Cultural Diversity in Corporate Boards and Firm Outcomes

Suresh Bhagavatula* Manaswini Bhalla†
IIM, Bangalore IIM, Bangalore

Manisha Goel‡ Balagopal Vissa§
Pomona College INSEAD

May 2019

Abstract

Using self-reported religions, castes and last names from Indian matrimonial websites and machine learning methods, we map individuals' last names to progressively finer levels of cultural identity. Using this mapping, we show that Indian corporate boards have systematically low cultural diversity over 1999-2015, however finely we measure identity. Greater board diversity accompanies better firm quality and governance. Causal analysis shows that lack of diversity on boards worsens firm performance, with strongest results for the finest measure of identity. Our findings demonstrate people?s desire to associate with those who share their cultural identity, even in economically important settings, with detrimental effects.

JEL Classifications: Z13, G34, M14, L14, J44

Keywords: Corporate Culture, Corporate Governance, Boards of Directors, Caste

*sureshbh@iimb.ernet.in

†manaswinib@iimb.ernet.in

‡manisha.goel@pomona.edu (Corresponding Author)

§Balagopal.vissa@insead.edu. We gratefully acknowledge the financial support of Pomona College and Indian Institute of Management, Bangalore. We thank Stephan Siegel, Jared Rubin, Latika Chaudhary, Sriya Iyer, Tahir Andrabi, Asim Khwaja, Rajiv Sethi, Simi Kedia, Anand Venkateshwaran and Michelle Zemel for helpful conversations and seminar participants at IIM Bangalore, Indian School of Business, and Indian Statistical Institute, and conference participants at AALIMS 2017 and Liberal Arts Macro and Finance Workshop 2018 for their comments. Thanks to VSK Teja Konduri, Timothy Howe, Charles Fries, Leena Kinger Hans, and Sharada PM for valuable research assistance.

"Moreover, we know that diversity is good for the economy; it improves corporate performance, drives growth and enhances employee engagement. Simply put, organizations with diverse teams perform better."

The Pledge, CEO Action for Diversity and Inclusion

1 Introduction

Social networks are not diverse. People associate with others who share their identity along traits such as race, gender, and ethnicity.[1] Shared cultural identity also influences economic interactions, such as between lender and borrower and hiring within firms.[2] As the above quote demonstrates, diversity is widely considered desirable. However, research finds that diversity negatively affects aggregate outcomes such as economic growth. The effects of diversity on firm outcomes have only recently been analyzed rigorously.[3] In this strand of literature, studies have examined diversity in firms' boards, focusing primarily on gender.[4] But group dynamics and decisions can be influenced by diversity along several attributes beyond just gender. In this paper, we examine a little-explored dimension of diversity, namely cultural identity.

Using data from India, we examine whether, and to what extent, boards of directors of large firms tend to be culturally diverse. We find that India's corporate boards are characterized by extremely low levels of diversity over time, across regions, and across industries. We further assess if the extent of diversity depends on how coarsely directors' cultural identity is defined, and find a striking lack of diversity at all aggregation levels. Finally, we ask how this lack of diversity impacts firm outcomes, and show that firm value and performance are negatively affected by high cultural homogeneity of boards. These negative effects are strongest for homogeneity along our narrowest measure of cultural identity.

A major challenge faced by studies focusing on shared cultural identity is the identification of individuals' cultural groups. As a result, studies have considered diversity along easily identifiable coarse groups such as gender, race, or country of origin. However, this

[1] A 2013 American Values Survey of the Public Religion Research Institute shows that 91% of whites' friends are white, 83% of blacks' friends are black and 64% of Hispanics' friends are Hispanic. See also McPherson et al. (2001).

[2] See, for example, Hegde and Tumlinson (2013), Fisman et al. (2017), Freeman and Huang (2015), Åslund et al. (2014), Giuliano et al. (2009), Giuliano and Ransom (2013), and Petersen et al. (2000).

[3] For growth effects of diversity, see Easterly and Levine (1997), Desmet et al. (2009), Esteban et al. (2012), De Luca et al. (2018), among others. For effects on firms, see Hjort (2014) and Glover et al. (2017).

[4] See, for example, Adams and Ferreira (2009), Ahern and Dittmar (2012), and Bertrand et al. (2018).

has two drawbacks. First, there is considerable heterogeneity in backgrounds and characteristics of individuals within these groups, so that the finding of homogeneity along a broad feature, such as gender, leaves considerable diversity still unmeasured. Measuring cultural identity more narrowly can help account for these otherwise overlooked dimensions of diversity. Second, some aspects of cultural identity may be more salient, so that individuals may feel stronger affinity towards each other along *those* traits. For example, in a group of people from several countries, two individuals who are from the same region of a country may feel stronger affinity towards each other than those from different parts of it. *These* shared traits may matter more than others in group dynamics, and again may go unaccounted for.

India offers a uniquely useful cultural context to overcome these challenges. The country has a multi-religious society, and Hindus, the dominant religious group, are divided into hundreds of subgroups by the caste system, a traditional institution that has endured since c. 1300 B.C.[5] The Hindu society is divided into four hierarchical *varnas* – *Brahmins, Kshatriyas, Vaishyas*, and *Shudras*, in that order, and a fifth, de facto, "*varna*" of *Dalits*.[6] Within the five *varnas*, there are hundreds of subgroups called *jatis*. An individual belongs to a certain *jati* and *varna* based on her lineage. Historically, *jatis* are endogamous and have been associated with occupations.[7] Besides marriages, *jatis* have also been shown to influence a range of outcomes such as where people reside, elections, public good provision, etc. Given that the operative cultural identity of individuals in India is primarily *jati*, it becomes imperative to understand diversity along this dimension in the context of firms.[8]

However, even when using caste as a measure of cultural identity, most research has been limited by relatively coarse classifications such as upper and lower castes since identifying individuals' *varnas* and *jatis* without subjective guesses has been difficult. In our paper, we use a novel computational methodology to develop a data driven mapping of last names into distinct cultural groups at varying levels of aggregation, i.e. *religion*, *varnas* and *jatis*. To identify the castes (*varnas* and *jatis*), we exploit the facts that individuals' last names are indicative of their religion and caste, and that Indian marriages

[5] Macdonell (1914).

[6] *Dalits* is a term used by the government to describe those disadvantaged groups who were considered "untouchables" and formally outside the caste system. The *Shudras* and *Dalits* historically faced significant discrimination. Today they are respectively designated "other backward castes (OBC)" and "scheduled castes (SC)" by the Indian government and are beneficiaries of several affirmative action programs. Affirmative actions are also aimed at uplifting some tribal populations that are classified as "Scheduled Tribes (ST)".

[7] There is some flexibility in the association of *jatis* to occupations. (Srinivas (1995)).

[8] Indian population is diverse along other cultural traits too, such as language, food, and attire. In this paper, we focus only on caste.

are overwhelmingly intra-religion, with Hindu marriages, in particular, being predominantly intra-*jati*. We obtain information from three prominent matrimonial websites on the first and last names of nearly six million registered users and their self-reported religions and *jatis*. Since the mapping between last name and caste/religion is not always one-to-one, we use these data to assign probabilities with which a last name belongs to each religion, and for Hindu names, each caste (*varna* and *jati*). We apply this mapping to directors of large public and private firms, whose names are taken from an annual firm level database. Thus, the cultural groups identified for each director range from a broad classification into eight religions to an extremely granular classification into 471 distinct *jatis*.

We use these data to develop a rich set of stylized facts about religion and caste homophily in Indian corporate boards during 1999-2015. We show that corporate boards in India lack cultural diversity. Figure 1 shows the proportions of boards constituted by directors belonging to the boards' dominant religion, *varna*, and *jati* using data for 1999. We see that the boards are characterized by high cultural homogeneity, with the dominant cultural group accounting for a large share of directors in a board. This share tends to be high, ranging from 55% (Parsis) to 90% (Hindus) for religion, and 52% (*dalits*) to 65% (*vaishyas*) for *varna*.[9] A similar pattern emerges when identity is measured as *jati* in Figure 1(c), which shows the top five most represented dominant *jatis* on boards in the sample.[10] Measuring cultural homophily (i.e., inverse of diversity) of a board as its cultural Herfindahl-Hirschman index (HHI), we find that cultural homophily is pervasively high across states and industries. It also persistently stays high throughout the sample period. However, it varies systematically across firm types indicating that higher performance and better corporate governance are associated with lower cultural homophily of boards.

Comparing our data to patterns in several simulated samples, we show that low cultural homophily of Indian firms' boards is not simply coincidental or driven by low diversity in the supply pool of directors themselves. However, we cannot isolate other reasons underlying the low cultural diversity of corporate boards. It may be that to find new board members, incumbent directors, managers, and owners rely on their social networks, which tend to be culturally homogeneous. Alternatively, high cultural homophily of boards may reflect incumbents' favoritism, in-group bias, or greater trust toward po-

[9] In addition to displaying low caste diversity, Figure 1(b) also shows that individuals tend to associate with those of their own *varna*, even when the *varna* is a socially-disadvantaged one.

[10] A full graphical representation of the shares of directors belonging to the numerous *jatis* in the data is difficult. A complete list is available upon request.

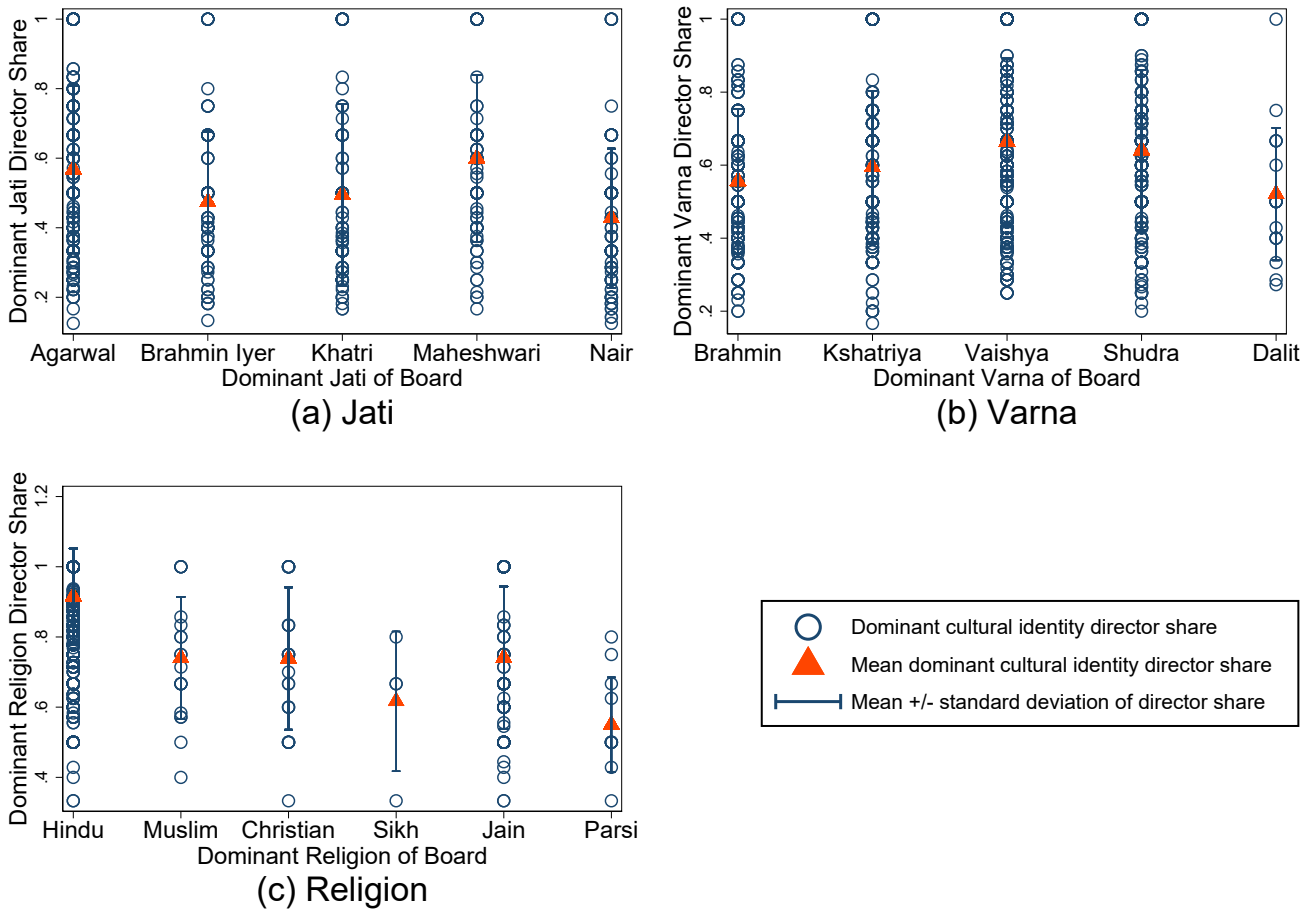


Figure 1: Shares of Directors Belonging to Boards' Dominant Cultural Group[a]

[a]Source: Prowess, matrimonial data. Figures (a), (b), and (c) show the proportions of directors that belong to the *jati*, *varna*, and religion, respectively, with the highest representation on a board, as well as the overall averages across boards. All figures use data for the year 1999.

tential new directors who share their cultural identity.[11] Regardless of the reason why boards lack diversity, the effects of diversity on firm performance are unclear, ex ante. One one hand, diverse board members may bring a wider range of experience and information to bear upon the decisions they make for the firm, improving their advisory

[11]Yet another possibility is that directors are hired solely on the basis of their human capital and they happen to be of specific castes or religions. We can rule this out, however, for two reasons. First, individuals serving as directors on firms presumably already have a high socio-economic status and human capital, even if they belong to historically disadvantaged cultural groups. Second, as shown in Figure 1, we observe homophily even among groups that we would expect to be socially disadvantaged so that some boards are dominated by disadvantaged groups. Yet, these firms' performance is not systematically any different from that of firms whose boards are dominated by socially privileged groups.

role. They may also be willing to monitor the management better. On the other hand, culturally homogeneous directors may have greater trust or fewer differences in opinions, reducing conflicts in the boardroom and improving firm performance. However, they may also be more prone to cronyism, hurting the firms they serve.

Regression analysis, therefore, provides us with estimates of the net effect of these mechanisms. We use several instrumental variable strategies to examine how cultural homophily of boards affects key measures of firm performance (operating income, operating cash flow, profitability) and market related variables (market to book ratio, Tobin's Q, and volatility). To instrument for homophily, we use the homophily of the firm's director supply pool, measured as the set of directors in the firm's state or industry. In a second approach, we additionally use as instruments the distance of a board's cultural 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. Results from these analyses show that higher cultural homophily on corporate boards negatively affects firm value and performance. Importantly, these results reveal that board diversity effects on firms depend on the granularity of the cultural identity measure. We find the strongest negative effects on firms when we consider homophily in boards along *jatis*, our narrowest measure of cultural identity. However, our broadest measure of homophily, along religion, does not significantly affect firms. In other words, *jati* diversity matters strongly for firms, but religion homophily does not. This is consistent with the reality of India's social fabric, wherein marriages, residence, occupations, voting patterns, public good provision, etc. are all influenced by *jati* (Joshi et al. (2018), Kumar et al. (2017), Beteille (1996), Srinivas (1995)). We take this finding to indicate that, as researchers investigating the effects of cultural diversity on economic outcomes, we need to consider that level of identity along which people in fact feel affinity toward others. It is that identity that would most strongly influence group dynamics, groupthink, trust, or conflicts.[12]

These findings are remarkable for two reasons. During the years in our sample period, 1999-2015, India's economy underwent a transformation. Annual growth rates were of the order of 8%, there were waves of privatization of state owned enterprises, and with international trade liberalization, the country rapidly modernized and integrated into the world economy. Our paper reveals, however, that despite such economic dynamism, those at the top echelons of corporate India continued to be influenced in their decisions

[12]Given the dominance of Hindus in firm boards, variation in religion homophily is less relative to *varna* or *jati* homophily. Nonetheless, in absolute terms, there is a fair amount of variation in religion homophily, so that we do not expect it to be the sole driver of the small effects of religion homophily on firms.

by the traditional institution of caste. Second, over these years, Indian firms increasingly engaged in cross-border mergers and acquisitions, a trend that is expected to continue in the future. This means that firms in other countries are affected by decisions taken in India's boardrooms whose memberships, as our results show, are affected by non-economic considerations.

Our paper relates to the rapidly growing literature on economic effects of culture. Many papers have examined how cultural attitudes and assimilation, religion, and family values impact economic growth (see Guiso et al. (2003), McCleary and Barro (2003, 2006), Noland (2005), Ashraf et al. (2007), Tabellini (2010), Fernández (2011), Alesina and Giuliano (2010), and Campante and Yanagizawa-Drott (2015)) and a range of other economic outcomes.[13] One strand of work focuses on how cultural identity shapes networks (see, for example, Currarini et al. (2009)), hiring (Åslund et al. (2014), Giuliano et al. (2009), Giuliano and Ransom (2013), and 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 (see Gompers et al. (2016), Glover et al. (2017), Shayo and Zussman (2011), Fisman et al. (2017), Bengtsson and Hsu (2015), Hegde and Tumlinson (2013)), Freeman and Huang (2015), Dee (2005) and Fairlie et al. (2014)). Our paper is different in that we analyze the effect of cultural diversity in *teams* of multiple agents and their joint decisions in high stakes economic settings.[14] Importantly, we are able to examine the firm performance effects of board diversity at three different cultural classifications and find that the coarseness of the identity measure matters for our understanding of the economic effects of diversity.

Several studies have analyzed the effects of board composition on corporate governance and firm outcomes along characteristics such as independence, experience in related industries, etc. (see, for instance, Knyazeva et al. (2013) and Dass et al. (2013)). An important strand of this literature focuses on board diversity, but the dimension examined is almost exclusively gender (see, for example, Adams and Ferreira (2009), Ahern and Dittmar (2012), Kim and Starks (2016), Sila et al. (2016), and Bertrand et al. (2018). Terjesen et al. (2009) provide an excellent review.). A few exceptions include Bernile et al. (2018) and Arnaboldi et al. (2018) who develop a multidimensional diversity index

[13]Other outcomes studied include institutions (see Alesina and Giuliano (2014) for an excellent review), public good provision (Alesina et al. (1999), Alesina et al. (2017), Benjamin et al. (2010)), management practices and organization of firms (Bloom et al. (2012, 2014)), trade and economic exchange (Anderson (2011)), nutrition (Atkin (2016)), female labor force participation (Alesina et al. (2013), Fernandez (2007); Fernández (2013), Fernández and Fogli (2006); Fernandez and Fogli (2009), among others), inequality (Alesina et al. (2016)) and political outcomes (Alesina and Giuliano (2011), Gorodnichenko and Roland (2015)).

[14]Hjort (2014) examines productivity effects of ethnic divisions in teams of three workers.

and Kramarz and Thesmar (2013) who show that social networks formed through shared alma maters strongly influence memberships on firm boards. We add to this body of work by considering the *cultural* proximity of directors, as measured by the traditional institutions of religion and caste. The impact of such traditional institutions on board composition and firm performance has not been previously explored. Note that cultural proximity, the focus of our paper, is inherently different from *social* ties. As described by Fisman et al. (2017), while social ties are formed endogenously by conscious choice, cultural identity is inherited at birth. Individuals may feel affinity to those who are culturally proximate, even if they have never met or interacted before. Indeed, we find that the traditional constructs of caste and religion continue to significantly influence the modern institution of corporate boards.

This paper also contributes to the literature examining the economic effects of caste. Previous studies mainly compare socio-economic outcomes of disadvantaged castes to those of advantaged upper castes (see, among others, Hnatkovska et al. (2012, 2013), Iyer et al. (2013), Ghani et al. (2014), Damodaran (2008), Thorat and Neuman (2012), Jodhka (2010), and 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. Fisman et al. (2017) show that when borrowers and loan officers belong to the same caste or religion, the likelihood of a loan being made and repaid increases. Munshi and Rosenzweig (2013) show that relying on caste networks increases intergenerational occupational mobility, and Munshi and Rosenzweig (2016) argue that these networks serve as informal insurance mechanisms.[15] 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, or traditional businesses, we show that cultural proximity shapes economic outcomes nationally, even in urban and elite corporate environments. 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, Damaraju and Makhija (2018) who consider caste proximity between CEOs and firm owners or chairpersons, and Vissa (2011) who shows that entrepreneurs form ties with those who belong to the same caste as themselves.[16]

[15]See also Anderson (2011), Munshi and Rosenzweig (2006) and Banerjee and Munshi (2004).

[16]To the best of our knowledge, only one previous study by Ajit et al. (2012) has examined the role of caste composition of Indian boards. But the authors assigned caste affiliations to directors based on their conjectures and information gathered online. They also did not attempt an investigation into the implications of caste concentration on boards for firm performance. Their caste groups are classified into *varnas* and SC/ST/OBC, instead of the much finer *jatis*. We also use a much larger sample of firms and years. Biswas et al. (2016) also analyzes linguistic homophily of Indian boards but does not explain how

The rest of the paper is organized as follows. In section 2, we describe our data sources. Section 3 presents stylized facts about cultural homophily in India. In section 4, we describe our empirical strategy to identify the causal effect of board cultural homophily on firm performance. Results are presented in section 5. Section 6 explores the mechanisms underlying our results. Section 7 concludes.

2 Building the Database

We use three main data sources: (1) Data from matrimonial websites (*Jeevansathi.com, Bharatmatrimony.com and Shaadi.com*): to probabilistically match last names to religion, *varna* and *jati*, (2) Prowess database: for information on boards of directors and other firm characteristics, and (3) Indian Boards Database: to document board interlocks and examine the relevance of our instrumental variable strategy.

2.1 Data from Matrimonial Websites

Social and cultural practices relating to naming customs in India are complex. In general, however, the last names adopted by Indians are indicative of their religion and caste (Dumont (1980)).

We use a unique data set of self reported castes from three popular matrimonial sites of India – *Shaadi*.com, *Jeevansathi*.com and Bharatmatrimony.com.[17] The historically endogamous institution of caste continues to be a predominant factor based on which marriages are determined in India. This importance of caste is reflected in such websites, where prospective brides/grooms self-report their castes (specifically, *jatis*). Moreover, since people want to marry within caste, users who do report their caste, have an incentive to report it truthfully.

Probabilistic Mapping of Last Names to Religion and Caste

The raw data obtained from the three matrimonial websites include over six million self-made profiles, which provide information on an individual's first and last names, native language, religion, and caste (*jati*). To build a robust mapping, we drop all last

she identifies linguistic affiliations of board members.

[17] In 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 the Associated Chambers of Commerce and Industry in India, the online matrimony business was expected to be worth $250 Million by 2017 (Titzmann (2013)).

names that appear only once in the database. After a considerable amount of cleaning, we are left with 5,447,129 profiles, spanning 16,637 unique last names, 8 unique religions, 5 unique *varnas*, and 471 unique *jatis*.

These data show that the same last name may be associated with more than one religion or caste, often depending on the geographical region. Moreover, the same last name can have different spelling variations. In building a concordance between last names and religion/caste, we take both of these factors into account. We describe the methodology below.

The vast majority of names are words in Indian languages (e.g. Hindi, Tamil, Marathi, etc.), whereas the websites from which the data are culled are written in English. Therefore, an English equivalent (not translation) of these names are represented in the data. Thus, in many cases, multiple English spellings of the same Indian last name are represented. To accurately map last names to castes, we need to collate all the alternative spellings of the same last name. To do so, we use two different word matching algorithms to predict the similarity of different last names. If the similarity predictions from both the algorithms are above a certain threshold, then the two last names are considered to be the same and their caste mappings are combined.

The first word matching algorithm is a modified Levenshtein distance algorithm. In this algorithm, a distance measure between two strings is calculated using a dynamic programming approach, with each replacement alphabet adding one to the distance measure. The standard Levenshtein distance measure is appropriate for English words, whereby each difference in alphabets 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 alphabets 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 word.

The last names are also matched using a modified version of the Ratcliff/Obershelp pattern matching algorithm. This algorithm looks for matches in the longest contiguous matching subsequence of 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 provided a match score of greater than 85%, the two last names were deemed to be matched.

As a final step, we consider two last names to be fully matched if both the 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.[18] This name matching results in groups of similar sounding last names that have different spellings. Thus, for example, one last name group consists of *Rathore*, *Rathor* and *Rathour*.

We observe that many last names have different self reported castes according to the geographical regions. We exploit the fact that different regions of India have different dominant languages. Therefore, the mapping from last name to caste is conditioned on the first language identified by the user. Following this approach, we count the total number of self reported profiles with same last name, that are associated with the same language. In case of a last name group, each name in the group is assigned the sum of occurrences of all last names in the group, subject to the language being the same.[19] We take all religions/*jatis* reported for each last name over all its (within-language) occurrences and count their respective occurrences. Again, in case of a group, we identify the superset of all religions/*jatis* reported across all spelling variations of the same last name (within-language) and count their occurrences. Dividing the number of times a last name is associated with a particular religion/*jati* 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 religion/*jati*. Doing so over all religions and *jatis* associated with a last name gives us the probabilistic mapping of each last name to religions/*jatis*. In case of a name group, each name variant in the group is assigned the same probability distribution over religions/*jatis*.[20]

Although the institution of caste originated in Hindu society, it has been adopted de facto in many non-Hindu religions too (Dumont (1980)), albeit weakened or incomplete. This is demonstrated in our data with some individuals that report themselves to be non-Hindu also indicating a *jati*. For our analysis, we assume that last names reported with non-Hindu religions (3070 in number) do not have a *jati* associated with them. Therefore, in the last name to caste (*jati*) mapping, we simply assign the corresponding non-Hindu religion to these names.

[18] Friedman et al. (2001)

[19] The resulting average total count for each last name is 362.68.

[20] In one region of South India (the state of Tamil Nadu), the last name of an individual is simply 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 *jati* instead of the last name. Although even first names vary by *jati*, we expect the mapping from first name to *jati* to be noisier when compared to mappings of last names that remain invariant across generations.

Table 1: Examples of Last Name to Religion and Caste Mapping

Last Name	**Total Occurrence**	**Religion**	***Varna***	***Jati***	
				Probabilities in Parentheses	
Wadhwani	335	Hindu (1)	Vaishya (1)	Sindhi (0.99)	Arora (0.01)
Dandriyal	11	Hindu (1)	Brahmin (1)	Brahmin Garhwali (0.82)	Brahmin Pandit (0.18)

Source: Matrimonial data. This table provides two examples of last names to religion and caste (*varna* and *jati*) mappings developed by the authors as described in Section 2.1.

We also map last names to *varnas*. For this, we rely on government of India reports, Wikipedia entries, and other sources to build a mapping from *jatis* to *varnas*. Thus, there is some subjectivity involved. This mapping yields five *varna* categories – *Brahmin*, *Kshatriya*, *Vaishya*, *Shudra*, *Dalit*. Additionally, we assign "unknown" to names for which we are unable to find a *jati* to *varna* mapping, or if a *jati* maps to multiple *varnas*.

In Table 1, we provide two examples of the resulting probabilistic mappings of last names to religion/*varna*/*jati*. Both last names are associated with a single religion and *varna* but have two possible *jatis*. More generally, the mappings have the following basic features. Each last name is 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 just the top two religions, with them jointly accounting for about 99% of the total likelihood, on average. As for *varnas (jatis)*, last names are associated with an average of 3.6 (10.4) and a max of 11 (138) *varnas (jatis)*. Table 2 shows the most likely religion, *varna*, and *jati* composition of names in the final matrimonial sample. As expected, 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. According 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 *varna* or *jati*. In the data, however, we see the maximum representation of the *shudra varna* followed by *brahmin*. For space considerations, the table only reports the percentages of last names for the top ten most frequently occurring *jatis* in the data.

We use these concordances to assign religions and castes to corporate directors serving on firms in our sample, described next.

Table 2: Religion, Varna and Jati Composition of Matrimonial Data

Religion	% Last Names	Varna	% Last Names	Jati	% 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, *varna*, and *jati* distribution of the last names included in the final mappings developed by the authors as described in Section 2.1. For space considerations, the table only shows the distribution for the top ten most frequently occurring *jatis* out of a total of 471 distinct *jatis* that we can identify.

2.2 Firm Level Data

Data on Indian firms is obtained from the Prowess database provided by the Centre for Monitoring the Indian Economy (CMIE). The data cover large listed, unlisted, and government owned firms over the period 1989-present. Detailed information is available on the identity (entity type, ownership, industry, age, etc.), and governance and financial aspects of these firms. We choose to use data for the period 1999-2015, as the number of firms covered by Prowess is much smaller prior to 1999.

Prowess covers firms for which it can obtain publicly available information. The data are sourced mainly from annual reports, quarterly financial statements, and profit and loss accounts of firms. Thus, information on all listed companies that are reasonably active on the major stock exchanges of India is available in the database. Though the database includes mostly publicly listed firms, a smaller number of unlisted firms are also included. The reason for smaller coverage of these firms is that they are not required to publicly disclose their financial statements.[21] Prowess does not cover informal firms since there is little publicly available information on them. Thus, the database does not cover the universe of all firms in India. Nonetheless, the firms 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 taxes and all excise taxes collected by the government.

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

Table 3: Religion, Varna and Jati Mapping for Names of Directors

Number of classifications	% last names with up to that number of classifications	Average cumulative 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
	Varna	
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
	Jati	
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, *varna*, and *jati* distribution of director last names. For space considerations, the table only shows the distribution for last names associated with up to five *jatis*.

While the database follows firms longitudinally, most firms appear in the data only for a few years. Thus, for most of our analyses, we treat these annual data as repeated cross sections of firms.

To identify the religions and castes of company 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 religion/caste of all board members.[22] We also consider only those firm-years

[22]There are also 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 simply assigned a religion, *varna*, and *jati* category of "NA."

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. [23]

Table 3 profiles the religion and caste mapping for directors. For religion, we see that 69% of directors' last names are associated with a single religion, and 92% names are associated with up to two religions. As expected, the mapping for caste is noisier, with 45% (40%) last names associated with a single *varna* (*jati*) and 85% (65.5%) last names associated with up to five *varnas* (*jatis*). Although the mapping is probabilistic, the probabilities are front loaded. Thus, the top two most likely religions account for 99.5% of the total likelihood, on average. Similarly, the top five most likely *varnas* (*jatis*) account for 99.3% and 89% of the total likelihood, on average. Table 4 provides the composition of directors for the first and last years (1999 and 2015) of our sample, as measured according to their most likely religion, *varna*, and *jati*.[24] In both years, *vaishyas* dominate the director sample, accounting for 26.6% and 28% of directors, respectively. Looking at *jatis*, *Agarwals*, who belong to the *vaishya varna* are the most represented, constituting just over 11% of directors in both years.

We also use information on several firm characteristics, including sales, total assets, annual profits, export status, industry, location of headquarters, etc. A few key characteristics of firms in our sample are provided in Table 5. We note that the highest proportions of firms at both the start and end of our sample period belong to manufacturing, and finance, insurance and real estate sectors. Under half of them are listed on India's stock exchanges and about half are members of business groups. The mean real assets of these firms were about Rs. 18 million in 1999 and Rs. 23.3 million in 2015. The average board size is about 5.4. Note that although we report the percentages of firms with dual CEOs and the mean percentages of independent directors, the data on these aspects is quite sparsely populated. Specifically, we can identify whether CEO are dual for only 46.14% of our sample, and calculate the percentage of independent directors on boards for only 21.38% of the sample. Therefore, we are unable to use these board characteristics in our regression analyses discussed in Sections 4 and 5.

Homophily Index: To measure the degree of cultural diversity, or lack of it (homophily) in a board, we calculate the Blau index for religion, *varna*, and *jati* for every board.[25] The Blau index for a board is the sum of squared shares of directors belonging to the various cultural identities represented on the board. For example, consider a board with

[23] As a percentage of firms and directors in the Prowess database over 1999-2015, we are able to match 63.52% of firm-year observations and 57.11% of all directorships.

[24] For space considerations, we only report the top ten most frequently observed *jatis* of directors.

[25] This measure is based on the Gini-Simpson index which is also known as the Blau or Hirschman-Herfindahl Index (Hirschman (1945), Herfindahl (1950)).

Table 4: Religion, Varna and Jati Composition of Corporate Directors

Religion	% Last Names	Varna	% Last Names	Jati	% Last Names
		1999			
Hindu	85.24	Brahmin	20.78	Agarwal	11.41
Jain	8.54	Kshatriya	14.51	Brahmin Iyer	4.80
Christian	2.47	Vaishya	26.60	Brahmin	4.64
Muslim	1.92	Shudra	18.06	Khatri	4.56
Parsi	1.15	Dalit	1.24	Nair	4.52
Sikh	0.68	Unknown Varna	3.14	Maheshwari	3.74
				Kayastha	3.74
				Vaishnav	3.31
				Arora	2.94
				Gupta	2.62
		2015			
Hindu	85.85	Brahmin	22.10	Agarwal	11.71
Jain	9.94	Kshatriya	15.01	Brahmin	6.51
Muslim	1.60	Vaishya	28.08	Khatri	4.50
Christian	1.26	Shudra	16.23	Nair	4.47
Sikh	0.80	Dalit	0.88	Kayastha	3.91
Parsi	0.54	Unknown Varna	2.54	Maheshwari	3.72
NA	0.01			Arora	3.39
				Brahmin Iyer	3.22
				Gupta	3.19
				Vaishnav	2.85

Source: Matrimonial and Prowess data. This table provides the religion, *varna*, and *jati* composition of corporate directors using the most likely religion and caste classifications assigned to their last names. For space considerations, the table only shows the composition for the top ten most frequently observed *jatis* in the data. *Varna* is assigned as unknown when we are unable to find the *varna* for a last name that does have a *jati* assignment. Religion is assigned as NA for all foreign sounding last names.

five board members – three Hindus and one each with the most likely religion as Muslim and Christian. The religion Blau index of this board is $0.44 (= (0.6)^2 + 2 * (0.2)^2)$. In our calculations of caste Blau index, the *varna* and *jati* of directors whose most likely religion is not Hindu is treated as their corresponding most likely religion. Continuing with our example above, suppose that of the three Hindu board members, the most likely *varna* of two of them is *Brahmin* and the other is *Kshatriya*. Then, the *varna* Blau index for this board is $0.28 = ((0.4)^2 + (0.2)^2 + 2(0.2)^2)$. A higher Blau index represents lower cultural diversity, or greater cultural homophily, in a board. In the paper, we present results using the Blau Index. In Appendix C, we present results for a more nuanced measure of homophily, which we call the "fuzzy" Blau index. The fuzzy Blau index incorporates the

Table 5: Summary Statistics

Panel A: Sectoral Distribution of Firms (Percentages)		
	1999	2015
Agriculture, Forestry and Fishing	2.46	1.50
Mining, Utilities and Construction	6.47	12.02
Manufacturing	41.42	23.32
Trade	11.74	12.86
Transport, Accommodation	3.41	4.66
Information & Communication	4.01	4.90
Finance, Insurance & Real Estate	23.02	26.25
Professional, Technical and Admin. Services	3.86	6.95
Education & Health	0.50	1.29
Arts, Recreation & Others	0.85	5.23
Diversified	2.26	1.02

Panel B: Firm Characteristics		
	1999	2015
Total Firms	1994	5597
% Listed	47.94	40.16
% Exporters	35.91	21.65
% Group Firms	57.37	48.92
Mean Age	20.22	19.73
Mean Assets (Rupees Millions)	17.99	23.33
Mean Profits (Rupees Millions)	2.16	2.41
Mean Sales (Rupees Millions)	13.34	15.20
Mean Net Tangible Asset Intensity	0.32	0.25
Mean Leverage	0.45	0.45
Mean Return on Assets	0.05	0.04
Mean Asset Turnover	0.97	0.83
Mean Tobin's Q	1.02	1.72
Mean Market to Book Ratio	1.41	2.77
Mean Risk	0.11	0.04

Panel C: Board Characteristics		
	1999	2015
% with CEO Duality	15.05	15.15
Mean % Independent Directors	20.26	2.22
Mean Board Size	5.44	5.44

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 cultural identity for all directors on the boards.

entire probabilistic mapping of religion, *varna* and *jati* associated with every director on the board.

2.3 Indian Boards Database and Other Data Sources

Prowess does not allow us to identify unique individuals serving as directors since we do not have unique director identification numbers. Since two individuals may have the same name, we do not rely on names to identify unique directors. This prevents us from examining the degree and nature of interlocks across boards, i.e., which individuals hold multiple directorships and the firms these positions are in. To do this, we instead use the Indian Boards Database, a database maintained by the Prime database group, which provides a unique identification code for each individual serving as a director for about 1,500 firms during 2006-2015, along with demographic information such as age, gender, nationality, educational qualifications, experience. Additional information on their directorial position is also available, including independent/non-independent status, remuneration, date of appointment, cessation date, and reason for cessation. The unique identification code for directors and information about all the boards they serve on allows us to measure the degree of interlocks. We use this information to examine the relevance of our instrumental variable strategy. The information on other board membership in Indian Boards Database is only available starting 2012. Since the time period of our study is 1999-2015, we only examine the data for 2012-2015 to calculate different measures of board interlocks within the same broad industries. For these years, we have information on 17,608 unique directors across 1,501 firms.

All nominal data are deflated by all-India CPI (2001=100).

3 Patterns of Cultural Homophily in India's Corporate Boards

3.1 Cultural Homphily in Corporate Boards is Systematically High

The average homophily index in Indian firms is high for all measures of cultural identity, at 0.87 for religion, 0.56 for *varna* and 0.45 for *jati*. To assess if boards' cultural homogeneity is systematically high or simply a result of the cultural composition of all directors potentially available to them, we compare the observed homophily levels to those in random simulated samples of the data. We create three different sets of randomly

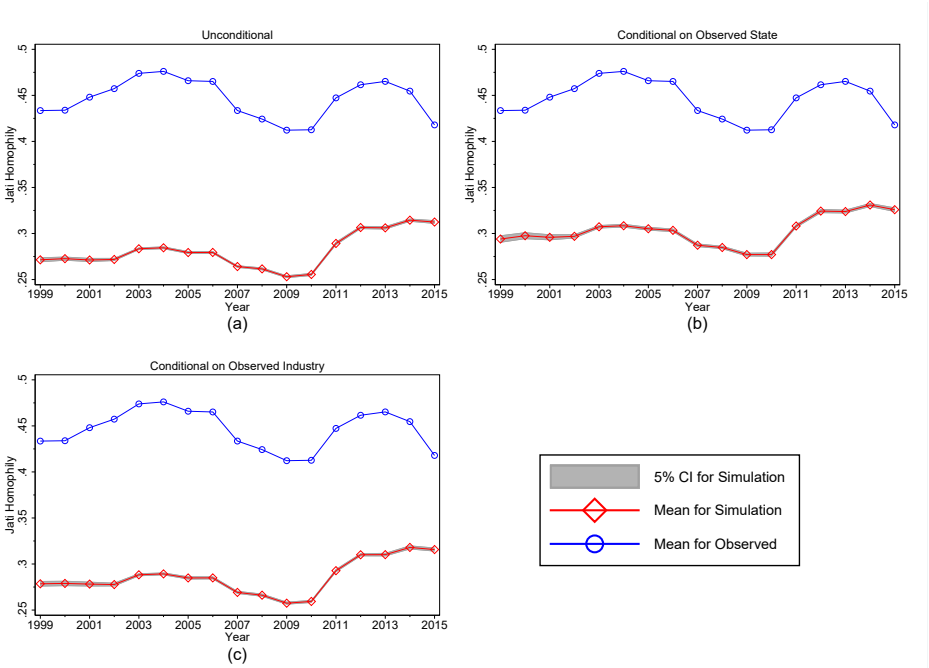


Figure 2: Observed vs. Simulated Average Jati Homophily[a]

[a]Source: Prowess, matrimonial data. The three graphs in the figure present the mean *jati* homophily (Blau index) 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 3.1.

simulated samples. In the first method, we consider all directors appearing in the data 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 the observed board size. For example, a firm with an observed board size of five is assigned five directors at random from the supply pool. We create 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 appearing across all firms in the same state (two-digit industry) as that firm.[26]

[26]As mentioned earlier, we cannot identify individual directors since we do not have unique numeric 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 are available upon request.

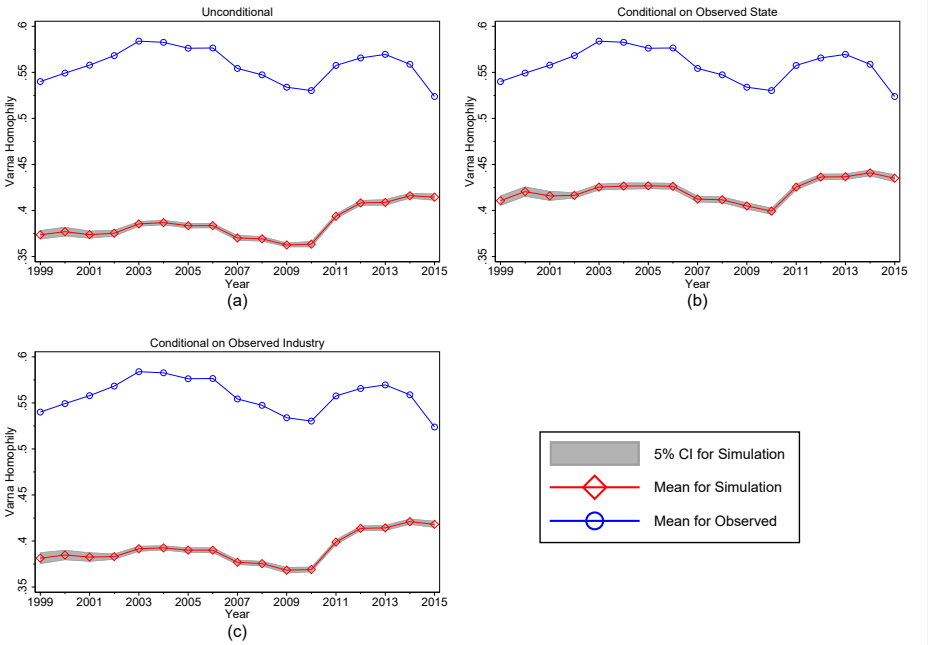


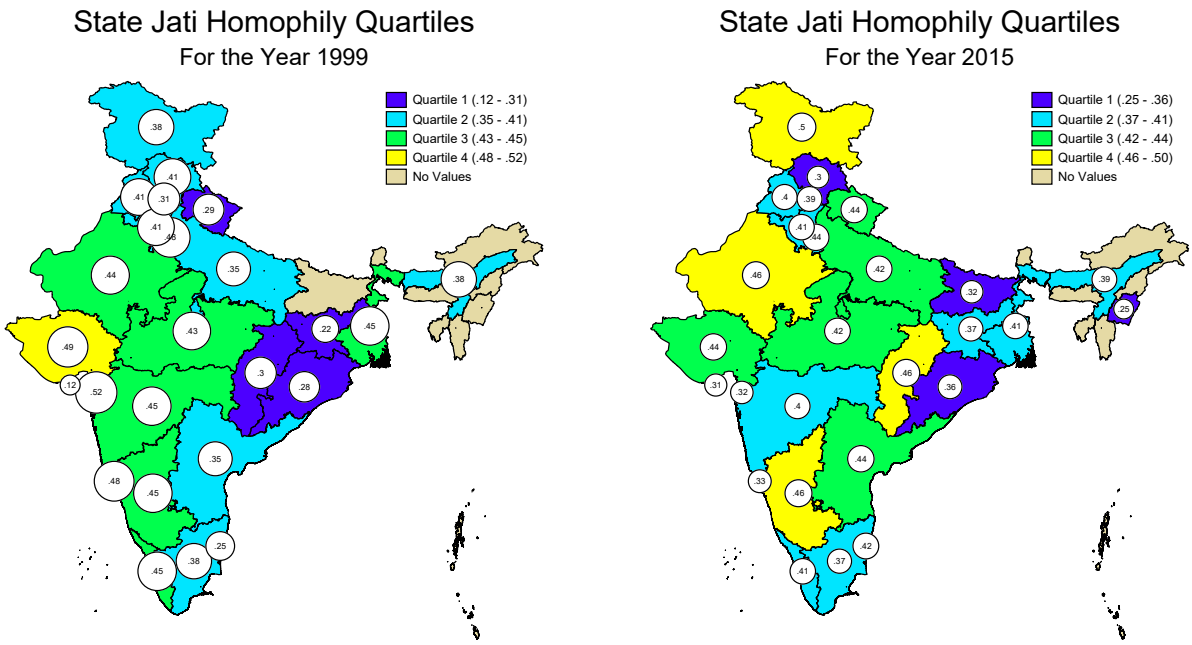
Figure 3: Observed vs. Simulated Average Varna Homophily[a]

[a]Source: Prowess, matrimonial data. The three graphs in the figure present the mean *varna* homophily (Blau index) 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 3.1.

Table 6: Simulation Results

Simulation Method	Cultural Identity	Mean Homophily in Observed Boards	Mean Homophily in Simulated Boards	Difference	t statistic
Unconditional	Jati	0.449	0.282	0.167***	-7110.636
Conditional on Sector	Jati	0.449	0.287	0.162***	-7844.739
Conditional on State	Jati	0.449	0.304	0.145***	-5558.132
Unconditional	Varna	0.56	0.386	0.174***	-4557.745
Conditional on Sector	Varna	0.56	0.392	0.168***	-4171.524
Conditional on State	Varna	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 (Blau index) of boards in observed and simulated data. Homophily is measured for three cultural identities: *jati*, *varna*, 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.



(a) (b)

Figure 4: Average Jati Homophily Across States[a]

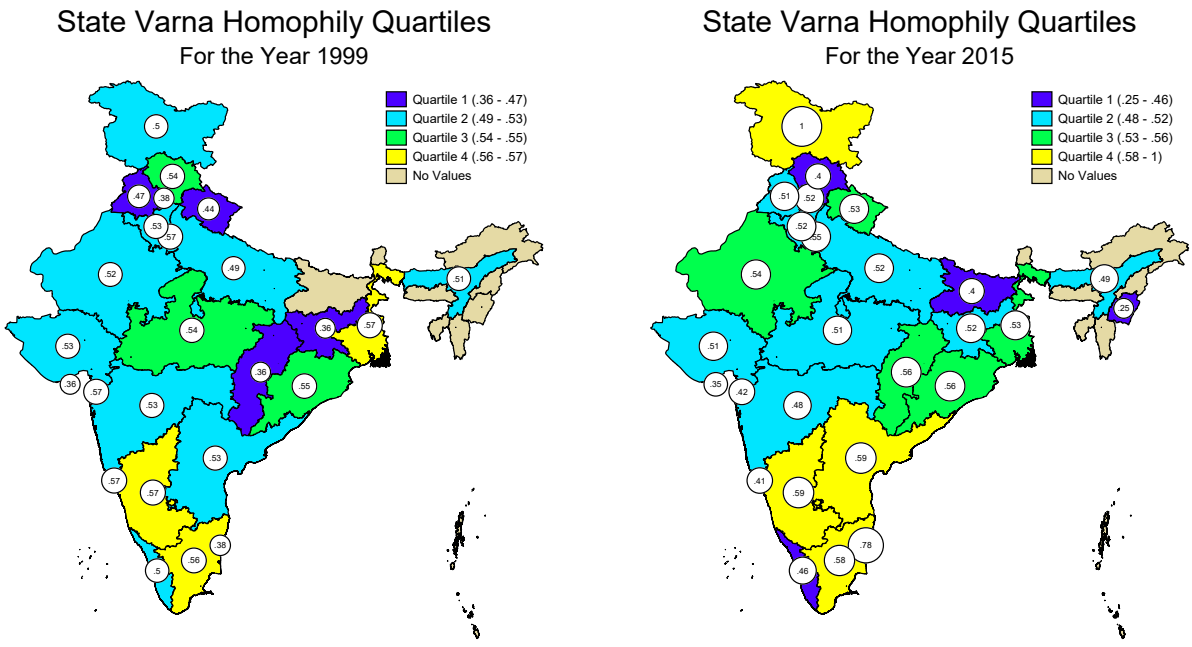
[a]Source: Prowess, matrimonial data. The map shows the average *jati* homophily (Blau index) for all firms in each state in the years 1999 and 2015. The color coding represents the quartile position of a state in the distribution of mean *jati* homophily levels.

Results for religion are presented in Appendix D. Results for *jati* and *varna* and are presented in Figures 2 and 3. Figure 2 shows the yearly means of firm boards' *jati* Blau indices 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 *jati* homogeneity of boards is significantly higher than the corresponding simulated mean in every year. A very similar picture emerges for *varna* in Figure 3. Table 6 presents hypothesis tests for comparisons of observed and simulated means for *jati* (*varna*) homophily. In all years, the t-statistics are large, indicating that the observed mean homohily is significantly different from the simulated means. On the basis of these results, we conclude that firm boards have systematically low caste and religion diversity.

We find that low cultural diversity of boards has persisted over time across states and sectors. Figures 4 and 5 respectively present state-wise mean *jati* and *varna* Blau indices 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. For instance, the state of Maharashtra had a jati homophily of 0.45 in 1999 and 0.4 in 2015. It changed its position in the state distributions of mean board *jati* homophily from the third to the second quartile. Figure 6 shows that in all sectors, the mean *jati* and *varna* homophily stayed persistently high across all years. Across all years and sectors, the *jati* homophily varies in the range of 0.32-0.58. The sectors of professional, technical, and administrative services, and arts and recreation have the highest homophily levels, while the information, communication, and real estate, diversified, and health and education sectors have the least homophily. There is no uniform secular trend in homophily across sectors; while some sectors witness a small decline, others see a modest rise. Such persistence in cultural homophily on boards is remarkable. Our sample period 1999-2015 witnessed unprecedented economic change and rapid growth, with the country attracting considerable amount of foreign investment and offshored activity, and becoming integrated more closely with the rest of the world. Yet, corporate boards of large firms that constitute our sample remained culturally homogeneous.

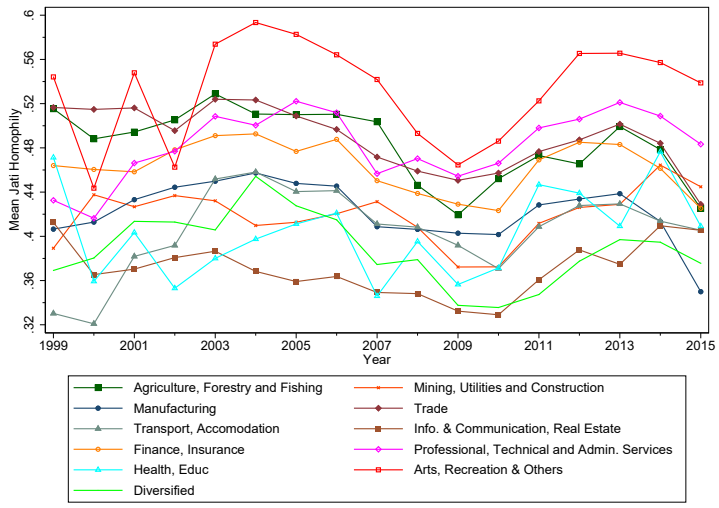
Results for religion homophily are presented in Appendix D.



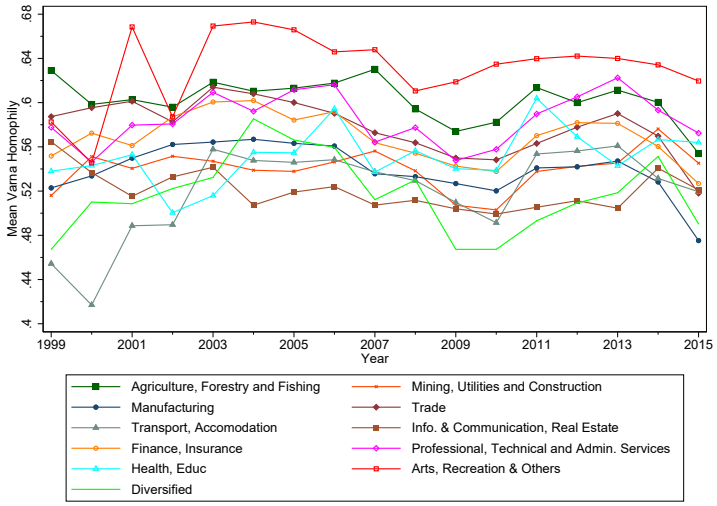
(a) (b)

Figure 5: Average Varna Homophily Across States[a]

[a]Source: Prowess, matrimonial data. The map shows the average *varna* homophily (Blau index) 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 *varna* homophily levels.



(a) Jati



(b) Varna

Figure 6: Average Caste Homophily By Sector[a]

[a]Source: Prowess, matrimonial data. Figures 6(a) and 6(b) show the mean *jati* and *varna* homophily (Blau index) for all firms in eleven broad sectors. Diversified includes all firms that could not be classified primarily into one industry.

3.2 Homophily is Lower in Higher Quality Firms

Next, we present homophily patterns across different firm characteristics. Figures 7 and 8 show that larger firms have more diverse boards. The three panels (a, b, and c) in these figures show the evolution over time of average *jati* and *varna* homophily in firms of different assets, sales, and profits quartiles, respectively. We see a systematic pattern

of lower homophily as we move from lower to higher quartiles of assets and sales. For profits, the second quartile firms have higher homophily, on average, than firms in the first quartile. However, both have a higher homophily than the average of firms in the third quartile which, in turn, have higher homophily on average than firms in the fourth quartile. As before, the differences across quartiles in all three panels are small in magnitude, and the fluctuations over time within each quartile are negligible.

Figure 9 demonstrates that older firms have board members from more diverse cultural backgrounds. As we go from lower to higher quartiles of firm age (measured as years since incorporation), the average caste homophily falls. Figure 10 shows that exporting firms have significantly more diverse boards, on average, than non-exporting firms. Finally, we observe higher caste homophily, on average, among firms that do not belong to business groups relative to those that do. This is noteworthy since one might expect that business groups in India, that are often dominated by a single extended family, would tend to hire directors from among their kin. In that case, family ties, as opposed to shared cultural identity, would drive the high caste homophily on the board. However, we see the opposite. This may be suggestive of the greater productivity, size, and prominence that is associated with business groups enabling or incentivizing these firms to have more diverse boards. We see little difference in average board homophily across other firm characteristics such as public versus private firms and government versus non-government firms. Results for religion homophily are presented in Appendix D.

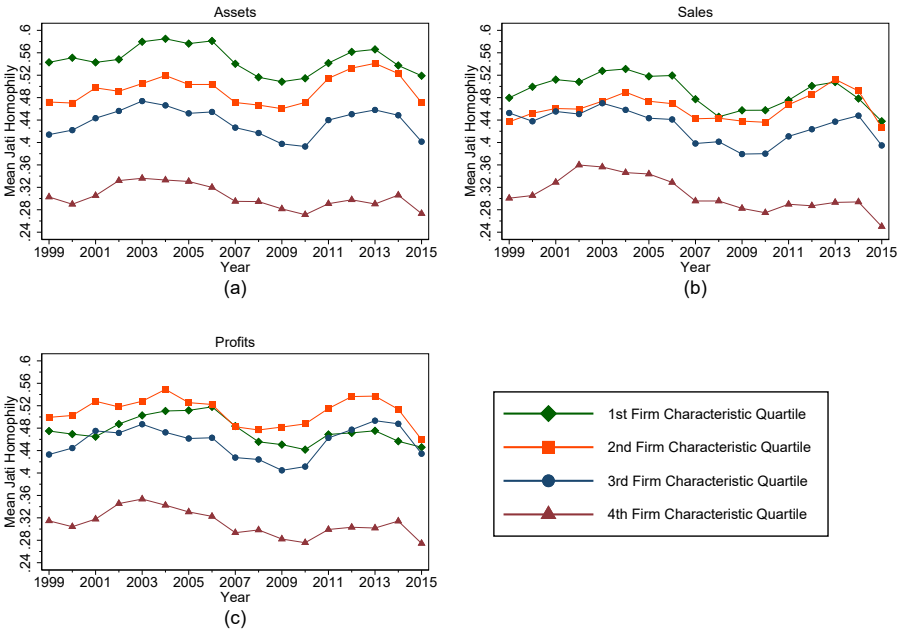


Figure 7: Average Jati Homophily by Firm Size[a]

[a]Source: Prowess, matrimonial data. Figures (a), (b), and (c) show the average *jati* homophily levels in firms falling in the four quartiles of assets, sales, and profits, respectively.

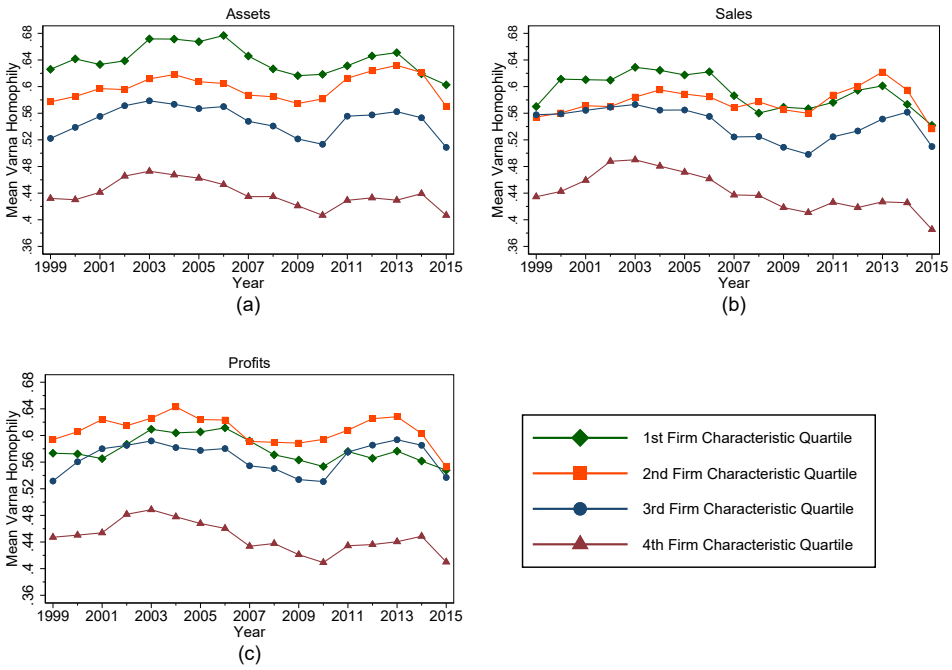
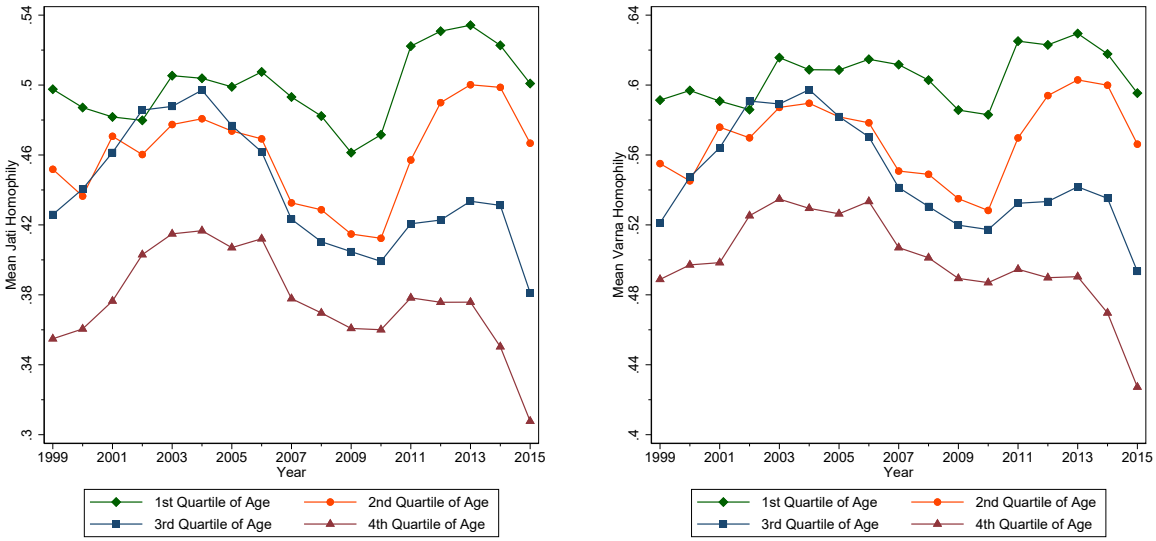


Figure 8: Average Varna Homophily by Firm Size[a]

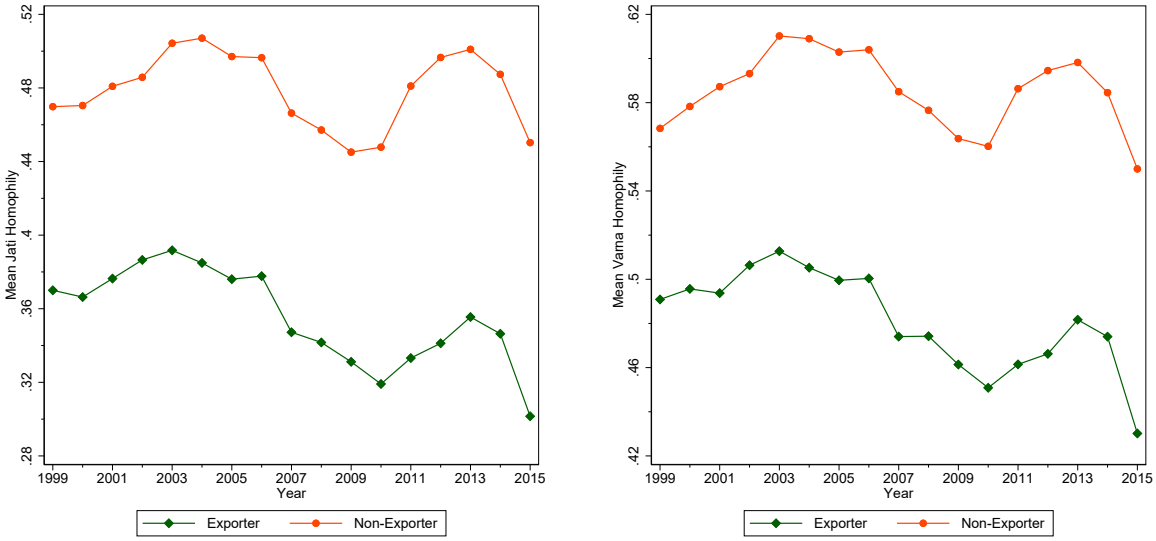
[a]Source: Prowess, matrimonial data. Figures (a), (b), and (c) show the average *varna* homophily levels in firms falling in the four quartiles of assets, sales, and profits, respectively.



(a) Jati (b) Varna

Figure 9: Average Cultural Homophily by Firm Age[a]

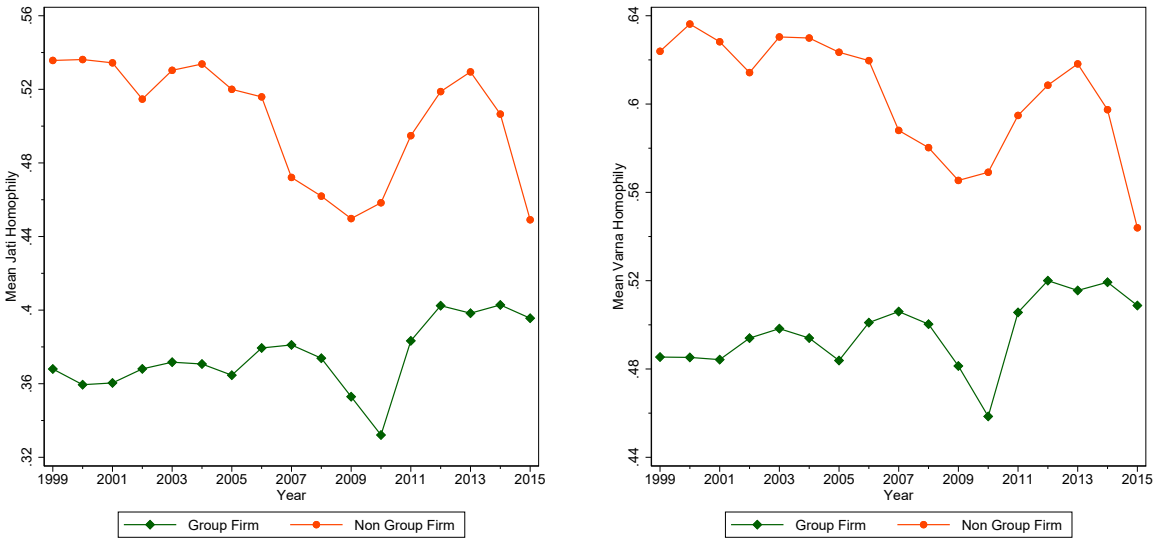
[a]Notes: Source: Prowess, matrimonial data. Homophily (Blau index) is averaged over all firms in each quartile of the age distribution. Firm age is defined as the number of years since the incorporation year.



(a) Jati (b) Varna

Figure 10: Average Cultural Homophily by Exporting Status[a]

[a]Notes: Source: Prowess, matrimonial data. Homophily (Blau index) is averaged over all exporting and non-exporting firms separately.



(a) Jati

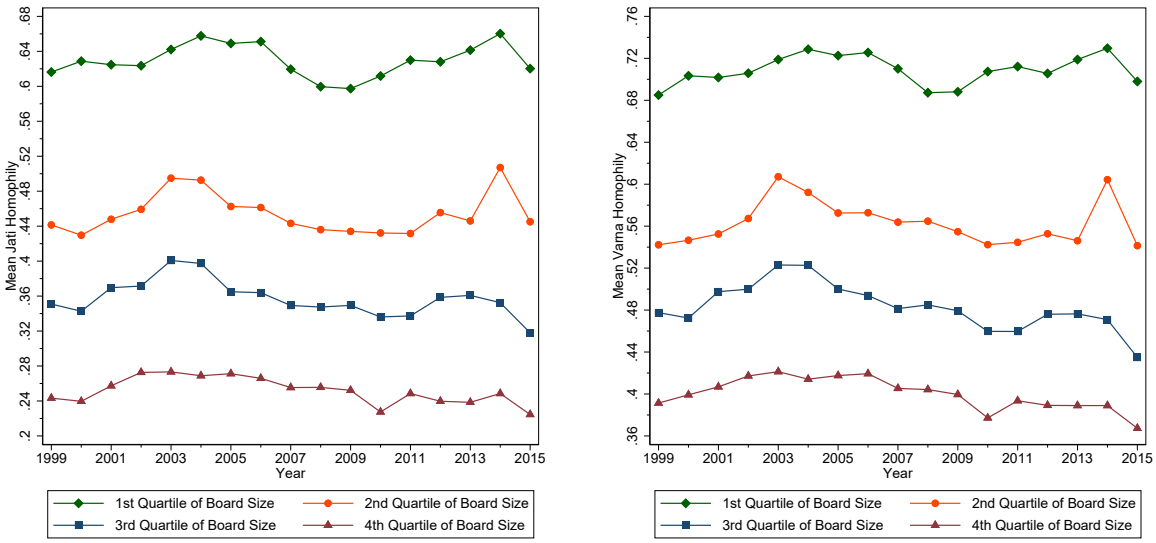
(b) Varna

Figure 11: Average Cultural Homophily by Business Group Membership[a]

[a]Notes: Source: Prowess, matrimonial data. Homophily (Blau index) is averaged over all firms that belong to business groups and all firms that do not.

3.3 Better Corporate Governance Accompanies Lower Homophily

We also find that boards that display lower levels of cultural homophily also have features indicative of superior corporate governance. Figures 12 and 13, respectively, show the association between *jati* and *varna* homophily and two features of corporate governance: size of the board and proportion of independent directors on the board. Figure 12 presents average cultural homophily for firms with different board sizes, grouped into four quartiles. We take larger board sizes as indicative of better governance. We see that firms with larger corporate boards have lower *varna* and *jati* homophily among their directors. Figure 13 shows the association between average board cultural homophily and the average proportion of independent directors across listed firms in one-digit sectors in the year 2015. Panel (a) shows negative correlation between mean *jati* homophily and the proportion of directors that are independent across sectors. Sectors with the highest *jati* homophily levels such as trade and finance also have the low proportions of independent directors in their corporate boards. The correlation between the mean percentage of board constituted by independent directors and *varna* homophily is positive, however. Both figures also show that all sectors have lower percentages of independent directors compared to international standards.

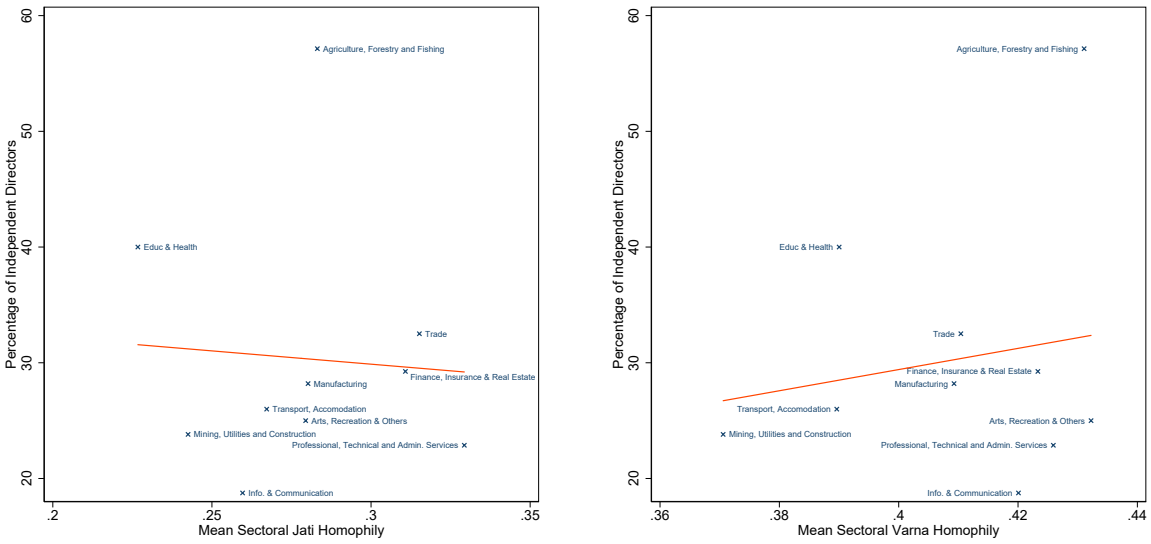


(a) Jati

(b) Varna

Figure 12: Average Cultural Homophily by Board Size[a]

[a]Notes: Source: Prowess, matrimonial data. Homophily (Blau index) is averaged over all firms in each quartile of the board size distribution.



(a) Jati (b) Varna

Figure 13: Average Cultural Homophily and Percentage of Independent Directors[a]

[a]Notes: Source: Prowess, matrimonial data. Homophily (Blau index) is averaged over all firms in each one-digit industry. This figure is made using data for year 2015.

4 Empirical Strategy

Section 3 demonstrates that although cultural diversity in corporate boards has remained low throughout the sample period, it systematically varies across firms such that it is higher in larger and better performing firms. Next, we examine whether this systematic pattern is causal. In particular, we investigate whether and to what extent higher religion and caste homophily on firm boards negatively impacts key measures of firm performance. In this section, we describe our empirical strategy for this analysis.

Consider the following regression equation:

$$P_{it} = \beta_0 + \beta_1 H_{it} + \beta_2 X_{it} + \beta_3 B_{it} + \delta_1 I_j + \delta_2 T_t + \epsilon_{it} \tag{4.1}$$

where P_{it} denotes firm i's value and performance in year t, H_{it} is the *varna*, *jati* or religion homophily (Blau index) 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_j denotes a vector of two-digit industry fixed effects, and T_t is a vector of year fixed effects. We cluster the standard errors by industry and correct them for arbitrary heteroskedasticity.

Our dependent variables are performance variables including operating income, operating cash flow, and profits (all in natural logs), and market based indicators including

market to book ratio, Tobin's Q, and firm volatility. Operating Income is defined as the difference between sales and operating expenses. Operating cash flow is the cash flow from operating activities before depreciation. Market to book ratio is defined as the ratio of market price per share to book value per share. Tobin's Q is calculated as the market value of a company divided by the replacement value of the firm's assets. A low Q (between 0 and 1) means that the cost to replace a firm's assets is greater than the value of its stock. This implies that the stock is undervalued. Conversely, a high Q (greater than 1) implies that a firm's stock is more expensive than the replacement cost of its assets, which implies that the stock is overvalued. Volatility is measured as the standard deviation of the returns on a firm's security over a year.[27]

The control variables include firm age, board size, firm size (measured by real assets), tangibility, and book leverage. We define firm's age as the number of years since incorporation of the firm. Board size is the number of directors on board of a firm in a year. Tangiblity (or tangible asset intensity) is defined as the fraction of tangible assets in the total assets of a company. Book leverage is calculated as the ratio of the total debt of a company and the total assets.[28] Other controls include indicators for whether the firm is listed on the stock market, part of a business group, exporting status, year and industry fixed effects. Standard errors are corrected for arbitrary heteroskedasticity and are clustered at the level of two-digit industries.

In the above regression, β_1 captures the association between cultural homophily and firm performance. However, this coefficient is not a causal estimate 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. Moreover, 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 the board composition and, hence, homophily.

To overcome this endogeneity, we employ three instrumental variable strategies. In

[27]We measure market to book ratio, Tobin's Q ,and firm volatility in two ways. One 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 for the latter method. Volatility is interpreted as the amount of uncertainty or risk about the size of changes in a security's value. A higher volatility means that a security's value can potentially be spread out over a larger range of values. This means that the price of the security can change dramatically over a short time period in either direction. A lower volatility means that a security's value does not fluctuate dramatically, but changes in value at a steady pace over a period of time. Majority of firms are traded on Bombay Stock Exchange (BSE), and some are traded on the National Stock Exchange(NSE). Through out the paper we only consider firms returns at BSE.

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

the first approach, we use two instruments: (1) the religion or caste (*jati* or *varna*) Blau index for all directors in the (two-digit National Industrial Classification (NIC) 2008) industry that the firm belongs to, (2) the religion or caste Blau index for all directors in the state where the firm is located.[29] [30] These two variables provide us a measure of the religion or caste 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 religion or caste composition from that of the full set of directors in the corresponding industry, and (4) the Euclidean distance of a board's religion or caste composition vector from that of the full set of directors in the corresponding state. In the third approach, we exploit 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 remaining part of this section, we discuss the plausible relevance and validity of all of these instruments.

The intuition for using the instruments in the first two approaches is that a firm's board composition may be similar to that of other firms in the same industry or geographical area. This suggests that if the group of all directors in the industry/region displays low levels of cultural diversity (i.e., high homophily), then firms in that industry/region may also be more likely to lack board diversity. Moreover, a firm may choose its directors from its local geographical region and/or industry. Previous studies have shown that both geography and industry influence the supply of directors that firms can choose from (see Knyazeva et al. (2013) and Dass et al. (2013)). We show that this holds in our setting too by documenting that (a) a non-negligible proportion of directors on a board are also directors of other firm(s) in the same industry and state, and (b) the religion and caste composition of directors on firm boards is very similar to that in the entire industry (even if directors on these boards have few or no members with additional directorships in the same industry) or in the entire region (identified by state). Table

[29] A more disaggregated classification is unsuitable for two reasons. First, directors may not serve on closely competing firms' boards due to conflicts 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 disaggregated classification level, on the other hand, is undesirable as it will not yield enough variation in the industry level homophily index.

[30] We 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/sector. In the paper, we present results using the first approach. Results using the second approach are extremely close.

Table 7: 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
Info. & 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
Educ & 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 other firms' boards, currently (2015) or in the past (2012-2014), such that these firms belong to the same one-digit industry.

7 documents within-industry board interlocks for one-digit industries for the year 2015. 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 as a director on a firm that 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. This gives us a firm's degree of within-industry interlocks. Key moments of these interlocks for firms in each 1-digit industry in the year 2015 are presented in Table 7. We observe that the average interlock ranges from 0% to 31% across these broad industries. But the maximum degree of interlocks can be as high as 100%. Looking at two-digit and three digit industries, we see that even at these narrower levels, there are interlocks, 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 religious/caste 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 the Kolmogorov-Smirnov test (K-S test).[31] The test statistic requires sample sizes of the two samples that it compares. Thus, it

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

does not allow us to compare the distribution of the top religion/caste of directors of an average firm with that of the entire industry since the sample size of directors for this distribution is not defined. Thus, we compare the distribution of the top religion/caste of unique directors in the set of firms that have below-median interlocks to that of the entire industry. We repeat this for firms with above-median interlocks. To look at this differently, we do another K-S test where the distribution of firms' dominant religion/caste for the set of firms with below (and above) median interlocks is compared with that of firms in the entire industry. Results for *jati* from these tests are presented in Table 8. The table shows that for each one-digit industry, we are unable to reject the null hypothesis that the samples of directors in firms below (and above) median 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 *jati*. We also reach the same conclusion for *varna* and religion.

In our second approach, we additionally use the distance between the firm and industry/region with regard to their cultural composition. Note that several different religion/caste compositions can yield the same homophily index. So whether a firm's board composition is similar to that in its industry/region can be determined not only by comparing its overall homophily index with that of the industry/region but also its underlying religion/caste composition. The larger this distance, the less similar is the firm's director composition to that in the industry. 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 demonstrates that Euclidean distances between firms and industry/state level *jati* 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. However, even beyond the 50th percentile, the distances remain small, so that at the 90th percentile, the distance ranges between 0.5-0.6 across years. 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 firms' and state director compositions (Panel B).

The validity of our set of 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 state and industry

Table 8: K-S Test Results Comparing Jati 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 *jati* 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.

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 the 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, to further ensure against this possibility, we also include several firm and board characteristics besides homophily in our regression. This accounts for mechanisms through which any one firm may influence the state or industry level homophily. Thus, we expect that given all the fixed effects and control variables included in the regres-

Table 9: Distance Between Firm and Industry/State Director Jati 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 *jati* composition of firms' directors and that of the set of directors in the same two-digit industry (Panel A) or state (Panel B).

sions, the state and industry level homophily only affect firm performance through their influence on the firm's board homophily.

The Euclidean distance between a board's religion or caste composition and that of the aggregate set of directors in the corresponding industry or state also meets the exclusion criterion. The three reasons described above for the validity of the state and industry-level homophily indices also apply to the distance measures. Further, there is an additional reason why this distance of a board's composition from that of industry or state composition is valid. Consider the following example. Suppose an industry's directors belong to three different religions – 50% are Hindus, 25% are Muslims, and another 25% are Christians. A firm in this industry may have ten directors, eight of whom are Hindu and two are Christians. Now, if this firm replaces two of its Hindu directors with two Muslims, then the firm's director composition becomes closer to the industry-level composition. The only way that this distance can affect firm outcomes is through the endogenous regressor (board homophily). There is no reason to expect, ceteris paribus, that simply replacing two Hindus on the board with two Muslims would have any independent effect on firm performance, i.e, there is no reason to expect a pure religion effect.

In a third approach, we exploit board membership changes induced by firms complying to 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 the largest firms (Group A companies listed on the Bombay Stock Exchange (BSE), and National Stock Exchange (NSE) S&P CNX Nifty Index companies), March 31, 2002 for 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 company's history, March 31, 2003 for firms with paid-up share capital of at least Rs. 30 million, and any newly listed or re-listed firms at the time they get listed (see Dharmapala and Khanna (2012) for more detail). To construct the instrument, we exploit the variation in the timing of the deadlines by when different groups of firms had to comply with the new requirements and whether they needed to change their boards in order to comply. Specifically, our instrument is defined as $I(\text{eligible}) * I(\text{below threshold})$, where a firm is considered eligible to comply with the Clause requirements if they fall into any of the above-described groups of firms and we observe them after February 2000. A firm is considered below threshold if in any year it has fewer than 50% of its directors who are non-executive. 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.

5 Homophily and Firm Performance

5.1 Homophily Is Negatively Associated with Firm Performance

We first present fixed effects results for the association between firm performance measures and *jati* and *varna* homophily of boards of directors in Tables 10 and 11, respectively. Corresponding results for religion homophily are in Appendix D. Columns (1)-(3) present results for the association between cultural homophily of boards and firm performance measures – log (operating income), log (operating cash flow), and log (profits), respectively. Columns (4)-(6) present analogous results for homophily and firm value measures – market to book ratio, Tobin's Q, and volatility, respectively. We observe that for both measures of caste homophily, firm performance and value are lower in firms with more homophilous boards. A one unit increase in *jati* homophily reduces operating income by 0.48 log points and profits by 0.41 log points, on average. Market to book

ratio falls by 0.97 points for a one unit increase in *jati* homophily. Tobin's Q is also negatively associated with homophily but the estimated coefficient is statistically insignificant. Higher board homophily is also correlated with greater stock market volatility for the firm. A one unit increase in *jati* homophily is associated with a 0.01 increase in the standard deviation of the firm's stock market returns, on average. The corresponding associations between *varna* homophily and firm outcomes are similar, but the estimated coefficients are smaller in magnitude in all cases. The results for religion homophily in Appendix D show that firm performance measures are positively associated with religion homophily of boards and firm value and volatility do not vary with it. Most coefficients are small and statistically indistinguishable from zero.

Table 10: Jati Homophily and Firm Outcomes: Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Log(Operating Income)	Log(Operating Cash Flow)	Log(Profits)	Market to Book Ratio	Tobin's Q	Volatility
Jati Homophily	-0.477***	-0.622***	-0.407***	-0.968***	-0.144	0.010***
	(0.098)	(0.090)	(0.081)	(0.360)	(0.123)	(0.002)
Firm Age	-0.003**	-0.002*	-0.001	0.007**	0.001	-0.000***
	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)	(0.000)
Board Size	0.182***	0.170***	0.198***	-0.039**	-0.016*	-0.001***
	(0.009)	(0.008)	(0.011)	(0.019)	(0.009)	(0.000)
Leverage	-0.115***	-0.109**	0.053	1.101***	1.190***	0.002***
	(0.038)	(0.051)	(0.071)	(0.303)	(0.059)	(0.001)
Assets	0.009***	0.009***	0.010***	0.001	-0.000	-0.000***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Tangibility	0.162	0.191	0.016	-0.069	-0.381***	0.004***
	(0.137)	(0.149)	(0.166)	(0.278)	(0.095)	(0.001)
I(Listed)	-0.261***	-0.349***	-0.240***			
	(0.069)	(0.046)	(0.070)			
I(Group Firm)	0.602***	0.668***	0.729***	0.339**	0.015	-0.005***
	(0.045)	(0.052)	(0.099)	(0.144)	(0.062)	(0.001)
I(Export Status)	1.162***	0.998***	1.240***	-0.390**	-0.135*	-0.006***
	(0.057)	(0.059)	(0.059)	(0.159)	(0.075)	(0.001)
Constant	1.402***	1.562***	-3.261***	1.485***	0.818***	0.126***
	(0.103)	(0.073)	(0.100)	(0.312)	(0.148)	(0.005)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	46,809	39,056	68,398	26,901	29,206	28,811
R-squared	0.420	0.409	0.424	0.032	0.181	0.474

Notes: This table presents results for fixed effects regressions of several firm outcomes on board jati homophily and other control variables. The dependent variables are: log(operating income), log(operating cash flow), log(profits), market to book ratio, Tobin's Q, and volatility. Control variables include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. Columns (4)-(6) omit listing status since the samples for those regressions include only listed firms. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Table 11: Varna Homophily and Firm Outcomes: Fixed Effects

Variables	(1) Log (Operating Income)	(2) Log (Operating Cash Flow)	(3) Log (Profits)	(4) Market to Book Ratio	(5) Tobin's Q	(6) Volatility
Varna Homophily	-0.440***	-0.382***	-0.343***	-0.502**	-0.071	0.007***
	(0.090)	(0.072)	(0.079)	(0.234)	(0.112)	(0.001)
Firm Age	-0.003**	-0.002*	-0.001	0.007**	0.001	-0.000***
	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)	(0.000)
Board Size	0.186***	0.180***	0.203***	-0.026	-0.014	-0.001***
	(0.008)	(0.009)	(0.011)	(0.018)	(0.009)	(0.000)
Leverage	-0.113***	-0.105**	0.055	1.118***	1.190***	0.002***
	(0.037)	(0.052)	(0.071)	(0.302)	(0.059)	(0.001)
Assets	0.009***	0.009***	0.010***	0.001	-0.000	-0.000***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Tangibility	0.172	0.198	0.023	-0.040	-0.377***	0.004***
	(0.136)	(0.150)	(0.166)	(0.279)	(0.095)	(0.001)
I(Listed)	-0.265***	-0.343***	-0.241***			
	(0.069)	(0.048)	(0.070)			
I(Group Firm)	0.611***	0.691***	0.740***	0.370**	0.020	-0.005***
	(0.045)	(0.055)	(0.100)	(0.142)	(0.060)	(0.001)
I(Export Status)	1.164***	1.004***	1.241***	-0.378**	-0.133*	-0.006***
	(0.057)	(0.059)	(0.059)	(0.156)	(0.075)	(0.001)
Constant	1.403***	1.424***	-3.291***	1.241***	0.780***	0.127***
	(0.098)	(0.091)	(0.102)	(0.329)	(0.149)	(0.005)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	46,809	39,056	68,398	26,901	29,206	28,811
R-squared	0.420	0.407	0.424	0.031	0.180	0.473

Notes: This table presents results for fixed effects regressions of several firm outcomes on board varna homophily and other control variables. The dependent variables are: log(operating income), log(operating cash flow), log(profits), market to book ratio, Tobin's Q, and volatility. Control variables include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. Columns (4)-(6) omit listing status since the samples for those regressions include only listed firms. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10

5.2 Homophily Negatively Affects Firm Performance

As explained in Section 4, cultural homophily of the board is an endogenous regressor. To examine the causal effects of homophily on firm performance, we instrument for it using the homophily among directors of all firms in the state or industry of the firm. In another specification, we also include the Euclidean distance between the cultural composition of firm boards and of all directors in the state or industry. In a third approach, we exploit board composition changes induced by firms' compliance to the Clause 49 requirement of having at least 50% of the board be constituted by non-executive directors.

Table 12: Instrumental Variable Regressions: First Stage

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	State and Industry Homophily				State and Industry Homophily and Euclidean Distance				Clause 49			
Variables	Jati Homophily		Varna Homophily		Jati Homophily		Varna Homophily		Jati Homophily		Varna Homophily	
State Jati Homophily	0.958***	0.765***			0.677***	0.599***						
	(0.070)	(0.144)			(0.077)	(0.092)						
Industry Jati Homophily	0.711***	0.228			1.163***	0.789***						
	(0.094)	(0.148)			(0.093)	(0.171)						
State Varna Homophily			0.528***	0.504***			0.263***	0.230***				
			(0.044)	(0.069)			(0.043)	(0.080)				
Industry Varna Homophily			0.482***	0.113			1.131***	0.704***				
			(0.076)	(0.168)			(0.040)	(0.147)				
Distance from State Director Jati Composition					0.369***	0.349***						
					(0.063)	(0.057)						
Distance from Industry Director Jati Composition					0.981***	0.853***						
					(0.065)	(0.058)						
Distance from State Director Varna Composition							0.106***	0.107***				
							(0.029)	(0.029)				
Distance from Industry Director Varna Composition							1.011***	0.899***				
							(0.031)	(0.034)				
Clause 49 Based Instrument									-0.011**	-0.020***	-0.015***	-0.017***
Observations	68,318	26,900	68,318	26,900	68,318	26,900	68,318	26,900	32,502	11,543	32,502	11,543
First stage F-statistic	128.25	17.15	106.84	27.95	8943.41	3881.86	8380.19	828.38	10.88	16.98	12.93	8.15

Notes: This table presents first stage results for instrumental variable regressions of log(profits) (columns 1, 3, 5, 7, 9, and 11) and market to book ratio (columns 2, 4, 6, 8, 10, and 12) on board caste (jati and varna) homophily and other control variables. The excluded instruments in columns 1 and 2 (3 and 4) are the *jati* (*varna*) homophily levels of the two director supply pools for a given firm. The supply pools are: the set of directors serving in all firms in the same industry as the given firm, and the set serving in all firms in the same state. The excluded instruments in columns 5 and 6 (7 and 8) are the *jati* (*varna*) homophily levels of the two director supply pools and the distance of *jati* (*varna*) composition of the firm's board from that of the two supply pools. The excluded instrument in columns 9-12 is the product of a variable indicating whether the firm is eligible to comply with Clause 49 and a variable indicating whether it is below the 50% non-executive directors threshold required under the Clause. Control variables (not shown) include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Table 12 presents the first stage results for all three instrumental variable (IV) specifications. There is a different first stage regression equation estimated for each of the six dependent variables we consider since the samples differ somewhat due to missing observations of the dependent variables. For space considerations, we do not show all six first stage equations for all measures of cultural identity across the three IV approaches. Instead, the table presents the first stage results for *jati* and *varna* for the three IV approaches for two dependent variables: log(profits) (columns 1, 3, 5, 7, 9, and 11) and market to book ratio (columns 2, 4, 6, 8, 10, and 12).[32] The table also shows coefficients only for the excluded instruments. The excluded instruments in columns 1 and 2 (3 and 4) are the *jati* (*varna*) homophily levels of the two director supply pools for a given firm. The supply pools are the set of directors serving in all firms in the same industry as the given firm, and the set serving in all firms in the same state. The excluded instruments in columns 5 and 6 (7 and 8) are the *jati* (*varna*) homophily levels of the two director supply pools and the distance of *jati* (*varna*) composition of the firm's board from that of the two supply pools. The excluded instrument in columns 9-12 is the product of a variable indicating whether the firm is eligible to comply with Clause 49 and a variable indicating whether it is below the 50% non-executive directors threshold required under the Clause.

The table shows that homophily of directors at the state and industry level are strongly positively associated with an average firm's board homophily. Greater distance from the state/industry caste composition of directors is also associated positively with an average firm's board homophily. Finally, the estimated coefficients on the instrument exploiting Clause 49 indicate that board membership changes induced by firms complying with the new requirements reduced caste homophily on an average firm's board. Except in column 12, all first stage F-statistics are well above 10, indicating that the instruments explain a significant proportion of the variation in the endogenous regressor, the homophily of firm boards.

Table 13 and 14 present second stage results for the first IV approach where the excluded instruments are the *jati* and *varna* homophily of the set of directors in the same two-digit industry or state as the firm. The estimated coefficients on *jati* and *varna* homophily show that higher homophily in boards leads to statistically significant declines in key measures of firm performance and firm value. A one unit increase in *jati* homophily leads to over 3 log points drop in all three balance sheet performance measures - operating income, operating cash flow and profits. Additionally, a one unit increase in *jati* homophily reduces an average firm's market to book ratio by nearly 9 points

[32]Other results are available upon request.

and Tobin's Q by 3 points. Volatility also increases, but the estimate is statistically insignificant. Increases in *varna* homophily similarly cause declines in firm performance and value, but note that all coefficients are smaller in absolute value. Comparing these results to those for religion homophily in boards (Table Appendix D.2), we find that changes in religion homophily does not have as large, consistent, or statistically significant effects on firm performance and value.

Next, we consider results from our second instrumental variable approach, in which the excluded instruments are the homophily of state and industry level director supply pools as well as the distance between the cultural composition of the supply pools and that of individual boards. Results are presented in Tables 15 and 16 for *jati* and *varna* homophily and in Table Appendix D.3 for religion homophily. We find similar results as in the first instrumental variable strategy. Specifically, we see that *jati* homophily reduces firm balance sheet performance and market value indicators by large and statistically significant magnitudes. Firm volatility also increases significantly due to an increase in *jati* homophily. *Varna* homophily also worsens firm outcomes to a slightly smaller extent. As before, religion homophily does not appear to affect firm value and volatility. However, firm balance sheet performance indicators increase with an increase in religion homophily.

Finally, we discuss results from the third IV approach, which exploits board composition changes resulting from firms' compliance with Clause 49 requirements. We again find that firm performance as measured by all three balance sheet variables worsens due to increases in *jati* and *varna* homophily. However, our coefficients appear inordinately large. Volatility also increases significantly when caste homophily increases. However, inconsistent with our previous findings, we see that market to book ratio and Tobin's Q increase.

On the basis of all our regression results, we conclude that lack of diversity has negative effects on key firm outcomes. An understanding of how important diversity may be for firms, however, depends on the granularity with which diversity is measured. Our results demonstrate the strongest negative effects when diversity along our narrowest measure of cultural identity is low.

As a final point, consider the mechanisms through which diversity in boards matters for firms. Directors with diverse backgrounds may enable the board as a whole to access larger amounts of information, making them more capable of experimentation and solving complex problems, thereby improving firm productivity. This is the *advisory channel*. Next, more diverse boards may be better able to monitor the management. A better monitored firm is less fraudulent and performs better. This is the *monitoring channel*.

While advising and monitoring may improve with board diversity, they may also worsen due to higher levels of interpersonal conflict in interactions of diverse directors leading to poor firm performance. We call this *board frictions channel*. Evidence shows that diversity can lead to increased conflicts within groups (OReilly et al. (1993);Smith et al. (1994)).

Unfortunately, Prowess does not provide us with data to test the presence or strength of these mechanisms. To test whether diverse boards take riskier projects or engender more innovative outcomes, studies have used measures of quantity and quality of firm innovation like number of patents, ratio of patents to R&D, patent citations etc. Prowess does not provide most of this information except for firms' annual R&D expenditures. But even this variable is sparsely populated. Of the firms in our sample, 92% do not have information on R&D rendering any meaningful analysis impossible. Literature typically measures the monitoring function of the board by looking at CEO compensation sensitivity to firm performance. In our data, in over 20,000 firms and CEOs, only 181 CEOs resign of which only 8 resign after poor firm performance (measured as firm sales below average of the industry in the past one or three years). Using our data, we are able to observe the relation between board diversity and several measures of frictions within the board: frequency of board meetings, attendance, director resignations, and board turnover. We see some evidence here that diversity increases frictions in the boardroom – more diverse boards meet less often, have lower attendance at meetings, see greater incidence of resignations and have more turnover. However, these associations are also based on a small subset of firms for which these data are available. So we are unable to make any conclusive assessments.

Table 13: Jati Homophily and Firm Outcomes: Second Stage, IV Approach 1

Variables	(1) Log (Operating Income)	(2) Log (Operating Cash Flow)	(3) Log (Profits)	(4) Market to Book Ratio	(5) Tobin's Q	(6) Volatility
Jati Homophily	-3.003***	-3.158***	-3.563***	-8.815***	-3.387***	0.008
	(0.714)	(0.757)	(0.547)	(2.827)	(0.999)	(0.018)
Firm Age	-0.004***	-0.002*	-0.002*	0.005	0.000	-0.000***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.000)
Board Size	0.092***	0.090***	0.080***	-0.213***	-0.089***	-0.001***
	(0.025)	(0.027)	(0.021)	(0.066)	(0.022)	(0.000)
Leverage	-0.130***	-0.131**	0.028	0.943***	1.150***	0.002***
	(0.044)	(0.052)	(0.072)	(0.296)	(0.058)	(0.001)
Assets	0.009***	0.009***	0.010***	0.000	-0.000	-0.000***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Tangibility	0.136	0.166	-0.015	-0.234	-0.426***	0.004***
	(0.152)	(0.150)	(0.180)	(0.320)	(0.099)	(0.001)
I(Listed)	-0.431***	-0.465***	-0.445***			
	(0.093)	(0.064)	(0.094)			
I(Group Firm)	0.385***	0.483***	0.431***	-0.047	-0.146	-0.005***
	(0.081)	(0.085)	(0.109)	(0.223)	(0.094)	(0.001)
I(Export Status)	1.082***	0.924***	1.153***	-0.602***	-0.225**	-0.006***
	(0.052)	(0.056)	(0.057)	(0.201)	(0.092)	(0.001)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	46,753	39,025	68,318	26,900	29,204	28,809
R-squared	0.362	0.363	0.334	-0.073	0.091	0.474

Notes: This table presents second stage results from instrumental variable regressions of several firm outcomes on board jati homophily and other control variables.The dependent variables are: log(operating income), log(operating cash flow), log(profits), market to book ratio, Tobin's Q, and volatility. The excluded instruments are the jati homophily levels of two director supply pools for a given firm. The supply pools are: the set of directors serving in all firms in the same industry as the given firm, and the set serving in all firms in the same state. Control variables include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. Columns (4)-(6) omit listing status since the samples for those regressions include only listed firms. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Table 14: Varna Homophily and Firm Outcomes: Second Stage, IV Approach 1

Variables	(1) Log (Operating Income)	(2) Log (Operating Cash Flow)	(3) Log (Profits)	(4) Market to Book Ratio	(5) Tobin's Q	(6) Volatility
Varna Homophily	-1.056	-1.252*	-2.294***	-7.896***	-2.239***	0.005
	(0.885)	(0.666)	(0.774)	(1.975)	(0.548)	(0.009)
Firm Age	-0.003**	-0.002*	-0.001	0.005	0.001	-0.000***
	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)	(0.000)
Board Size	0.169***	0.158***	0.147***	-0.155***	-0.052***	-0.001***
	(0.027)	(0.016)	(0.021)	(0.039)	(0.015)	(0.000)
Leverage	-0.111***	-0.106**	0.053	1.077***	1.163***	0.002***
	(0.037)	(0.051)	(0.072)	(0.288)	(0.061)	(0.001)
Assets	0.009***	0.009***	0.010***	0.000	-0.000	-0.000***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Tangibility	0.177	0.196	0.036	0.086	-0.333***	0.004***
	(0.138)	(0.150)	(0.172)	(0.356)	(0.100)	(0.001)
I(Listed)	-0.315***	-0.396***	-0.398***			
	(0.083)	(0.064)	(0.073)			
I(Group Firm)	0.568***	0.640***	0.582***	0.124	-0.053	-0.005***
	(0.075)	(0.058)	(0.082)	(0.193)	(0.065)	(0.001)
I(Export Status)	1.145***	0.979***	1.185***	-0.571***	-0.187**	-0.006***
	(0.048)	(0.056)	(0.063)	(0.189)	(0.077)	(0.001)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	46,753	39,025	68,318	26,900	29,204	28,809
R-squared	0.417	0.401	0.389	-0.082	0.132	0.473

Notes: This table presents second stage results from instrumental variable regressions of several firm outcomes on board varna homophily and other control variables.The dependent variables are: log(operating income), log(operating cash flow), log(profits), market to book ratio, Tobin's Q, and volatility. The excluded instruments are the varna homophily levels of two director supply pools for a given firm. The supply pools are: the set of directors serving in all firms in the same industry as the given firm, and the set serving in all firms in the same state. Control variables include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. Columns (4)-(6) omit listing status since the samples for those regressions include only listed firms. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Table 15: Jati Homophily and Firm Outcomes: Second Stage, IV Approach 2

Variables	(1) Log (Operating Income)	(2) Log (Operating Cash Flow)	(3) Log (Profits)	(4) Market to Book Ratio	(5) Tobin's Q	(6) Volatility
Jati Homophily	-0.609***	-0.748***	-0.497***	-0.882**	-0.137	0.010***
	(0.106)	(0.103)	(0.096)	(0.413)	(0.140)	(0.002)
Firm Age	-0.003**	-0.002*	-0.001	0.007**	0.001	-0.000***
	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)	(0.000)
Board Size	0.177***	0.166***	0.195***	-0.037*	-0.016*	-0.001***
	(0.009)	(0.009)	(0.011)	(0.019)	(0.009)	(0.000)
Leverage	-0.113***	-0.108**	0.055	1.103***	1.190***	0.002***
	(0.039)	(0.051)	(0.070)	(0.300)	(0.059)	(0.001)
Assets	0.009***	0.009***	0.010***	0.001	-0.000	-0.000***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Tangibility	0.158	0.186	0.014	-0.067	-0.381***	0.004***
	(0.138)	(0.149)	(0.165)	(0.275)	(0.094)	(0.001)
I(Listed)	-0.271***	-0.355***	-0.246***			
	(0.067)	(0.046)	(0.069)			
I(Group Firm)	0.590***	0.660***	0.720***	0.343**	0.015	-0.005***
	(0.043)	(0.052)	(0.100)	(0.145)	(0.063)	(0.001)
I(Export Status)	1.157***	0.995***	1.237***	-0.388**	-0.135*	-0.005***
	(0.057)	(0.059)	(0.059)	(0.159)	(0.075)	(0.001)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	46,753	39,025	68,318	26,900	29,204	28,809
R-squared	0.420	0.409	0.424	0.032	0.181	0.474

Notes: This table presents second stage results from instrumental variable regressions of several firm outcomes on board jati homophily and other control variables. The dependent variables are: log(operating income), log(operating cash flow), log(profits), market to book ratio, Tobin's Q, and volatility. The excluded instruments are the jati homophily levels of two director supply pools for a given firm, and the Euclidean distances of the firm's board jati composition from those of the two supply pools. The supply pools are: the set of directors serving in all firms in the same industry as the given firm, and the set serving in all firms in the same state. Control variables include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. Columns (4)-(6) omit listing status since the samples for those regressions include only listed firms. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Table 16: Varna Homophily and Firm Outcomes: Second Stage, IV Approach 2

Variables	(1) Log (Operating Income)	(2) Log (Operating Cash Flow)	(3) Log (Profits)	(4) Market to Book Ratio	(5) Tobin's Q	(6) Volatility
Varna Homophily	-0.670***	-0.570***	-0.501***	-0.503*	-0.055	0.008***
	(0.106)	(0.089)	(0.092)	(0.262)	(0.111)	(0.001)
Firm Age	-0.003**	-0.002*	-0.001	0.007**	0.001	-0.000***
	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)	(0.000)
Board Size	0.180***	0.175***	0.199***	-0.026	-0.014	-0.001***
	(0.009)	(0.009)	(0.011)	(0.019)	(0.009)	(0.000)
Leverage	-0.110***	-0.104**	0.057	1.118***	1.191***	0.002***
	(0.037)	(0.052)	(0.070)	(0.299)	(0.059)	(0.001)
Assets	0.009***	0.009***	0.010***	0.001	-0.000	-0.000***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Tangibility	0.172	0.194	0.022	-0.040	-0.378***	0.004***
	(0.136)	(0.151)	(0.165)	(0.277)	(0.094)	(0.001)
I(Listed)	-0.284***	-0.355***	-0.254***			
	(0.067)	(0.048)	(0.069)			
I(Group Firm)	0.595***	0.681***	0.726***	0.370***	0.020	-0.005***
	(0.044)	(0.054)	(0.100)	(0.140)	(0.060)	(0.001)
I(Export Status)	1.157***	0.999***	1.237***	-0.378**	-0.133*	-0.006***
	(0.057)	(0.058)	(0.059)	(0.156)	(0.075)	(0.001)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	46,753	39,025	68,318	26,900	29,204	28,809
R-squared	0.420	0.407	0.424	0.031	0.180	0.473

Notes: This table presents second stage results from instrumental variable regressions of several firm outcomes on board varna homophily and other control variables. The dependent variables are: log(operating income), log(operating cash flow), log(profits), market to book ratio, Tobin's Q, and volatility. The excluded instruments are the varna homophily levels of two director supply pools for a given firm, and the Euclidean distances of the firm's board varna composition from those of the two supply pools. The supply pools are: the set of directors serving in all firms in the same industry as the given firm, and the set serving in all firms in the same state. Control variables include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. Columns (4)-(6) omit listing status since the samples for those regressions include only listed firms. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Table 17: Jati Homophily and Firm Outcomes: Second Stage, IV Approach 3

Variables	(1) Log (Operating Income)	(2) Log (Operating Cash Flow)	(3) Log (Profits)	(4) Market to Book Ratio	(5) Tobin's Q	(6) Volatility
Jati Homophily	-48.974**	-60.531	-48.753***	11.377**	4.774**	0.548***
	(22.796)	(38.532)	(18.785)	(5.792)	(1.958)	(0.185)
Firm Age	-0.009	0.003	-0.007	0.012***	0.001	0.000
	(0.009)	(0.011)	(0.008)	(0.004)	(0.002)	(0.000)
Board Size	-1.566*	-1.717	-1.646**	0.270*	0.116**	0.011**
	(0.824)	(1.217)	(0.718)	(0.146)	(0.052)	(0.005)
Leverage	-0.836*	-0.970	-0.656*	1.969***	0.997***	0.014***
	(0.442)	(0.691)	(0.395)	(0.408)	(0.050)	(0.005)
Assets	0.016***	0.012***	0.016***	0.002***	0.001	-0.000***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.000)
Tangibility	-0.562	-1.043	-0.705	-0.328	-0.102	0.013
	(0.852)	(1.093)	(0.719)	(0.441)	(0.137)	(0.011)
I(Listed)	-2.847**	-1.985*	-2.488***			
	(1.238)	(1.139)	(0.949)			
I(Group Firm)	-3.561*	-3.708	-3.657**	1.075***	0.342***	0.020*
	(1.941)	(2.886)	(1.734)	(0.350)	(0.125)	(0.011)
I(Export Status)	-0.410	-0.930	0.014	0.323	0.139	0.007
	(0.781)	(1.297)	(0.517)	(0.228)	(0.089)	(0.007)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	22,381	17,737	32,502	11,543	12,750	12,494

Notes: This table presents second stage results from instrumental variable regressions of several firm outcomes on board jati homophily and other control variables. The dependent variables are: log(operating income), log(operating cash flow), log(profits), market to book ratio, Tobin's Q, and volatility. The excluded instrument is the product of indicators for whether a firm is required to comply with Clause 49 in a year and whether it is above or below the 50% threshold for non-executive directors. Control variables include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. Columns (4)-(6) omit listing status since the samples for those regressions include only listed firms. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Table 18: Varna Homophily and Firm Outcomes: Second Stage, IV Approach 3

Variables	(1) Log (Operating Income)	(2) Log (Operating Cash Flow)	(3) Log (Profits)	(4) Market to Book Ratio	(5) Tobin's Q	(6) Volatility
Varna Homophily	-32.694***	-39.489**	-35.550***	13.789**	5.988**	0.703**
	(12.077)	(17.196)	(9.271)	(6.960)	(3.052)	(0.277)
Firm Age	-0.008	0.001	-0.006	0.013***	0.002	0.000
	(0.006)	(0.007)	(0.005)	(0.004)	(0.002)	(0.000)
Board Size	-0.726**	-0.755*	-0.809***	0.247*	0.109*	0.011**
	(0.349)	(0.415)	(0.277)	(0.129)	(0.058)	(0.005)
Leverage	-0.447**	-0.453**	-0.222	1.691***	1.018***	0.016***
	(0.201)	(0.230)	(0.212)	(0.464)	(0.057)	(0.006)
Assets	0.011***	0.010***	0.013***	0.002***	0.001	-0.000
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.000)
Tangibility	0.542	-0.079	0.078	-0.695	-0.269*	-0.008
	(0.658)	(0.687)	(0.573)	(0.496)	(0.154)	(0.012)
I(Listed)	-2.561***	-2.139***	-2.623***			
	(0.815)	(0.792)	(0.623)			
I(Group Firm)	-1.720**	-1.535	-2.000***	0.966***	0.298**	0.015
	(0.873)	(1.031)	(0.735)	(0.281)	(0.130)	(0.012)
I(Export Status)	0.121	-0.376	0.201	0.449	0.180	0.013
	(0.401)	(0.600)	(0.299)	(0.329)	(0.128)	(0.010)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	22,381	17,737	32,502	11,543	12,750	12,494

Notes: This table presents second stage results from instrumental variable regressions of several firm outcomes on board varna homophily and other control variables. The dependent variables are: log(operating income), log(operating cash flow), log(profits), market to book ratio, Tobin's Q, and volatility. The excluded instrument is the product of indicators for whether a firm is required to comply with Clause 49 in a year and whether it is above or below the 50% threshold for non-executive directors. Control variables include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. Columns (4)-(6) omit listing status since the samples for those regressions include only listed firms. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10

6 Mechanisms

Results in Section 5 demonstrate that firm performance is worsened by lack of diversity on boards. In this section we explore the mechanisms through which diversity affects firm performance. Given data limitations, we are only able to make some conjectures and provide suggestive evidence.

A diverse board may perform its advisory role better. Directors with diverse backgrounds may be able to access to more information, enabling experimentation and complex problem solving. To test whether diverse boards encourage management to undertake riskier projects or engender more innovative outcomes, studies have used measures such as number of patents, ratio of patents to R&D, and patent citations. However, we

are unable to examine similar indicators since Prowess does not provide most of this information except for firms' annual R&D expenditures. But even this variable is sparsely populated. Of the firms in our sample, 92% do not have information on R&D, rendering any meaningful analysis impossible.

More diverse boards may also be better able to monitor the management. Better monitored managers may be less fraudulent and perform better. The literature typically measures the monitoring function of the board by looking at CEO turnover and compensation sensitivity to firm performance. In our data, of over 20,000 firms and CEOs, only 181 CEOs resign, of which only 8 resign after poor firm performance (measured as firm sales below average of the industry in the past one or three years). We also do not have compensation data for the majority of CEOs in our sample.

Diversity (or lack of it) among directors on a board also affects boardroom group dynamics, some indicators of which we can observe in our data. We expect these dynamics to, in turn, affect firm performance. On one hand, culturally homogeneous directors may get along better and have fewer conflicts (OReilly et al. (1993); Smith et al. (1994)). These are positive effects and should improve decision making. However, negative consequences are also possible if directors with shared caste identity engage in cronyism, meeting often as friends, or those belonging to caste groups with large shares on the board getting disproportionate importance. We present our findings below.

We consider board meeting attendance. As Figure 14 shows, we find that boards with high homophily have higher meeting attendance rates. In Figure 15a, we also find that directors that share their *jati* with the dominant *jati* of the board have greater attendance than the directors that do not. Figure 15b indicates that the difference in attendance of the dominant and non dominant *jati* directors increases with boards homophily. We take these data patterns as suggestive of directors belonging to dominant castes being more willing to attend meetings since they may be friendly with others of their own caste and may get along better.

We also find that boards where at least one director has resigned have lower homophily than the ones where no director has resigned (Figure 16(a)). Also, firms with higher homophily see a smaller fraction of directors resigning (Figure 16(b)). Both indicate that more homogeneous boards continue to serve with their culturally proximate peers, suggestive of cronyism.

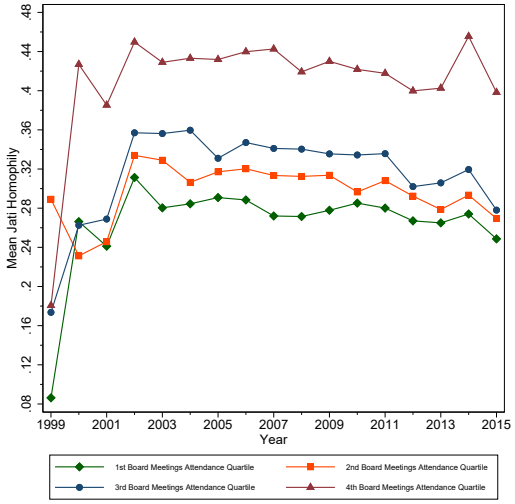
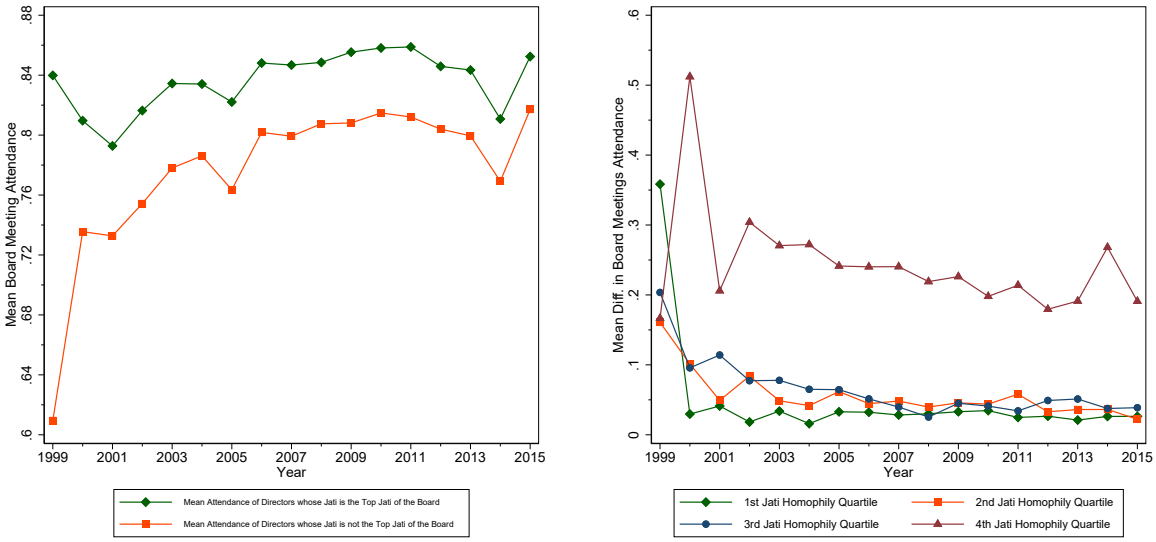


Figure 14: Jati Homophily and Attendance of Board Meetings[a]

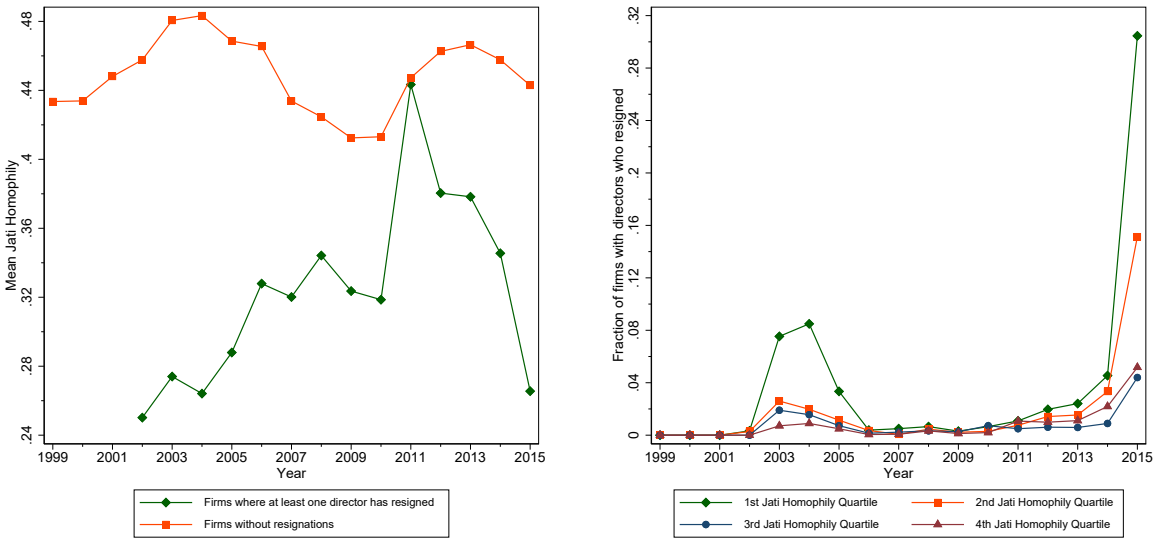
[a]Notes: Source: Prowess, authors' last name to religion *jati* mapping using matrimonial data. Homophily is averaged over all firms in each quartile of the attendance of Board meetings distribution. Attendance of board meetings in a year is calculated as the average attendance of board members across all board meetings of a board in a year.



(a) (b)

Figure 15: Meeting Attendance by Jati [a]

[a]Notes: Source: Prowess, authors' last name to religion*(jati)* mapping using matrimonial data. In Figure (a) mean board meeting attendance is averaged over all directors with same top jati as the board and ones that do not. In Figure (b), difference of board meetings attendance between directors with same top jati as the board and ones the ones that do not, is averaged over all firms in each quartile of the *jati* homophily of the board.



(a) (b)

Figure 16: Jati Homophily and Resignation of Board Members [a]

[a]Notes: Source: Prowess, authors' last name to religion*(jati)* mapping using matrimonial data. In Figure 18(a) *jati* homophily is averaged over all firms where at least one director resigned and ones where no director resigned.

Next, we assess if directors of castes with large representations on boards serve disproportionately on important sub-committees of directors. This would be indicative of cronyism. Table 19 shows this to be indeed the case. Panel A is for *jati* and Panel B is for *varna*. In both panels, the first row shows the percentage share of directors that belong to the most, second most, and third most dominant caste on the board, averaged over all firms for the sample period. The subsequent rows depict if these castes, by their dominance status, are represented disproportionately in important roles on the board. For example, if we consider the chairs of all committees, we see that the fraction of committee chairs that belong to the most dominant *jati* relative to the share of directors on the full board that belong to this *jati* is 0.99. Since the fraction is not greater than 1, this indicates that, on average, dominant caste directors do not hold disproportionately more committee chair positions. However, this fraction is greater than 1 for the second and third dominant *jatis*, indicating that these *jatis* are over-represented in chair positions. We find a similar over-representation when we consider a few important committees in particular – audit, remuneration, and shareholder grievance.

Table 19: Committee Chair Proportions by Caste Status on Board [a]

Panel A: Jati

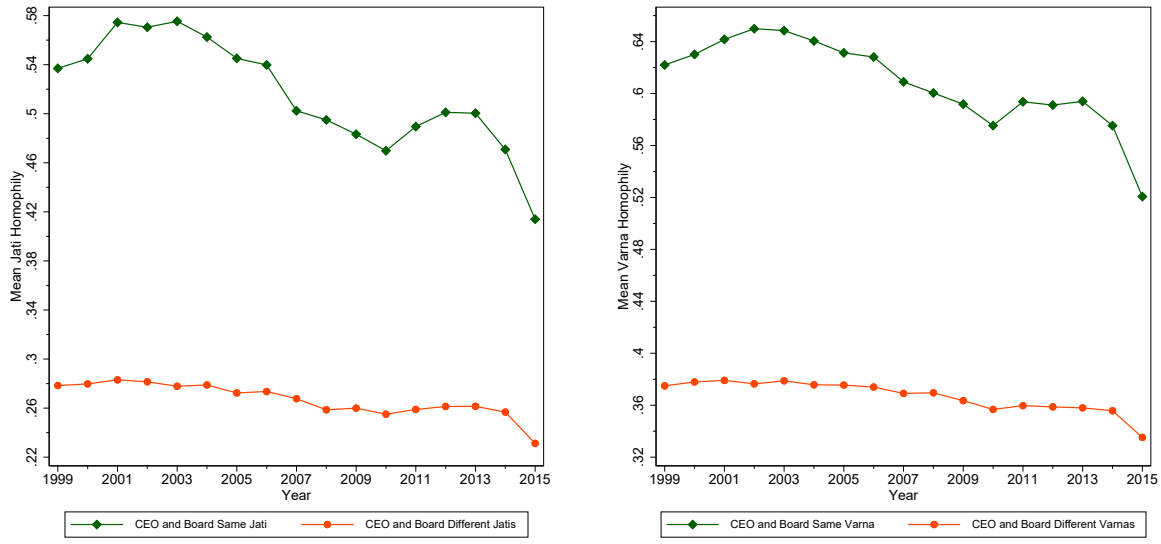
	Dominant Jati	Second Most Dominant Jati	Third Most Dominant Jati
% of directors in nth dominant jati of board (denominator for subsequent rows)	52.29	22.15	11.93
% of new directors from nth dominant jati on board	0.96	0.85	0.87
% of committee chairs from nth dominant jati of board	0.99	1.29	1.32
% of audit committee members from nth dominant jati of board	0.88	1.11	1.15
% of remuneration committee members from nth dominant jati of board	0.84	1.17	1.22
% of shareholder grievance committee members from nth dominant jati of board	1.02	1.02	1.03

Panel B: Varna

	Dominant Varna	Second Most Dominant Varna	Third Most Dominant Varna
% of directors in nth dominant varna of board (denominator for subsequent rows)	63.43	22.28	8.48
% of new directors from nth dominant varna on board	0.87	1.03	1.19
% of committee chairs from nth dominant varna of board	0.97	1.15	1.16
% of audit committee members from nth dominant varna of board	0.93	1.1	1.13
% of remuneration committee members from nth dominant varna of board	0.91	1.13	1.17
% of shareholder grievance committee members from nth dominant varna of board	1.01	1.01	1

[a]Notes: Source: Prowess, matrimonial data. The table shows, by dominance status of a caste, the proportions of particular positions occupied by directors of that caste, relative to their share in the full board. A fraction greater than 1 indicates over-representation.

Finally, we also observe that high homophily firms are also more likely to have the CEO belong to the same caste as the dominant caste represented among the directors. This is evident in Figure 17 which shows that firms that have the CEO belonging to the same caste (*jati* or *varna*) as the dominant caste of the rest of the board also have less caste diversity on their boards. This is strongly suggestive of crony behavior of the majority caste group on the board.



(a) Jati

(b) Varna

Figure 17: Homophily by CEO Caste Relative to Dominant Caste of Board[a]

[a]Notes: Source: Prowess, matrimonial data. Homophily (Blau index) is averaged over all firms in a year that fall into either of two groups: those that have their CEO belonging to the same caste as the dominant caste in the rest of the board, and those where the CEO belongs to a different caste.

7 Conclusion

We build a unique dataset that allows us to map Indian last names to religion and caste. Combining these data with data on large public and private firms in India, we show that boards of directors of large Indian firms are characterized by persistently low levels of cultural diversity. This is the case at both coarse (religion) and really fine (*jati*) definitions of cultural identity. Rigorous analysis demonstrates that such lack of diversity has a detrimental effect on key measures of firm performance.

Our findings highlight the importance of cultural diversity on corporate boards. Norway in 2008, several other European countries in the following years, and California in 2019, passed laws requiring firms to have women on their boards. Similarly, in India, a new act in 2013 made it mandatory for the top 1000 listed firms to have at least one female director. However, this paper informs us that diversity along traits beyond gender may be beneficial to firms. Indeed, in the United Kingdom, the Financial Reporting Council's 2018 guidelines on board effectiveness recognizes that diversity should include ethnicity, background, and personal attributes including judgment, courage, and the ability to listen, forge relationships, and develop trust. Many of these attributes are likely

to vary with people's cultural identity and backgrounds, and in the Indian context, with their caste identity.

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.

Ajit, D., Donker, H., and Saxena, R. (2012). Corporate boards in india: blocked by caste? *Economic and Political Weekly*, pages 39–43.

Alesina, A., Baqir, R., and Easterly, W. (1999). Public Goods and Ethnic Divisions. *The Quarterly Journal of Economics*, 114(4):1243–1284.

Alesina, A. and Giuliano, P. (2010). The Power of the Family. *Journal of Economic growth*, 15(2):93–125.

Alesina, A. and Giuliano, P. (2011). Family Ties and Political Participation. *Journal of the European Economic Association*, 9(5):817–839.

Alesina, A. and Giuliano, P. (2014). Family Ties. In *Handbook of economic growth*, volume 2, pages 177–215. Elsevier.

Alesina, A., Giuliano, P., and Nunn, N. (2013). On the Origins of Gender Roles: Women and the Plough. *The Quarterly Journal of Economics*, 128(2):469–530.

Alesina, A., Michalopoulos, S., and Papaioannou, E. (2016). Ethnic Inequality. *Journal of Political Economy*, 124(2):428–488.

Alesina, A., Reich, B., and Riboni, A. (2017). Nation-building, nationalism and wars. Technical report, National Bureau of Economic Research.

Anderson, S. (2011). Caste as an impediment to trade. *American Economic Journal: Applied Economics*, 3(1):239–63.

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.

Ashraf, Q., Galor, O., et al. (2007). *Cultural Assimilation, Cultural Diffusion and the Origin of the Wealth of Nations*. Centre for Economic Policy Research.

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.

Atkin, D. (2016). The caloric costs of culture: Evidence from indian migrants. *American Economic Review*, 106(4):1144–81.

Banerjee, A. and Munshi, K. (2004). How efficiently is capital allocated? evidence from the knitted garment industry in tirupur. *The Review of Economic Studies*, 71(1):19–42.

Bengtsson, O. and Hsu, D. H. (2015). Ethnic matching in the us venture capital market. *Journal of Business Venturing*, 30(2):338–354.

Benjamin, D. J., Choi, J. J., and Fisher, G. W. (2010). Religious Identity and Economic Behavior. Technical report, National Bureau of Economic Research.

Bernile, G., Bhagwat, V., and Yonker, S. (2018). Board Diversity, Firm Risk, and Corporate Policies. *Journal of Financial Economics*, 127(3):588–612.

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.

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.

Biswas, S. et al. (2016). Promoter homophily in boards: Does it really matter?-an analysis of indian firms. *Economics Bulletin*, 36(1):237–252.

Bloom, N., Liang, J., Roberts, J., and Ying, Z. J. (2014). Does Working from Home Work? Evidence from a Chinese Experiment. *The Quarterly Journal of Economics*, 130(1):165–218.

Bloom, N., Sadun, R., and Van Reenen, J. (2012). The Organization of Firms Across Countries. *The quarterly journal of economics*, 127(4):1663–1705.

Campante, F. and Yanagizawa-Drott, D. (2015). Does Religion Affect Economic Growth and Happiness? Evidence from Ramadan. *The Quarterly Journal of Economics*, 130(2):615–658.

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.

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.

De Luca, G., Hodler, R., Raschky, P. A., and Valsecchi, M. (2018). Ethnic favoritism: An axiom of politics? *Journal of Development Economics*, 132:115–129.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2):158–165.

Desmet, K., Weber, S., and Ortuño-Ortín, I. (2009). Linguistic diversity and redistribution. *Journal of the European Economic Association*, 7(6):1291–1318.

Dharmapala, D. and Khanna, V. (2012). Corporate governance, enforcement, and firm value: Evidence from india. *The Journal of Law, Economics, & Organization*, 29(5):1056–1084.

Dumont, L. (1980). *Homo Hierarchicus: The Caste System and Its Implications*. University of Chicago Press.

Easterly, W. and Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. *The quarterly journal of economics*, 112(4):1203–1250.

Esteban, J., Mayoral, L., and Ray, D. (2012). Ethnicity and conflict: An empirical study. *American Economic Review*, 102(4):1310–42.

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.

Fernandez, R. (2007). Women, Work, and Culture. *Journal of the European Economic Association*, 5(2-3):305–332.

Fernández, R. (2011). Does Culture Matter? In *Handbook of social economics*, volume 1, pages 481–510. Elsevier.

Fernández, R. (2013). Cultural Change as Learning: The Evolution of Female Labor Force Participation over a Century. *American Economic Review*, 103(1):472–500.

Fernández, R. and Fogli, A. (2006). Fertility: The Role of Culture and Family Experience. *Journal of the European Economic Association*, 4(2-3):552–561.

Fernandez, R. and Fogli, A. (2009). Culture: An Empirical Investigation of Beliefs, Work, and Fertility. *American Economic Journal: Macroeconomics*, 1(1):146–77.

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.

Ghani, E., Kerr, W. R., and O'connell, S. (2014). Spatial Determinants of Entrepreneurship in India. *Regional Studies*, 48(6):1071–1089.

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.

Gorodnichenko, Y. and Roland, G. (2015). Culture, institutions and democratization. Technical report, National Bureau of Economic Research.

Guiso, L., Sapienza, P., and Zingales, L. (2003). People's Opium? Religion and Economic Attitudes. *Journal of monetary economics*, 50(1):225–282.

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*.

Herfindahl, O. C. (1950). *Concentration in the Steel Industry*. PhD thesis, Columbia University New York.

Hirschman, A. O. (1945). *National Power and the Structure of Foreign Trade*. Univ of California Press.

Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics*, 129(4):1899–1946.

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., Somanathan, R., et al. (2017). Caste connections and government transfers: The mahadalits of bihar. Technical report.

Macdonell, A. A. (1914). The early history of caste. *The American Historical Review*, 19(2):230–244.

McCleary, R. and Barro, R. (2003). Religion and Economic Growth Across Countries.

McCleary, R. M. and Barro, R. J. (2006). Religion and Economy. *Journal of Economic perspectives*, 20(2):49–72.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444.

Munshi, K. and Rosenzweig, M. (2006). Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *American Economic Review*, 96(4):1225–1252.

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.

Noland, M. (2005). Religion and Economic Performance. *World development*, 33(8):1215–1232.

OReilly, C., Snyder, R., and Boothe, J. (1993). Effects of executive team demography on organizational change. *Organizational change and redesign*, pages 147–175.

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.

Smith, K. G., Smith, K. A., Olian, J. D., Sims Jr, H. P., O'Bannon, D. P., and Scully, J. A. (1994). Top management team demography and process: The role of social integration and communication. *Administrative science quarterly*, pages 412–438.

Srinivas, M. N. (1995). *Social Change in Modern India*. Orient Blackswan.

Tabellini, G. (2010). Culture and Institutions: Economic Development in the Regions of Europe. *Journal of the European Economic association*, 8(4):677–716.

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.

Titzmann, F.-M. (2013). Changing patterns of matchmaking: The indian online matrimonial market. *Asian Journal of Women's Studies*, 19(4):64–94.

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.

Appendices

Appendix A Other Sources for Caste Association

We compare the last name to caste mapping introduced in this paper with other data sources available on caste identity. There are few datasets that record self reported or otherwise castes – *varnas* and *jatis* of respondents. 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.

Records of lists of communities in the colonial period have been made on an extensive scale since 1806. The process gathered momentum over the Censuses conducted between 1881 and 1941. 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 mentioned. While comparing our mapping with PoI data, we consider each main community to be a unique *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 like "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 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.

Appendix B Variable Definitions

Table Appendix B.1: Variable definitions

Variables	Definitions
Panel A: Firm Variables	
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 digit NIC-2008 sector
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)
Operating cash flow	Cash flow from operating activities before depreciation
Leverage	Book value of debt over book value of total assets
Operating income	Sales less operating expenses
Tangibility	Net Property plant 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 (Latest) b. the mean market value over the entire reporting period (Full Period)
Market to book ratio	Market price per share/book value per share. The market to book ratio is measured in two ways - a. the latest market to book ratio available on or before the reporting date (Latest) b. the mean market to book ratio over the entire reporting period (Full Period)
Volatility	The standard deviation of stock returns of a firm in the entire reporting period
Cumulative Abnormal Returns (CAR)	The difference between the return on the stock over the announcement window and the corresponding return on the market index for the firms who participated in M&As as acquirers

Panel B: Board Characteristics

Board size	Number of directors in the board
Frequency of board meetings	Number of board meetings per year
Board meeting attendance	Mean number of board meetings attended by all members of a board over total number of board meetings
Resignation	Indicator variable: one if a board member resigns, zero otherwise
Board turnover 1Y	% of directors in a board who were not present in the previous year
Board turnover 3Y	% of directors in a board who were not present in the board three years prior to the current year
CEO duality	Indicator variable: one if the at least one CEO of a firm is also the chair.
Clause 49 based instrument	Indicator variable: I(Eligibility) $\times$ I(Below Threshold % of Non-Executive Directors)

Panel C: Measures of Cultural Homophily

Dominant *jati (varna, religion)* of a board	The *jati (varna, religion)* of the maximum number of directors of a board. In case of ties, dominant *jati (varna,religion)* is chosen randomly from the tie
Board *jati (varna, religion)* homophily	*jati (varna, religion)* HHI, i.e., sum of squared shares of all *jatis (varnas, religions)* represented on the board.

Sector *jati (varna, religion)* homophily	*jati (varna, religion)* HHI, i.e., sum of squared shares of all *jatis (varnas, religions)* represented 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 *jati (varna, religion)* homophily	*jati (varna, religion)* HHI, i.e., sum of squared shares of all *jatis (varnas, religions)* represented 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 *jati (varna, religion)* euclidean distance of a board	Distance between the vector representing the *jati (varna, 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 *jati (varna, religion)* euclidean distance of a board	Distance between the vector representing the *jati (varna, religion)* composition of directors in the state 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.

Appendix C Fuzzy Blau Index

Fuzzy Blau Index is a more nuanced measure of cultural homophily of a board than the Blau index. It incorporates the full probabilistic mapping of religion, *varna* and *jati* associated with every director of the board. Consider a board j with N members. Each director i is associated with a mapping of its last name to a probability distribution over k unique cultural categories. Thus, each director i is identified with a probability distribution,

$$p_i = (p_{im})_{m=1}^{k}$$

where p_{im} is the probability that director i is of cultural category m.

If each director is associated with a unique cultural identity then the probability asso-

Table Appendix C.1: Fuzzy Blau Examples

		Board A			
Member	Dalit	Brahmin	Kshatriya	Vaishya	Shudra
B1	0	1	0	0	0
B2	0	1	0	0	0
B3	1	0	0	0	0

		Board B			
Member	Dalit	Brahmin	Kshatriya	Vaishya	Shudra
B1	0	1	0	0	
B2	0	1	0	0	
B3	0	1	0	0	

ciated with each board member is degenerate. Consider the examples in Table Appendix C.1, Board A has three members, i.e. $N = 3$ and each of them is associated with a probability distribution over five *varnas*, i.e. $k = 5$. Two directors are uniquely identified as *Brahmins* and the third as *Dalit*. In this case, the probability distribution associated with each director is degenerate. For the *Dalit* directors, the probability distribution $p_i = (p_{im})_{m=1}^{k}$, where $p_{im} = 1$ for m= *Dalit* and 0 otherwise. Analogous probability distribution holds for the *Brahmin* director. The Blau index is the same as the fuzzy Blau index for Board A $=(\frac{1}{3})^2 + (\frac{2}{3})^2 = 0.553$. Similarly, for Board B, where all the members are Brahmins, the fuzzy Blau index is 1.

However, last names may not be uniquely associated with a cultural identity. In this case, the probability mapping associated with each last name is not degenerate. For example, consider Board C with three board members in Table Appendix C.2. Two of them are uniquely identified as *Brahmin*. The third board member's last name maps to *Brahmin* with probability 0.6 and a *Dalit* with probability 0.4. This board can be thought

Table Appendix C.2: Fuzzy Blau Examples

		Board C			
Member	Dalit	Brahmin	Kshatriya	Vaishya	Shudra
B1	0	1	0	0	0
B2	0	1	0	0	0
B3	0.4	0.6	0	0	0

;=		Combination 1: Probability 0.4		Combination 2: Probability 0.6
Member	Dalit	Brahmin	Dalit	Brahmin
B1	0	1	0	1
B2	0	1	0	1
B3	1	0	0	1

of as a composite of Board A and Board B with 0.6 and 0.4 probability respectively. In such a case, we measure the fuzzy Blau index of the board as a weighted sum of the Blau Index of Boards A and B. The fuzzy Blau index of Board C is equal to 0.6∗ Blau index of Board A +0.4∗ Blau index of Board B. Let a combination of a board be a degenerate mapping of each board member to a caste. A board of size N, where each board member has a non degenerate mapping over k unique social categories has N^k distinct possible such combinations of its board members. We define the Fuzzy Blau index to be a weighted sum of Blau indices of all such combinations of the board.

Fuzzy Blau Index for a board j is defined as:

$$FB_j = \sum_{c=1}^{N^k} p_c B_c$$

where, p_c is the probability of the cth combination of the board and B_c is the Blau index of the cth combination.

$$B_c = \sum_{m=1}^{k} (n_m/N)^2$$

We assume that the probability distribution of any board member is independent of that of others.

Figure Appendix C.1 shows the kernel density plots for firms' Blau and fuzzy Blau indices in the year 2015. It is clear that the two distributions are quite similar. Further, the yearly as well as across all years, the correlation between the two indices is about 0.9 and highly statistically significant. Thus, we feel confident that even though the Blau index omits some information about the possible cultural identities of directors, our results would still go through with the fuzzy Blau index too. Indeed, the patterns in the data presented in Section 3 are robust to using the fuzzy Blau index. These patterns, for *jati* fuzzy Blau index, are presented in Figures C.2 through C.8. Patterns are also similar for *varna* and religion fuzzy Blau index.

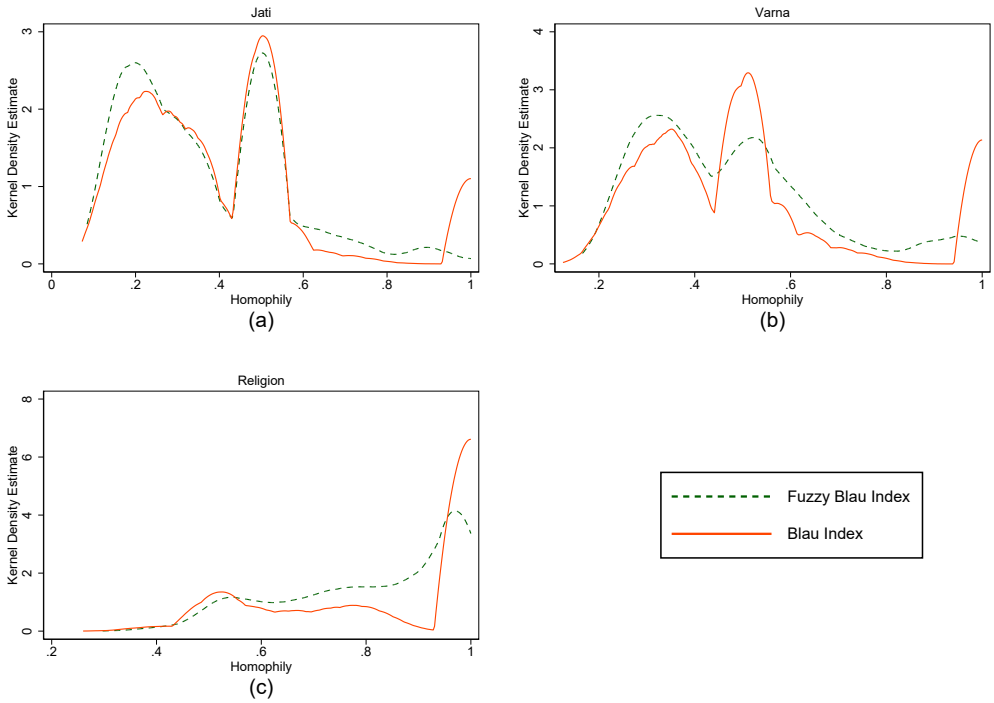
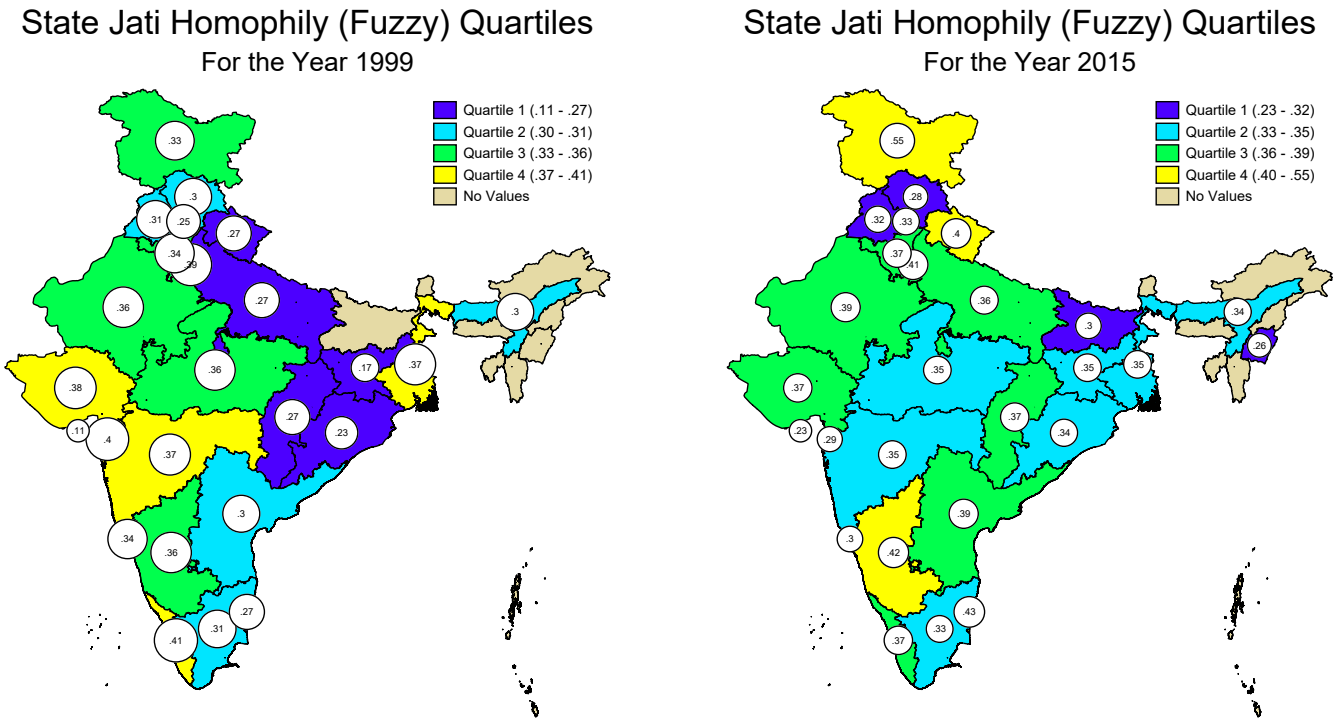


Figure Appendix C.1: Blau and Fuzzy Blau Index Comparison[a]

[a]Source: Prowess, matrimonial data. The figure shows the kernel density plots for Blau and fuzzy Blau indices for all firms in the year 2015, and for all measures of cultural identity – religion, *varna* and *jati*.



(a) (b)

Figure Appendix C.2: Average Jati Fuzzy Blau Index Across States[a]

[a]Source: Prowess, matrimonial data. The map shows the average *jati* homophily (fuzzy Blau index) 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 *jati* homophily levels.

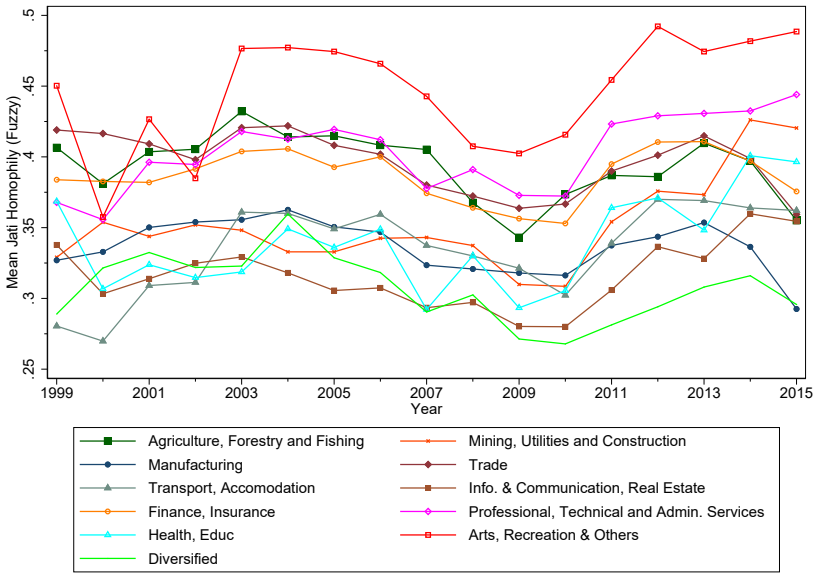


Figure Appendix C.3: Average Caste Fuzzy Blau Index By Sector[a]

[a]Source: Prowess, matrimonial data. The figure shows the mean *jati* homophily (fuzzy Blau index) for all firms in eleven broad sectors. Diversified includes all firms that could not be classified primarily into one industry.

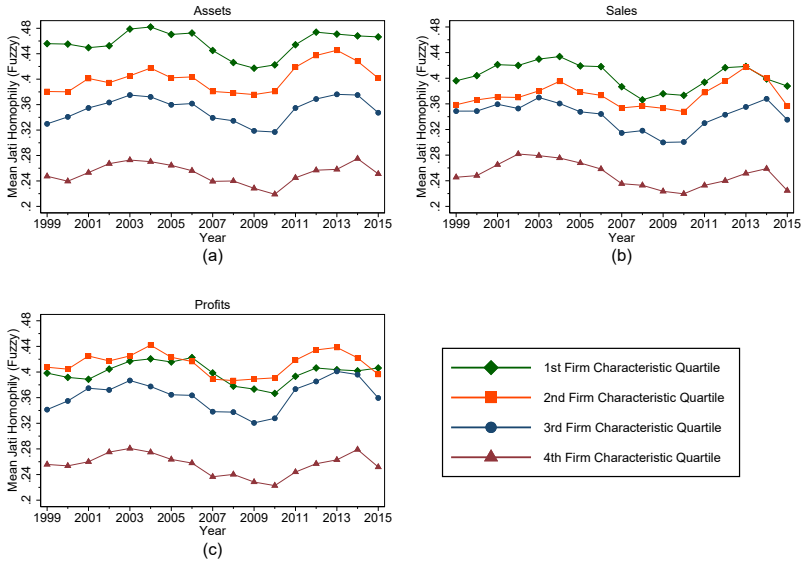
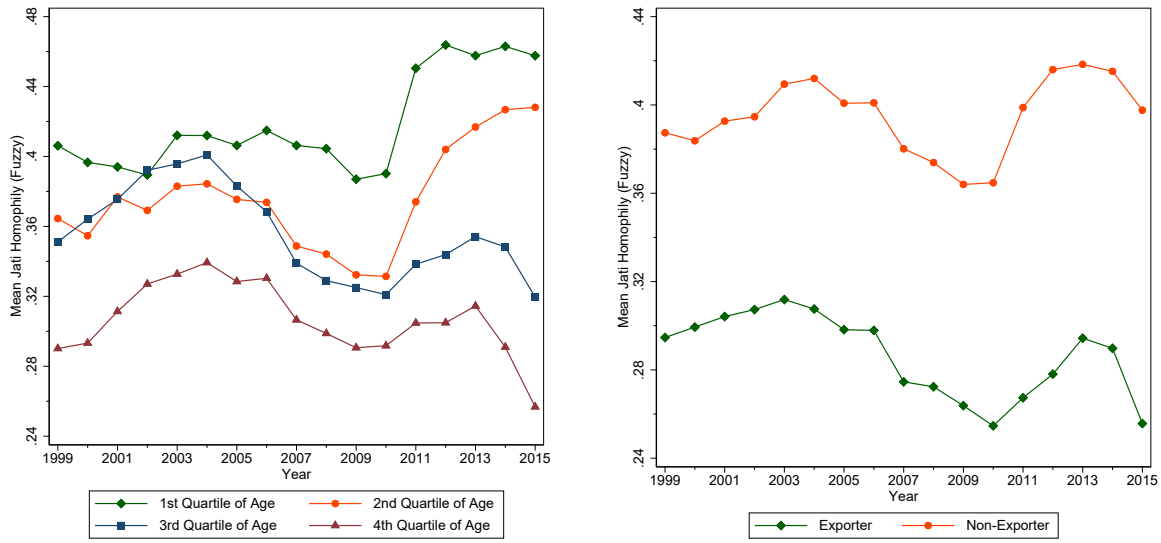


Figure Appendix C.4: Average Jati Fuzzy Blau Index by Firm Size[a]

[a]Source: Prowess, matrimonial data. Homophily (fuzzy Blau index) is averaged over all firms in a broad sector in each year.



(a) (b)

Figure Appendix C.5: Average Jati Fuzzy Blau Index by Firm Age and Export Status[a]

[a]Notes: Source: Prowess, matrimonial data. Firm age is defined as the number of years since the incorporation year. Homophily (fuzzy Blau index) is averaged over all firms in each quartile of the age distribution in figure (a) and over all exporting and non-exporting firms separately in figure (b).

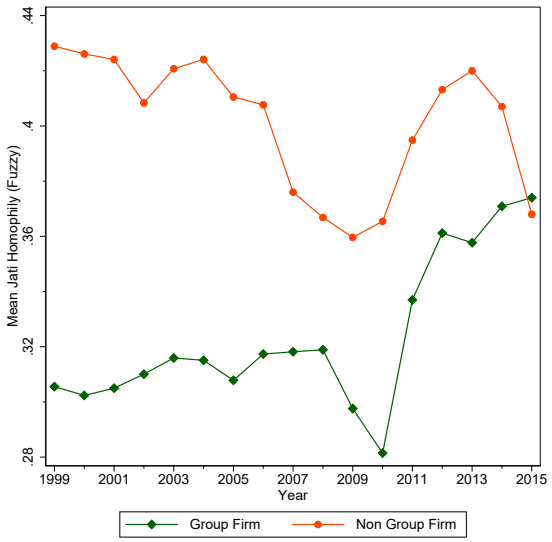
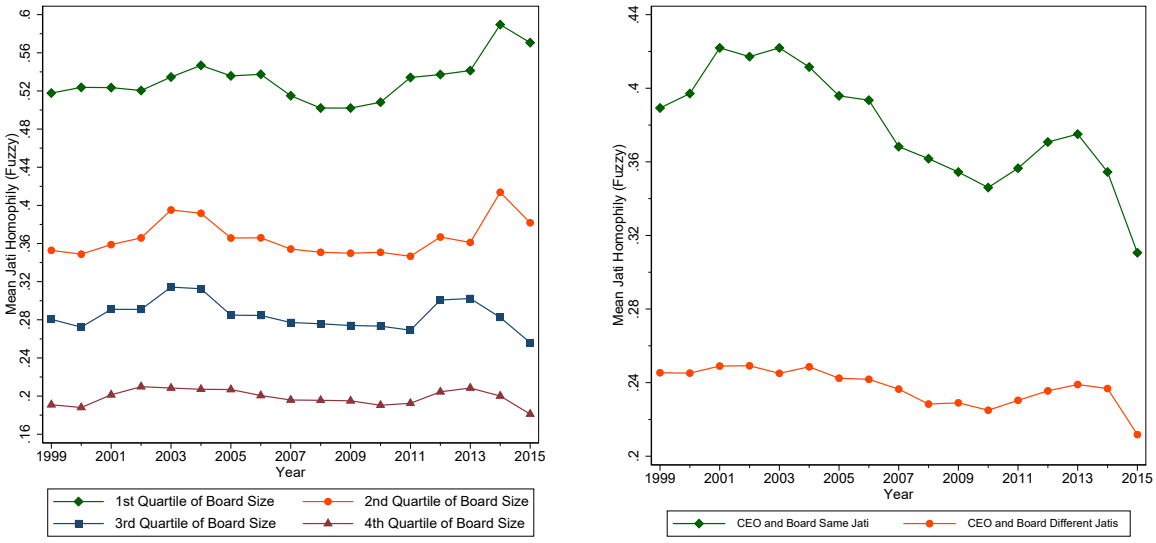


Figure Appendix C.6: Average Jati Fuzzy Blau Index by Group Membership [a]

[a]Notes: Source: Prowess, matrimonial data. Homophily (fuzzy Blau index) is averaged over all firms that belong to business groups and all firms that do not.



(a) (b)

Figure Appendix C.7: Average Jati Fuzzy Blau Index by Board Size and CEO Caste[a]

[a]Notes: Source: Prowess, matrimonial data. Homophily (fuzzy Blau index) is averaged in figure (a) over all firms in each quartile of the board size distribution and in figure (b) over all firms in a year that fall into either of two groups: those that have their CEO belonging to the same caste as the dominant caste in the rest of the board, and those where the CEO belongs to a different caste.

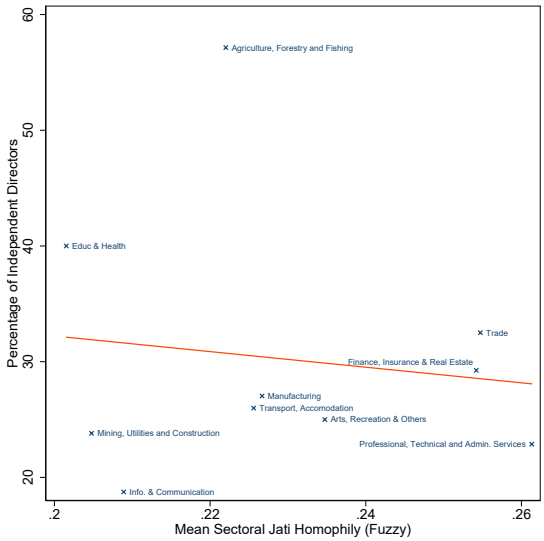


Figure Appendix C.8: Average Jati Fuzzy Blau Index vs. % Independent Directors [a]

[a]Notes: Source: Prowess, matrimonial data. Homophily (Fuzzy Blau index) is averaged over all firms in each one-digit industry. This figure is made using data for year 2015.

Appendix D Religion Homophily

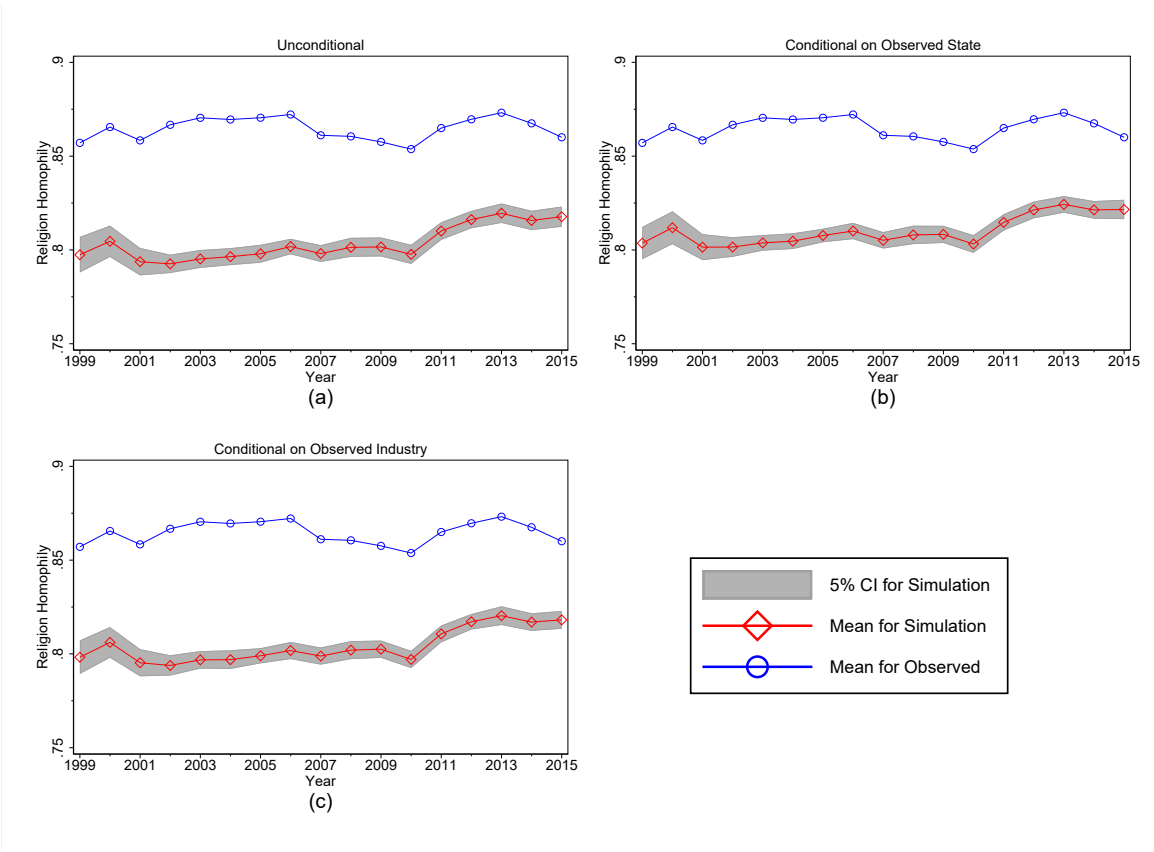
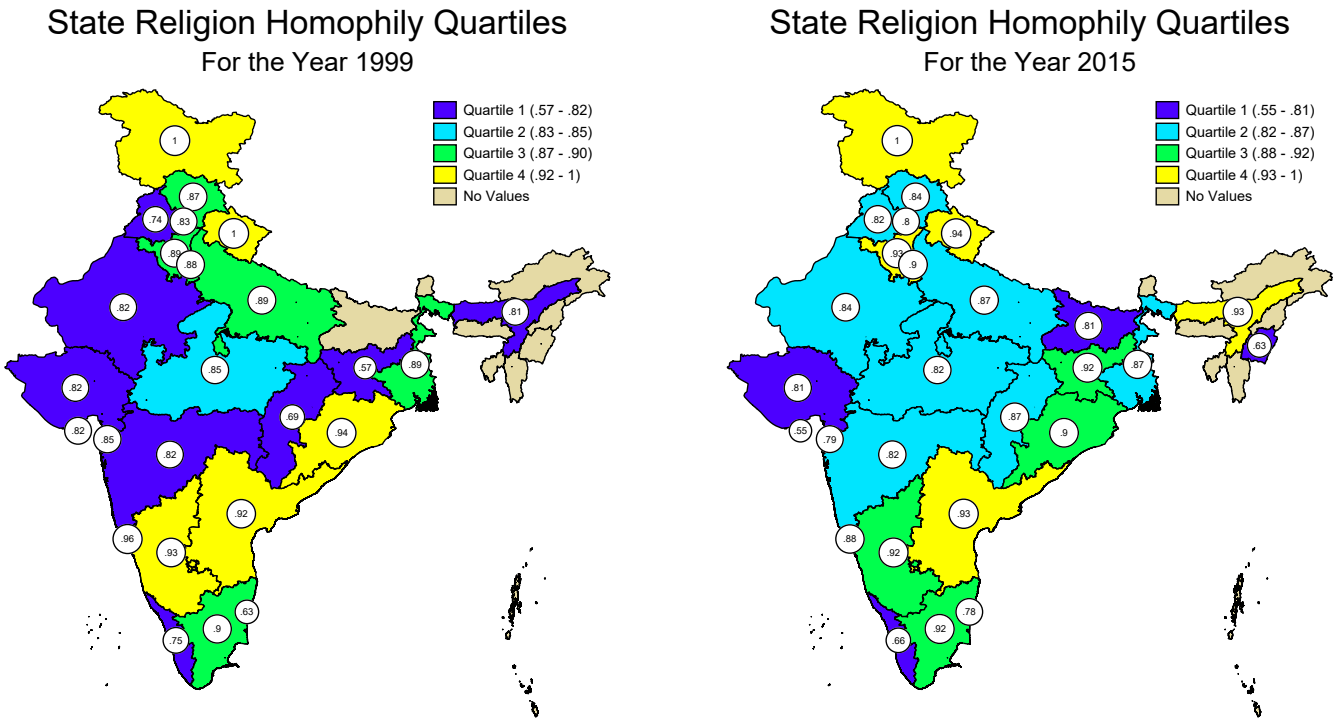


Figure Appendix D.9: Observed vs. Simulated Average Religion Homophily[a]

[a]Source: Prowess, matrimonial data. The three graphs in the figure present the mean religion homophily (Blau index) 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 3.1.



(a) (b)

Figure Appendix D.10: Average Religion Homophily Across States[a]

[a]Source: Prowess, matrimonial data. The map shows the average religion homophily (Blau index) 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.

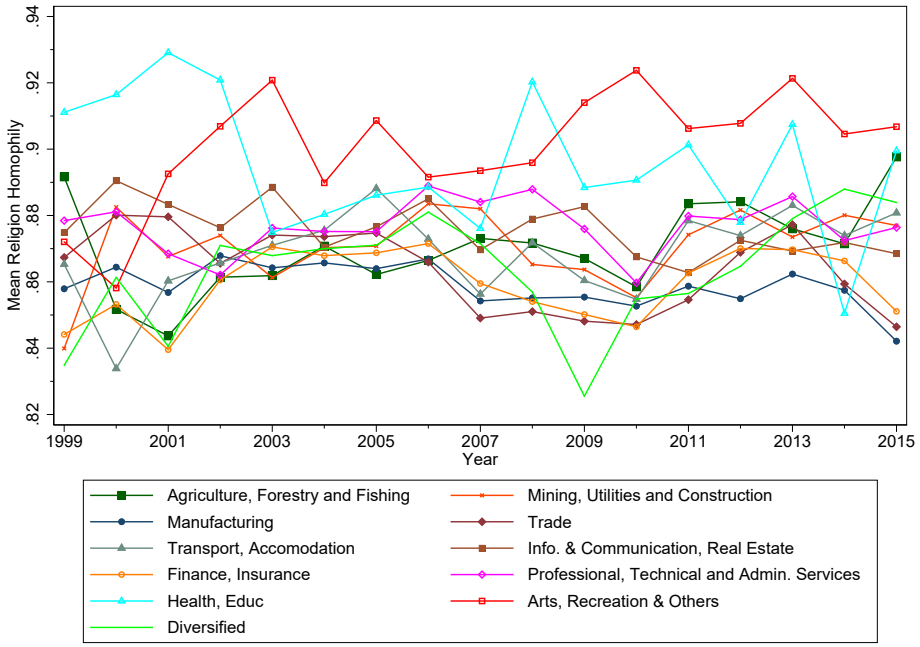
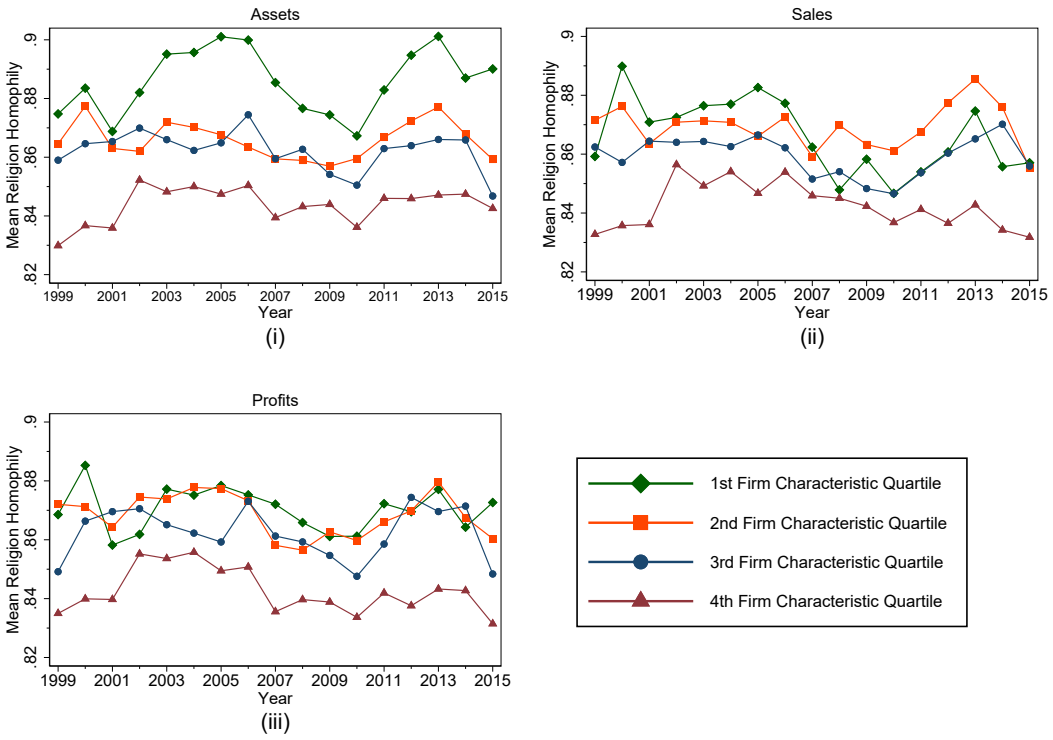
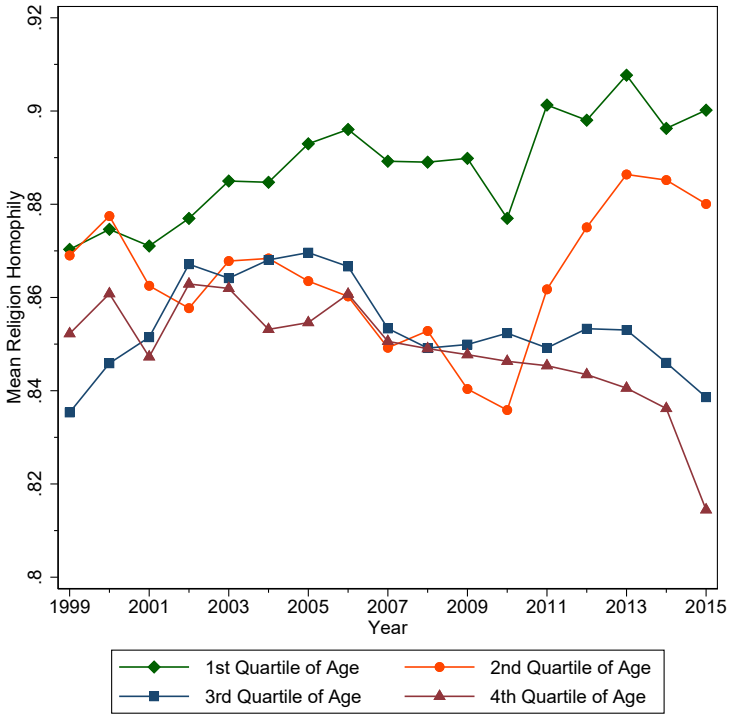


Figure Appendix D.11: Average Religion Homophily By Sector[a]

[a]Source: Prowess, matrimonial data. The figure shows the mean religion homophily (Blau index) for all firms in eleven broad sectors. Diversified includes all firms that could not be classified primarily into one industry.



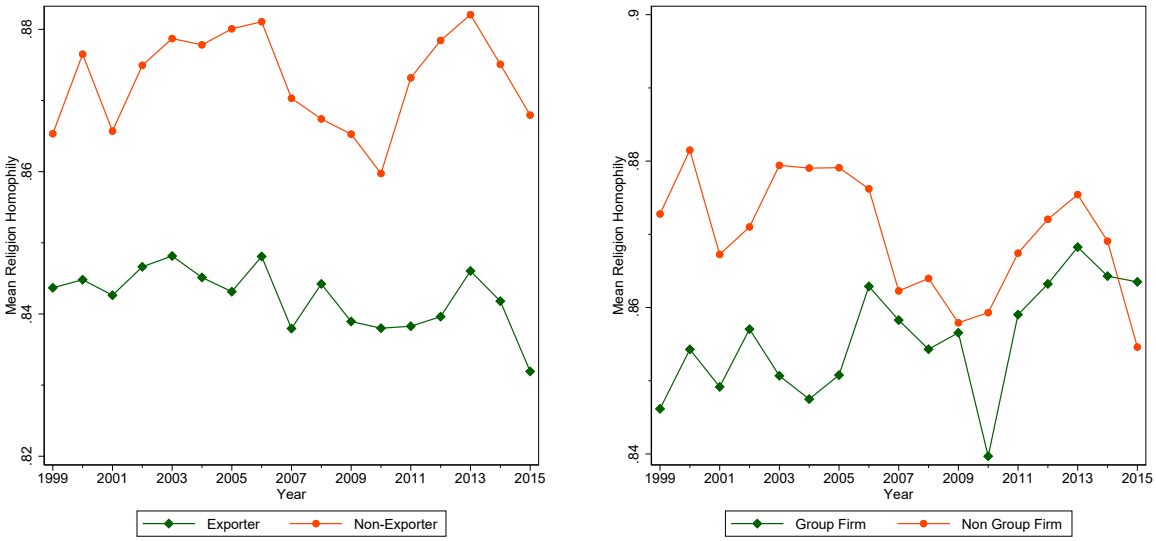
(a) Jati



(b) Varna

Figure Appendix D.12: Average Religion Homophily by Firm Size and Age[a]

[a]Source: Prowess, matrimonial data. Homophily (Blau index) is averaged over all firms in each quartile of the size and age distribution. Firm age is defined as the number of years since the incorporation year.

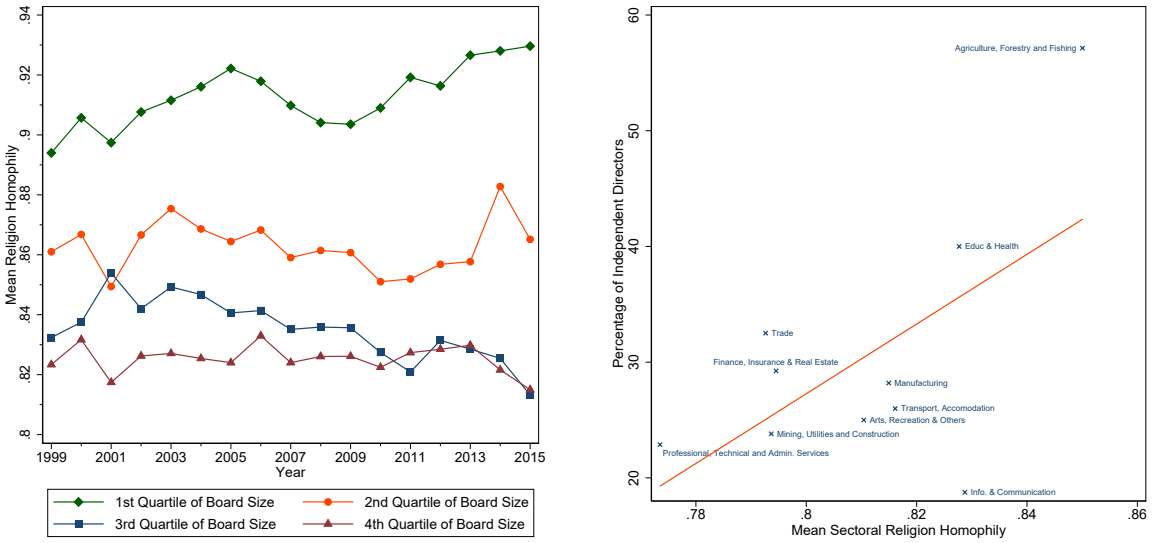


(a) Exporting Status

(b) Business Group Membership

Figure Appendix D.13: Average Religion Homophily by Export and Group Status[a]

[a]Source: Prowess, matrimonial data. Homophily (Blau index) is averaged over all exporting and non-exporting firms separately (figure (a), and over all firms that belong to business groups and all firms that do not (figure (b).



(a) Board Size

(b) Percentage Independent Directors

Figure Appendix D.14: Average Religion Homophily by Board Size & % Independent[a]

[a]Source: Prowess, matrimonial data. Homophily (Blau index) is averaged over all firms in each quartile of the board size distribution. Figure (b) is made using data for year 2015.

Table Appendix D.1: Religion Homophily and Firm Outcomes: Fixed Effects

Variables	(1) Log (Operating Income)	(2) Log (Operating Cash Flow)	(3) Log (Profits)	(4) Market to Book Ratio	(5) Tobin's Q	(6) Volatility
Religion Homophily	0.086	0.144**	0.161*	0.184	0.030	0.002
	(0.088)	(0.069)	(0.092)	(0.196)	(0.088)	(0.001)
Firm Age	-0.003*	-0.002*	-0.001	0.007**	0.001	-0.000***
	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)	(0.000)
Board Size	0.199***	0.190***	0.215***	-0.017	-0.013	-0.001***
	(0.007)	(0.008)	(0.010)	(0.019)	(0.009)	(0.000)
Leverage	-0.112***	-0.105**	0.056	1.112***	1.191***	0.002***
	(0.037)	(0.051)	(0.070)	(0.302)	(0.060)	(0.001)
Real Assets	0.009***	0.009***	0.010***	0.001	-0.000	-0.000***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Tangibility	0.166	0.195	0.018	-0.053	-0.380***	0.004***
	(0.134)	(0.150)	(0.165)	(0.275)	(0.095)	(0.001)
I(Listed)	-0.226***	-0.315***	-0.207***			
	(0.071)	(0.048)	(0.072)			
I(Group Firm)	0.643***	0.714***	0.769***	0.386***	0.022	-0.006***
	(0.045)	(0.054)	(0.099)	(0.139)	(0.059)	(0.001)
I(Export Status)	1.178***	1.017***	1.253***	-0.364**	-0.131*	-0.006***
	(0.057)	(0.059)	(0.058)	(0.155)	(0.075)	(0.001)
Constant	0.970***	1.003***	-3.716***	0.769**	0.710***	0.130***
	(0.098)	(0.100)	(0.109)	(0.385)	(0.151)	(0.005)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	46,809	39,056	68,398	26,901	29,206	28,811
R-squared	0.418	0.406	0.423	0.031	0.180	0.472

Notes: This table presents results for fixed effects regressions of several firm outcomes on board religion homophily and other control variables. The dependent variables are: log(operating income), log(operating cash flow), log(profits), market to book ratio, Tobin's Q, and volatility. Control variables include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. Columns (4)-(6) omit listing status since the samples for those regressions include only listed firms. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Table Appendix D.2: Religion Homophily and Firm Outcomes: Second Stage, IV Approach 1

Variables	(1) Log (Operating Income)	(2) Log (Operating Cash Flow)	(3) Log (Profits)	(4) Market to Book Ratio	(5) Tobin's Q	(6) Volatility
Religion Homophily	1.101**	0.263	0.410	-1.442*	-0.301	0.006
	(0.456)	(0.498)	(0.521)	(0.858)	(0.340)	(0.005)
Firm Age	-0.003*	-0.002*	-0.001	0.007**	0.001	-0.000***
	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)	(0.000)
Board Size	0.207***	0.191***	0.217***	-0.023	-0.014	-0.001***
	(0.009)	(0.008)	(0.009)	(0.018)	(0.009)	(0.000)
Leverage	-0.114***	-0.103**	0.057	1.186***	1.192***	0.002***
	(0.035)	(0.052)	(0.070)	(0.316)	(0.059)	(0.001)
Assets	0.009***	0.009***	0.010***	0.001	-0.000	-0.000***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Tangibility	0.158	0.190	0.013	-0.010	-0.371***	0.004***
	(0.131)	(0.148)	(0.160)	(0.271)	(0.097)	(0.001)
I(Listed)	-0.187***	-0.311***	-0.196***			
	(0.065)	(0.053)	(0.069)			
I(Group Firm)	0.651***	0.716***	0.771***	0.390***	0.023	-0.006***
	(0.046)	(0.052)	(0.094)	(0.137)	(0.059)	(0.001)
I(Export Status)	1.186***	1.019***	1.255***	-0.370**	-0.131*	-0.006***
	(0.055)	(0.057)	(0.057)	(0.156)	(0.075)	(0.001)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	46,753	39,025	68,318	26,900	29,204	28,809
R-squared	0.411	0.406	0.423	0.024	0.179	0.471

Notes: This table presents second stage results from instrumental variable regressions of several firm outcomes on board religion homophily and other control variables.The dependent variables are: log(operating income), log(operating cash flow), log(profits), market to book ratio, Tobin's Q, and volatility. The excluded instruments are the religion homophily levels of two director supply pools for a given firm. The supply pools are: the set of directors serving in all firms in the same industry as the given firm, and the set serving in all firms in the same state. Control variables include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. Columns (4)-(6) omit listing status since the samples for those regressions include only listed firms. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Table Appendix D.3: Religion Homophily and Firm Outcomes: Second Stage, IV Approach 2

Variables	(1) Log (Operating Income)	(2) Log (Operating Cash Flow)	(3) Log (Profits)	(4) Market to Book Ratio	(5) Tobin's Q	(6) Volatility
Religion Homophily	1.007***	0.778***	0.825***	0.427	0.104	-0.001
	(0.196)	(0.142)	(0.181)	(0.293)	(0.098)	(0.002)
Firm Age	-0.003*	-0.002	-0.001	0.007**	0.001	-0.000***
	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)	(0.000)
Board Size	0.207***	0.194***	0.220***	-0.016	-0.013	-0.001***
	(0.008)	(0.008)	(0.010)	(0.019)	(0.009)	(0.000)
Leverage	-0.113***	-0.105**	0.056	1.101***	1.191***	0.002***
	(0.035)	(0.050)	(0.068)	(0.302)	(0.059)	(0.001)
Assets	0.009***	0.009***	0.010***	0.001	-0.000	-0.000***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Tangibility	0.158	0.185	0.008	-0.060	-0.382***	0.004***
	(0.132)	(0.150)	(0.163)	(0.273)	(0.096)	(0.001)
I(Listed)	-0.191***	-0.293***	-0.178**			
	(0.067)	(0.048)	(0.069)			
I(Group Firm)	0.650***	0.717***	0.776***	0.385***	0.022	-0.006***
	(0.046)	(0.053)	(0.098)	(0.138)	(0.059)	(0.001)
I(Export Status)	1.185***	1.022***	1.259***	-0.363**	-0.132*	-0.006***
	(0.056)	(0.058)	(0.056)	(0.154)	(0.075)	(0.001)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	46,753	39,025	68,318	26,900	29,204	28,809
R-squared	0.412	0.404	0.420	0.031	0.180	0.472

Notes: This table presents second stage results from instrumental variable regressions of several firm outcomes on board religion homophily and other control variables. The dependent variables are: log(operating income), log(operating cash flow), log(profits), market to book ratio, Tobin's Q, and volatility. The excluded instruments are the religion homophily levels of two director supply pools for a given firm, and the Euclidean distances of the firm's board religion composition from those of the two supply pools. The supply pools are: the set of directors serving in all firms in the same industry as the given firm, and the set serving in all firms in the same state. Control variables include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. Columns (4)-(6) omit listing status since the samples for those regressions include only listed firms. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10

Table Appendix D.4: Religion Homophily and Firm Outcomes: Second Stage, IV Approach 3

Variables	(1) Log (Operating Income)	(2) Log (Operating Cash Flow)	(3) Log (Profits)	(4) Market to Book Ratio	(5) Tobin's Q	(6) Volatility
Religion Homophily	-256.512	258.149	-169.337	-104.645	-24.030	-2.268
	(802.867)	(969.554)	(305.414)	(351.488)	(47.275)	(3.335)
Firm Age	-0.060	0.049	-0.030	-0.031	-0.010	-0.001
	(0.185)	(0.184)	(0.060)	(0.121)	(0.017)	(0.001)
Board Size	-2.014	1.831	-1.197	-0.514	-0.117	-0.012
	(6.859)	(6.237)	(2.503)	(1.646)	(0.211)	(0.015)
Leverage	0.397	-1.155	0.443	6.589	0.885***	0.000
	(2.163)	(3.933)	(1.127)	(16.191)	(0.241)	(0.025)
Assets	0.025	-0.004	0.028	0.011	0.002	0.000
	(0.045)	(0.054)	(0.030)	(0.031)	(0.004)	(0.000)
Tangibility	4.274	-0.492	2.905	2.782	0.595	0.082
	(12.837)	(4.442)	(5.004)	(11.513)	(1.454)	(0.117)
I(Listed)	-9.469	7.527	-6.913			
	(28.945)	(29.221)	(12.088)			
I(Group Firm)	-1.350	0.854	-0.865	0.696	0.238	0.006
	(6.887)	(1.460)	(3.382)	(1.199)	(0.386)	(0.031)
I(Export Status)	-1.868	4.493	-1.119	-1.481	-0.217	-0.027
	(9.903)	(12.806)	(4.621)	(5.136)	(0.465)	(0.030)
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Observations	22,381	17,737	32,502	11,543	12,750	12,494

Notes: This table presents second stage results from instrumental variable regressions of several firm outcomes on board religion homophily and other control variables. The dependent variables are: log(operating income), log(operating cash flow), log(profits), market to book ratio, Tobin's Q, and volatility. The excluded instrument is the product of indicators for whether a firm is required to comply with Clause 49 in a year and whether it is above or below the 50% threshold for non-executive directors. Control variables include firm age, leverage, real assets, tangibility, listing status, export status, whether the firm belongs to a business group, and board size. Columns (4)-(6) omit listing status since the samples for those regressions include only listed firms. All financial variables are winsorized at the 1st and 99th percentiles over the sample period (1999-2015). All variables are defined in Section 4 and Appendix B. All regressions include two-digit industry and year fixed effects. Robust standard errors, clustered by industry, are in parentheses. *** p<0.01, ** p<0.05, * p<0.10