compositions can yield the same homophily index. So whether a firm's board composition is similar to that in its industry/state can be determined not only by comparing its overall homophily

²⁷Results for coarse-grained caste homophily and religion are available upon request.

index with that of the industry/state but also its underlying composition. The larger this distance, the less similar is the firm's director composition to that in the industry/state. Since these additional Euclidean distance based measures vary across firms (and over time), instead of only across industries or states, the relevance of our set of instruments also increases. Table 9 of the Online Appendix demonstrates that Euclidean distances between firms and industry/state level fine-grained caste composition of directors vary considerably, but are generally quite small. Panel A of the table shows key moments of the distances between firms' director composition and industry director composition for four years over the sample period. We see that the distribution of these distances is quite stable over time. In all years, the mean distance is slightly larger than the median, indicating that the distribution has a heavier right tail. Relative to the magnitudes of these distances, the standard deviation is quite large, suggesting considerable variation within years. Similar patterns are evident for distances between firm and state director compositions (Panel B).

The Euclidean distance between a board's caste/religion composition and that of the aggregate set of directors in the corresponding industry or state also meets the exclusion criterion. The reasons described above for the validity of the industry and state-level homophily indices also apply to the distance measures. We do an additional robustness check to further account for the possibility that firm performance itself may influence Euclidean distance of its board from the supply pools. We split the sample into firms that are above and below the industry median of a performance measure and estimate the same regression for these sub-samples. We also assess firm performance association with caste homophily for different dominant caste sub samples. This allows us to account for the possibility that different castes may have different levels of cohesion.

Results: We present results from this second instrumental variable approach, in which the excluded instruments are the homophily of industry and state level director supply pools as well as the distance between the caste composition of these supply pools and that of individual boards. Table 5 Panel (B) (Columns (1), (2) and (3)), shows the first stage results. We observe that greater homophily in both industry and state supply pools is significantly positively correlated with an

average firm's board homophily. Greater distance from the industry/state caste composition of directors is also associated positively and significantly with an average firm's board homophily. The F statistics are also large, indicating that these excluded instruments are highly correlated with our endogenous regressor.

Second stage results are presented in Table 5 Panel (A) for fine-grained caste homophily and in Online Appendix Table 10 Panel (A), for coarse-grained caste and religion homophily, respectively. We find similar results as in the first IV strategy. Specifically, we see that fine-grained caste homophily is associated with significantly worse accounting and market based measures of firm performance and significantly greater firm risk. Greater coarse-grained caste homophily is also associated with worse firm outcomes. Similar results are observed for religion homophily and firm performance, although they are less precisely estimated.

Industry Adjusted Regressions: Like in our first approach, we conduct an alternative robustness check wherein we use industry-adjusted firm performance measures as our dependent variables. Table 5 (Columns (4), (5) and (6)) presents the first and second stage results for the second IV approach after industry adjustment for fine-grained caste homophily. From Panel (B) of the table we observe that homophily of industry and state supply pools of directors are strongly and significantly positively associated with an average firm's fine-grained caste board homophily. The excluded instruments are significantly positively associated with firm's fine-grained caste homophily. Table 5, Panel (A) shows the second stage results. The estimated coefficients on fine-grained caste homophily show that all performance measures are worse when boards are less diverse. For example, Tobin's Q falls by 0.256 points when fine-grained caste homophily increases by one unit.

Above and Below Median Firms: We include several control variables to ensure exclusion restriction of the Euclidean distance based IVs. It is still possible that poorly performing firms are unable to hire directors from beyond their already represented castes/religions, thereby affecting

the distance between the boards' and supply pools' caste compositions. In that case, the exclusion restriction would not hold. To address this possibility, we split our sample into two sub-groups of high- and low-performing firms identified as those above and below the industry (NIC two digit) median for that performance measure in that year. Table 6 presents second stage results for firms above and below the median for the second IV strategy. The table also shows the F statistics from the first stage regressions. We observe that the excluded IVs are highly correlated with board caste homophily in both sub-groups (above and below median). We also find that for both high- and low-performing firms, all measures of performance are worse when boards are less diverse. The magnitude of coefficients on board homophily are higher for high-performing firms. These findings demonstrate that our results hold even after accounting for the possibility that our excluded instruments may be correlated to both performance and diversity.

Caste Sub-Samples: We recognize the possibility that different social groups may have different levels of cohesion or sense of affinity. This has two implications for our measure of board diversity. First, the same HHI number may reflect different levels of social diversity since more cohesive caste or religious groups may have less social diversity than others. Second, greater cohesion within a caste group may result in less variation in caste homophily measures for boards with that as dominant caste. It is then possible that the results that we obtain may be driven by firms dominated by those castes that are less cohesive.

To address both possibilities, we divide our sample into eleven sub-samples of firms dominated by various coarse-grained castes (for Hindus) and religions (for non-Hindus). We present results for four coarse-grained castes and one non-Hindu religion (Jain). These groups have substantial numbers of firms in our data. The remaining groups do not have a large enough number of firms to yield precise estimates or high first-stage F statistics in our second instrumental variable strategy. Table 7 shows the first and second stage results for this strategy for the five sub-samples. We observe that the excluded instruments are positively and significantly associated with board fine-grained caste homophily across all sub-samples and firm outcomes. For each sub-sample, we

also observe that operating income and Tobin's Q are significantly negatively and volatility is significantly positively associated with fine-grained caste homophily. These results demonstrate that even though there may be heterogeneously cohesive caste and religious groups, our main results are not driven by any particular group(s).

5.3 Approach 3: Clause 49

In a third approach, we exploit board membership changes induced by firms complying with Clause 49 of a new set of corporate governance regulations announced by the Securities and Exchange Board of India (SEBI) that went into effect in February 2000. Among other things, the new requirement was for firms to have at least 50% of their board be comprised of non-executive members. The compliance deadlines differed for different groups of firms; March 31, 2001 for Group 1 – the largest (“Group A”) companies listed on the Bombay Stock Exchange (BSE) and National Stock Exchange (NSE) and S&P CNX Nifty Index companies, March 31, 2002 for Group 2 – other companies with paid-up share capital of at least Rs. 100 million, or net worth of at least Rs. 250 million, at any time in the firm's history, March 31, 2003 for Group 3 – firms with paid-up share capital of at least Rs. 30 million, and Group 4 – any newly listed or re-listed firms at the time they get listed. Clause 49 did not apply to older firms with lower net worth or paid-up capital. To construct the instrument, we closely follow Helmers et al. (2017) and use a regression discontinuity design where we compare firms just below and above the threshold of 50% board members being non-executive along with whether they were required to comply with the clause requirements. Therefore, the treatment depends on the interaction of whether a firm was required to comply and whether a firm was below the threshold of 50% non-executive directors. Similar to Helmers et al. (2017), our first stage equation is:

$$\begin{aligned}
H_{it} = & \alpha_0 + \alpha_1 BT_{it} * I(Group1)_i + \alpha_2 BT_{it} * I(Group2)_i + \alpha_3 BT_{it} * I(Group3)_i + \alpha_4 BT_{it} * I(Group4)_i \\
& + \alpha_5 BT_{it} + \alpha_6 I(Group1)_i + \alpha_7 I(Group2)_i + \alpha_8 I(Group3)_i + \alpha_9 I(Group4)_i \\
& + \beta_1 NonExecProp_{it} + \beta_2 (NonExecProp_{it})^2 + \beta_3 ShareCapital_{it} + \beta_4 (ShareCapital_{it})^2 \\
& + \beta_5 NetWorth_{it} + \beta_6 (NetWorth_{it})^2 + \beta_7 (NonExecProp_{it} * ShareCapital_{it}) \\
& + \beta_8 (NonExecProp_{it} * ShareCapital_{it})^2 + \beta_9 (NonExecProp_{it} * NetWorth_{it}) \\
& + \beta_{10} (NonExecProp_{it} * NetWorth_{it})^2 + \delta_1 X_{it} + \delta_2 I_{jt} + \delta_3 I_j + \delta_4 T_t + \epsilon_{ijt} \tag{5.1}
\end{aligned}$$

where BT_{it} denotes a binary variable that equals 1 if firm i is below the 50% non-executive directors threshold in year t and 0 otherwise, $I(GroupN)$ are binary variables for whether a firm i is in a group of firms required to comply with Clause 49, for $N = 1, 2, 3, 4$, $NonExecProp_{it}$ is a continuous variable measuring the proportion of non-executive members on firm i 's board in year t , $NetWorth_{it}$ is the net worth of firm i in year t and $ShareCapital_{it}$ is firm i 's paid up share capital in year t .²⁸ The remaining control variables are as in equation (4.1). Note that the coefficients on the excluded instruments indicate the intent-to-treat since there may not be perfect compliance with the regulation. To estimate this regression, we build a longitudinal sample of firms for the period 1999-2007 such that we can observe a firm for at least two consecutive years. We cut off the sample period in 2007 since in 2008, a new requirement around independent directors was included in the amended Clause 49 which would induce other changes in board membership that would be correlated with the changes we are focusing on. Following Lee and Lemieux (2010) and Helmers et al. (2017), we include all firms in the sample but additionally include polynomials of the continuous running variables and their interactions as shown in the equation above. We report results for a quadratic polynomial.²⁹

²⁸Note that Group 2 firms were those with at least Rs. 250 million in net worth at any time in their histories to comply. Thus, for Group 2 firms, for the first year in our sample period, we look into each firm's records in all previous years and take the maximum net worth across those years.

²⁹We also estimate results for firms falling in narrow bandwidths around the cutoff of 50% non-executive directors. The main coefficient of interest (on board homophily) remains qualitatively similar but the instruments are weak for

The regulation only specified the percentage of non-executive directors that a firm must have. So the firms were free to choose directors of any caste or religion. The directors that they hired to comply with Clause 49 would influence the board caste/religion homophily, our endogenous regressor. Exclusion restriction on this set of instruments is also plausible given the regression discontinuity design. We further include controls to ensure the validity of the instrument. The choice of who to hire may be correlated to the industry caste composition, industry competition, and other observable or unobservable industry characteristics which also matter for firm performance. Thus, we include industry fixed effects, industry HHI, proportion of business group firms in the industry and shares of firms dominated by various coarse-grained castes/religions. Board hires may also be impacted by several firm characteristics that correlate with performance. We also control for these firm characteristics as described earlier for equation (4.1).

Results: We present first stage F statistics and second stage results from this third IV strategy in Table 8. The regression discontinuity based set of excluded instruments is highly correlated with the board's fine-grained caste homophily, as evident from the high F statistics for all first stages corresponding to each second-stage dependent variable. Estimates of our main coefficients of interest show that accounting and market-based measures of firm performance are worse when boards are less diverse. For the same change in homophily, Tobin's Q falls and volatility increases significantly by 4 and 0.2 points, respectively.

5.4 Approach 4: Exposure to Diversity Outside the Board

In our last strategy, we construct an instrument similar in spirit to the one developed by Adams and Ferreira (2009). In that study, the authors examine the firm performance effects of gender diversity on the board, an endogenous regressor. They instrument for the latter by calculating the fraction of a firm's male directors that serve on other boards that have female directors. The intuition is that connections of a firm's male directors with female directors on other boards is likely positively

some of these smaller samples for some performance variables.

correlated with gender-diversity on the firm's board. Following this logic, we measure the fraction of a firm's directors who serve on at least one other board whose dominant fine-grained caste is different from their own. We expect such exposure to caste diversity to be negatively correlated with caste homophily on the firm's board, our endogenous regressor.

Several control variables help us ensure the validity of this instrument. Directors' exposure to other castes may itself be influenced by the caste composition of the industry or other characteristics of the industry, both of which correlate with firm performance. To account for these possibilities, we include industry fixed effects, industry's HHI, the share of group firms in the industry, and a vector of shares of firms in the industry dominated by various coarse-grained castes (for Hindus) and non-Hindu religions. Several firm characteristics that influence firm performance may also drive the connectedness of the board. Thus, we include a vector of several firm characteristics as described earlier for equation (4.1) and, additionally, control for the total number of other directorships held by the directors of the firm.

Note that construction of this instrument relies on information in the Indian Boards Database. Using these data we can identify which other firms' boards a director serves on and then measure the dominant castes of those boards. Based on the information available, we are able to calculate this instrument only for 936 publicly traded firms.

Results: Table 9 (Panel B) presents the first stage results. The table shows that the instrument is strongly negatively correlated with firms' board homophily and that the first-stage F statistic is also large across all regressions corresponding to the various second stage dependent variables. Table 9 (Panel A) presents the second stage results for this IV approach. We observe that operating income and Tobin's Q are negatively associated with caste homophily of the board, although the coefficients are imprecisely estimated. Firm volatility significantly increases by 0.047 points with a one unit increase in board caste homophily.

On the basis of our detailed and rigorous empirical analyses, we conclude that lack of diversity on firm boards is negatively associated with key firm outcomes.

6 Possible Mechanisms

Our results demonstrate that social diversity of a firm's directors is associated with poor firm performance. This may happen because such diversity can affect directors' viewpoints, information sets, or group dynamics, which ultimately have a bearing on their monitoring and advisory functions. Homogeneous directors may have overlapping networks, perspectives and information sets, which can impair their advisory role. Socially homogeneous boards may also be characterized by cronyism, with directors not challenging each other and the management sufficiently. This can limit their ability to monitor and advise well. These phenomena can worsen board decisions and, in turn, firm outcomes. However, homogeneous boards can also prove beneficial for the firm since socially proximate directors may have more agreeable group dynamics and fewer interpersonal conflicts leading them to function efficiently. The relative strengths of these channels determine the net effect of board homogeneity on firm performance. In this section, we use available data to explore some of these channels.

Role of Board during slow economic growth years: During periods of challenging macroeconomic conditions, directors' advisory role, such as providing novel perspectives to address business problems as well as opening doors to new business opportunities via their networks, becomes particularly important for firm performance. Because of similar or overlapping viewpoints and networks, we expect that boards with high levels of caste homogeneity are less likely to provide effective guidance to the firm just when it is sorely needed. This reasoning implies that firms with homogeneous boards are likely to suffer more in terms of performance during economic slowdown.

We explore this possibility by classifying our sample years into high- and low-growth years based on whether the GDP growth rate in that year was above or below the average growth rate during the sample period, 1999-2015. We augment our baseline regression with an indicator for below-average growth year and its interaction with board fine-grained caste homophily. Results are presented in Table 10. We find that while higher homophily is always associated with poor performance, firms with higher board homophily perform even worse during years of slow growth. In such years, a one unit increase in board caste homophily is associated with 0.356 log points

lower operating income relative to sales and 0.07 points lower Tobin's Q. This finding is consistent with our reasoning that high caste homophily on the board, in part, adversely affects firm performance because it impedes the board's advisory function. This pattern in the data is thus consistent with a mechanism whereby board's social homogeneity is associated with poor firm performance due to a paucity of novel perspectives on the board to tackle thorny business problems faced by the firm.

Representation on Key Committees: Next, we examine the mechanism that socially homogeneous boards are less effective in monitoring management – such as challenging the firm's management on the formulation and execution of strategy and resultant firm performance outcomes. If this putative mechanism is indeed operative in our data, we expect that firms with homogeneous boards will perform even worse when key board committees – audit, shareholder grievance, and nomination and compensation committees - that impact the board's monitoring role have an over-representation of the dominant caste. For this purpose, we calculate the ratio of the proportion of dominant caste directors on the committee to that in the full board. A ratio higher than one indicates over-representation of dominant-caste directors on key committees.

We observe that the level of over-representation (defined as the ratio of the proportion of dominant caste directors on the committee to that in the full board) is somewhat more likely in more homogeneous boards. For instance, at the beginning of our sample period, the correlations between board homophily and the level of dominant-caste over-representation on audit, shareholder grievance and compensation and nomination committees are 0.1, 0.02 and 0.3, respectively. Next, we augment our baseline regression with indicators for whether these key committees have over-representations of dominant-caste directors and their interactions with overall board homophily. Results are presented in Table 10. We find that over-representation of dominant-caste directors on these important committees is associated with lower operating income and Tobin's Q and higher firm risk, although the coefficients are not always statistically significant. Looking at the interactions of over-representation incidences with board homophily, we find that the negative association of board caste homophily with firm outcomes is further compounded by such over-

representation on committees. For example, a one unit higher homophily is associated with 0.637 log points lower operating income relative to sales, 0.607 points lower Tobin's Q, and 0.075 points higher volatility when the compensation and nomination committee has over-representation of dominant-caste directors. Similarly, a unit higher homophily is associated with 0.834 points lower Tobin's Q for boards whose audit committee has over-representation of dominant-caste directors. These results suggest that the negative association of board homophily with firm performance is more pronounced when directors belonging to the boards' dominant caste are over-represented on key committees.³⁰ Such over-representation is suggestive of cronyism and absence of diverse viewpoints, both of which may impede the monitoring and advising work of the board.

CEO caste: CEO appointment is one of the most important tasks of the board. A socially homogeneous board characterised by cronyism is more likely to appoint a CEO of the same caste as that dominant on the board. This may reduce the care or intensity with which the board monitors the actions of the CEO, thereby worsening firm performance. Our empirical evidence is consistent with this possibility. We observe that firms with less diverse boards are more likely to have the CEO belong to the dominant fine-grained caste of the board, with a correlation of about 0.5 at the beginning and end of our sample period. We also see in Figure 2 that throughout the sample period, firms with their CEO belonging to the dominant fine-grained caste of the board have substantially higher average caste homophily than others.

Further, this is associated with worse firm outcomes. To establish this, we regress key measures of firm performance on an indicator variable for whether the CEO is of the same fine-grained caste as the one dominant on the board and its interaction with board fine-grained caste homophily, in addition to all other regressors included in our baseline regression. Results presented in Table 10 show that an average firm with a CEO belonging to the same caste as the one dominant on the board has lower operating income relative to sales, Tobin's Q and higher firm risk. Additionally, we note from the coefficients on the interaction variable of the indicator whether dominant caste CEO

³⁰All other coefficients on the interactions between board caste homophily and over-representation of dominant caste directors on committees are negative but not precisely estimated.

is same as the dominant caste with board caste homophily that the perverse association of board homophily with firm outcomes is worse when the CEO belongs to the board's dominant caste. A unit increase in caste homophily of the board is associated with 0.004 points higher volatility for firms whose CEO belongs to the board's dominant caste. However, the coefficients on this interaction term do not indicate a statistically significant compounding of homophily associations with operating income/sales and Tobin's Q.

These results provide suggestive evidence that shared caste identity between a large number of directors and the CEO entails that the directors do not effectively monitor the CEO or do not sufficiently challenge his/her decisions. Such impairment of the board's monitoring and advising functions ultimately harms the firm.^{31 32}

Taken together, our findings paint a picture wherein more homophilous boards are likely to have overlapping perspectives, networks and information sets, and engage in cronyism. In essence, greater caste homogeneity at the board level has a deleterious effect on firm performance in our data due, in part, to the impairment of a homogeneous board's monitoring as well as advisory role.

7 Conclusion

We build a unique dataset that allows us to map Indian last names to caste and religion and use it to measure caste diversity among corporate directors. We show that Indian corporate boards persistently and systematically lack in diversity. Our results demonstrate that the homogeneity at the board level has a negative association with key measures of firm value and performance. Through our moderating tests of mechanisms, we infer that homogeneous directors are likely to have overlapping perspectives, networks and information sets, and possibly engage in cronyism.

³¹Table 10 presents fixed effect results for the three mechanisms we explore. We find similar results for instrumental variable approaches 1 – homophily of state and industry director supply pools – and 2 – homophily of state and industry director supply pools and their Euclidean distance with the caste composition of a firm's board.,

³²We also profile two behavioral outcomes of directors – their attendance rates of board meetings and tenure. Data show that board caste homophily is positively correlated with meeting attendance rates of dominant caste directors. While this may indicate cronyism or more agreeable boardroom dynamics, we recognize that the reporting of meeting attendance by firms is likely to be incomplete and selective. Data also show that more homogeneous boards have higher tenure for dominant caste directors.

Consequently, boards' monitoring and advising functions are impaired, to the detriment of the firm.

Besides developing a potentially highly useful dataset for future studies and highlighting the importance of board diversity for firm outcomes, our paper has another key takeaway. While much research, corporate governance laws, and recommendations emphasize gender diversity in boards, our results demonstrate that socio-cultural aspects other than gender influence firm outcomes. Recent proposals that urge firms to hire diverse directors along dimensions such as race and sexual orientation (NASDAQ, 2020), and ethnicity and backgrounds (UK's Financial Reporting Council, 2018) appear to recognize this.

References

- Adams, R. B. and Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of financial economics*, 94(2):291–309.
- Ahern, K. R. and Dittmar, A. K. (2012). The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation. *The Quarterly Journal of Economics*, 127(1):137–197.
- Arnaboldi, F., Casu, B., Kalotychou, E., and Sarkisyan, A. (2018). The performance effects of board heterogeneity: what works for eu banks? *The European Journal of Finance*, pages 1–28.
- Åslund, O., Hensvik, L., and Skans, O. N. (2014). Seeking similarity: How immigrants and natives manage in the labor market. *Journal of Labor Economics*, 32(3):405–441.
- Bengtsson, O. and Hsu, D. H. (2015). Ethnic matching in the us venture capital market. *Journal of Business Venturing*, 30(2):338–354.
- Bertrand, M., Black, S. E., Jensen, S., and Lleras-Muney, A. (2018). Breaking the glass ceiling? the effect of board quotas on female labour market outcomes in norway. *The Review of Economic Studies*, 86(1):191–239.

- Beteille, A. (1996). Varna and jati. *Sociological Bulletin*, 45(1):15–27.
- Chen, G., Chittoor, R., and Vissa, B. (2015). Modernizing without westernizing: Social structure and economic action in the indian financial sector. *Academy of Management Journal*, 58(2):511–537.
- Claes, K. and Vissa, B. ((2020)). ”does social similarity pay off? homophily and venture capitalists’ deal valuation, downside risk protection, and financial returns in india.”. *Organization Science*.
- Coles, J., Naveen, D., and Naveen, L. (2008). Boards: Does one size fit all?
- Currarini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045.
- Damaraju, N. L. and Makhija, A. K. (2018). The Role of Social Proximity in Professional CEO Appointments: Evidence from Caste/Religion-Based Hiring of CEOs in India. *Strategic Management Journal*, 39(7):2051–2074.
- Damodaran, H. (2008). *India’s New Capitalists: Caste, Business, and Industry in a Modern Nation*. Springer.
- Dass, N., Kini, O., Nanda, V., Onal, B., and Wang, J. (2013). Board expertise: Do directors from related industries help bridge the information gap? *The Review of Financial Studies*, 27(5):1533–1592.
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2):158–165.
- Dumont, L. (1980). *Homo Hierarchicus: The Caste System and Its Implications*. University of Chicago Press.

- Fairlie, R. W., Hoffmann, F., and Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review*, 104(8):2567–91.
- Fisman, R., Paravisini, D., and Vig, V. (2017). Cultural Proximity and Loan Outcomes. *American Economic Review*, 107(2):457–92.
- Freeman, R. B. and Huang, W. (2015). Collaborating with People like Me: Ethnic Coauthorship Within the United States. *Journal of Labor Economics*, 33(S1):S289–S318.
- Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The Elements of Statistical Learning*, volume 1. Springer series in statistics New York.
- Gennaro, B., Bhagwat, V., and Yonker, S. (2018). Board diversity, firm risk, and corporate policies. *Journal of Financial Economics*, 127(3):588–612.
- Ghani, E., Kerr, W. R., and O’connell, S. (2014). Spatial Determinants of Entrepreneurship in India. *Regional Studies*, 48(6):1071–1089.
- Giannetti, M. and Zhao, M. (2019). Board ancestral diversity and firm performance volatility. *Journal of Financial and Quantitative Analysis*, 54(3):1117–1155.
- Giuliano, L., Levine, D. I., and Leonard, J. (2009). Manager race and the race of new hires. *Journal of Labor Economics*, 27(4):589–631.
- Giuliano, L. and Ransom, M. R. (2013). Manager ethnicity and employment segregation. *ILR Review*, 66(2):346–379.
- Glover, D., Pallais, A., and Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: Evidence from french grocery stores. *The Quarterly Journal of Economics*, 132(3):1219–1260.
- Gompers, P. A., Mukharlyamov, V., and Xuan, Y. (2016). The cost of friendship. *Journal of Financial Economics*, 119(3):626–644.

- Hegde, D. and Tumlinson, J. (2013). Can Birds of a Feather Fly Together? the Payoffs of Ethnic Proximity in Us Venture Capital. *Unpublished working paper. New York University and University of Munich.*
- Helmers, C., Patnam, M., and Rau, R. (2017). Do board interlocks increase innovation? evidence from a corporate governance reform in india. *Journal of Banking and Finance*, 80:51–70.
- Hnatkovska, V., Lahiri, A., and Paul, S. (2012). Castes and Labor Mobility. *American Economic Journal: Applied Economics*, 4(2):274–307.
- Hnatkovska, V., Lahiri, A., and Paul, S. B. (2013). Breaking the Caste Barrier Intergenerational Mobility in India. *Journal of Human Resources*, 48(2):435–473.
- Iyer, L., Khanna, T., and Varshney, A. (2013). Caste and entrepreneurship in india. *Economic and Political Weekly*, pages 52–60.
- Jodhka, S. S. (2010). Dalits in business: Self-employed scheduled castes in north-west india. *Economic and Political Weekly*, pages 41–48.
- Joshi, S., Kochhar, N., and Rao, V. (2018). Jati inequality in rural bihar. *The World Bank.*
- Kim, D. and Starks, L. T. (2016). Gender diversity on corporate boards: Do women contribute unique skills? *American Economic Review*, 106(5):267–71.
- Knyazeva, A., Knyazeva, D., and Masulis, R. W. (2013). The supply of corporate directors and board independence. *The Review of Financial Studies*, 26(6):1561–1605.
- Kramarz, F. and Thesmar, D. (2013). Social networks in the boardroom. *Journal of the European Economic Association*, 11(4):780–807.
- Kumar, H. and Somanathan, R. (2017). Caste connections and government transfers: The mahadalits of bihar. *Working Paper, Centre for Development Economics, Delhi School of Economics.*

- Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.
- Munshi, K. (2019). "caste and the indian economy.". *Journal of Economic Literature*, 57(4):781–834.
- Munshi, K. and Rosenzweig, M. (2013). Networks, commitment, and competence: Caste in indian local politics. Technical report, National Bureau of Economic Research.
- Munshi, K. and Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(1):46–98.
- Petersen, T., Saporta, I., and Seidel, M.-D. L. (2000). Offering a job: Meritocracy and social networks. *American Journal of Sociology*, 106(3):763–816.
- Shayo, M. and Zussman, A. (2011). Judicial ingroup bias in the shadow of terrorism. *The Quarterly Journal of Economics*, 126(3):1447–1484.
- Sila, V., Gonzalez, A., and Hagendorff, J. (2016). Women on board: Does boardroom gender diversity affect firm risk? *Journal of Corporate Finance*, 36:26–53.
- Srinivas, M. N. (1995). *Social Change in Modern India*. Orient Blackswan.
- Terjesen, S., Sealy, R., and Singh, V. (2009). Women directors on corporate boards: A review and research agenda. *Corporate governance: an international review*, 17(3):320–337.
- Thorat, S. and Neuman, K. S. (2012). *Blocked by Caste: Economic Discrimination in Modern India*. Oxford University Press.
- Varshney, A., Chatterjee, P., and Katznelson, I. (2012). Two Banks of the Same River? Social Order and Entrepreneurialism in India. *Anxieties of Democracy: Tocquevillean Reflections on India and the United States*, pages 225–56.

Vissa, B. (2011). A matching theory of entrepreneurs' tie formation intentions and initiation of economic exchange. *Academy of Management Journal*, 54(1):137–158.

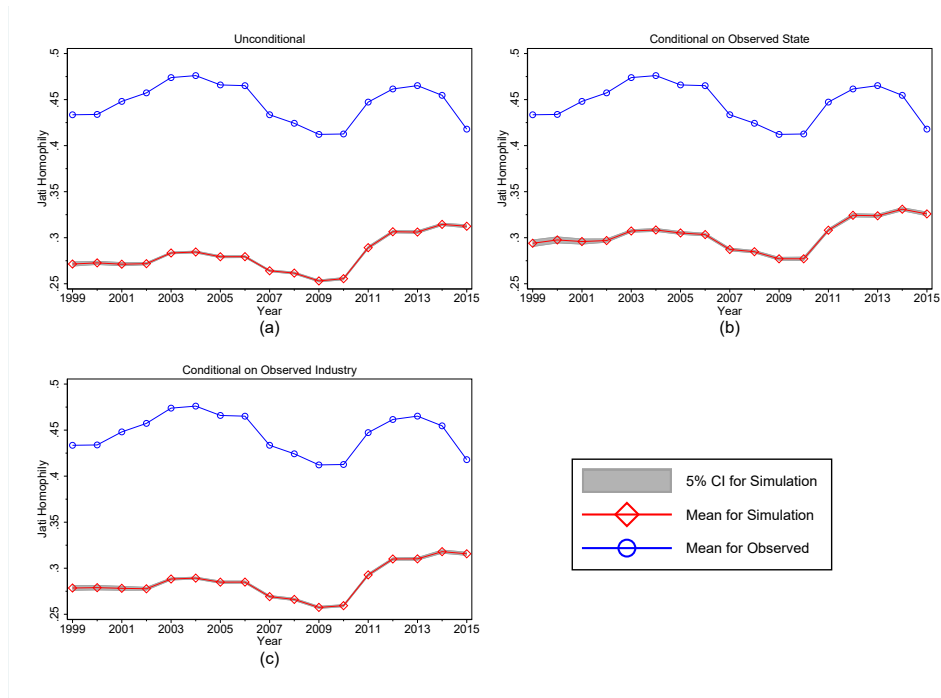


Figure 1: Observed vs. Simulated Average Fine-Grained Caste Homophily^a

^aSource: Prowess, matrimonial data. The three graphs in the figure present the mean fine-grained caste homophily across firms each year in the observed and simulated samples for three distinct simulation criteria: unconditional, conditional on firm's state and on firm's industry. Details about the simulation methods are provided in Section 3.

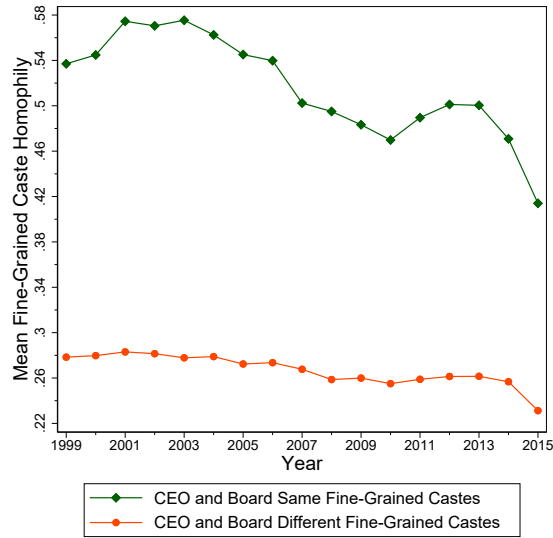


Figure 2: Caste Homophily for Firms with & without CEO of Board’s Dominant Fine-Grained Caste^a

^aNotes: Source: Prowess, matrimonial data. Homophily is averaged over all firms in a year that fall into either of two groups: those that have their CEO belonging to the same fine-grained caste as that dominant in the rest of the board, and those where the CEO belongs to a different fine-grained caste.

Table 1: Variable Definitions

Variables	Definitions
<i>Panel A: Firm Variables</i>	
Age of firm	Number of years since incorporation of firm
Export status	Indicator variable: one for exporting firms, zero otherwise
State of registration	The Indian state in which the firm is registered
Industry	Two or three digit industry; National Industrial Classification (2008)
Listing status	Indicator variable: one for firms listed either in the Bombay Stock Exchange (BSE) or the National Stock Exchange (NSE) at that point in time, zero otherwise
Assets	Book value of total assets in rupees million deflated by the all-India CPI (2001=100)
Sales	Total value of sales in rupees million deflated by the all-India CPI (2001=100)
Profits	Total value of profits in rupees million deflated by the all-India CPI (2001=100)
Leverage	Book value of debt over book value of total assets
Operating income	Sales less operating expenses
Tangibility	Net property, plant, and equipment over book value of total assets
Tobin's Q	Sum of book value of debt, book value of preferred stock and market value of common stock over book value of assets. The market value of common stock is measured in two ways – a. the latest market value available on or before the reporting date b. the mean market value over the entire reporting period
Volatility	The standard deviation of stock returns of a firm in the entire reporting period
<i>Panel B: Board Characteristics</i>	
Board size	Number of directors in the board
Board attendance	Mean number of board meetings attended by all members of a board over total number of board meetings

CEO duality Indicator variable: one if the at least one CEO of a firm is also the chair.

Panel C: Measures of Homophily

Dominant coarse- and fine-grained caste and religion of a board The coarse- and fine-grained caste and religion of the maximum number of directors of a board. In case of ties, the dominant coarse- and fine-grained caste and religion is chosen randomly from the tie.

Board coarse- and fine-grained caste and religion homophily Coarse- and fine-grained caste and religion HHI, i.e., the sum of squared shares of all fine-grained castes (coarse-grained castes, religions) represented on the board.

Sector coarse- and fine-grained caste and religion homophily Coarse- and fine-grained caste and religion HHI, i.e., the sum of squared shares of all fine-grained castes (coarse-grained castes, religions) of directors in an industry. The baseline approach considers each name as a distinct directorship, even if the name is same. The alternative approach considers all occurrences of the same name as one unique director.

State coarse- and fine-grained caste and religion homophily Coarse- and fine-grained caste and religion HHI, i.e., the sum of squared shares of all fine-grained castes (coarse-grained castes, religions) of directors in a state. The baseline approach considers each name as a distinct directorship, even if the name is same. The alternative approach considers all occurrences of the same name as one unique director.

Sector coarse- and fine-grained caste and religion Euclidean distance of a board Distance between the vector representing the coarse- and fine-grained caste and religion composition of directors in the industry and the corresponding vector for the firm board. The baseline approach considers each name as a distinct directorship, even if the name is same. The alternative approach considers all occurrences of the same name as one unique director.

State coarse- and Distance between the vector representing the coarse- and fine-grained fine-grained caste and caste and religion composition of directors in the state and the religion Euclidean corresponding vector for the firm board. The baseline approach distance of a board considers each name as a distinct directorship, even if the name is same. The alternative approach considers all occurrences of the same name as one unique director.

Table 2: Summary Statistics

	Above Median Homophily	Below Median Homophily
% Exporters	17.31	34.67
%Group Firms	27.06	41.76
Mean Age	18.35	23.52
Mean Assets (Rupees Millions)	5.3	100.82
Mean Profits (Rupees Millions)	0.53	9.92
Mean Net Tangible Asset Intensity	0.28	0.28
Mean Leverage	1.4	0.96
Mean Return on Assets	0.02	0.04
Mean Tobin's Q	0.53	0.82
Mean Risk	0.06	0.05
Mean % Independent Directors	2.06	5.58
Mean Board Size	3.98	6.86

Source: Prowess. This table provides basic summary statistics for firms in our sample that have at least two directors and for which we can assign a caste or religion identity for all directors on the boards.

Table 3: Fine Grained Homophily and Firm Outcomes: Fixed Effects

Variables	(1) Log(Operating Income/Sales)	(2) Tobin's Q	(3) Volatility
Fine-Grained Caste Homophily	-0.263*** (0.039)	-0.186*** (0.046)	0.008*** (0.002)
Board Size	-0.030** (0.014)	0.047** (0.019)	-0.003*** (0.001)
Board Size Squared	0.002*** (0.001)	-0.001 (0.001)	0.000*** (0.000)
Industry HHI	0.029 (0.059)	0.059 (0.071)	0.001 (0.002)
I(Listed)	-0.028 (0.025)	-1.390** (0.686)	-0.004 (0.003)
I(Group Firm)	0.051** (0.025)	0.180*** (0.034)	-0.005*** (0.001)
Proportion of Group Firms in Industry	0.354 (0.219)	-0.026 (0.103)	-0.006* (0.004)
Observations	46,808	27,163	28,809
R-squared	0.116	0.379	0.465

Notes: This table presents fixed effects results for firm performance variables – log of operating income relative to sales, Tobin's Q, and volatility – regressed on board fine-grained caste homophily and other control variables, and year and two-digit industry fixed effects. Control variables include age, board size, board size square, leverage, real assets, tangibility, listing status, export status, HHI measured at NIC3 level, whether the firm belongs to a business group, dominant caste of the board, shares of firms with various dominant coarse-grained castes in the industry, and proportion of group firms in the industry. All variables are defined in Table 1. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). Robust standard errors, clustered by industry and year, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 4: Fine-Grained Caste Homophily and Firm Outcomes: IV Approach 1

Variables	(A) Second Stage Results					
	No Industry Adjustment			Industry Adjusted		
	(1)	(2)	(3)	(4)	(5)	(6)
	In(Operating Income/Sales)	Tobin's Q	Volatility	In(Operating Income/Sales)	Tobin's Q	Volatility
Fine-Grained Homophily	-2.311** (0.998)	-2.320*** (0.891)	0.007 (0.027)	-1.729** (0.872)	-3.010*** (0.974)	0.011 (0.026)
Board Size	-0.228** (0.103)	-0.082 (0.055)	-0.003 (0.002)	-0.140** (0.059)	-0.002 (0.002)	-0.168* (0.091)
Board Size Squared	0.010** (0.004)	0.003 (0.002)	0.000 (0.000)	0.006** (0.002)	0.000 (0.000)	0.008** (0.004)
Industry HHI	-0.008 (0.064)	0.028 (0.071)	0.001 (0.002)	0.032 (0.075)	-0.003 (0.003)	-0.019 (0.057)
I(Listed)	-0.136** (0.061)	-1.393* (0.102)	-0.004 (0.066)	-1.248 (0.944)	-0.001 (0.003)	-0.094* (0.057)
Prp'n of Group Firms in Industry	0.267 (0.221)	-0.107 (0.104)	-0.006* (0.004)	-0.388*** (0.134)	0.010*** (0.004)	-0.292 (0.179)
	(B) First Stage Results					
	State Fine-Grained Caste Homophily	0.477*** (0.109)	0.447*** (0.125)	0.507*** (0.109)	0.479*** (0.12)	0.477*** (0.123)
	Industry Fine-Grained Caste Homophily	0.464*** (0.153)	0.275** (0.138)	0.334*** (0.081)	0.197 (0.144)	0.186 (0.131)
	First Stage F-statistic	15.13	10.65	21.32	10.85	11.52
	Observations	46,753	28,809	46,754	27,164	28,811

Notes: This table presents first and second stage IV - Approach 1 results for firm performance variables (log operating income relative to sales, Tobin's Q and volatility) regressed on board fine-grained caste homophily and other control variables, and year and two-digit industry fixed effects for instrumental variable approach 1. Panel (B) presents the first stage results where the excluded instruments are the homophily of state and industry level director supply pools. Panel (A) presents second stage results. Control variables include firm age, leverage, real assets, tangibility, export status, three digit industry HHI, whether the firm belongs to a business group, whether the firm is listed or not, dominant caste of the board, shares of firms with various dominant coarse-grained castes in three digit industry, and the proportion of group firms in the industry. All variables are defined in Table 1. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). The first stage F statistic is the Kleibergen-Paap rk Wald F statistic. Robust standard errors, clustered by industry and year, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 5: Fine-Grained Caste Homophily and Firm Outcomes: IV Approach 2

Variables	No Industry Adjustment			(A) Second Stage Results		
	(1)	(2)	(3)	(4)	(5)	(6)
	In(Operating Income/ Sales)	Tobin's Q	Volatility	In(Operating Income/ Sales)	Tobin's Q	Volatility
Fine-Grained Homophily	-0.242*** (0.043)	-0.182*** (0.045)	0.008*** (0.002)	-0.204*** (0.049)	-0.256*** (0.050)	0.007*** (0.002)
Board Size	-0.028* (0.105)	0.048** (0.020)	-0.003 (0.001)	-0.023 (0.015)	0.026 (0.017)	-0.003*** (0.001)
Board Size Squared	0.002** (0.001)	-0.001 (0.001)	0.000*** 0	0.002** (0.001)	-0.000 (0.001)	0.000** (0.000)
Industry HHI	-0.032 (0.058)	0.059 (0.071)	0.001 (0.002)	0.011 (0.052)	0.105 (0.079)	-0.003 (0.002)
I(Listed)	-0.027 (0.025)	-1.390 (0.686)	-0.004 (0.003)	-0.011 (0.025)	-1.245 (0.822)	-0.001 (0.003)
Prp'n of Group Firms in Industry	0.346 (0.219)	-0.026 (0.103)	-0.006* (0.004)	-0.228 (0.157)	-0.290** (0.137)	0.010*** (0.003)
(B) First Stage Results						
State Fine-Grained Caste Homophily	0.482*** (0.071)	0.419*** (0.075)	0.429*** (0.073)	0.490*** (0.074)	0.425*** (0.074)	0.433*** (0.072)
Industry Fine-Grained Caste Homophily	1.134*** (0.107)	0.815*** (0.175)	0.816*** (0.170)	1.177*** (0.116)	0.931*** (0.15)	0.934*** (0.151)
Distance from State Dir	0.316** (0.058)	0.286*** (0.048)	0.286*** (0.046)	0.326*** (0.059)	0.296*** (0.049)	0.296*** (0.047)
Distance from Industry Dir	0.987***	0.872***	0.880***	0.976***	0.862***	0.869***
Fine-Grained Caste Comp	(0.061)	(0.052)	(0.050)	-0.062	-0.053	-0.051
First Stage F-statistic Observations	8276 46,753	1967 27,163	1841 28,809	8211.87 46,754	1887.93 27,164	1803.63 28,811

Notes: This table presents first and second stage IV -Approach 2 results for firm performance variables(log operating income relative to sales, Tobin's Q and volatility) regressed on board fine-grained caste homophily and other control variables, and year and two-digit industry fixed effects for instrumental variable approach 2. Panel (B) presents the first stage results where the excluded instruments are the homophily of state and industry level director supply pools as well as the distance between the caste composition of these supply pools and that of individual boards. Panel (A) presents second stage results. Control variables include firm age, leverage, real assets, tangibility, export status, three digit industry HHI, whether the firm belongs to a business group, whether the firm is listed or not, dominant caste of the board, shares of firms with various dominant caste-grained castes in three digit industry, and the proportion of group firms in the industry. All variables are defined in Table 1. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). The first stage F statistic is the Kleibergen-Paap rk Wald F statistic. Robust standard errors, clustered by industry and year, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 6: Fine-Grained Caste Homophily and Firm Outcomes of firms Above and Below Median Performance : IV Approach 2

(A) Second Stage Results, Below Median					
Variables	(1)	(2)	(3)	(5)	
	ln(Operating Income/ Sales)	ln(Operating Cash Flow/Sales)	ln(Profits/Sales)	Tobin's Q	
Fine-Grain Homophily	-0.032 (0.055)	-0.052 (0.079)	-0.010 (0.044)	-0.027*** (0.008)	0.002** (0.001)
First Stage F-statistic Observations	11816.15 23,753	6106.25 17,382	7841.93 29,063	1970.31 13,590	960.76 14,410
(B) Second Stage Results, Above Median					
Fine-Grained Caste Homophily	-0.122*** (0.045)	-0.019 (0.069)	-0.170* (0.096)	-0.267*** (0.081)	0.003** (0.001)
First Stage F-statistic Observations	4820.48 23,000	3199.22 16,604	6170.64 28,441	961.19 13,570	2335.64 14,398

Notes: This table presents first and second stage IV- Approach 2 results for all firm performance variables (log of operating income relative to sales, Tobin's Q and volatility), regressed on board fine-grained caste homophily and other controls, and year fixed effects for two sets of firms. Panel (A) presents the first and second stage results for firms with performance variable below the two digit industry median. Panel (B) presents the first and second stage results for firms with performance variable above the two digit industry median. In both panels, the excluded instruments are the fine-grained caste homophily levels of the two director supply pools and the distance of fine-grained caste composition of the firm's board from that of the two supply pools. Control variables include age, board size, board size square, leverage, real assets, tangibility, listing status, export status, three digit industry HHI, whether the firm belongs to a business group, dominant caste of the board, shares of firms with various dominant coarse-grained castes in three digit industry, and the proportion of group firms in the industry. All variables are defined in Table 1. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). The first stage F statistic is the Kleibergen-Paap rk Wald F statistic. Robust standard errors, clustered by industry and year, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 7: Fine Grained Homophily and Firm Outcomes: Coarse-Grained Caste Subsamples, IV Approach 2

Variables	(A) Brahmin		(D) Jain	
	In(Operating Income/Sales)	Tobin's Q	In(Operating Income/Sales)	Tobin's Q
Fine-Grain Homophily	-0.317** (0.147)	-0.276* (0.151)	-0.224*** (0.085)	-0.240*** (0.088)
First Stage F-Statistic	2077.39	415.99	4681.08	648.97
Observations	7,701	4,916	4,560	3,010
	(B) Kshatriya		(E) Shudra	
Fine-Grain Homophily	-0.162* (0.090)	-0.260*** (0.090)	-0.131 (0.132)	0.048 (0.173)
First Stage F-statistic	3753.43	504.99	1686.22	688.21
Observations	6,003	3,110	8,623	4,115
	(C) Vaishya			
Fine-Grain Homophily	-0.317*** (0.058)	-0.183** (0.077)	0.007*** (0.003)	
First Stage F-Statistic	5377.72	915.96	1004.04	
Observations	16,526	10,465	11,116	

Notes: This table presents first and second stage IV- Approach 2 results for firm performance variables (log operating income relative to sales, Tobin's Q and volatility), regressed on board fine-grained caste homophily and other control variables, and year and industry fixed effects for five set of firms. Panel (A), (B), (C), (D) and (E) present the first and second stage results for firms whose dominant coarse-grained caste is *Brahmin*, *Kshatriya Vaishya*, *Jain* and *Shudra* respectively. In all panels the excluded instruments are the fine-grained caste homophily levels of the two director supply pools and the distance of fine-grained caste composition of the firm's board from that of the two supply pools. Control variables include age, board size, board size square, leverage, real assets, tangibility, listing status, export status, three digit industry HHI, whether the firm belongs to a business group, dominant caste of the board, shares of firms with various dominant coarse-grained castes in three digit industry, and the proportion of group firms in the industry. All variables are defined in Table 1. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). The first stage F statistic is the Kleibergen-Paap rk Wald F statistic. Robust standard errors, clustered by industry and year, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 8: Fine-Grained Caste Homophily and Firm Outcomes: IV Approach 3

(A) Second Stage Results			
Variables	(1) Log(Operating Income/Sales)	(2) Tobin's Q	(3) Volatility
Fine-Grained Homophily	-2.912*** (0.381)	-4.079*** (0.613)	0.180*** (0.041)
(B) First Stage Results			
First Stage F-statistic	20.10	10.25	10.92
Observations	94,167	113,676	115,157

Notes: This table presents first and second stage IV- Approach 3 results for firm performance variables(log operating income relative to sales, Tobin's Q and volatility) regressed on board fine-grained caste homophily and other control variables, and year and two-digit industry fixed effects for instrumental variable approach 3 (Clause 49). Panel (B) presents the first stage results where the excluded instruments are constructed based on a regression discontinuity design as shown in equation (4.2). Panel (A) presents second stage results. Control variables include firm age, leverage, real assets, tangibility, export status, three digit industry HHI, whether the firm belongs to a business group, whether the firm is listed or not, dominant caste of the board, shares of firms with various dominant coarse-grained castes in three digit industry, and the proportion of group firms in the industry. All variables are defined in Table 1. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). The first stage F statistic is the Kleibergen-Paap rk Wald F statistic. Robust standard errors, clustered by industry and year, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 9: Fine-Grained Caste Homophily and Firm Outcomes: IV Approach 4

(A) Second Stage Results			
Variables	(1) Log(Operating Income/Sales)	(2) Tobin's Q	(3) Volatility
Fine Grained Homophily	-0.636 (1.202)	-0.284 (1.604)	0.047*** (0.014)
Board Size	-0.026 (0.047)	0.054 (0.051)	-0.001 (0.001)
Board Size Squared	0.001 (0.002)	-0.003 (0.002)	0.000 (0.000)
Industry HHI	0.487*** (0.159)	0.546** (0.272)	-0.001 (0.002)
I(Group Firm)	-0.043 (0.054)	0.033 (0.060)	-0.001 (0.001)
Propn of Group Firms in Industry	-0.174 (0.262)	-0.149 (0.294)	0.002 (0.003)
R-squared	0.097	0.094	0.060
(B) First Stage Results			
Share exposed to other castes	-0.061*** (0.011)	-0.062*** (0.010)	-0.061*** (0.011)
First stage F-statistic	28.66	34.96	33.78
Observations	4,381	5,456	5,458

Notes: This table presents first and second stage IV- Approach 4 results for firm performance variables(log operating income relative to sales, Tobin's Q and volatility) regressed on board fine-grained caste homophily and other control variables, and year and two-digit industry fixed effects. Panel (B) presents the first stage results. Excluded instrument is the fraction of a firm's directors that serve on the board of at least one other firm whose dominant fine-grained caste is different from directors' own fine-grained caste. Panel (A) presents second stage results. Control variables include age, board size, board size square, leverage, real assets, tangibility, export status, three digit industry HHI, whether the firm belongs to a business group, dominant caste of the board, shares of firms with various dominant coarse-grained castes in three digit industry, proportion of group firms in the industry, and the total number of other directorships in the board. All variables are defined in Table 1. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). The first stage F statistic is the Kleibergen-Paap rk Wald F statistic. Robust standard errors, clustered by industry and year, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 10: Firm Outcomes and Mechanism Variables: Fixed Effects

Variable	Log(Operating Income/Sales)	Tobin's Q	Volatility
Role of Board during slow economic growth years			
Fine-grained caste homophily	-0.306*** (0.066)	-0.227*** (0.024)	0.011* (0.006)
I(Below-average growth year)	-0.072 (0.049)	1.066*** (0.039)	-0.073*** (0.005)
I(Below-average growth year)* Homophily	-0.356*** (0.048)	-0.07* (0.020)	0.007 (0.012)
Observations	27,164	46,754	28,810
R-squared	0.384	0.120	0.466
Representation on Key Committees			
Fine-grained caste homophily	-1.280*** (0.046)	-2.690*** (0.024)	0.000*** (0.000)
I(Over-representation in audit committee)	-0.151 (0.145)	-0.177 (0.149)	0.007* (0.004)
I(Over-representation in shareholder grievance Committee)	-0.061 (0.142)	-0.021 (0.129)	0.004** (0.002)
I(Over-representation in compensation and nomination Committee)	-0.140 (0.136)	-0.217** (0.054)	0.002 (0.002)
I(Over-representation in audit committee) * Fine-grained caste homophily	-0.689 (0.891)	-0.834* (0.25)	0.011 (0.024)
I(Over-representation in shareholder grievance committee) * Fine-grained caste homophily	-1.848 (1.193)	-1.806 (2.087)	0.029 (0.027)

Continued on next page

Table 10 – Continued from previous page

Variable	Log(Operating Income/Sales)	Tobin's Q	Volatility
I(Over-representation in compensation and nomination committee) * Fine-grained caste homophily	-0.637*** (0.014)	-0.607*** (0.129)	0.075*** (0.002)
Observations	160	141	160
R-squared	0.682	0.717	0.740
CEO caste			
Fine-grained caste homophily	-0.150*** (0.003)	-0.203** (0.053)	0.006*** (0.002)
I(CEO caste same as dominant caste of board)	-0.012 (0.021)	-0.007 (0.018)	0.001** (0.000)
I(CEO caste same as dominant caste of board) * Fine-grained caste homophily	-0.104 (0.118)	-0.150 (0.136)	0.004*** (0.001)
Observations	29,504	21,513	22,567
R-squared	0.133	0.332	0.472

Notes: This table presents fixed effects results for firm performance variables - log of operating income relative to sales, Tobin's Q, and volatility regressed on variables that showcase possible mechanisms, board fine-grained caste homophily, other control variables, and year and two-digit industry fixed effects. Control variables include age, board size, board size square, leverage, real assets, tangibility, listing status, export status, HHI measured at NIC3 level, whether the firm belongs to a business group, dominant caste of the board, shares of firms with various dominant coarse-grained castes in the industry, and proportion of group firms in the industry. Variables that showcase possible mechanisms are interaction variables of variable indicating if the year sees a below average growth year, variable indicating over representation of dominant caste of the board in major board committees, and variable indicating the caste of the board is same as that of the dominant caste of the board with board homophily. All variables are defined in Table 1. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). Robust standard errors, clustered by industry and year, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Online Appendix: Diversity in Corporate Boards and Firm Outcomes

A Other Sources for Caste Mapping

There are few datasets that record castes for names. Most datasets like the decennial Census, National Sample Survey (NSS) and National Family Health Survey (NFHS) record the government recognized social categories: Scheduled Caste, Scheduled Tribes, Other Backward Castes and General. However, they do not identify the Hindu caste identity – *varnas* and *jatis* – of individuals. The Socio-Economic and Caste Census 2011 is the first-ever caste-based census conducted since the 1931 Census of India. However, its data are not publicly available.

The People of India (PoI) project, undertaken by the Anthropological Society of India in 1985, records the lists of communities of India drawing upon ethnographic surveys and various official lists and records. 4635 communities were identified and studied. For each community, a list of associated last names is available. While comparing our mapping with PoI, we consider each main community to be a *jati*. These data have multiple shortcomings rendering them incomparable with our mapping. First, multiple last names are associated with multiple main communities (*jatis*), and the data do not provide any weights or the likelihood of these last names being associated with a particular *jati*. For example, the last name *Apte* is associated with main communities *Konkanastha* and *Brahmin*. However, we do not know the likelihood of *Apte* being associated with *Brahmin*. Next, many identified last names in PoI are not in fact last names; for example- “lodhi rajputa”, “adi gaura”, “adi-hindua”, or “dasa bhuiryaa”. Further, unlike the matrimony dataset, we do not

find many popular last names like Muthuswami, Premji, and Shanghvi (Sanghvi) in the PoI dataset. Of the last names that are common in both datasets, the community associations provided in many cases seem inappropriate in PoI. For example, in PoI, the last name *Mukherjee* is associated with the following communities: *Bengali, Christian, Jogi, Manipuri*. But *Bengali* and *Manipuri* indicate regional origin, as opposed to *jati* or *varna*. In our mapping, we find *Mukherjee* to be associated with *jatis* including *Brahmin Kulin, Brahmin, Brahmin Rarhi, Barendra* and *Rudraj*.

The Indian Human Development Survey collected information about the major *jatis* in each village from key informants in the village. These data may have informant bias as information is not self-reported and not collected or coded at the household level. Finally, Bharathi et al. (2018) do a comprehensive survey of *jatis* in rural Karnataka. Our last name to *jati* mapping is based on the profiles from the top three matrimonial websites, which are primarily used by urban people. Hence, comparison of our *jati* mapping with that of Bharathi et al. (2018) may not be appropriate.

B Figures

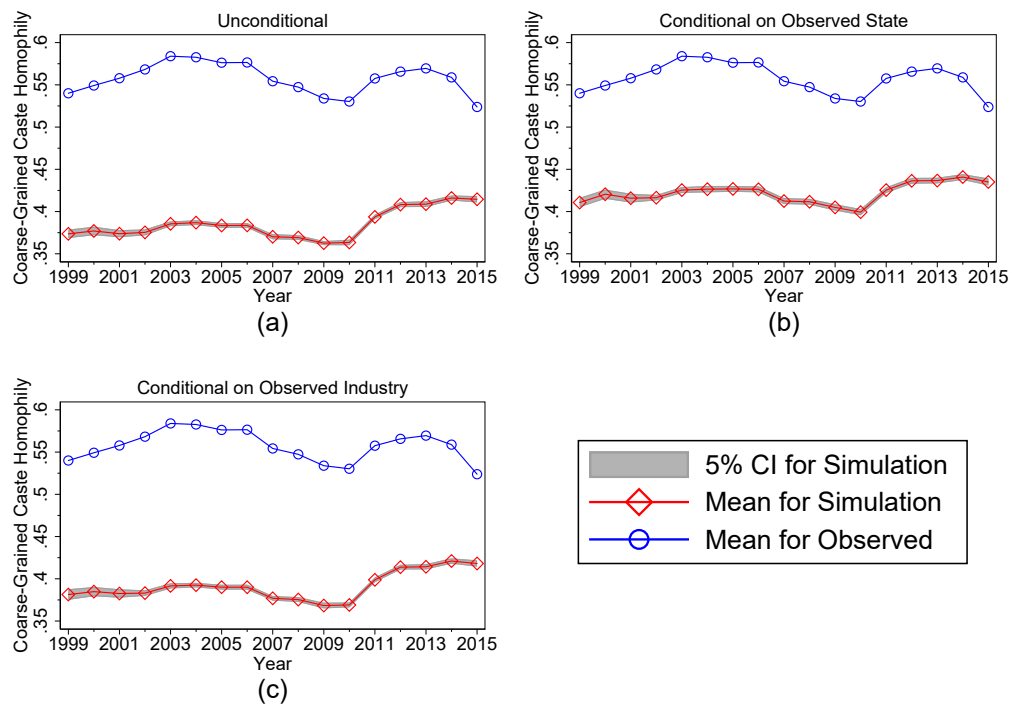


Figure 1: Observed vs. Simulated Average Coarse-Grained Caste Homophily^a

^aSource: Prowess, matrimonial data. The three graphs in the figure present the mean coarse-grained caste homophily across firms each year in the observed and simulated samples for three distinct simulation criteria: unconditional, conditional on firm's state and on firm's industry. Details about the simulation methods are provided in Section 3 of the main manuscript.

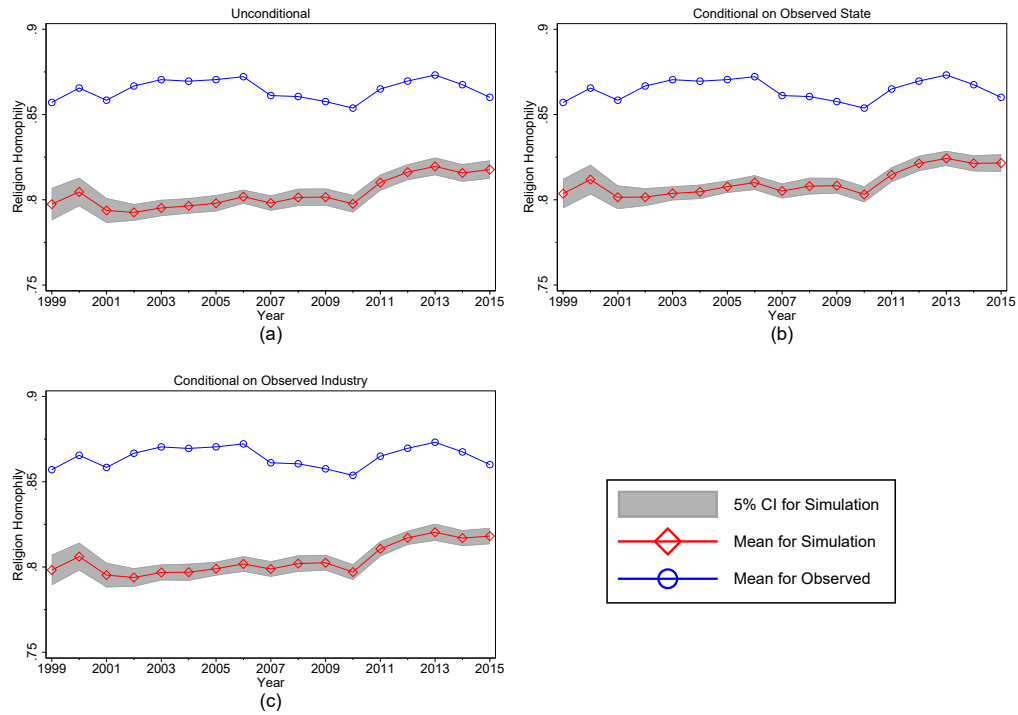


Figure 2: Observed vs. Simulated Average Religion Homophily^a

^aSource: Prowess, matrimonial data. The three graphs in the figure present the mean religion homophily across firms each year in the observed and simulated samples for three distinct simulation criteria: unconditional, conditional on firm’s state and on firm’s industry. Details about the simulation methods are provided in Section 3 of the main manuscript.

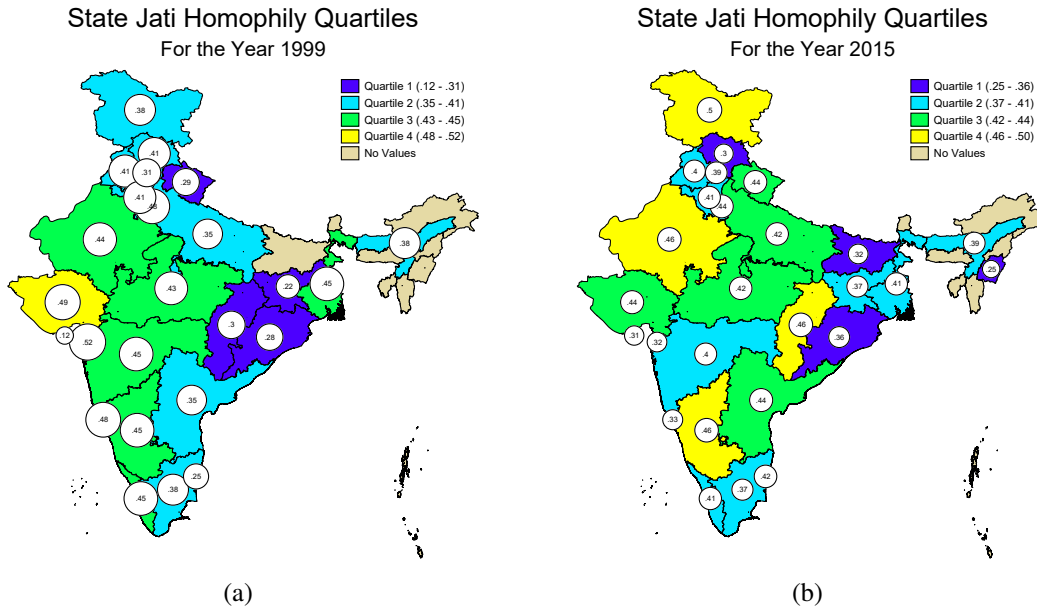


Figure 3: Average Fine-Grained Caste Homophily Across States^a

^aSource: Prowess, matrimonial data. The map shows the average fine-grained caste homophily for all firms in each state in the years 1999 and 2015. The color coding represents the quartile position of a state in the distribution of mean fine-grained caste homophily levels. Higher quartiles indicate higher homophily.

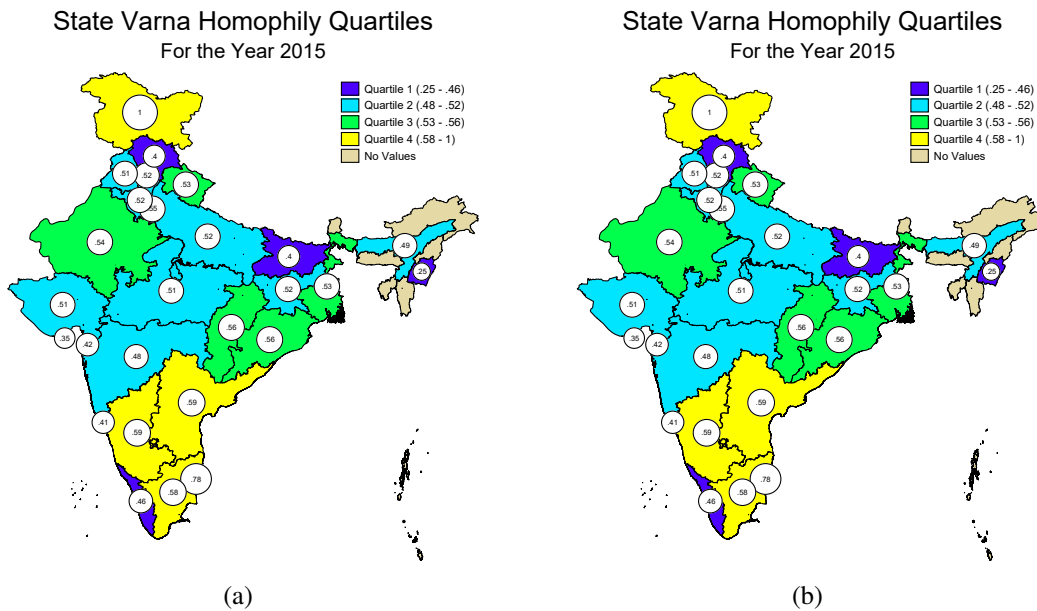


Figure 4: Average Coarse-Grained Caste Homophily Across States^a

^aSource: Prowess, matrimonial data. The map shows the average coarse-grained caste homophily for all firms in each state in the years 1999 and 2015. The color coding represents the quartile position of a state in the distribution of mean coarse-grained caste homophily levels. Higher quartiles indicate higher homophily.

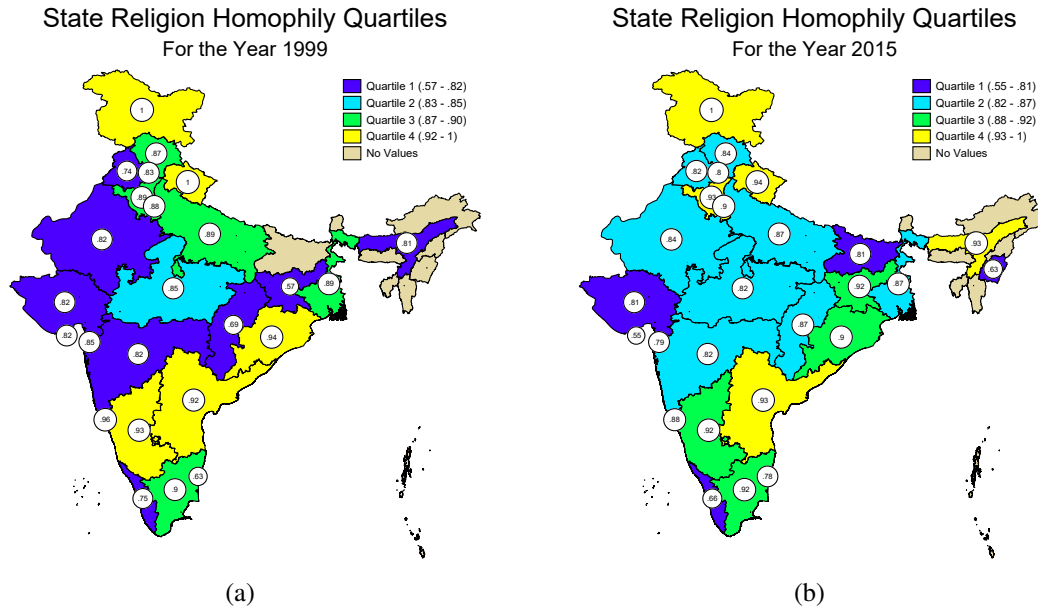
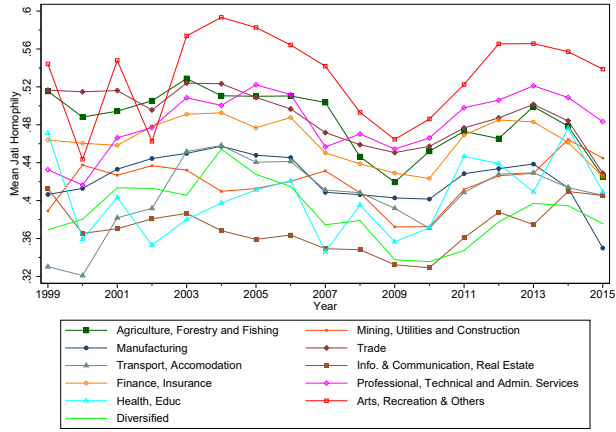
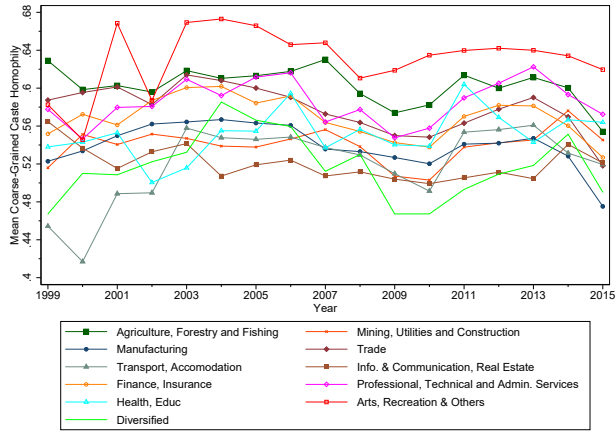


Figure 5: Average Religion Homophily Across States^a

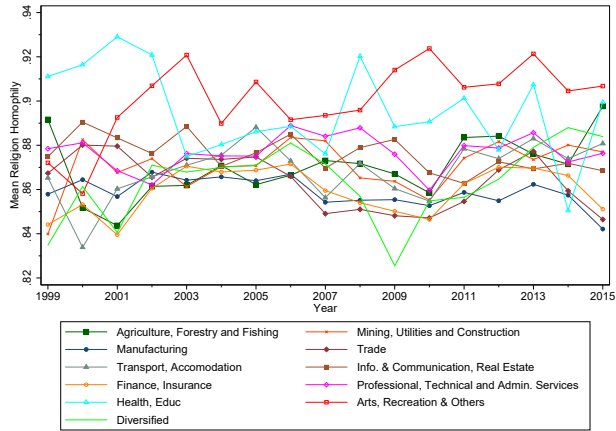
^aSource: Prowess, matrimonial data. The map shows the average religion homophily for all firms in each state in the years 1999 and 2015. The color coding represents the quartile position of a state in the distribution of mean religion homophily levels. Higher quartiles indicate higher homophily.



(a) Fine-Grained Caste



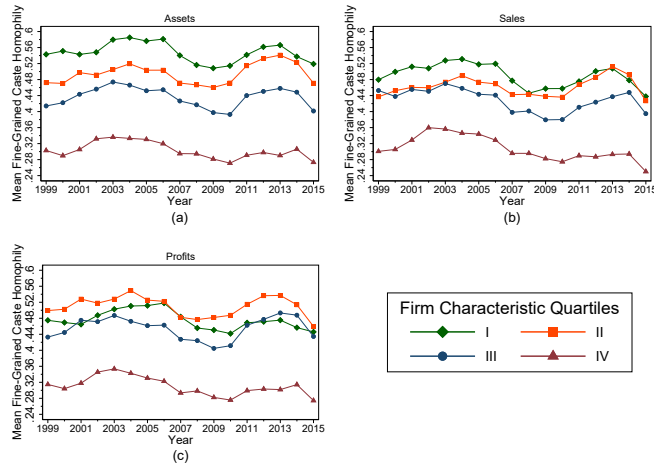
(b) Coarse-Grained Caste



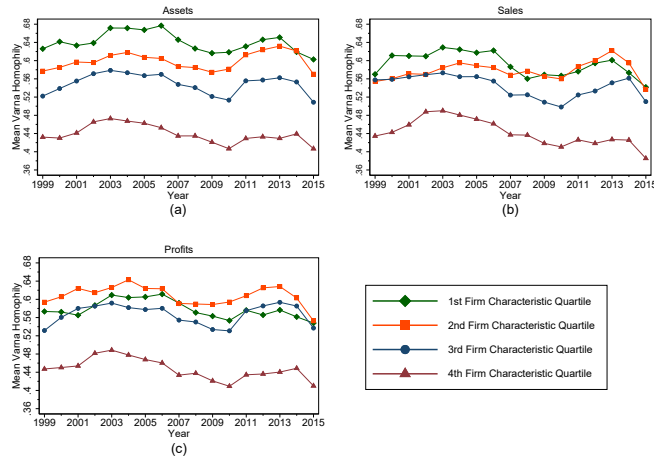
(c) Religion

Figure 6: Average Caste and Religion Homophily By Sector^a

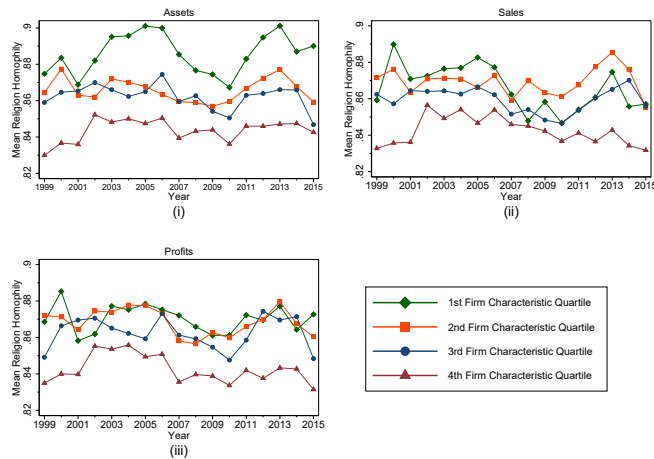
^aSource: Prowess, matrimonial data. Figures (a), (b) and (c) show the mean fine-grained caste, coarse-grained caste and religion homophily for all firms in eleven broad sectors. Diversified includes all firms that could not be classified primarily into one industry.



(a) Fine-Grained Caste



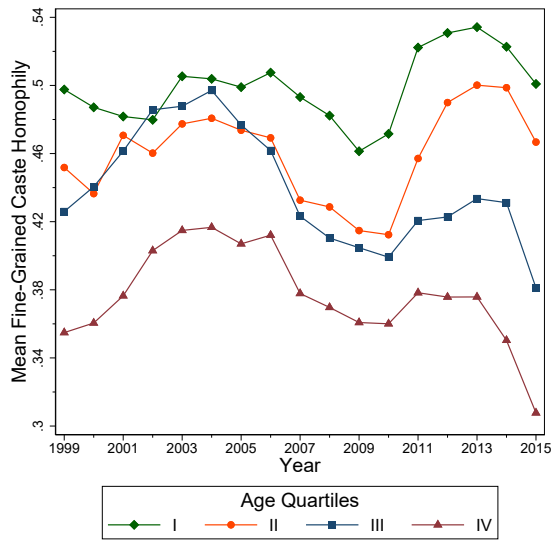
(b) Coarse-Grained Caste



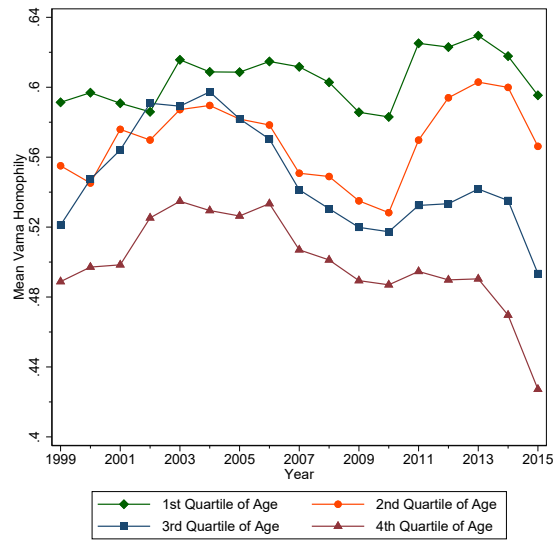
(c) Religion

Figure 7: Average Caste and Religion Homophily by Firm Size^a

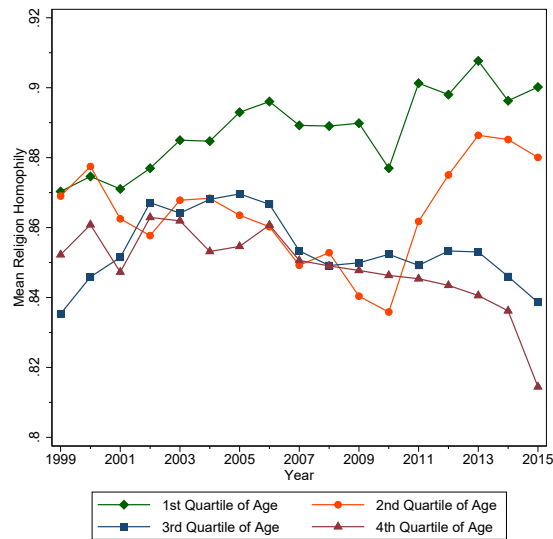
^aSource: Prowess, matrimonial data. Figures (a), (b) and (c) show the average fine-grained caste, coarse-grained caste and religion homophily levels in firms falling in the four quartiles of assets, sales, and profits, respectively. Higher quartiles indicate higher levels of the firm characteristic.



(a) Fine-Grained Caste



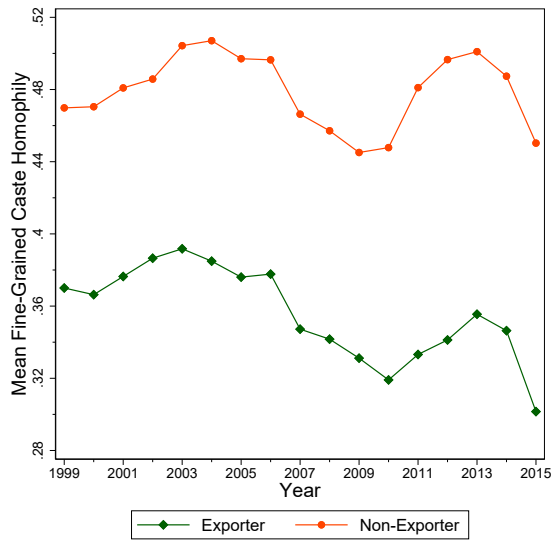
(b) Coarse-Grained Caste



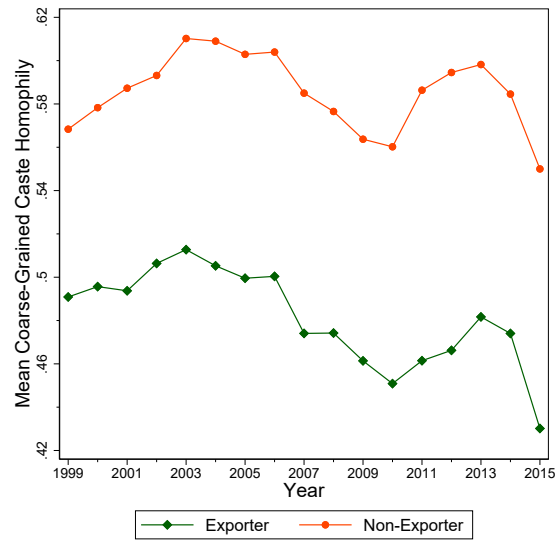
(c) Religion

Figure 8: Average Caste and Religion Homophily by Firm Age^a

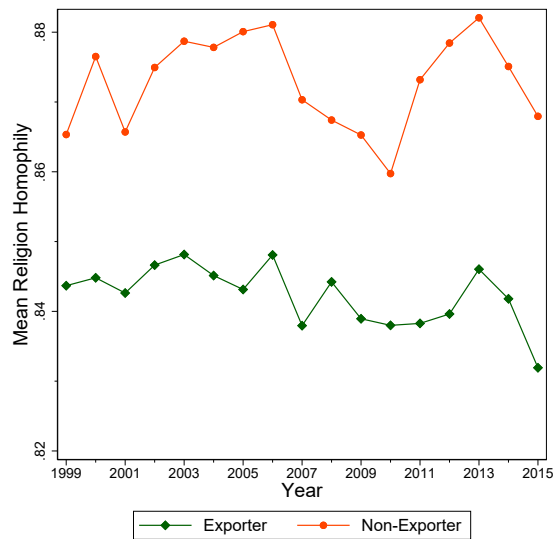
^aSource: Prowess, matrimonial data. Figures (a), (b) and (c) show the average fine-grained caste, coarse-grained caste and religion homophily, respectively, falling in the four quartiles of firm age. Higher quartiles indicate higher firm age.



(a) Fine-Grained Caste



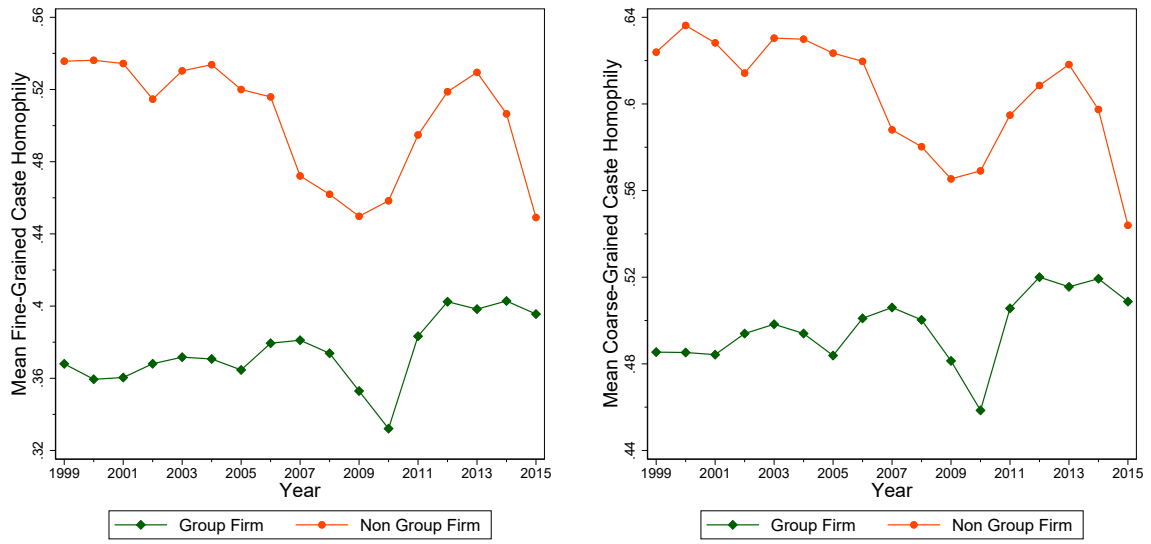
(b) Coarse-Grained Caste



(c) Religion

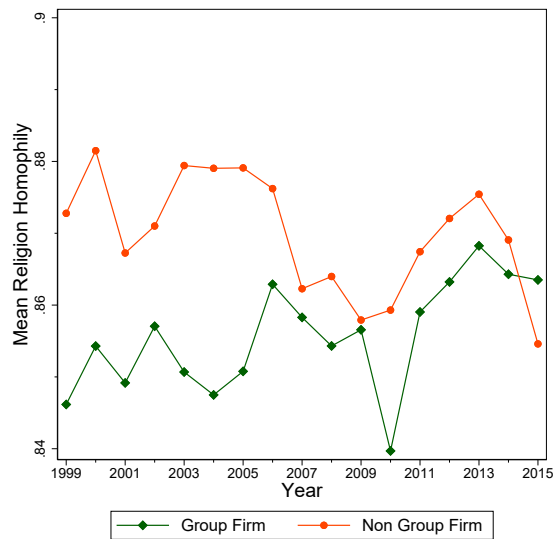
Figure 9: Average Caste and Religion Homophily by Exporting Status^a

^aSource: Prowess, matrimonial data. In Figures (a), (b), and (c), respectively, the average fine-grained caste, coarse-grained caste and religion homophily is averaged over all exporting and non-exporting firms separately.



(a) Fine-Grained Caste

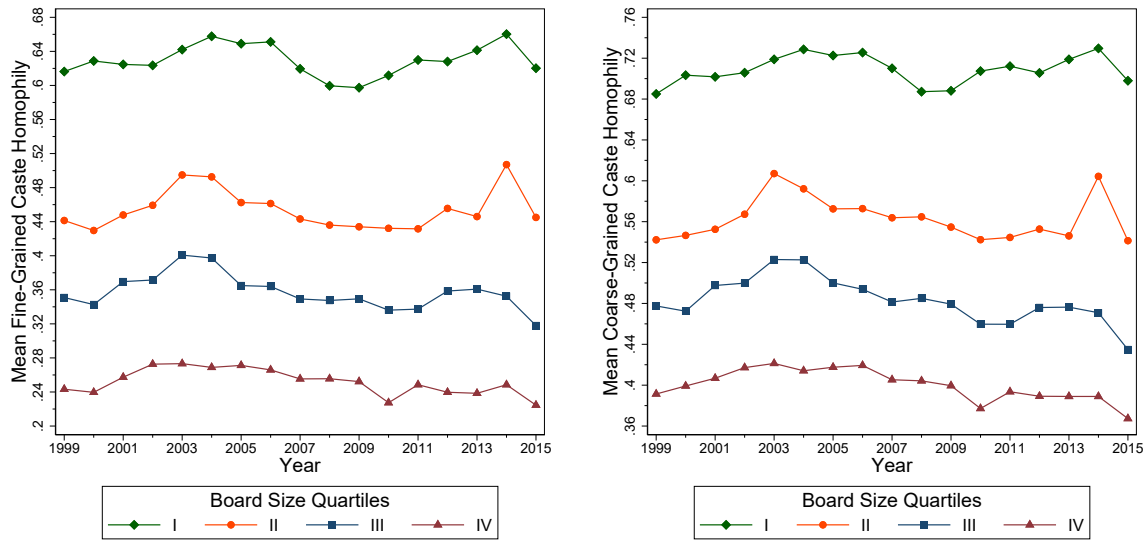
(b) Coarse-Grained Caste



(c) Religion

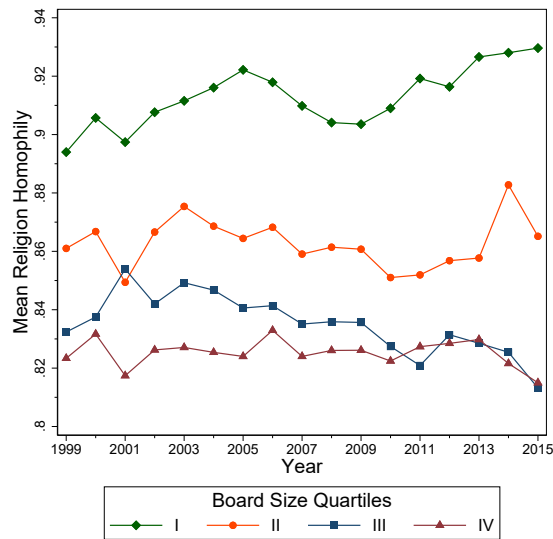
Figure 10: Average Caste and Religion Homophily by Business Group Status^a

^aSource: Prowess, matrimonial data. Figures (a), (b) and (c) show the average fine-grained caste, coarse-grained caste and religion homophily levels, respectively, in firms that belong to a business group and ones that do not.



(a) Fine-Grained Caste

(b) Coarse-Grained Caste



(c) Religion

Figure 11: Average Caste and Religion Homophily by Board Size^a

^aSource: Prowess, matrimonial data. Figures (a), (b) and (c) show the average fine-grained caste, coarse-grained caste and religion homophily levels in firms falling in the four quartiles of board size. Higher quartiles indicate larger board sizes.

C Tables

Table 1: Religion, Coarse- and Fine-Grained Caste Composition of Caste Mapping

Religion	% Last Names	Coarse-Grained Caste	% Last Names	Fine-Grained Caste	% Last Names
Hindu	80.69	Brahmin	18.67	Maratha	4.10
Muslim	8.00	Kshatriya	12.67	Brahmin Iyer	3.88
Christian	6.42	Vaishya	12.83	Brahmin	3.20
Jain	2.30	Shudra	32.05	Sindhi	3.02
Sikh	1.32	Dalit	1.78	Nair	2.84
Parsi	1.15	Unknown Varna	1.35	Arya Vysya	2.46
Buddhist	0.04			Agarwal	2.30
Jewish	0.01			Khatri	1.97
NA	0.08			Vannia Kula Kshatriyar	1.93
				Brahmin Deshastha	1.89
				Ezhava	1.89

Source: Matrimonial data. This table provides the religion, coarse- and fine-grained caste distribution of last names included in the final mappings developed by the authors, as described in Section 2. For space considerations, the table only shows the distribution for the top ten most frequently occurring fine-grained castes out of a total of 471 distinct fine-grained castes that we can identify.

Table 2: Religion, Coarse- and Fine-Grained Caste Mapping for Names of Directors

Number of classifications	% last names with up to that number of classifications	Average probability associated with classifications
Religion		
1	68.6	95.3
2	92.2	99.5
3	98.0	99.9
4	99.6	100
5	100	100
Coarse-Grained Caste		
1	45.1	78.5
2	58.3	90.6
3	68.5	95.8
4	76.9	98.1
5	85.1	99.3
Fine-Grained Caste		
1	39.7	67.0
2	49.8	77.8
3	56.5	83.2
4	61.5	86.5
5	65.5	88.8

Source: Matrimonial and Prowess data. This table provides the religion, coarse- and fine-grained caste distribution of director last names. For space considerations, the table only shows the distribution for last names associated with up to five fine-grained castes.

Table 3: Simulation Results

Simulation Method	Cultural Identity	Mean Homophily in Observed Boards	Mean Homophily in Simulated Boards	Difference	t statistic
Unconditional	Fine-Grained Caste	0.449	0.282	0.167***	-7110.636
Conditional on Sector	Fine-Grained Caste	0.449	0.287	0.162***	-7844.739
Conditional on State	Fine-Grained Caste	0.449	0.304	0.145***	-5558.132
Unconditional	Coarse-Grained Caste	0.56	0.386	0.174***	-4557.745
Conditional on Sector	Coarse-Grained Caste	0.56	0.392	0.168***	-4171.524
Conditional on State	Coarse-Grained Caste	0.56	0.423	0.137***	-2880.585
Unconditional	Religion	0.866	0.804	0.062***	-1015.249
Conditional on Sector	Religion	0.866	0.804	0.062***	-958.03
Conditional on State	Religion	0.866	0.81	0.056***	-895.616

Source: Prowess, matrimonial data. This table shows the mean homophily of boards in observed and simulated data. Homophily is measured for fine-grained caste, coarse-grained caste, and religion. Simulations have been conducted under three criteria: unconditional random sampling of directors, random sampling conditional on observed firm's sector (two digit industry), and random sampling conditional on observed firm's headquarter state. The t-statistics are for the null hypothesis that mean homophily levels in the observed and simulated samples are equal.

Table 4: Coarse-Grained Caste Homophily and Firm Outcomes: Fixed Effects

Variables	(1) Log(Operating Income/Sales)	(2) Tobin's Q	(3) Volatility
Coarse-Grained Caste Homophily	-0.238*** (0.062)	-0.120*** (0.042)	0.006*** (0.002)
Board Size	-0.024 (0.016)	0.052** (0.021)	-0.004*** (0.001)
Board Size Square	0.001** (0.001)	-0.002 (0.001)	0.000*** (0.000)
Industry HHI	0.030 (0.059)	0.058 (0.072)	0.001 (0.002)
I(Listed)	-0.028 (0.025)	-1.384** (0.680)	-0.004 (0.003)
I(Group Firm)	0.054** (0.025)	0.184*** (0.034)	-0.005*** (0.001)
Share of Group Firms	0.353 (0.216)	-0.022 (0.103)	-0.006* (0.004)
Observations	46,808	27,163	28,809
R-squared	0.116	0.378	0.464

Notes: This table presents fixed effects results for firm performance variables – operating income relative to sales, Tobin's Q, and volatility – regressed on board coarse-grained caste homophily and other control variables, year and two-digit industry fixed effects. Control variables include age, board size, board size square, leverage, real assets, tangibility, listing status, export status, HHI index measured at NIC3 level, whether the firm belongs to a business group, dominant caste of the board, shares of firms with various dominant coarse-grained castes in an industry, and proportion of group firms in the industry. All variables are defined in the main manuscript. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). Robust standard errors, clustered by industry and year, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5: Religion Homophily and Firm Outcomes: Fixed Effects

Variables	(1) Log(Operating Income/Sales)	(2) Tobin's Q	(3) Volatility
Religion Homophily	-0.094* (0.055)	-0.141*** (0.045)	0.003* (0.002)
Board Size	-0.007 (0.013)	0.057*** (0.021)	-0.004*** (0.001)
Board Size Square	0.001 (0.001)	-0.002 (0.001)	0.000*** (0.000)
Industry HHI	0.035 (0.059)	0.061 (0.072)	0.001 (0.002)
I(Listed)	-0.016 (0.026)	-1.370** (0.680)	-0.004 (0.003)
I(Group Firm)	0.068** (0.027)	0.186*** (0.034)	-0.005*** (0.001)
Share of Group Firms	0.364* (0.219)	-0.017 (0.103)	-0.007* (0.004)
Observations	46,808	27,163	28,809
R-squared	0.115	0.378	0.464

Notes: This table presents fixed effects results for firm performance variables – operating income relative to sales, Tobin's Q, and volatility – regressed on board religion homophily and other control variables, year and two-digit industry fixed effects. Control variables include age, board size, board size square, leverage, real assets, tangibility, listing status, export status, HHI index measured at NIC3 level, whether the firm belongs to a business group, dominant caste of the board, shares of firms with various dominant coarse-grained castes in an industry, and proportion of group firms in the industry. All variables are defined in Appendix A. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). Robust standard errors, clustered by industry and year, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 6: Within-Industry Board Interlocks

NIC	Within-Industry Board Interlocks			
	Mean	Minimum	Maximum	% Firms with interlocks
Agriculture, Forestry and Fishing	0.05	0.00	0.20	39.29
Mining, Utilities and Construction	0.10	0.00	0.83	44.53
Manufacturing	0.31	0.00	1.00	76.60
Trade	0.09	0.00	1.00	27.08
Transport, Accomodation	0.17	0.00	0.80	63.16
Information & Communication	0.13	0.00	1.00	46.67
Finance, Insurance & Real Estate	0.20	0.00	1.00	66.07
Professional, Technical and Admin. Services	0.00	0.00	0.11	4.17
Education & Health	0.03	0.00	0.31	11.76

Source: Indian Boards database. This table presents proportions of directors of firms that also serve on at least one other firm's board in 2015 or in the period of 2012-2014, with that firm belonging to the same one-digit industry.

Table 7: K-S Test Results Comparing Fine-Grained Caste Distributions of Directors in Firms to Industries

	NIC	Above Median Interlocks			Below Median Interlocks		
		Test Stat	Critical Value	Decision	Test Stat	Critical Value	Decision
Directors	1	0.05	0.23	Not Rejected	0.07	0.25	Not Rejected
	2	0.07	0.1	Not Rejected	0.06	0.09	Not Rejected
	3	0.05	0.04	Rejected	0.04	0.04	Rejected
	4	0.12	0.13	Not Rejected	0.05	0.1	Not Rejected
	5	0.05	0.16	Not Rejected	0.03	0.14	Not Rejected
	6	0.08	0.14	Not Rejected	0.07	0.13	Not Rejected
	7	0.06	0.08	Not Rejected	0.1	0.1	Rejected
	8	0.35	0.47	Not Rejected	0.04	0.21	Not Rejected
	9	0.15	0.37	Not Rejected	0.04	0.22	Not Rejected
	10				0	0.32	Not Rejected
Firms	1	0.11	0.65	Not Rejected	0.15	0.72	Not Rejected
	2	0.1	0.29	Not Rejected	0.08	0.27	Not Rejected
	3	0.04	0.12	Not Rejected	0.03	0.11	Not Rejected
	4	0.14	0.37	Not Rejected	0.06	0.28	Not Rejected
	5	0.16	0.47	Not Rejected	0.11	0.42	Not Rejected
	6	0.05	0.37	Not Rejected	0.05	0.36	Not Rejected
	7	0.1	0.26	Not Rejected	0.12	0.27	Not Rejected
	8	0.67	1.42	Not Rejected	0.06	0.57	Not Rejected
	9	0.75	1.44	Not Rejected	0.11	0.7	Not Rejected
	10				0	0.79	Not Rejected

Source: Matrimonial data, Prowess, Indian Boards database. This table presents results from Kolmogorov-Smirnoff tests comparing the fine-grained caste distribution of firms' directors to that of the entire one digit industry, separately for firms that have above- and below-median interlocks in that industry.

Table 8: Coarse-Grained Caste, Religion Homophily and Firm Outcomes: IV Approach 1

Variables	Coarse- Grained Caste			Religion		
	(1)	(2)	(3)	(4)	(5)	(6)
	In(Operating Income/ Sales)	Tobin's Q	Volatility	In(Operating Income /Sales)	Tobin's Q	Volatility
Coarse-Grained Caste /Religion Homophily	-0.415 (0.399)	-0.822* (0.438)	0.010 (0.010)	-0.141 (0.315)	-0.408 (0.225)	0.015* (0.009)
	(B) First Stage Results					
State Coarse Grained Caste /Religion Homophily	0.454*** (0.042)	0.484*** (0.052)	0.469*** (0.052)	0.290*** (0.023)	0.353*** (0.038)	0.348*** (0.038)
Industry Coarse-Grained /Religion Homophily	0.313*** (0.093)	0.156 (0.156)	0.170 (0.142)	0.222*** (0.041)	0.207*** (0.068)	0.215*** (0.065)
First Stage F-statistic	60.61	44.68	43.36	103.9	43.97	46.29
Observations	46,753	27,163	28,809	46,753	27,163	28,809

Notes: This table presents first and second stage IV results for firm performance variables(log operating income relative to sales, and Tobin's Q and volatility) regressed on board coarse-grained caste, religion homophily and other control variables, and year and two-digit industry fixed effects for instrumental variable approach 2. Columns (1)-(3) present results for coarse-grained caste homophily and Columns (4)-(6) present results for religion homophily. Panel (B) presents the first stage results where the excluded instruments are the homophily of state and industry level director supply pools. Panel (A) presents second stage results. Control variables include firm age, leverage, real assets, tangibility, export status, three digit industry, HHI, whether the firm belongs to a business group, whether the firm is listed or not, dominant caste/religion of the board, and the proportion of group firms in the industry. All variables are defined in Table 1. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). The first stage F statistic is the Kleibergen-Paap rk Wald F statistic. Robust standard errors, clustered by industry and year, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table 9: Distance Between Firm and Industry/State Director Fine-Grained Caste Composition

Year	10th Percentile	50th Percentile	90th Percentile	Mean	Standard Deviation
Panel A: Distance Between Firm and Industry Director Composition					
1999	0.32	0.32	0.87	0.54	0.19
2004	0.34	0.34	0.88	0.58	0.2
2009	0.32	0.32	0.82	0.54	0.18
2015	0.31	0.31	0.83	0.54	0.19
Panel B: Distance Between Firm and State Director Composition					
1999	0.3	0.3	0.84	0.52	0.19
2004	0.32	0.32	0.88	0.55	0.2
2009	0.3	0.3	0.77	0.5	0.18
2015	0.29	0.29	0.77	0.52	0.19

Source: Matrimonial data, Prowess. This table shows moments for the Euclidean distances between the fine-grained caste composition of firms' directors and that of the set of directors in the same two-digit industry (Panel A) or state (Panel B).

Table 10: Coarse-Grained Caste, Religion Homophily and Firm Outcomes: IV Approach 2

Variables	Coarse- Grained Caste					
	(A) Second Stage Results			Religion		
	(1)	(2)	(3)	(4)	(5)	(6)
	In(Operating Income/ Sales)	Tobin's Q	Volatility	In(Operating Income/ Sales)	Tobin's Q	Volatility
Coarse-Grained /Religion Homophily	-0.226*** (0.071)	-0.091* (0.046)	0.006*** (0.002)	-0.062 (0.243)	-0.065 (0.153)	0.007 (0.004)
	(B) First Stage Results					
State Coarse-Grained Caste /Religion Homophily	0.193*** (0.030)	0.182*** (0.037)	0.180*** (0.036)	0.285*** (0.028)	0.347*** (0.043)	0.341*** (0.042)
Industry Coarse-Grained Caste /Religion Homophily	0.904*** (0.048)	0.530*** (0.145)	0.557*** (0.140)	0.144*** (0.053)	0.134* (0.072)	0.139** (0.070)
Distance from State Director Coarse-Grained/ Religion Composition	0.212*** (0.025)	0.183*** (0.026)	0.182*** (0.027)	0.065 (0.079)	0.059 (0.084)	0.054 (0.084)
Distance from Industry Director Coarse-Grained/ Religion Composition	1.027*** (0.031)	0.954*** (0.036)	0.962*** (0.037)	-0.229** (0.114)	-0.296** (0.121)	-0.295** (0.121)
First Stage F-statistic	4709	1842	1797	66.03	27.59	31.44
Observations	46,753	27,163	28,809	46,753	27,163	28,809

Notes: This table presents first and second stage IV results for firm performance variables(log operating income relative to sales, and Tobin's Q and volatility) regressed on board coarse-grained caste, religion homophily and other control variables, and year and two-digit industry fixed effects for instrumental variable approach 2. Columns (1)-(3) present results for coarse-grained caste homophily and Columns (4)-(6) present results for religion homophily. Panel (B) presents the first stage results where the excluded instruments are the homophily of state and industry level director supply pools as well as the distance between the caste composition of these supply pools and that of individual boards. Panel (A) presents second stage results. Control variables include firm age, leverage, real assets, tangibility, export status, three digit industry HHI, whether the firm belongs to a business group, whether the firm is listed or not, dominant caste/religion of the board, and the proportion of group firms in the industry. All variables are defined in Table 1. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). The first stage F statistic is the Kleibergen-Paap rk Wald F statistic. Robust standard errors, clustered by industry and year, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

References

Bharathi, N., Malghan, D. V., and Rahman, A. (2018). More heat than light: census-scale evidence for the relationship between ethnic diversity and economic development as a statistical artifact. *IIM Bangalore Research Paper*, (575).