Group Identity and Agency Frictions: Evidence using Big Data

July 2025

Abstract

Does group identity alleviate agency frictions? Using a novel database of 3.78 million Indian last names with multiple identities to perceive the managers' and board members' identities, I provide evidence that the manager with a similar social identity as the board members relatively earns higher compensation. A firm benefits from paying the cost of in-group favoritism as the manager's reciprocation increases the long-term value and reduces agency frictions. An increase in compensation 1% due to group identity increases the value of the firm by 1.8%. The results are robust to alternate definitions of group identity and other checks. These findings suggest that firms should use group identity between managers and the board to mitigate agency frictions rather than solely focusing on board diversity. Nevertheless, the efficiency of group identity declines for complex firms and during crises.

KEYWORDS: Group Identity, In-group Favoritism, Managerial Compensation, Firm Value, Statistical Discrimination.

JEL Codes: D21, D22, G30, G40, J71

1. INTRODUCTION

"The relationship of agency is one of the oldest and commonest codified modes of social interaction."

-Stephen Ross (1973)

The separation of firm ownership and control give rise to agency conflict and challenges such as moral hazard, information asymmetry, and similar others (Ross, 1973; Jensen and Meckling, 1976; Fama, 1980; Fama and Jensen, 1983; Hermalin and Weisbach 1991; Jensen 1993; Shleifer and Vishny 1997). The corporate governance literature suggests the changes in the manager's incentive structure as a solution (Yermack 1995; Hall and Liebman 1998; Mehran and Tracy 2001); however, the problem persists.[1] Motivated by a massive body of literature on identity economics and social psychology, this article provides a behavioral solution to reduce agency frictions. I show that the similarity of social identity (or group identity) between a manager and the board members leads to higher managerial compensation, but those firms also have higher long-run value. This implies that group identity reduces agency frictions despite the cost of in-group favoritism. These results also suggest that the statistical discrimination can be efficient and Pareto optimal as group identity provides informational gains on manager's specialization. These findings have wider implementation as the agency problem is not limited to the manager-board duo but exists in various other corporate and non-corporate decisions.

At first glance, group identity might seem anti-diversity. However, group identity and diversity can co-exist. Two hypothetical situations can elaborate on it. If a German board has a choice between a German manager and a Vietnamese manager, both with equal merit,

[1]The agency problem is well-discussed subject in various disciplines such as finance (Fama, 1980), economics (Spence and Zeckhauser, 1971), accounting (Fan and Wong, 2002), management (Douma, George, and Kabir, 2006; Peng and Jiang, 2010), and organizational studies(Schulze, Lubatkin, Dino, and Buchholtz, 2001). Due to the agency frictions, the corporate governance is required.

whom will the board hire? Due to cultural proximity with the board and maintaining ease of communication, the preference would be the German manager. Similarly, if in a corporate board of eight members and a male manager, there is only one woman, then the woman board member's hesitation in putting her viewpoint would be higher than if the manager is a woman too. Taking these hypothetical situations as a medium, I suggest that the board's diversity concerning the manager's characteristics is a solution. However, group identity between a manager and the board is necessary to mitigate the friction between them, which can help the firm in the long run.

For decades, the standard neoclassical models in finance posit that corporate decisions are made virtually in a vacuum and independently of the agents' social identities. Since the seminal work of Akerlof and Kranton (2000), the interest in social identity research in economics and finance has increased. The development of identity research illustrates that the similarity (dissimilarity) of social identity between two individuals culminates in favoritism (discrimination).[2] The discrimination literature mainly emphasizes two forms - taste-based (Becker, 1957) and statistical discrimination (Arrow, 1972; Phelps, 1972). Taste models include in the utility functions of employers, fellow workers or customers a desire to avoid members of certain groups. Under such an approach, discrimination cannot be characterized as either efficient or inefficient. On the other side, statistical discrimination, by contrast, can affect efficiency. It differs from taste models in assuming no prejudice or invidious motive by employers or employees but rather that employers use average group characteristics to predict individual worker attributes (Schwab, 1986). Inspired by the literature on group identity and statistical discrimination, this study begins by providing an intuitive stance on

[2]Social identity models are widely applied to understand discrimination in the workplace (Akerlof and Kranton, 2000; Bertrand and Mullainathan, 2004), including the economics of poverty, social exclusion, and the household division of labor (Akerlof and Kranton, 2000), contract theory (Akerlof and Kranton, 2005), economic development (Basu, 2006), public goods provision (Croson, Marks, and Snyder, 2008), team identity in the workplace (Eckel and Grossman, 2005), banking (Fisman, Paravisini, and Vig, 2017), and investment decisions (Morse and Shive, 2011; Grinblatt and Kelohraju, 2001).

how the group identity can change the equilibrium outcomes of executive compensation and firm value and the endogenous relation between them. By borrowing Akerlof and Kranton's (2005) identity-based agent's utility function in Holmstrom's (1979) model, I show that the manager gets a higher compensation due to the in-group bias and that firm value increases as the manager exerts an additional effort for the firm for the same reason. These theoretical findings are similar to Norman's (2003) work on statistical discrimination and efficiency. As theoretical predictions can differ from empirical findings, it is crucial to verify whether empirical analyses sustain the predictions. Therefore, I test these predictions using India as a empirical setting.

Past papers in finance show how group identity affects various corporate and investment decisions without verifying whether biases based on those identities exist or not. The psychology and behavioral studies lab experiments use frequency magnetic resonance images (fMRIs) or implicit association tests (IATs) to capture the implicit biases.[3] However, most of these findings are based on lab experiments, therefore, hard to scale for cross-industry analysis. To work around this shortcoming, I followed Fershtman and Gneezy's (2001) empirical design and chose a setting where stereotypes based on specific identities exist, and the individual's last name reveals those identities.

In India, a surname (i.e., a family name or last name) provides information on individual's family lineage, native language, place of origin, and caste.[4] As a result, surnames are likely

[3]The IAT relies on the test-taker's speed of response to represent the strength of their unconscious mental associations. IATs are used to measure a wide range of implicit attitudes about social groups, products, or self-identity (Greenwald et al. 1998).

[4]According to the ancient literature, castes are mainly divided into four - the Brahmins (priestly people), the Kshatriyas (also called Rajanyas, who were rulers, administrators, and warriors), the Vaishyas (artisans, merchants, tradesmen, and farmers), and the Shudras (laboring classes). People who do not come into these castes are called 'Harijans' (untouchables). The Brahmins, Kshatriyas, and Vaishyas are collectively known as the upper caste and, the Shudras and Harijans are of the lower castes. After establishing the Indian constitution in 1950, caste was redefined based on the reservations provided to the under-privileged in relation to government jobs, education, and other services. As per this new categorization, the Brahmins, Kshatriyas, and Vaishyas are known as the 'other' caste, the Shudras are the 'scheduled tribes' (ST), and the Harijans as 'scheduled caste' (SC). The caste in relation to the Socio-Economic Caste Census (SECC) data is according to the new categorization. I include caste as an identity because caste remains an enduring

an important basis of social ties. This is also a source of favoritism (or discrimination). The main reason for using surnames as a source of social identity is that people who belong to a particular surname live in a state of tight kinship and provide effective cooperation to the other group members, while people outside of the group are distrusted (Alesina and Giuliano 2013; Moscona, Nunn, and Robinson 2017).

As no readily available database provides the managers' and directors' identities, a surname dataset was developed using the information of 474 million Indians from the Socio-Economic Caste Census (SECC) and the Linguistic Survey of India (LSI) data. On merging these databases using machine learning algorithms, I develop a novel database of 3.78 million Indian last names with multiple identities. This database has information on Indian surnames, native language, caste, and native place. [5] By discerning the identities of managers and directors using this database, group identity (or homophily) variables are computed based on native language, native place, and caste.

To examine how group identity impacts agency frictions, I analyzed the panel data of 2,324 non-financial firms listed on two leading Indian stock exchanges - Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE) from 2004-2018. The main empirical results indicate that the manager gets a 6-8 % higher compensation if the manager belongs to the same identity as at least one of the board members. The magnitude of the incremental compensation can varies across the different identities but the direction remains same. For instance, the results are stronger for identities such as native language and native place than

predictor of economic status in India. It is correlated with occupation and employment (Prakash, 2015; Ito, 2009; Thorat and Attewell, 2007), income and expenditure (Deshpande, 2000), and capital more generally (Kijima, 2006).

[5]These social identities are used because social psychology and neuroscience studies have shown that favoritism or discrimination based on natural identities (such as gender, race, native language, native place, caste and similar others) exists and that it is stronger than induced identities (such as club membership, alumni, and similar others).The surname dataset provides a composition of all the Indian surnames at the state level, language level, and caste level. As there was no significant intra-country movement of any community in India except Kashmiri pandits in 1990, this surname database correctly provides managers' and directors' identities.

the caste. This incremental compensation due to group identity is construed as a cost of the in-group favoritism borne by the shareholders. Firm value (measured as market to book (MB) ratio) is also 11-12% higher for such firms compared to their counterparts.[6] Collectively, these results suggest that every 1% increase in executive compensation due to group identity increases firm value by 1.8 %. This result implies that the benefit from group identity overcomes its cost and complements the existing incentive structure in reducing agency frictions. This finding that informational gain on manager's specialization by group identity reduce frictions confirms that statistical discrimination (favoritism) can be efficient and Pareto optimal.

For robustness, I verify the results using various other definitions of group identity, such as the manager's group identity with the majority of board members and using the continuous group identity measure. The manager's group identity with the majority of board members scaled by board size defines the latter. Rather than using the MB ratio, I alternatively use sales growth, return on assets (ROA), return on equity (ROE), and the logarithm of market cap plus book value of debt as alternate measures of firm value. The results are robust to these changes.

As managerial compensation and firm value are determined endogenously, therefore, to know the impact of group identity on firm value conditioned on compensation, I decompose the firm value. One part explains the previous year's managerial compensation and the second part, residual firm value. The underlying theory for using the residual firm value as a proxy of managerial reciprocation is that the determinants of residual firm value should only be correlated with firm and governance characteristics. If homophily explains a part of it. It is due to the manager's extra effort put into the firm. I find that firms with more homophily between a manager and board have a 5.6-9.5% higher residual value, i.e., higher

[6]I use the market to book ratio because it is well established measure for firm value (see, Ritter (1981) and Downes and Heinkel (1981)) as well as for firm performance (see, McDonald, Khanna, Westphal, 2008; Morck, Shleifer, Vishny, 1989, and similar others).

firm value excluding the cost of in-group favoritism. However, one can argue that the residual firm value may fail to capture the simultaneity issue. To mitigate such concerns, I followed the work of Palia (2000) and confirmed the results using the two-stage least square (2SLS) method. The latter results imply that the manager exerts an extra effort due to the shared group identity, which reduces agency frictions. These results are robust to firm fundamentals and corporate governance characteristics. The result that individuals put extra effort into in-group members compared to out-group members is in line with the literature on group identity in social psychology and social neuroscience (Cikara and Bavel, 2014; Allport, 1954).

As the board can select a manager having the same identity to reduce the frictions in corporate decision-making (Damaraju and Makhija, 2018), this causes self-selection bias. To address this problem, I followed the work of Fracassi and Tate (2012) and used the director's death or non-voluntary retirement as a shock to the homophily constructs. The results are still consistent using this exogenous shock. [7]

Next, I check whether these results differ due to a firm's organizational structure and manager's group identity with directors with different fiduciary duties. For the impact of organizational structure, I check the results for business group firms, government firms, and complex firms. For the role of manager's group identity with directors having different duties, I check the role of manager's group identity with independent directors and with non-executive directors. The test for organizational structure is crucial as Indian institutions differ from developed countries for various reasons. For instance, there is a significant presence of concentrated ownership in the form of business groups.[8] Khanna and Palepu (2000)

[7] One can argue that a director's death or retirement may or may not change the board identity and hence, the homophily constructs. To mitigate such concerns, I conducted an additional test. Rather than directly using a director's death or retirement as a shock, a change in homophily due to it is considered a shock. The additional test results suggest that the coefficient of language and the place homophily are positively associated with managerial compensation, firm value, and residual firm value; it confirms the paper's main findings.

[8] The business groups are collections of publicly traded firms spread over different industries. These groups have a significant amount of common ownership and control, and are mainly family-owned groups.

compare the performance of business group firms with the non-business group firms of India. They show that the business group firms outperform the non-business group firms. They provide the channel showing that the group-affiliated firms have disproportionately good access to international sources of capital due to the group's reputation. The current study proposes an alternative channel for the better performance of group affiliated firms', i.e., the manager's reciprocation towards a firm due to group identity. Although managerial compensation due to in-group favoritism is higher in non-business group firms, the residual firm value is higher in business group firms. This result signifies that managerial reciprocation toward a firm due to group identity is higher in group-affiliated firms than in others.

As complex organizations are challenging to govern, the requirement of managerial skillset dominates the group identity in such firms. Therefore, the latter factor would not lead to favoritism. To test this argument, I measure firm complexity as the square root of the number of subsidiaries and find that the manager is overpaid to handle such firms; however, the group identity negatively affects the compensation. This finding implies the significance of merit over group identity in complex firms. I also find that firm value and residual firm value are higher, suggesting that it is beneficial for these firms to hire a manager with group identity with board members as it increases future firm value and the manager is paid less than their counterparts. I find similar results for firms in higher market competition.

Some organization types (such as government firms) have different selection criteria and payment processes for the manager. Therefore, it would be interesting to know whether group identity works in this organization type. For simplicity, I focus on public sector undertakings (PSUs) which are government commercial enterprises but hold autonomous power to take decisions up to the non-management level. I find no impact of group identity on managerial compensation and firm value in PSUs. These results are intuitive as government involvement in managerial recruitment is considered unbiased from any form of favoritism (discrimination).

In their paper, Hwang and Kim (2009) argue that independent board members should be conventionally and socially independent from the manager. Otherwise, it affects the monitoring and advisory roles of the board. They use induced identity measures in their paper. For generalization, it is imperative to check whether their findings hold with natural identity measures. Considering that, I examine how the manager's group identity with independent directors affects his compensation and future firm value. Findings suggest that the manager's social ties with independent directors do not influence the compensation but negatively affect the future firm value. These results confirm Hwang and Kim (2009) findings in a different setting and validate Bebchuk and Fried's (2003) argument that independent directors do not influence managerial compensation.

The battery of robustness checks is conducted to confirm the results. First, I external validate the results using alternate measures of group identity, compensation, and firm value. I reconstruct the identity variables using an alternative dataset from "Indianchildnames.com." This website features the most extensive collection of names (roughly 125,000) and surnames (approximately 15,000) of both Indian and international origins and their mapping with native language and native place. I compute the group identity variables using this dataset. The results are robust to these alternative group identity constructs. For external validity of main dependent variables, the logarithm of total compensation is replaced with three alternative variables - compensation relative to the industry peers, log salary, and log of salary plus bonus plus sitting fees. The sales growth, return on equity (ROE), and log of the market value of equity plus book value of debt are used as alternate firm value measures. The main results are robust to these changes.

Second, following Bhagat and Bolton (2008), I use current profitability and future profitability as the dependent variables. I found that group identity measures are positively associated with profitability. Third, the impact of geographical variation on the homophily constructs is studied in connection to compensation, firm value, and residual firm value.

Although culture in south India is different from north India, I found no difference in results if a manager and the socially tied board member(s) are from north India compared to if both belong to south India. Fourth, there is a possibility that migration from one state to another affects the census data, which impacts the homophily construct. To mitigate such concerns, I exclude surnames in the census data which affected by the migration. Even with this change, I found the same results. Fifth, to mitigate the possibility that increase in managerial compensation is due to his influence in the firm rather than the homophily. I control the 'powerful CEO' measure. Adams et al.(2005) show that firms have a higher performance variability if the CEO has a greater power to influence decisions. I control this measure and show that my results still hold. Sixth, using social tie measures based on induced identities, Farcassi and Tate (2012) show that a powerful CEO is more likely to appoint known directors and that the CEO-director ties reduce firm value. From the measures used in Farcassi and Tate (2012), the most common is a social tie based on past employment, and the least common is education. To mitigate the concern that the results are not driven by induced identities based ties, I show the results controlling these measures. The results are still robust. The coefficient of 'powerful CEO' measures and past experience ties align with the work of Adams et al. (2005) and Farcassi and Tate (2012).

Lastly, I check the robustness of the results using different ways to measure group identity. Rather than using group identity as the social tie of a manager with at least one board member, I use group identity as the tie of a manager with a maximum number of board members and the percentage of managerial and social relations with the board. In both scenarios, I find the same results: a manager gets higher compensation due to in-group favoritism. Still, the firm value also increases due to the same reason signifying that group identity can reduce agency frictions. These results support the notion that statistical discrimination can be efficient as group identity provides informational gain on manager's specialization.

Contribution: This paper contributes to the literature on agency theory (Jensen and Meck-

ling, 1976; Fama, 1980), especially behavioral agency theory as propounded by Wiseman and Gomez-Mejia (1998), Sanders and Carpenter (2003), and Pepper and Gore (2012). Decades of research on corporate governance and executive compensation put the foundation of the different theories on agency conflict and its existence. Arrow (1971) and Wilson (1968) postulate that agency conflict occurs due to risk-sharing between the principal and the agent with opposite risk preferences. Either the principal or the owners invest their capital and expect to gain economic benefits, whereas the agents are risk-averse and concerned with maximizing their private benefits. Ross (1973) regards the agency problem as the problem of incentives, while Mitnick (1975) considers the problem to occur due to the institutional structure. These theories have a similar central idea. Grossman and Hart (1983) construed that an agent's effort shapes the principal's consumption; therefore, the latter desire a higher level of effort from the agent. The principal should trade off the agent's behavior with a proper payment structure. The agents' attitude towards the principals' risk and information quality affects the incentive structure. No incentive problem arises if the agent is risk-neutral. Differing from the past literature, rather than using incentive mechanisms to resolve agency problem, I argue in favor of a behavioral solution. Using the identity economics literature, I show, theoretically and empirically, that the group identity between an agent (manager) and the principal (board) can reduce agency frictions.

Agency problem also has exhilarated a debate among academics and practitioners on executive pay and how it affects the firm value. Edmans, Gabaix, and Jenter (2017) literature review paper discourse three views on executive compensation. The current study binds the "rent extraction view" (executives themselves set contracts to maximize their rents) with the "shareholder view" (compensation contracts are the outcome of shareholder value maximizing the firms that compete with each other in an efficient market for managerial talent (Fama, 1980). I argue that rent extraction by a manager (hidden pay due to similar social identities as the board members) is not a cost to the firm if the manager increases the

firm value more than the compensation paid.[9] The recent literature on corporate governance has focused on how board diversity with respect of either gender or ancestry affects firm performance, firm value, board effectiveness, and similar other variables.[10] However, the results are mixed. The current study proposes that rather than focusing on board diversity, the emphasis should be given to group identity between board members and the manager to reduce the friction between them.

This paper also contributes to the literature on identity economics and discrimination in corporate finance and corporate governance. First, by providing a theoretical perspective and empirical evidence on how homophily changes managerial compensation, firm value, and agency frictions, and second, by providing the supporting evidence that statistical discrimination can be Pareto efficient. Identity economics literature mainly discusses how identities can change outcomes due to implicit biases (Bertrand et al., 2005), taste-based discrimination (Becker, 1971), or statistical discrimination (Phelps (1992)). Taste-based discrimination cannot be characterized as either efficient or inefficient. On the other side, statistical discrimination, by contrast, can affect efficiency (Schwab, 1986). It differs from taste models in assuming no prejudice or invidious motive by employers or employees but rather that employers use average group characteristics to predict individual worker attributes. Most of the studies in this area focused on how taste-based discrimination (favoritism) changes the economic outcomes but less on whether (statistical) discrimination can be efficient; this

[9]For different views on executive compensation and a literature review, see the work of Edmans, Gabaix, and Jenter (2017).

[10]Delis et al.(2017) shows the impact of genetic diversity in the country of origin of the firms' board members on corporate performance. It is shown that adding a board of directors from countries with different levels of genetic diversity (either higher or lower) can increase firm performance. On the contrary, using a similar kind of data, Giannetti and Zhou (2019) show that greater ancestral diversity may lead to inefficiencies in the decision-making process and conflicts in the boardroom. It also increases the firm's performance volatility. Using a sample of US firms and with gender representing identity, Adams and Ferreira (2009) show that there is a negative impact on firm performance due to having women on the board, despite the better attendance records and the more effective monitoring of firms with more gender-balanced boards. On the contrary, Gregory-Smith et al. (2014) show no evidence that the gender composition of the board affects firm performance. These papers imply that diversity in the boardroom is still a mootable topic.

paper fills that gap. For instance, Fisman et al. (2018) show how a candidate's hometown ties with fellow selection committee members can increase the selection probability in the Chinese Academies of Sciences and Engineering by 39%. Fisman, Paravisni, and Vig (2017) show how the group identity between the loan officer and the borrower based on identities like religion and caste increases the level of credit access and loan size dispersion, as well as reducing the collateral requirements and inducing better repayments even after the in-group officer leaves. Similarly, related to in-group favoritism and its impact, Gompers et al. (2016) show that venture capitalist with the same ethnic, educational, or career background are more likely to syndicate with each other. This homophily reduces the probability of investment success, and the detrimental effect is most prominent for early-stage investments. The focus of all these papers is on economic distortions due to taste-based discrimination; the current study show that statistical discrimination can be efficient when there is a information gain from it about the agent's (manager's) specialization. The theoretical model by Norman (2003) show that statistical discrimination can be efficient and welfare improvement. The findings in this study confirms Norman(2003) predictions.

There is also a literature discussing the meritocracy versus group identity (homophily). The theoretical work of Moisson and Tirole (2020) explores the trade-off between meritocracy and homophily in the selection and promotion process in the organization setup. It also checks the entrenchment and welfare properties of an organization. Using the sub-sample on managerial demographics, I show that group identity results remain even controlling for manager's gender, education, and age. Nevertheless, further research is required to compare the meritocracy with group identity in different organizational setup or at different time periods.

Lastly, this paper also contributes to emerging literature on big data in finance especially the applications in corporate governance. Unlike, the big data application in selecting the directors (Erel et al., 2021) or limiting the insider trading (Zhu, 2021), this paper shows how

it can be used to learn the additional characteristics of a manager and the directors which can assist in curbing the agency frictions.[11]

This paper is close to Cai et al. (2013), Fracassi and Tate (2012), and Huber, Lindenthal, and Waldinger (2021). Cai et al. (2013) investigated how the family ties with the firm owners affect managerial compensation and job assignment in Chinese private firms. They show that family managers earn higher salaries and receive more bonuses, hold higher positions, and are given more decision rights and job responsibilities than non-family managers in the same firm. However, unlike Cai et al. (2013), this paper shows that favoritism due to in-group bias (not familial bias) also leads to higher managerial compensation. The firms also benefited from the bias, increasing the firm value and reducing agency frictions. In Fracassi and Tate (2012), the CEO-director ties based on shared experiences reduce firm value without other governance mechanisms to substitute for board oversight. I show that shared identities between the CEO and the board increase firm value under strong governance, such as high promoters (founders) holding. Cai et al. (2013) and Fracassi and Tate (2012) focus on the cost of favoritism. On the other side, Huber et al. (2021) examined the opportunity cost of discrimination; they found that the discrimination against qualified managers of a particular ethnic group leads to persistent reductions in stock prices, dividends, and asset returns. Rather than the opportunity cost of discrimination, I check the cost, benefit, and net benefit (net cost) of favoritism in corporate decisions. I show that the shared group identity between a manager and the board members can increase the level of executive compensation due to in-group favoritism. The firms with group identity have a higher value than the other firms even when conditioned according to managerial compensation. Furthermore, the residual firm value (i.e., the firm value conditional on executive compensation of the last year) is higher for such firms. This result implies that the manager puts in an extra effort when there is a shared group identity between them and the board. These results indicate that although

[11]For review of papers on big data applications in finance, refer Goldstein, Spatt, and Ye (2021).

in-group favoritism is a cost, bearing this cost is beneficial to a firm as it reduces the agency frictions and can increase the long-term firm value through the manager's reciprocation.

Section 2 discusses the theoretical model and its predictions. Section 3 elaborates on the data and variables descriptive. Section 4 explains the methodology and empirical specification. Section 5 discusses the results. Section 6 concludes.

2. THEORITICAL MODEL

The model's motivation comes from the work of Ross (1973), Holmstrom (1979), and Edmans, Gabaix, and Jenter (2017).[12] According to this model, the board of directors hires a manager to run the firm. The firm value is indicated by $V(a, S, \epsilon)$, which increases in the manager's action a and firm size S. The function b(S) measures the effect of effort on firm value for a firm of size S. Suppressing the dependence on S and ϵ for simplicity, firm value is a function of managerial effort only.

$$V(a) = S + b(S)a + \epsilon$$

The manager is paid a compensation $c(V)$ that is contingent upon firm value. The board's limited liability is assumed, which means $c(V) \leq V$; that is manager pay should not be more than firm value. Similarly, the limited liability of the manager is assumed which means $c(V) \geq 0$ and reservation wage, $w \geq 0$.

The model is modified by introducing group identity into the preference function by providing different weights to the managerial effort component. It defines the agent utility function using the identity. This kind of agent utility function is suggested by Akerlof and Kranton (2005), and a similar method of incorporating group identity into preferences is adopted in the works of Chen and Li (2009), and Chen and Chen (2011). This theoretical

[12]In the first-order model of principal and agent used by Holmstrom (1979), the agent's utility maximization problem is replaced with the first order constraint and it is based on the assumption that an optimum exists and is differentiable. The principal and agent both are risk-neutral in this setting. See Mirrlees (1976) and Rogerson (1985) to know more about first order and other related models.

result is in line with the literature on minimum-effort coordination games.[13] The complete model for this study is provided in **Appendix B**.

The following model setup is from Holmstrom (1979),[14]

$max E(V - c)$ (Principal's Problem)

s.t. $E(u(c - g(a))) \geq E(u(w))$ (Participatory Constraint)

$c = \phi + \theta V$ (Managerial Compensation)

$a^* \epsilon$ argmax $E(u(c - g(a)))$ (Agent's Problem)

$V = S + b(S)a + \epsilon$ (Firm Value)

g(a): The cost function of the manager's efforts and it is increasing and weakly convex

ϕ : Fixed component of the manager's compensation (Salary)

θ: Variable component of the manager's compensation

w: Reservation wage

ϵ: Random variable with a normal distribution. $\epsilon \sim N(0, \sigma^2)$

Agent utility function from Akerlof and Kranton (2005)[15]

$U(c, a, m) = c\text{-}a + I_m - t_m|\hat{a} - a|$

$\hat{a}$: Ideal effort

I_m : Additional utility due to group identity (or homophily)

Complete Homophily (or in-group): When a manager and board have the social identity

[13] For more details on this topic, refer to Camerer (2003).

[14] For simplicity, I consider risk neutrality and additive preferences and therefore, $u(x) = x$ and $v(c) = c$). For details, refer to Edmans, Gabaix, and Jenter (2017, Pp-39).

[15] I_m is identity utility, which is considered as 0 for simplicity (see, Akerlof and Kranton (2008), pg-213), and $t_m|\hat{a} - a|$ is disutility for diverging from the ideal effort level, denoted by $\hat{a}$. For in-group member, $|\hat{a} - a| > 0$ and for out-group member, $|\hat{a} - a| < 0$.

In-complete Homophily (or out-group): When a manager and board have different social identities.

m: Variable to categorize different social identities such as native language, native place, or caste.

Consider the first-best benchmark where effort is observable. Let a^* be the effort level that the principal wants to implement. Principal can simply direct the CEO to exert effort a^*, and so we can ignore the agent's problem. It is easy to show that the CEO is given a constant wage $c(V) = \bar{c}$, as this leads to efficient risk-sharing. The participatory constraint would become $\bar{c} \geq w + g(a^*)$. Therefore, the principal maximizes,

$$E(V(a^*) - g(a^*) - w)$$

This defines the first-best effort level as,

$$g'(a^*_{FB} = b(S)$$

This shows that the principal trades off the marginal increase in firm value from effort, b(S), with the CEO's marginal cost, $g'(a^*_{FB})$. Nevertheless, usually the effort is unobservable, and therefore, first best is not possible. Hence, I focus on the second-best solution.

Proposition. *The managerial efforts are higher in the case of complete homophily than incomplete homophily.*

Proof:At equilibrium, the managerial efforts for complete homophily is,

$a^*_{in-group} = \frac{1}{1-t_m}[\frac{\theta b(s)}{k} - t_m \hat{a}]$

Similarly, for incomplete homophily case,

$a^*_{out-group} = \frac{1}{1+t_m}[\frac{\theta b(s)}{k} + t_m \hat{a}]$

$a^*_{in-group}$ and $a^*_{out-group}$ depends on the managerial effort which depends on three factors (- sensitivity of effort to identity (t_m), sensitivity of the firm value to effort ($b(S)$), and sensitivity of the managerial compensation (variable component) to firm value(θ)) and their dynamics.[16] Next, I simulate the model with restricting the parameters to compare the $a^*_{in-group}$ and $a^*_{out-group}$. The sensitivity parameters i.e, (t_m), $b(S)$, and θ are restricted between 0 and 1. These restrictions are intuitive considering that firm value could not be more than effort, compensation could not be more than firm value, and effort is not completely driven by group identity. On combining the above two equations and rearranging,

$$(1 - t_m)a^*_{in-group} + (1 + t_m)a^*_{out-group} = \frac{2}{k}(b\theta)$$

As b is sensitivity of the firm value to effort and θ is sensitivity of compensation to firm value. Following the chain rule in derivatives, above equation can be seen as,

$$(1 - t_m)a^*_{in-group} + (1 + t_m)a^*_{out-group} \cong \frac{dc}{da}$$

This shows that impact of a manager's effort on compensation depends on identity overlap between a manager and the board and the sensitivity of manager's utility to the identity i.e. (t_m). In other terms, the impact of a manager's effort on compensation is weighted average of manager's in-group efforts and out-group efforts weighted by sensitivity of manager's utility to the identity.

The $\frac{dc}{da}$ has upper bound of 1 (complete effort convert into compensation) and lower bound of 0 (effort does not convert into compensation at all).

[16] Simulation is performed using the Desmos.com.Graph link:https://www.desmos.com/calculator/eywticjr34

$$0 \leq (1 - t_m)a^*_{in-group} + (1 + t_m)a^*_{out-group} \leq 1$$

Lower bound does provide the real solution as in complete homophily and for incomplete homophily, both $a^*_{in-group}$ and $a^*_{out-group}$ would be 0 and it cannot be the solution. Therefore, from the upper bound,

$$(1 - t_m)a^*_{in-group} + (1 + t_m)a^*_{out-group} = 1$$

For complete homophily, $a^*_{in-group}$ is $\frac{1}{1-t_m}$ and for incomplete homophily, $a^*_{out-group}$ is $\frac{1}{1+t_m}$. As t_m is positive and lies between 0 and 1, this shows,

$$a^*_{in-group} > a^*_{out-group}$$

This implies that in the case of complete homophily, the manager's effort at equilibrium is higher than it is for incomplete homophily.

Lemma 1. *Higher managerial efforts lead to higher firm value at equilibrium for complete homophily compare to incomplete homophily case.*

Proof. As,

$$a^*_{in-group} > a^*_{out-group}$$

Therefore, at equilibrium (for same b(S)),

$$V^*_{in-group} > V^*_{out-group}$$

This implies that a firm in which there is complete homophily has a higher value than a firm with incomplete homophily, this derives from the difference in man-

agerial efforts in both cases.

Lemma 2. *Higher managerial efforts and firm value for complete homophily lead to higher managerial compensations at equilibrium compared to incomplete homophily.*

Proof. From the participation constraint,

$$E[c] = \begin{cases} E[c]_{in-group} = w + \frac{1}{2}k(a^*_{in-group})^2 - \frac{\nu}{2}\theta^2_{in-group}\sigma^2, \text{ For complete homophily} \\ E[c]_{out-group} = w + \frac{1}{2}k(a^*_{out-group})^2 - \frac{\nu}{2}\theta^2_{out-group}\sigma^2, \text{ For in-complete homophily} \end{cases} \tag{3}$$

$$\text{Principal's utility} = E[V - c] = \begin{cases} S + b(S)a^*_{in-group} - E[c]_{in-group} \\ S + b(S)a^*_{out-group} - E[c]_{out-group} \end{cases}$$

$\theta^*_{out-group} > \theta^*_{in-group}$

As, $a^*_{in-group} > a^*_{out-group}$

Therefore, from eq(3), $E[c]^*_{in-group} > E[c]^*_{out-group}$

This theoretical prediction implies that the manager's total compensation is higher for a firm with group identity. This incremental compensation is the cost of the in-group favoritism to the firm.[17]

There is one limitation of this model, such that, the cost of managerial effort can be pecuniary and non-pecuniary. Following the work of Holmstrom and Milgrom (1987), only the pecuniary cost of effort is considered.

In sum, the firm value and the manager's total compensation (and salary) are higher if the manager and the board share the same identity (i.e., group identity or homophily). Although the additional managerial compensation is a cost to the firm due to the in-group favoritism, this cost is compensated for by the manager's extra efforts to increase the firm

[17]The simulation graph is available at following link: https://www.desmos.com/calculator/m7sscqeyie

value. In the next section, I verify whether these theoretical predictions sustain the empirical analysis.

3. DATA AND DESCRIPTIVES

3.1. Data Sources

The surname data is invoked as the key to knowing the identities of both the managers and the directors. I follow on from the sociology, anthropology, and social neuroscience literature. The identity database, which has Indian surnames linked with the native place, native language, and caste, is constructed using the Socio-Economic and Caste Census (SECC) 2011 data of 474 million Indians and the Linguistic Survey of India (LSI) data. Internet Appendix Table IA2 Panel A and Panel B provide a preview of the SECC data with the procedure to extract the surnames. The SECC data, post filtering, includes information on the Indian surnames, their native place (district (county) and state), and caste. The LSI data provides information on the most populous language in the district.[18] On merging these databases using machine learning algorithms, I develop a novel database of 3.78 million Indian last names with multiple identities.[19] This database has information on Indian surnames, native language, caste, and native place. Figure 1 shows the state-wise, language-wise, and caste-wise distribution of 3.78 million Indian last names.

[Insert Figure 1 Here]

To externally validate the surname database developed using the SECC and the LSI, I use an alternate surname dataset from the website "Indianchildnames.com." This dataset

[18] The Appendix provides details on constructing the identity database using the SECC and LSI datasets.

[19] This is inline with Ancestory.com database which suggests there are almost 4 million unique last names in India. See, https://blogs.ancestry.com/cm/indian-surnames-a-mix-of-culture-caste-and-religion/. However, Ancestory.com lacks the identities of these last names.

contains information on 2,225 unique common Indian surnames with their native language and native place.

ProwessDx (CMIE), Global Compustat, and the BoardEx databases are used for the manager and board member names, executive compensation data, firm fundamentals, board characteristics, and managerial demographics. Extracting the surnames from the managers' and board members' names was not straightforward because of the various designations and their patterns, such as Mr./MR/Mr/(MR)/(Mr.)/(Mr), Dr./Dr/(Dr), Prof., Sh./Shri, I.A.S., in addition to the army designations (Maj. (Major), Gen. (General), Col. (Colonel), navy designations (Vice Admiral, Commodore) and many others. An algorithm employed in this work differs from those used in extracting the surnames from the SECC database. Table IA2 Panel C shows the 200 most common Indian surnames among the managers with their ranking. For instance, 'Gupta', 'Jain', 'Agarwal', 'Patel', and 'Shah' are the top five surnames in India managers. I merge prowessDx and identity database using unsupervised machine learning using the surname as an identifier.[20] The Section 4.1. provides a detail discussion of the methodology.

The sample comprises the panel of non-financial firms listed on two leading Indian stock exchanges- the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE), spanned from 2004-2018. There are 2,324 firms in the sample and 28,026 firm-year observations. Table 1 Panel A shows the filtration of observations from the ProwessDx database to form the sample data. As the concept of CEO is new in India, therefore, CEO-equivalent positions such as president, managing director, and chairman are also part of the sample.[21] The yearly distribution of the observations is provided in Appendix Table IA1 and shows

[20]Unsupervised machine learning algorithm is used because of some unknown designations in the ProwessDx and the human error in the census data as some of the surnames in the census data are given as "Kuma r", "Sharm a", "Ra i", "S harma", and similar others.

[21]Bertrand and Hallock (2001) use the CEO, chairman, Managing Director positions for the top corporate jobs. One can argue that the chairman is also a board position; therefore, including it can make identification complicated. To resolve such a concern, an additional robustness test is performed by excluding the observations where the manager is chairman (see Table IA10).

that the sample is balanced.

The main dependent variables are managerial compensation, firm value, and residual firm value (i.e., firm value conditioned on last year's managerial compensation). The compensation data from ProwessDx includes total compensation, salary, bonus, perks, contribution to provident fund (PF), and sitting fees. Figure 3 illustrates the distribution of total compensation and its components. [22]

The main managerial compensation variable is the logarithm of total compensation. However, for external validity, I also have to use the additional set of compensation variables such as the logarithm of salary, logarithm of salary, bonus, and sitting fees, and total compensation relative to industry average. Logarithm forms are used due to the skewness in the distribution. The firm value is measured as the ratio of market price of the equity by the book value of the equity (*Market to Book ratio*). I use the market-to-book ratio because it is a well-established measure for firm value (see Ritter (1981) and Downes and Heinkel (1981)) as well as for firm performance (see McDonald, Khanna, Westphal, 2008; Morck, Shleifer, Vishny, 1989, and similar others). To show how firm value is affected by homophily even after paying the cost of in-group favoritism, the residual firm value is computed. The residual firm value (*Residual Firm Value*) is measured as a residual term by regressing the firm value on last year's managerial compensation. The underlying theory for using it as a proxy of managerial reciprocation is that the determinants of residual firm value should only be correlated with firm and governance characteristics. If homophily explains a part of it. It is due to the manager's extra effort put into the firm. For the robustness check, sales growth (*Sales Growth*) and return on equity (*ROE*) are used as another proxies of firm value. *Sales Growth* is calculated as the growth in sales compared to the previous year. *ROE*

[22] Total compensation paid to an executive can be either salary or a combination of salary and bonus or a combination of salary, bonus and sitting fees. The main component of the total compensation is salary, but salary with bonus and sitting fees cover 90 % of total compensation.

is the ratio of the net income to the total book value of equity.[23] The main independent variables are three homophily (group identity) measures - homophily language (*Homophily Language*), homophily place (*Homophily Place*), and homophily caste (*Homophily Caste*). The homophily measures are dummy variables that take on the value of 1 if the manager and at least one of the board members share the same social identity. Otherwise, it is 0.

For the controls, firm size (*Size*), firm age (*Firm age*), institutional ownership (*Institutional Ownership*), promoters ownership (*Promoters Ownership*), leverage(*Leverage*), volatility (*Volatility*), big 5 auditing firms (*BIG5*), board size (*Board Size*), the number of board meetings (*Number of Meetings*), and return on assets (*ROA*) are used. The data to construct these variables is from ProwessDx (CMIE). *Size* is measured as a log of market capitalization. Market capitalization is in USD. Nevertheless, the way firm size is measured can affect the association between compensation (or firm value) with firm Size (Luo, Liu, and Tripathy (2021)); I also conduct a robustness check using log of assets as firm Size (refer Table IA13 for results with this change). *Firm Age* is the firm's age from the year of incorporation. *Promoters Ownership* is the percentage of promoters' equity holding to the total equity issued by the firm. *Institutional Ownership* is the percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *ROA* is measured as the ratio of net income to total assets. The total debt and total assets are in millions of USD. *Board Size* is the total number of directors on board. *Volatility* is the return volatility. *BIG5* is a dummy variable which takes on the value of 1 if the firm's auditor is from one of the BIG 5 auditing firms - Deloitte & Touche, Ernst & Young, Pricewatehouse Coopers, Grant Thornton, or KPMG; otherwise, 0.[24] *Number of Meetings* is the number of board meetings in a year. The institutional ownership, promoter ownership, board size, the number of board meetings, and the BIG 5 are the internal and

[23]Results of external validity tests are provided in Internet Appendix Table IA 8.

[24]In India, the Big 5 accounting firms also work through their affiliated firms. The information on the affiliated firms has been gathered from the BIG5 auditing firms websites.

external governance characteristics.

Because managerial characteristics like age (*Age*), gender (*Gender*), and education (*Education*) are not available in ProwessDx, this data is gathered from the BoardEx database. Due to the constraints matching these two databases, these characteristics are unavailable for all managers. Therefore, this constraint limits the number of observations for the sub-sample analysis to 3,542.

3.2. Descriptives

Figure 2 show maps showcasing the managers and directors' distribution per their native place in India. These maps exhibit that although directors and managers are from all over India, most of them are originally from southern India.[25]

[Insert Figure 2 Here]

The pie chart in Figure 2 shows that most of the managers and directors in India belong to the upper caste, and only 19-21 % are from a lower caste. This distribution is more skewed at a managerial level than at the director's level. Figure 2 shows the distribution of managers and directors as per their native language and conveys that most of the managers and directors speak Marathi, Tamil, or Western Hindi. This group comprises 14-18% of the sample. Other popular native languages spoken by managers and directors are Malayalam, Gujarati, and Kannada. The native place, native language, and caste form the identities of the managers and the directors. [26]

[25]Indian Media has also pointed out that half of the top ten Indian billionaires figuring in the list of Bloomberg Billionaires are Gujaratis (a state in northern India) and one in three CFOs of Nifty index companies is south Indian. For details check, https://economictimes.indiatimes.com/news/politics-and-nation/stereotypes-dominate-indias-rich-list/articleshow/46857906.cms?from=mdr

[26]Appendix Table IA3 shows the industry-based distribution of the homophily variables. Certain industries such as metal and mining, consumer durables, media print, wine cables, wood products, fertilizers, and similar others have more homophily between a manager and the board than other industries. There can be a concern that homophily constructs might be time-invariant, but it is not valid, as these constructs change with time due to manager or director turnover.

[Insert Figure 3 Here]

Table 1 Panel B reports the summary statistics of different components of managerial compensation in millions of USD. The average (maximum) total compensation of a CEO or CEO equivalent position in India is 192,000 USD (2,461,000 USD).[27] The average (maximum) salary, bonus, sitting fees, and perks are 132,000 (1,116,000), 262,000 (3,663,000), 3,000 (23,000), and 17,000 (138,000) USD. Unlike in the US, the mean variable pay in India is 17% of total remuneration. Summary statistics and distribution of managerial compensation are compared with Hill and Thomas (2014).

Table 1 Panel C show the summary statistics of the variables used for the analysis. The mean (max) of the log total compensation (*Log Compensation*) and industry relative compensation (*Industry Relative Compensation*) are 10.611 (13.634) and 0.003 (3.752). The mean(max) of log salary (*Log Salary*) is 10.4 (13.493). Log compensation, log salary, the industry relative compensation, and log of salary, bonus, and sitting fees are used to demonstrate how executive compensation is affected by in-group favoritism.

[Insert Table 1 Here]

To show how firm value is affected by group identity (homophily), market to book ratio (*Market to Book ratio*) is used. However, for external validation, sales growth (*Sales Growth*) and the sum of the market value of equity and book value of debt (*Log Market Value*) are used as proxies for firm value. The mean (max) value of *Market to Book ratio*, *Sales Growth*, *Log Market Value* are 2.238 (19.260), 0.145 (7.067), and 4.180 (9.610). The mean (max) of *Residual Firm Value* is -0.014 (8.470). The residual firm value captures the impact of the managerial reciprocation of in-group favoritism.

[27] To provide a sense of these numbers compared to US executive compensation, for the period 1982 through 1988, the average salary and bonus for the CEOs of large public companies were 843,000 USD. (Jensen and Murphy, 1990).

The three main independent variables are homophily language, homophily place, and homophily caste. These variables are dummy variables and take on the value of 1 if the board and the manager share the same identity, i.e., native language, native place, or caste. In 40-42% of the sample's observations, the board and the manager speak the same native language or come from the same native place. This factor indicates that the sample is balanced regarding these identities. However, there is skewness in the homophily caste variable. In 89% of the observations, the board and the manager belong to the same caste. As upper-caste individuals dominate the board and managerial positions, the skewness in homophily caste results from it. Table 1 Panel D shows the distribution of Homophily variables in detail. The control variables used are firm size (*Size*), firm age (*Firm age*), institutional ownership (*Institutional Ownership*), promoters ownership (*Promoters Ownership*), leverage(*Leverage*), volatility (*Volatility*), the big 5 auditing firms (*BIG5*), board size (*Board Size*), the number of board meetings (*Number of Meetings*), and return on assets (*ROA*). In the sample, the average firm size, age, and leverage are 16.982, 34 years, and 15.9% respectively. In India, the promoters (founders) own most shares, and the ownership distribution clearly reflects that. The average for promoter ownership and institutional ownership are 54% and 9.5% . The definitions of the variables are provided in Appendix A.

. Regarding the board characteristics, the average board size is 10 members. On average, 27.5% of observations involve BIG5 auditors, and Indian boards conduct 8 meetings per year. Regarding the managerial demographics, the minimum (average) age (*Age*) of a manager is 40 (64) yrs old, and the average qualification (*Education*) is a masters degree. The average number of women on the board (*Gender Diversity*) is only 9-10 %. Panel D shows that the sample is balanced for homophily language and homophily place. Homophily caste is skewed as most managers and directors in India belong to the upper caste. This implies that either lower caste individuals do not get opportunities to reach such posts or they are crowded out due to lobbying by the upper caste managers and directors. Testing the

implications is beyond the scope of this paper but it can be examined in future research. The univariate analysis of the total compensation, firm value, and residual firm value based on the homophily constructs is provided in Online Appendix Table IA4. The univariate analysis supports the main argument that there is a significant difference in executive compensation, firm value, and residual firm value when a manager shares the same identity as the board rather than not. The covariate balance between the control variables for the three homophily constructs is provided in Online Appendix Table IA5. It confirms that there is not much difference in covariates between the comparison groups.

4. METHODOLOGY

4.1. Surname Matching Using Unsupervised Machine Learning

The lack of a common identifier between the identity database and ProwessDx generates the usage of a text similarity algorithm from the unsupervised machine learning literature. It is hard to exclude less-known designations from the names of the managers and directors because banning those designations can also lead to a drop in observations and would generate a measurement error.[28] Therefore, to avoid that, a text similarity algorithm is used for merging. Jaccard algorithm, a well-known text similarity algorithm, is used to discern a manager's identity and the remaining board members.[29] For matching, surnames are first divided into bi-grams, which are used for matching and computing similarity scores. The Jaccard similarity score is the ratio of the number of common bi-grams in two surnames

[28] For instance, 'Sharma' is a well-known Indian surname. However, the data has various other forms of this surname such as'S harma', 'sharm a', 'Sharmams' etc. These erroneous surnames have formed due to man-made errors when doing the entry in the census report.

[29] Another common text similarity algorithm is the Cosine algorithm. These two algorithms give different results if the two documents or sentences in question are too different in length. In this case, both give similar results as the comparison is between the two surnames rather than between the two documents.

divided by the union of bi-grams across the two surnames.

$$\text{Jaccard Similarity} = (A \cap B)/(A \cup B) \tag{1}$$

Here, A and B are two surnames.

It can be understood by computing the Jaccard similarity score of 'Smith' and 'Thomas'.For instance, the bi-grams of surnames such as 'Smith' and 'Thomas' are $\{sm, mi, it, th\}$ and $\{th, ho, om, ma, as\}$. As there is only one common bi-gram –'th', between these two surnames, therefore, $A \cap B$ is 1. The total unique bi-grams between them –'$A \cup B$', is 8. This means that the Jaccard similarity score is 0.125. This is quite low because the surnames in this example are too discrete. For exact matching, the threshold of the similarity score is kept at 0.9 and above. To confirm the matching, a manual check was also conducted.

4.2. Empirical Specification

To test the role of group identity (or homophily) in the executive compensation setting, the regression of the log of total compensation on homophily constructs ran, controlling for firm characteristics, governance characteristics, and managerial demographics. [30] The empirical specification for this test is:

$$LogCompensation_{i,t} = \alpha_{i,t} + \beta_1 Homophily_{i,t} + \beta_2 X_{i,t} + \beta_3 L_{i,t} + \beta_4 Z_{i,t} + \gamma_{firm} + \gamma_{year} + \gamma_{designation} + \epsilon_{i,t} \tag{2}$$

For the external validation, in place of a log of total compensation, I also use log of salary, total compensation relative to the industry peers, and log of salary plus bonus plus sitting

[30] As the number of observations available for the managerial demographics is limited; a sub-sample analysis is provided controlling for the vectors of the managerial demographics.

fees. Homophily (or group identity) is a dummy variable that takes a value 1 if a manager and at least one board member share the identity; otherwise, 0. I separately tested three natural identity traits such as native language, same caste, and native place. For external validation, I also measure homophily differently. This paper uses a simple baseline homophily measure to estimate in-group favoritism.

The detailed literature surveys on executive compensation by Core, Guay, and Larcker (2002), Frydman and Jenter (2010), and Edmans, Gabaix, and Jenter (2017) suggest that the main determinants of executive pay are the firm fundamentals, governance characteristics, and manager demographics. Following that suggestion, X is a vector of firm characteristics such as firm size, firm age, return on assets (ROA), leverage, and volatility. L is the vector of the governance characteristics such as board size, the number of board meetings, one of the BIG 5 auditors, institutional ownership, and promoter ownership. Z is the vector of the manager's demographics (such as education, age, gender, and others). To address the impact of the unobservables, firm fixed effects γ_{firm} and year fixed effects γ_{year} are used. As the results can be affected due to the managerial position, $\gamma_{designation}$ fixed effects are also used. As the managerial demographics are not available for the full sample, the subsample results control for the manager's age, gender, and education. These results are provided in Internet Appendix Table IA6. The following empirical specification is implemented to check the impact of homophily on firm value:

$$FirmValue_{i,t} = \alpha_i + \beta_1 Homophily_{i,t} + \beta_2 X_{i,t} + \beta_3 L_{i,t} + \gamma_{firm} + \gamma_{year} + \nu_{i,t} \qquad (3)$$

The coefficient of interest is β_1. As homophily is based on language, place, and caste, β_1 can be seen as $\beta_{Language}$, β_{Place}, and β_{Caste}. Firm Value is measured as the market-to-book ratio. X and L are the firm characteristics and governance characteristics listed earlier. Based on the previous literature, all βs of the firm fundamentals and governance are expected

to be positive except for volatility and leverage (Hovakimian, Hovakimian, and Tehranian (2004)). Next, I decompose the firm value explaining last year's managerial compensation and the residual firm value. The impact of homophily on the residual firm value is examined to check the reciprocation of the manager towards in-group favor. The residual firm value is an error term of the regression of firm value on the last year's managerial compensation.

$$FirmValue_{i,t} = \omega_i + \phi_i LogCompensation_{i,t-1} + \mu_{i,t} \tag{4}$$

The residual firm value (*Residual Firm Value*) is μ. It captures the firm value conditioned on managerial compensation. An increase in residual firm value proxies a higher managerial effort to increase the firm value. To test whether the residual firm value (*Residual Firm Value*) is positively associated with the homophily variables or not, the following regression specification is used,

$$ResidualFirmValue_{i,t} = \alpha_i + \beta_1 Homophily_{i,t} + \beta_2 X_{i,t} + \beta_3 L_{i,t} + \gamma_{firm} + \gamma_{year} + \nu_{i,t} \tag{5}$$

The interpretation and expectations of the βs are the same as for the firm value regression. If the firm characteristics and governance characteristics only explain the residual firm value, we expect β_1 to be statistically insignificant. However, if β_1 is positive and significant, the manager exerts an extra effort to increase the firm value more than what he is paid for. This change happens due to the shared group identity with the board.

4.3. Endogeneity Concerns and Identification Strategy

Two econometric issues with the above described empirical specifications are essential to be addressed. First, the executive compensation and firm value are endogenously determined, and therefore, there is a possibility that the described empirical specifications do not capture

it completely. Second, as the firm hires a particular manager due to an unobservable reason, there is a self-selection problem. Following Palia (2000) example, I use the two-stage least square method to mitigate the simultaneity issue. Similarly, to minimize the self-selection bias, following the example of Fracassi and Tate (2012), I use death and the retirement of a director as an exogenous shock to board identity and to the homophily constructs.

To implement the two stage least square method, the fitted value of $LY_{i,t}$ from eq (6) is used in eq (7).

$$LogCompensation_{i,t} = \alpha_{i,t} + \beta_1 Homophily_{i,t} + \beta_2 FirmValue_{i,t} + \beta_3 X_{i,t} + \beta_4 L_{i,t} + \beta_5 Promoter\ Dummy + \gamma_{firm} + \gamma_{year} + \gamma_{designation} + \epsilon_{i,t} \quad (6)$$

$$FirmValue_{i,t} = \alpha_i + \beta_1 Homophily_{i,t} + \beta_2 LogCompensation_{i,t} + \beta_3 X_{i,t} + \beta_4 L_{i,t} + \gamma_{firm} + \gamma_{year} + \epsilon_{i,t} \quad (7)$$

The coefficient of interest is β_1. The promoter dummy (Promoter Dummy) is used as an instrument that takes value 1 if the manager is also a firm's promoter; otherwise, 0. In the first stage, the estimated regression is the logarithm of the total managerial compensation based on the promoter dummy as an instrument and all of the firm and governance level controls. In the second stage, the estimated regression looks at firm value concerning three group identity measures controlling for the fitted value from the first stage regression with other controls.

As the board can select a manager of the same identity to reduce the frictions in the corporate decision-making (Damaraju and Makhija, 2018), this can affect the causal inference of the results. I use a similar identification strategy to mitigate this concern in Fracassi and Tate (2012). The director's death or retirement is used as an exogenous shock as these events are unrelated to the firm's conditions. Voluntary retirement is not included as this can be due to firm-related issues.[31] As ProwessDx does not have information on the director's date of

[31]There is a possibility that even after the director's death or retirement, the board identity remains the

death, this information is collected from the BoardEx database. The empirical specification for this test is,

$$Y_{i,t} = \alpha_i + \beta_1 Shock + \beta_2 Homophily_{i,t} + \beta_3 Shock\text{x}Homophily_{i,t} + \beta_4 X + \beta_5 L + \gamma_{firm} + \epsilon_{i,t} \quad (8)$$

Here Y is the log compensation (*Log Compensation*), firm value (*Firm Value*), or firm value conditioned on managerial compensation (*Residual Firm Value*). *Shock* is a dummy variable that takes a value 1 if any director has died or retired in that firm year; otherwise, 0. X is a vector of firm fundamentals and L is a vector of governance characteristics. The coefficient of interest is β_3. The firm (γ_{firm}) fixed effects is used to control for the firm-level unobservables. As time-fixed effects can affect the coefficient of the *Shock* dummy variable, it is not included in this specification.

5. RESULTS AND DISCUSSION

5.1. Group Identity, Executive Compensation, and Firm Value

Table 2 Columns (1) - (3) show the baseline results of how group identity (or homophily) affects managerial compensation. The results show that when a manager and the board share the same native language or come from the same native place, it helps the manager to earn higher compensation. The impact of group identity is stronger for native language than the native place as $\beta_{Language} > \beta_{Place}$. The $\beta_{Language}$ is 0.066, and the β_{Place} is 0.071, but the difference is not statistically significant. When the board and the manager are of the same caste, homophily negatively affects the total compensation. However, the latter result is insignificant post endogeneity test. Nevertheless, this result is similar to Freshtman and Gneezy (2001) work, where the behavioral experiment's predictions found that Eastern

same. Therefore, I ran the additional test as diff-in-diff between two groups where, in one group, board identity (and hence, the homophily) changed after the exogenous shock compared to the group where board identity didn't change. The results for this test are available in Internet Appendix.

Jews discriminate against their people. However, this result can also be due to the skewness in the caste variable towards the upper caste. Managerial designation, firm, and year fixed effects are used to control the impact of the unobservables.

[Insert Table 2 Here]

Table 2 Columns (4)-(6) show the impact of group identity on executive compensation with controls. These results confirm that if the manager and the board speak the same native language or belong to the same native place, the manager earns higher compensation than their peers. $\beta_{Language}$ and β_{Place} are 0.078 and 0.077, and both coefficients are significant at 1 % level. This result implies that in a firm where the board is more homophilic to the manager due to native language or native place, the manager earns a 7- 8% higher level of compensation. For external validation, I replace total compensation with salary, total current compensation (salary, bonus, and sitting fees), and compensation relative to industry peers, and confirm the main results. Detailed results for external validation are provided in Table IA7 Panel (A)-(C).The coefficient of the homophily caste shows that the variable is negative and significant, but it is hard to infer as the main variable is highly skewed. However, in untabulated results, I check the results for the upper caste and lower caste group separately and found that the coefficient of homophily language and place are positive and significant. Among the controls, *Size*, *ROA*, *Institutional Ownership*, *Promoters Ownership*, *Number of Meetings*, *Firm Age*, and *BIG5* are significant and positively associated with managerial compensation. As expected, the coefficients of *Leverage* and *Volatility* are significant and negatively associated with compensation. The direction and magnitude of the controls' coefficients are in line with the findings in the executive compensation literature. I confirm these findings using a different database (based on "Indianchildnames.com") to measure homophily constructs and using different definitions of homophily. Detailed results for the same are provided in Table IA7 Panel D and Table IA8 Panels (A)-(B).

Table 3 Columns (1)-(3) exhibit the impact of homophily on firm value. The results show that homophily in the native language is positively associated with firm value, and the results are significant at the 5% level. Similar to the executive compensation results, homophily based on native language is stronger than homophily based on native place, i.e., $\beta_{Language} > \beta_{Place}$. $\beta_{Language}$ is 0.125, and it is significant at the 1% level. β_{Place} is 0.058, but it is significant at 12%. This result implies that firms in which there is more homophily between a manager and the board have a 12.5 % higher value. Collectively, the 1% increase in executive compensation due to group identity leads to 1.8% increase in the firm value. Among the controls- *Size*, *Institutional Ownership*, *Promoters Ownership*, *Number of Meetings*, and *BIG5*, are significant and positively associated with firm value. The coefficient of *Leverage* and *Volatility* are significant at the 5 % level and negatively associated with firm value. To avoid the time trend and firm-level unobservables, firm and year fixed effects are used.For external validation, I replace market to book value of equity with two other firm value measures - sales growth and market value of equity plus book value of debt, and confirm the results. Detailed results for external validation are provided in Table IA7 Panel (A)-(C).

[Insert Table 3 Here]

Table 3 Columns (4)-(6) show the results of group identity on residual firm value. The results indicate that homophily due to language is positively associated with residual firm value. Similar to the previous results, the impact of homophily based on native language is stronger than homophily based on native place, i.e., $\beta_{Language} > \beta_{Place}$. $\beta_{Language}$ is 0.095, which is significant at 1 % level. β_{Place} is 0.056, which is significant at 15 %. This result implies that firms with more homophily between a manager and board have a 5.6-9.5% higher residual value, i.e., higher firm value excluding the cost of in-group favoritism. As managerial demographics are unavailable for the full sample, I show the results (Table IA6)

using sub-sample analysis. I find that male manager earns higher compensation than a same-level female manager, which is significant at 1% level. The manager's age and education do not affect the total compensation significantly.

As there is a possibility that the results appeared due to the specific dependent variables or the way homophily variables are constructed. To mitigate such concerns, I replace log total compensation with log salary, compensation relative to the industry average, and log salary plus bonus plus sitting fees. Market to book value is replaced with sales growth, ROE, log of market value of equity plus book value of debt. I also estimated new homophily constructs based on 'indianchildnames.com' database. The external validity test confirms the main findings of the paper. The results for the external validity tests are provided in Internet Appendix Table IA7.

5.2. Endogeneity Tests Results

The endogenous relationship between executive compensation and corporate governance is always an issue (Hermalin and Weisbach, 1998; Wintoki et al., 2012) for this research question. As executive compensation and firm value are simultaneously determined (Palia (2001)), I use two-stage least square (2SLS) regression to show the relation between firm value conditioned on compensation with the homophily constructs. The results for the second stage regression are provided in Table 4 Panel A. The fitted value of log compensation from first stage is negatively associated with firm value, which supports the notion that managerial compensation is a cost to the firm. $\beta_{Language}$ is 0.124 and it is significant at 1%. β_{Place} is 0.096 and it is significant at 10%. The coefficient of the fitted value of the total compensation is 0.570, which is significant at 1 %. These results confirm the paper's main findings that homophily increases the firm value even when the cost of managerial compensation is considered. Internet Appendix Table IA12 provides the test statistics and p-values of the Durbin (score) and Wu-Hausman tests. These statistics confirm that 2SLS is more efficient

than the ordinary least square (OLS) method in the case of the self-selection problem. [32]

[Insert Table 4 Here]

There is a possibility that the board chooses a manager that matches their identity just to avoid conflict regarding corporate decision-making (Damaraju and Makhija, 2018). Following the example of Fracassi and Tate (2012) to mitigate self-selection bias, the director's death or retirement is used as a shock to the board identity and hence to the homophily variables.

Table 4 Panel B reports that the $\beta_{Language}$ for log compensation and residual firm value are 0.045 (significant at 5 %) and 0.065 (significant at 1 %) and the same for β_{Place} are 0.062 (significant at 5 %) and 0.042 (significant at 10.6 %).This confirms the paper's main results that the group identity between a manager and the board based on native language or native place leads to higher managerial compensation. Although the increment in managerial compensation is a cost to the firm, group identity also incentives the manager to put more effort toward the firm's growth and thus increase its value.

5.3. Additional Hypotheses

Powerful CEO

A powerful CEO influences the level of compensation and increases the firm performance variability (Adams et al. 2005). To mitigate the possibility that results are driven by the position of the CEO rather than homophily. I also control for the manager's influence in the organization. Following Adams et al. (2005), I measure a powerful manager (Powerful CEO) as a dummy variable that takes value 1 if either manager is also one of the founders

[32]The difference between the Durbin (1954) and Wu–Hausman ('T-2' statistics of Wu(1974)) tests of endogeneity is that the former uses an estimate of the error term's variance based on the model assuming the variables being tested are exogenous, while the latter uses an estimate of the error variance based on the model assuming the variables being tested are endogenous. Under the null hypothesis that the variables being tested are exogenous, both estimates of the error variance are consistent.

(promoters) as well as the chairman of the board. Results are provided in Table 5, and it shows that even controlling for the manager's influence in the organization, homophily due to language and place are still positively associated with executive compensation and residual firm value. These findings support the main results of the paper. The $\beta_{Language}$ for log compensation and residual firm value are 0.078 (significant at 1 %) and 0.095 (significant at 1 %) and the same for β_{Place} are 0.077 (significant at 1 %) and 0.056 (significant at 9.95 %). I also find a positive and statistically significant relation between powerful manager and executive compensation but no relation between powerful manager and firm value. This result implies rent extraction due to the manager's influence on the organization's decision-making process.

[Insert Table 5 Here]

College Ties and Past Experience

Farcassi and Tate (2012) show that powerful CEOs are likelier to appoint directors with ties to the CEO and that CEO-director ties reduce the firm value. There is a possibility that an increase in compensation, firm value, and residual firm value is not due to the homophily based on native language or native place but due to the induced identities such as college ties or past employment between a manager and the directors. I control for college ties and the past employment connections between a manager and the directors to address this concern. Using BoardEx's education database and the network database, I created the variables based on college ties (*College Ties*) and past experiences (*Experience Ties*). *College Ties* is a dummy variable that takes the value 1 if there is at least one director on the board from the same college as the manager; otherwise, 0. Similarly, *Experience Ties* is a dummy variable that takes the value 1 if the manager has experience working with at least one board director; otherwise, 0.

[Insert Table 6 Here]

The results for this test are shown in Table 6. Panel A shows the results when controlling for college ties. $\beta_{Language}$ is 0.131, and β_{Place} is 0.126 for compensation, and both are statistically significant. $\beta_{Language}$ is also positive and significant for residual firm value. It implies that homophily based on the native language between a manager and the board increases the executive compensation and residual firm value, even controlling for college ties. The college ties between a manager and the board negatively impact compensation and residual firm value. Panel B shows the results controlling for ties from past employment. $\beta_{Language}$ and β_{Place} are positive and statistically significant for executive compensation and residual firm value. This confirms the paper's main results that homophily based on native language and native place increases executive compensation and reduces agency frictions by increasing the residual firm value. The results also show that ties due to past employment reduce total compensation, firm value, and residual firm value. These results are in line with the work of Farcassi and Tate (2012).

Promoter as a Manager

In India, the promoters (or founders) have a stronger influence on firms as they are the major shareholders. Therefore, it is crucial to check whether the results are different for firms where one of the promoters is a manager. As this information is not available for the full sample, a sub-sample test is undertaken. The results are provided in Table 7. Column (1) and (2) show that there is no impact due to homophily on compensation. These results are intuitive as promoters have no incentive to earn extra compensation based on their identity as they hold the majority of shares in the firm. Column (4)-(5) show that homophily based on native language and native place increases firm value by 18.6 % and 15.6 %. This result is higher compared to the full sample results, implying that the promoter being a manager, governs the firm better and undertakes constructive investment decisions. Column (7)-(8) show that the homophily due to language and place increases the residual firm value by 14.6

% and 11.2 %. This implies that even after paying the cost of in-group favoritism, the firm value for these homophily variables increases. This result is prominent in the cases where the manager is also one of the promoters.

[Insert Table 7 Here]

Business Group vs. Non-Business Group Firms

India's business groups are collections of publicly traded firms with a presence in different industries and a significant amount of common ownership and control, usually by a family. Tarun and Palepu (2000) compare the performance of business group firms with the non-business group firms in India. They show that business group firms outperform non-business group firms. They argue that the group-affiliated firms have disproportionately good access to international sources of capital due to the business group's reputation.

The current study proposes an alternative channel for the better performance of group affiliated firms', i.e., manager's reciprocation towards a firm due to homophily. The results are provided in Table 8. I do not find that homophily affects managerial compensation differently in business group firms than in others. However, its impact on firm value and residual firm value is more prominent in business group firms. The coefficient of interests, $\beta_{BGxHomophilyLanguage}$ and $\beta_{BGxHomophilyPlace}$, for residual firm value are 7.3 % (significant at 10%) and 4.9 % (significant at 25 %). In business group firms, the managers mainly hire from the same family or someone known to the promoter (founder). Therefore, the manager's reciprocation towards the firm increases its value. Still, the increment in compensation due to homophily is not different from non-business group firms.[33]

[Insert Table 8 Here]

[33]The empirical specification for these regressions does not include firm fixed effects as BG dummy variable is used as a firm fixed effect.

Social Independence of Board

Hwang and Kim (2009) show that social ties between a manager and independent board members affect how directors monitor and discipline the CEO. Using the US data, they show that the board which is conventionally independent but socially connected with the manager provides higher compensation to him. However, unlike this paper, authors concentrated only on independent board members. I replicate their test for Indian corporate and find that social ties between a manager and the independent board members do not affect the managerial compensation significantly but negatively affect the future firm value and residual firm value. The coefficient of $\beta_{Language}$ and β_{place} for firm value are 10.5 % and 11.2 %, and both are significant at 5 %. This supports the conclusion of Hwang and Kim (2009) that the independent board members should be both convenient and socially independent to monitor and discipline the manager. Results are provided in Table 9.

[Insert Table 9 Here]

Group Identity in Complex Firms

The complex firms are difficult to govern; there is a possibility that the requirement of a managerial skillset dominates the group identity. In that case, the latter factor would not lead to favoritism. Following Craswell, Stokes, and Laughton (2002) test this theory, I measure firm complexity as the square root of the number of subsidiaries. The regression results of managerial compensation, firm value, and residual firm value on firm complexity and its interaction with group identity variables are provided in Table 10.

[Insert Table 10 Here]

The coefficient of *Complexity* is positive for log compensation, implying that the manager is overpaid to handle the complex firms. However, the coefficient of *Firm Complexity X*

Homophily Language is negative 0.129 at 1% significance level signifying that the group identity in complex firms reduces managerial compensation. However, the firm value and residual firm value increase by 12.8 % and 8.4 % suggesting that it is beneficial for these firms to hire a manager having a group identity with board members as it increases future firm value and the manager is paid less than their counterparts.

Group Identity in Government Firms

A manager's recruitment process and working culture are different in government firms than in others.[34] Therefore, it would be interesting to know whether group identity works in this type of organization. For simplicity, I focus on public sector undertakings (PSUs) which are government commercial enterprises with autonomous power to take decisions up to the non-management level. Results are provided in Table 11. *PSU* is a dummy variable that takes a value 1 if a firm is center or state government's commercial enterprise; otherwise, 0. I find that PSUs pay less to managers, and their firm value is also less than the other firms. It is intuitive as PSUs provide many non-pecuniary benefits to the managers, and it is hard to fire them from the post. A stable job and less incentive to take a risk by the manager reduces the future firm value. I find no impact of group identity on managerial compensation and firm value in PSUs. These results align with intuition as government involvement in the managerial recruitment process is unbiased from any implicit stereotypes.

[Insert Table 11 Here]

Robustness Checks

Additional tests are conducted to explore the robustness of the results. First, I examine whether the cultural differences between the southern and northern region of India influences

[34] In India, the manager and the directors are hired by the government and the recruitment process is considered to be unbiased.

the homophily between a manager and directors. I have not found evidence that the results will be different if a manager and the board members are from the southern region compared to a situation when both are from the northern region. Second, I examine whether the migration of communities (group belonging to a particular surname) from one region of India to another affects the results. Only one community, 'Kashmiri Pandits', migrated from the Kashmir region to other parts of India in 1991 due to riots. I exclude the surnames belonging to this community and run the same regressions. Results are robust to this test (refer to Table IA14). Third, as the principal and an agent have multiple identities and bias can exist due to any of it. Therefore, rather than using a particular homophily construct, I use a homophily dummy variable which takes value 1 if all the three identities match between a manager and at least one of the board members (refer Table IA8 Panel A). Results suggest that the managerial compensation increases by 2.9 % due to the homophily index, which implies the cost of in-group favoritism. However, the net firm value also increases by 2.7 %. Fourth, I examine whether the homophily provides a higher return to equity holders or not by using return on equity *ROE* as a dependent variable. $\beta_{Language}$ and β_{Place} are positive and significant for ROE and Residual ROE which are in line with the main findings of the paper (refer Table IA9). Lastly, following the example of Bhagat and Bolton (2008), I use current and future return on assets (ROA) as dependent variables to check the impact of homophily on profitability. The results (Internet Appendix Table IA11) suggest that homophily language and homophily place are positively associated with current and future ROA. These findings are in contrast with Cai et al. (2013), which shows that profitability reduces if a manager and owner are from the same family. I presume that these differences in the results arise due to the differences in the governance system. The high promoter ownership maintains a higher level of scrutiny over the manager, which leads to higher profitability in the case of homophily between the manager and board members. The results table for the robustness tests is provided in the Internet Appendix (IA).

6. HOMOPHILY CONSTRUCTS AND FIRM VALUE: A CHANNEL

Managers and directors have multiple identities, such as age, sex, religion, race (or caste), affiliation with a native place, native language, and many others. Why, then, is homophily based on some identities affect the firm decision (such as selecting the manager (or director), executive compensation, job assignments, etc.) and firm performance (or value) and not all of them? The reason for this is 'trust' and supervision (governance). As the underlying element of homophily is trust, homophily based on different identities reflects different levels of trust. If a firm hires a trustworthy manager under the proper supervision (or governance), the manager exerts an additional effort to increase the firm value. The firm recruits this manager even if the firm has to pay extra compensation to do so.

An alternative view of trust is the reduction in information asymmetry. Individuals who are similar in their personal or social characteristics and share a common identity (e.g., race, gender, culture) tend to have common interests and worldviews, which explains the formation of ties based on interpersonal attraction (McPherson, Smith-Lovin, and Cook, 2001). For this reason, the board members understand the manager well, leading to better monitoring and a higher likelihood of the manager's refraining from shirking because of possible sanctions from the board. In other words, trust and reduced information asymmetry imply lower agency and monitoring costs. Using a survey of 600 Chinese private family firms, Cai et al. (2013) show that family managers earn 18-22 % higher salaries and receive more bonuses but face weaker incentives. Using the demographic similarity between the CEO and the directors of US firms, Westphal and Zajac (1995) provides evidence that similarity results in more generous CEO compensation contracts. Fracassi and Tate (2012) show that the CEO-director ties based on shared experiences (such as past employment) reduce the firm value, particularly in the absence of other governance mechanisms to substitute for board oversight. The main results of this paper are that manager gets a 6-8% higher level of compensation based on

homophily such as native language and a native place, and even after paying a higher level of compensation, such firms have an 8-9% more increased value. Managerial efforts can drive the latter result due to the promoters' intense supervision as they are major equity holders in Indian firms (Khanna and Palepu (2000)).

In sum, I argue that as homophily based on different identities reflects different levels of trust relationships, the level of association with the executive compensation can be different. Most lead to high executive compensation due to favoritism. However, different shared identities are differently associated with the firm outcomes due to weaker or stronger governance.

7. CONCLUSION

This paper examines how group identity due to implicit bias affects executive compensation, firm value, and agency frictions. I begin by providing an intuitive stance on how identities can influence the disagreements between a manager and the board by introducing Akerlof and Kranton's (2005) suggested identity-based agent's utility function in the first-order principal-agent model of Holmstrom (1979). The change in the Holmstrom (1979) framework shows that the manager gets a higher level of compensation due to in-group bias. The same factor motivates the manager to exert more effort for the firm. This additional effort increases the higher future firm value compared to its counterparts.

To test this empirically, I use the information on 474 million Indians from the socio-economic caste census (SECC) and the linguistic survey of India (LSI) databases to develop a novel database of 3.78 million Indian surnames with multiple identities. This database has information on last names, native language, native place, and the caste. The Indian surnames provide information on the native language, native place, caste, and other identity facets of an individual; therefore, it is a single source of multiple identities. Mapping this data with the directors' and managers' surnames reveal their identities. Using this data, I analyze

whether manager gains total compensation due to group identity and whether such firms are better-off (or worse-off) for the same reason. I found that the manager gets 6-8% higher compensation if he speaks the same native language or belongs to the same native place as the board members. This result implies the cost of in-group favoritism. Furthermore, I also show that such firms have an 11-12% higher value than their counterparts. Even adjusting the cost of in-group favoritism, the firms have a 5-8% higher value (measured as residual firm value. This outcome demonstrates that the manager exerts an additional effort to increase the firm value due to the shared group identity. These results are robust to external validity tests, endogeneity tests, geographical variation in homophily, induced identities (such as college-level and past employment ties), manager's influence on the firm, and different homophily measures and various other checks.

To check further, I explore these findings in different organization types (such as business group firms, complex firms, and government firms). I found that the increment in managerial compensation due to in-group favoritism is insignificantly lower in business group firms than its counterparts, but the residual firm value is higher. This result signifies that managerial reciprocation toward a firm due to homophily is higher in group-affiliated firms than in others. As complex organizations are challenging to govern, I find that the firm overpays a manager for this. Surprisingly, group identity negatively affects the compensation in such firms. This result suggests the significance of merit over group identity in complex firms. The firm value and residual firm value are higher, indicating that it is beneficial for these firms to hire a manager with a group identity with board members as it increases the future firm value, and the manager is paid less than their counterparts. Lastly, public sector undertakings (PSUs) are government commercial enterprises with autonomous power to take decisions up to the non-management level. I find no impact of group identity on managerial compensation and firm value in PSUs. These results are intuitive as government involvement in managerial recruitment is considered unbiased from any implicit stereotypes.

In India, the ownership structure in many firms is skewed as promoters own the majority of shares. This institutional arrangement suggests that a promoter is crucial to the organization. Therefore, I also explore the effect of group identity on compensation and firm value when a manager is a promoter. I find no difference in promoter's compensation due to group identity but firm value and residual firm value increases. This result is intuitive as the promoter is the major owner of the firm; therefore, for him, the firm value is more crucial than an increment in compensation.

According to Hwang and Kim (2009), the independent board members should be socially independent of the manager to execute the monitoring and advisory roles of the board without any hindrance. They use induced identity measures to test their theory. It is imperative to check their results in a different setting with natural identity measures for generalization. Therefore, I examine how the manager's group identity with independent directors affects his compensation and future firm value. I find that the manager's social ties with independent directors do not influence the compensation but negatively affect the future firm value. These results align with Hwang and Kim (2009), which shows that their results are valid in different settings.

In essence, this article provides evidence of higher managerial compensation due to in-group favoritism and the increase in future firm value due to the manager's reciprocation. The latter results support the notion that group identity is not just a hidden cost but can reduce agency frictions. It supports the notion that statistical discrimination can be efficient and welfare improvement where group identity reflects information gained on a manager's specialization. These results have comprehensive implementation as the principal-agent problem exists in various institutions.

References

Adams, Renée B., Heitor Almeida, and Daniel Ferreira. "Powerful CEOs and their impact on corporate performance." The Review of Financial Studies 18.4 (2005): 1403-1432.

Akerlof, George A., and Rachel E. Kranton. "Economics and identity." The Quarterly Journal of Economics 115.3 (2000): 715-753.

Akerlof, George A., and Rachel E. Kranton. "Identity and the Economics of Organizations." Journal of Economic Perspectives 19.1 (2005): 9-32.

Akerlof, George A., and Rachel E. Kranton. 2008. "Identity, Supervision, and Work Groups." American Economic Review, 98 (2): 212-17.

Akerlof, George A., and Rachel Kranton. "Identity economics" The Economists' Voice 7.2 (2010).

Allport. "The nature of prejudice." Addison-Wesley Press, (1954).

Arrow, Kenneth. "The Theory of Discrimination." Discrimination in Labor Markets (1973): 3-33.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. "Social connections and incentives in the workplace: Evidence from personnel data." Econometrica 77, no. 4 (2009): 1047-1094.

Becker, Gary S. "The economics of discrimination."The University of Chicago Press, (1957).

Bertrand, Marianne, and Sendhil Mullainathan. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." American Economic Review 94.4 (2004): 991-1013.

Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan. "Implicit discrimination." American Economic Review 95.2 (2005): 94-98.

Bertrand, Marianne, and Kevin F. Hallock. "The gender gap in top corporate jobs." ILR Review 55.1 (2001): 3-21.

Bhagat, Sanjai, and Brian Bolton. "Corporate governance and firm performance." Journal of corporate finance 14, no. 3 (2008): 257-273.

Brown, Rupert. Group processes: Dynamics within and between groups. Basil Blackwell,(1988).

Cai, Hongbin, et al. "Family ties and organizational design: Evidence from Chinese private firms." Review of Economics and Statistics 95.3 (2013): 850-867.

Carlana, Michela. "Implicit stereotypes: Evidence from teachers' gender bias." The Quarterly Journal of Economics 134.3 (2019): 1163-1224.

Cikara, Mina, and Jay J. Van Bavel. "The neuroscience of intergroup relations: An integrative review." Perspectives on Psychological Science 9.3 (2014): 245-274.

Chen, Yan, and Sherry Xin Li. "Group identity and social preferences." American Economic Review 99.1 (2009): 431-57.

Chen, Roy, and Yan Chen. "The potential of social identity for equilibrium selection." American Economic Review 101.6 (2011): 2562-89.

Craswell, Allen, Donald J. Stokes, and Janet Laughton. "Auditor independence and fee dependence." Journal of accounting and Economics 33, no. 2 (2002): 253-275.

Downes, David H., and Robert Heinkel. "Signaling and the valuation of unseasoned new issues." The Journal of Finance 37, no. 1 (1982): 1-10.

Zhu, Christina. "Big data as a governance mechanism." The Review of Financial Studies 32, no. 5 (2019): 2021-2061.

Edmans, Alex, Xavier Gabaix, and Dirk Jenter. "Executive compensation: A survey of theory and evidence." The Handbook of the Economics of Corporate Governance. Vol. 1. North-Holland, 2017. 383-539.

Fama, Eugene F. "Agency problems and the theory of the firm." Journal of political economy 88, no. 2 (1980): 288-307.

Fershtman, Chaim, and Uri Gneezy. "Discrimination in a segmented society: An experimental approach." The Quarterly Journal of Economics 116, no. 1 (2001): 351-377.

Fisman, Raymond, et al. "Social ties and favoritism in Chinese science." Journal of Political Economy 126.3 (2018): 1134-1171.

Fisman, Raymond, Daniel Paravisini, and Vikrant Vig. "Cultural proximity and loan outcomes." American Economic Review 107.2 (2017): 457-92.

Fracassi, Cesare, and Geoffrey Tate. "External networking and internal firm governance." The Journal of finance 67, no. 1 (2012): 153-194.

Giannetti, Mariassunta, and Mengxin Zhao. "Board ancestral diversity and firm-performance volatility." Journal of Financial and Quantitative Analysis 54, no. 3 (2019): 1117-1155.

Goldstein, Itay, Chester S. Spatt, and Mao Ye. "Big data in finance." The Review of Financial Studies 34, no. 7 (2021): 3213-3225.

Gompers, Paul A., Vladimir Mukharlyamov, and Yuhai Xuan. "The cost of friendship." Journal of Financial Economics 119.3 (2016): 626-644.

Grinblatt, Mark, and Matti Keloharju. "How distance, language, and culture influence stockholdings and trades." The Journal of Finance 56.3 (2001): 1053-1073.

Hill, Jennifer G., and Randall S. Thomas, eds. Research Handbook on Executive Pay. Edward Elgar Publishing, 2012.

Holmstrom, Bengt. "Moral hazard and observability." The Bell Journal of Economics. (1979): 74.

Hovakimian, Armen, Gayane Hovakimian, and Hassan Tehranian. "Determinants of target capital structure: The case of dual debt and equity issues." Journal of Financial Economics 71, no. 3 (2004): 517-540.

Hwang, Byoung-Hyoun, and Seoyoung Kim. "It pays to have friends." Journal of financial economics 93, no. 1 (2009): 138-158.

Jensen, Michael C., and William H. Meckling. "Theory of the firm: Managerial behavior, agency costs and ownership structure." Journal of financial economics 3, no. 4 (1976): 305-360.

Khanna, Tarun, and Krishna Palepu. "Is group affiliation profitable in emerging markets?

An analysis of diversified Indian business groups." The Journal of Finance 55.2 (2000): 867-891.

Kumar, Alok, Alexandra Niessen-Ruenzi, and Oliver G. Spalt. "What's in a name? Mutual fund flows when managers have foreign-sounding names." The Review of Financial Studies28.8 (2015): 2281-2321.

Lacina, Bethany. Rival claims: Ethnic violence and territorial autonomy under Indian federalism. University of Michigan Press, 2017.

McDonald, Michael L., Poonam Khanna, and James D. Westphal. "Getting them to think outside the circle: Corporate governance, CEOs' external advice networks, and firm performance." Academy of Management Journal 51, no. 3 (2008): 453-475.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. "Birds of a feather: Homophily in social networks." Annual Review of Sociology 27, no. 1 (2001): 415-444.

Moscona, Jacob, Nathan Nunn, and James A. Robinson. "Keeping it in the family: Lineage organization and the scope of trust in sub-Saharan Africa." American Economic Review 107.5 (2017): 565-71.

Norman, Peter. "Statistical discrimination and efficiency." The Review of Economic Studies 70, no. 3 (2003): 615-627.

Palia, Darius. "The endogeneity of managerial compensation in firm valuation: A solution." The Review of Financial Studies 14, no. 3 (2001): 735-764.

Ritter, Jay R. Insider holdings and the pricing of initial public offerings. Rodney L. White Center for Financial Research, The Wharton School, University of Pennsylvania, 1981.

Ross, Stephen A. "The economic theory of agency: The principal's problem." The American economic review 63, no. 2 (1973): 134-139.

Schwab, Stewart. "Is statistical discrimination efficient?." The American Economic Review 76, no. 1 (1986): 228-234.

Westphal, James D., and Edward J. Zajac. "Who shall govern? CEO/board power, demographic similarity, and new director selection." Administrative Science Quarterly (1995): 60-83.

Wiseman, Robert M., and Luis R. Gomez-Mejia. "A behavioral agency model of managerial risk taking." Academy of management Review 23, no. 1 (1998): 133-153.

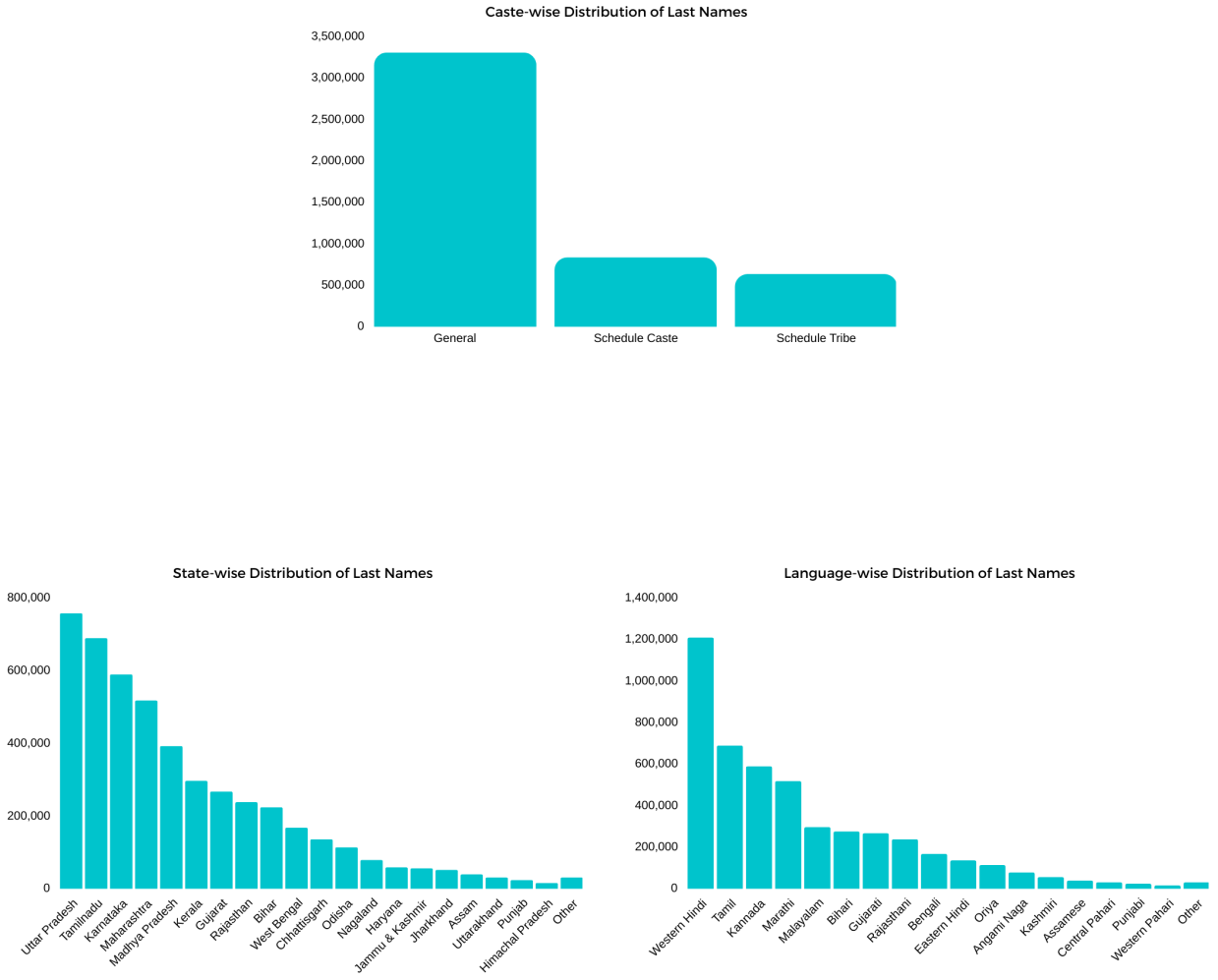
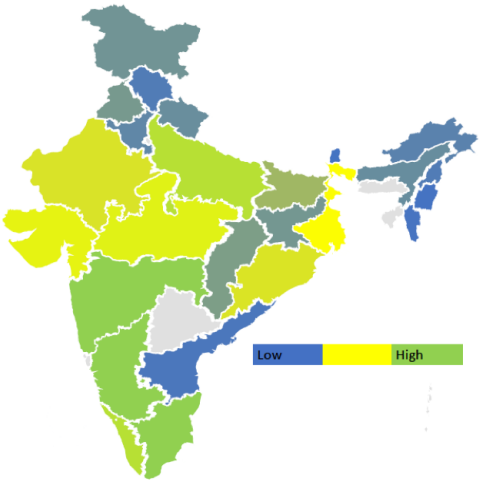
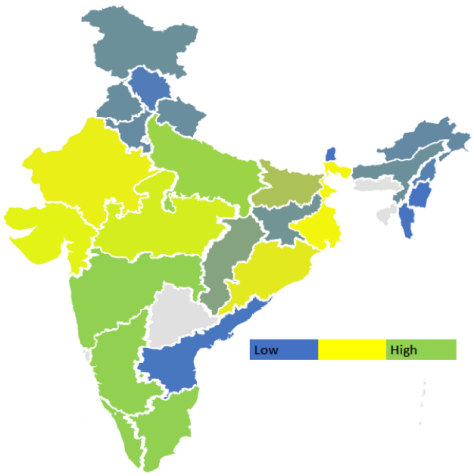


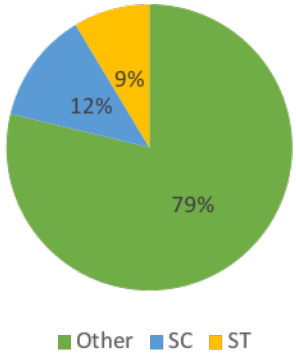
Figure 1: These figures show the state-wise, language-wise, and caste-wise distribution of 3.78 million Indian last names. The caste classification is based on administrative data which classifies caste as schedule caste (SC), schedule tribe (ST), and other. The native language is from the 22 schedule languages registered under the Indian constitution.



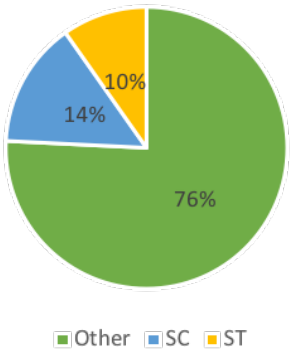
(a) **Manager's Native Place**



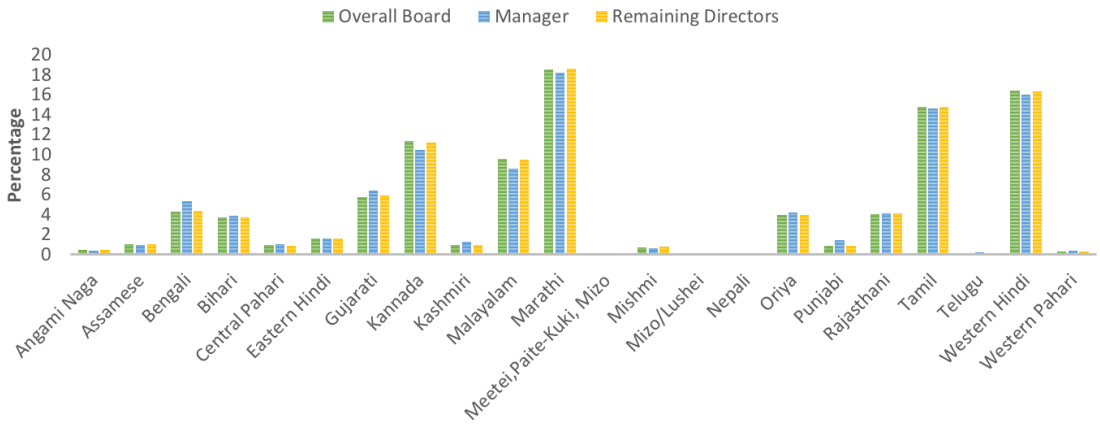
(b) **Director's Native Place**



(c) **Manager's Caste**

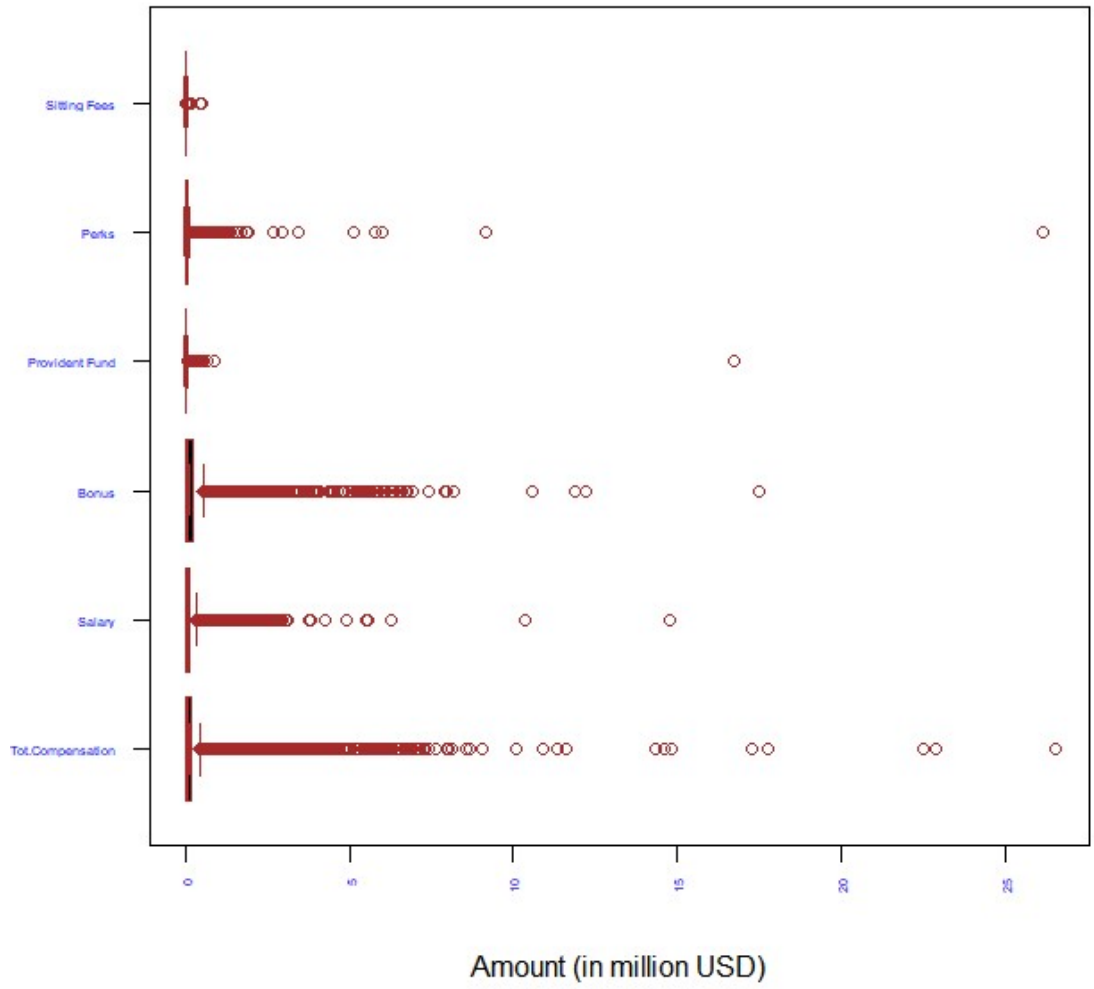


(d) **Director's Caste**



(e) **Native Language**

Figure 2: These figures show the distribution of managers and directors based on their identities. (a) Distribution of managers based on native place (Indian states). (b) Distribution of directors based on native place (Indian states). (c) Distribution of managers based on caste. (d) Distribution of directors based on caste. The caste classification is based on administrative data which classifies caste as schedule caste (SC), schedule tribe (ST), and other. (e) Distribution of managers and directors based on native language. The native language is from the 22 schedule languages registered under the Indian constitution.



Managerial Compensation and Its Components

Figure 3: This figure shows the distribution of total managerial compensation and its components - salary, bonus, sitting fees, and contribution to provident fund. All are in million USD.

Table 1

Panel A: Sample Selection

This table describes the sample selection procedure.

	Firms	Observations (Manager and Board Members)
Merging with surname and social identities data	24,326	1,473,836
Total remuneration is given	13,226	285,159
BSE and NSE listed non- financial firms	3,668	208,275
BSE and NSE listed non- financial firms (Manager)	3,668	41,507
Balanced Sample	2,324	28,026

Panel B: Managerial Compensation - Distribution and Summary Statistics

This table shows the components of managerial compensation (not in log form) and their summary statistics. Total Compensation (*Total Compensation*) is total managerial compensation. Salary (*Salary*) is salary of a manager. The other components of total compensation can be bonus (*Bonus*), contribution to provident fund (*Contribution to PF*), sittings fees (*Sitting Fees*), and perks (*Perks*). Components of total managerial compensation can vary depends on the position of a manger such as if manager is also a promoter or manager is also a chairman of the board. Total compensation and its components are in million USD. All variables are winsorized at 1 %.

Variable	Obs	Mean	Std. Dev.	Min	Max
Total Compensation	28,026	0.192	0.373	0.000	2.461
Salary	21,328	0.132	0.183	0.002	1.116
Bonus	7,831	0.262	0.545	0.000	3.663
Perks	9,471	0.042	0.079	0.000	0.530
Sitting Fees	6,705	0.003	0.004	0.000	0.023
Contribution to PF	5,047	0.017	0.023	0.000	0.138

Panel C: Summary Statistics

This table provides the summary statistics of the variables used in the analyses. *Log Compensation* is log of total compensation. *Industry Relative Compensation* is compensation relative to industry average. *Firm value* is ratio of market value to the book value of equity. *Residual Firm Value* is residual of regression of firm value on manager's total compensation of last year. *Homophily Language* is a dummy variable whose value equal to 1 if the board (at least one director) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same caste otherwise 0. All other variables are defined in variable definition sheet in Appendix. All variables are winsorized at 1 %.

Variable	Obs	Mean	Std. Dev.	Min	Max
Identity Variables					
Homophily Caste	28,026	0.892	0.310	0	1
Homophily Place	28,026	0.397	0.489	0	1
Homophily Language	28,026	0.421	0.494	0	1
HIL	28,026	0.224	0.379	0	1
HIP	28,026	0.182	0.351	0	1
HIC	28,026	0.874	0.302	0	1
Experience Ties	2,921	0.755	0.430	0	1
College Ties	3,299	0.267	0.443	0	1
Firm Characteristics					
Firm Value (MB Ratio)	28,026	2.238	3.073	-3.07	19.26
Residual Firm Value	27,908	-0.014	1.754	-5.143	8.470
Market Value (Debt + Market Cap)	28,026	627.36	2018.18	0.70	14899.99
Size	28,026	17.326	1.521	14.103	21.463
Firm Age	28,026	33.872	20.413	4	102
Institutional Ownership	28,026	9.540	12.006	0	48.365
Promoters Ownership	28,026	54.471	16.282	5.960	88.580
ROA	28,026	0.038	0.099	-1.435	0.416
Leverage	28,026	0.159	0.179	0	2.372
BG	28,026	0.489	0.500	0	1
PSU	28,026	0.024	0.154	0	1
Complexity	28,026	0.419	0.493	0	1
BIG 5	28,026	0.273	0.445	0	1
Board Characteristics					
Number of Board Meetings	28,026	7.855	3.155	1	29
Board Size	28,026	10.476	3.252	3	23
Shock	28,026	0.307	0.461	0	1
CEO-Chair Dummy	28,026	0.256	0.437	0	1
Manager Characteristics					
Log Compensation	28,026	10.611	2.158	3.412	13.634
Industry Relative Compensation	28,026	0.003	2.070	-5.699	3.752
Age	3,542	64.566	11.282	33	98
Gender	3,598	.982	0.132	0	1
Qualification	3,430	2	1.103	1	9
Powerful CEO	28,026	0.363	0.481	0	1
Promoter_Manager	10,802	0.933	0.249	0	1

Panel D: Homophily Distribution (For Managers)

This table provides the distribution of observations based on homophily variables. The homophily variables are based on native language (*Homophily Language*), native place (*Homophily Place*), and caste (*Homophily Caste*). The variable definitions are provided in Appendix .

	Language		Place		Caste	
Homophily	Freq.	Percent	Freq.	Percent	Freq.	Percent
0	16,240	57.95	16,897	60.29	3,018	10.77
1	11,786	42.05	11,129	39.71	25,008	89.23
Total	28,026	100	28,026	100	28,026	100

Table 2: Group Identity and Executive Compensation

This table provides regressions results of executive compensation on three measures of group identity - homophily language, homophily place, and homophily caste. *Log Compensation* is log of total compensation. The total compensation is in million USD. *Homophily Language* is a dummy variable whose value equal to 1 if the board (at least one director) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same caste otherwise 0. All other variables are defined in variable definition sheet in Appendix. The standard errors are in parentheses. All variables are winsorized at 1 %. In all panels, *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Log Compensation	Log Compensation	Log Compensation	Log Compensation	Log Compensation	Log Compensation
Homophily Language	0.066***			0.078***		
	(0.021)			(0.021)		
Homophily Place		0.071***			0.077***	
		(0.022)			(0.022)	
Homophily Caste			-0.141***			-0.138***
			(0.041)			(0.041)
Size				0.094***	0.095***	0.093***
				(0.017)	(0.017)	(0.017)
Firm Age				0.032*	0.032*	0.032*
				(0.017)	(0.017)	(0.017)
ROA				1.133***	1.130***	1.126***
				(0.099)	(0.099)	(0.099)
Institutional Ownership				0.015***	0.015***	0.015***
				(0.001)	(0.001)	(0.001)
Promoters Ownership				0.007***	0.007***	0.007***
				(0.001)	(0.001)	(0.001)
Leverage				-0.148**	-0.146**	-0.149**
				(0.068)	(0.068)	(0.068)
Volatility				-5.924***	-5.915***	-5.883***
				(0.961)	(0.961)	(0.961)
Big 5				0.096**	0.094**	0.096**
				(0.038)	(0.038)	(0.038)
Number of Meetings				0.015***	0.015***	0.015***
				(0.004)	(0.004)	(0.004)
Board Size				-0.005	-0.005	-0.005
				(0.004)	(0.004)	(0.004)
Constant	10.583***	10.583***	10.737***	7.431***	7.414***	7.593***
	(0.011)	(0.011)	(0.037)	(0.668)	(0.668)	(0.668)
Observations	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.758	0.758	0.758	0.762	0.762	0.762
Managerial Designation	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Table 3 : Group Identity, Firm Value, and Residual Firm Value

This table provides regressions results of firm value and residual firm value on three measures of group identity - homophily language, homophily place, and homophily caste. *Firm value* is ratio of market value of equity to the book value of equity. Market value and book value of equity are in million USD. *Residual Firm Value* is residual of regression of firm value on manager's total compensation of last year. Total compensation is in million USD. *Homophily Language* is a dummy variable whose value equal to 1 if the board (at least one director) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same caste otherwise 0. All other variables are defined in variable definition sheet in Appendix. The dependent variable in Columns (1)-(3) is firm value and in Columns (4)-(6) is residual firm value in a year. The standard errors are in parentheses.All variables are winsorized at 1 %. In all panels, *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

VARIABLES	(1) Firm Value	(2) Firm Value	(3) Firm Value	(4) Residual Firm Value	(5) Residual Firm Value	(6) Residual Firm Value
Homophily Language	0.124***			0.095***		
	(0.040)			(0.034)		
Homophily Place		0.058			0.056	
		(0.041)			(0.034)	
Homophily Caste			-0.122			-0.095
			(0.076)			(0.065)
Size	0.240***	0.240***	0.239***	0.225***	0.226***	0.224***
	(0.033)	(0.033)	(0.033)	(0.028)	(0.028)	(0.028)
Firm Age	0.053*	0.053*	0.053*	0.044	0.044	0.044
	(0.032)	(0.032)	(0.032)	(0.027)	(0.027)	(0.027)
ROA	4.243***	4.237***	4.234***	3.953***	3.948***	3.946***
	(0.185)	(0.185)	(0.185)	(0.157)	(0.157)	(0.157)
Institutional Ownership	0.048***	0.047***	0.047***	0.045***	0.045***	0.045***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Promoters Ownership	0.042***	0.042***	0.042***	0.037***	0.037***	0.037***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Leverage	0.136	0.138	0.135	0.153	0.154	0.152
	(0.127)	(0.128)	(0.128)	(0.108)	(0.108)	(0.108)
Volatility	-9.150***	-9.116***	-9.089***	-9.159***	-9.136***	-9.112***
	(1.803)	(1.803)	(1.803)	(1.526)	(1.527)	(1.526)
Big 5	0.317***	0.315***	0.316***	0.263***	0.261***	0.262***
	(0.071)	(0.071)	(0.071)	(0.060)	(0.060)	(0.060)
Number of Meetings	0.044***	0.044***	0.044***	0.043***	0.043***	0.043***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Board Size	0.013*	0.013*	0.013*	0.011*	0.011*	0.011*
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Constant	-7.021***	-6.985***	-6.835***	-8.351***	-8.335***	-8.208***
	(1.253)	(1.253)	(1.254)	(1.060)	(1.061)	(1.061)
Observations	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.586	0.585	0.585	0.084	0.084	0.084
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Table 4: Resolving Simultaneity and Self-Selection Biases

This table provides results on resolving the simultaneity between firm value and executive compensation, and self-selection issues in executive hiring. *Log Compensation* is log of total compensation. *Firm value* is ratio of market value of equity to the book value of equity. *Residual Firm Value* is residual of regression of firm value on manager's total compensation of last year. Total compensation, market value of equity, and book value of equity are in million USD. *Homophily Language* is a dummy variable whose value equal to 1 if the board (at least one director) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same caste otherwise 0. Panel A reports the regression results of firm value on group identity measures considering the simultaneity between firm value and executive compensation. Panel B reports the regression results of group identity and self-selection bias in hiring of a manager. All other variables are defined in variable definition sheet in Appendix. The standard errors are in parentheses. All variables are winsorized at 1 %. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Panel A: Group Identity and Simultaneity of Firm Value and Managerial Compensation

Panel A provides results of two stage least square (2SLS) regression used to resolve the simultaneity of firm value and managerial compensation . In first stage, the estimated regression is logarithm of total managerial compensation on promoter dummy as a intrument and all the firm and governance level controls. In second stage, estimated regression is firm value on three group identity measures controlling for fitted value from first stage regression with other controls. The three measures of group identity are homophily language, homophily place, and homophily caste.

VARIABLES	(1) Firm Value	(2) Firm Value	(3) Firm Value
Log Compensation	-0.570***	-0.562***	-0.540***
	(0.125)	(0.125)	(0.118)
Homophily Language	0.124***		
	(0.047)		
Homophily Place		0.096*	
		(0.049)	
Homophily Caste			-0.050
			(0.057)
Size	0.276***	0.274***	0.272***
	(0.019)	(0.018)	(0.018)
Firm Age	-0.003***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)
ROA	7.226***	7.206***	7.158***
	(0.361)	(0.361)	(0.346)
Institutional Ownership	0.045***	0.045***	0.044***
	(0.004)	(0.004)	(0.004)
Promoters Ownership	-34.334***	-34.219***	-33.786***
	(2.653)	(2.658)	(2.559)
Leverage	0.025***	0.024***	0.024***
	(0.002)	(0.002)	(0.001)
Volatility	-0.832***	-0.828***	-0.834***
	(0.111)	(0.111)	(0.110)
Big 5	0.772***	0.769***	0.755***
	(0.049)	(0.049)	(0.047)
Number of Meetings	0.015	0.015	0.013
	(0.011)	(0.011)	(0.010)
Board Size	-0.029***	-0.029***	-0.030***
	(0.007)	(0.007)	(0.007)
Constant	2.435**	2.397**	2.334**
	(0.947)	(0.947)	(0.913)
Observations	28,026	28,026	28,026
R-squared	0.057	0.061	0.071
Industry	Yes	Yes	Yes
Year	Yes	Yes	Yes

Panel B: Group Identity and Self-Selection in Hiring of a Manager

This panel provides regression results of managerial compensation, firm value, and residual firm value on three measures of group identity considering the self-selection bias in hiring of manager. The selection bias is mitigated using the death or the non-voluntary retirement of a director as a exogenous shock to group identity measures. In Columns (1) - (3), the dependent variable is logarithm of total compensation (*Log Compensation*) of a manager in that firm-year. In Columns (4)-(6), the dependent variable is firm value (*Firm Value*). In Columns (7)-(9), the dependent variable is residual firm value (*Residual Firm Value*). All other variables are defined in variable definition sheet in Appendix. The standard errors are in parentheses.All variables are winsorized at 1 %. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

VARIABLES	(1) Log Compensation	(2) Log Compensation	(3) Log Compensation	(4) FirmValue	(5) FirmValue	(6) FirmValue	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual FirmValue
Shock	-0.094***	-0.099***	-0.064	-0.116***	-0.108**	-0.109	-0.106***	-0.093**	-0.074
	(0.019)	(0.020)	(0.064)	(0.034)	(0.040)	(0.069)	(0.034)	(0.039)	(0.059)
Homophily Language	0.063			0.105**			0.069*		
	(0.050)			(0.049)			(0.038)		
Shock x Homophily Language	0.045**			0.059***			0.065***		
	(0.022)			(0.019)			(0.016)		
Homophily Place		0.057			0.043			0.041	
		(0.039)			(0.055)			(0.047)	
Shock x Homophily Place		0.062**			0.043*			0.042	
		(0.023)			(0.023)			(0.026)	
Homophily Caste			-0.134***						-0.095
			(0.034)						(0.092)
Shock x Homophily Caste			-0.014						-0.005
			(0.066)						(0.050)
Size	0.093***	0.094***	0.093***	0.238***	0.239***	0.238***	0.225***	0.225***	0.223***
	(0.023)	(0.022)	(0.022)	(0.022)	(0.021)	(0.022)	(0.020)	(0.019)	(0.019)
Firm Age	0.032	0.032	0.032	0.054	0.054	0.054	0.042	0.045	0.045
	(0.062)	(0.061)	(0.060)	(0.054)	(0.054)	(0.055)	(0.049)	(0.049)	(0.050)
ROA	1.131***	1.127***	1.123***	4.241***	4.234***	4.230***	3.956***	3.946***	3.943***
	(0.089)	(0.086)	(0.084)	(0.315)	(0.315)	(0.314)	(0.284)	(0.282)	(0.281)
Institutional Ownership	0.015***	0.015***	0.015***	0.047***	0.047***	0.047***	0.045***	0.045***	0.045***
	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Promoters Ownership	0.007***	0.007***	0.007***	0.042***	0.042***	0.042***	0.037***	0.037***	0.037***
	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
Leverage	-0.144	-0.141	-0.144*	0.141	0.143	0.141	0.155	0.158	0.156
	(0.085)	(0.086)	(0.085)	(0.245)	(0.244)	(0.246)	(0.180)	(0.176)	(0.178)
Volatility	-5.882***	-5.871***	-5.830***	-9.103***	-9.059***	-9.027***	-9.090***	-9.089***	-9.059***
	(1.119)	(1.103)	(1.094)	(2.500)	(2.504)	(2.515)	(1.721)	(1.717)	(1.722)
Big 5	0.098***	0.096***	0.097***	0.318***	0.316***	0.318***	0.263***	0.263***	0.264***
	(0.015)	(0.016)	(0.015)	(0.082)	(0.080)	(0.080)	(0.066)	(0.063)	(0.063)
Number of Meetings	0.016***	0.016***	0.016***	0.044***	0.044***	0.044***	0.043***	0.044***	0.044***
	(0.002)	(0.002)	(0.002)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Board Size	-0.001	-0.001	-0.000	0.018	0.018	0.018	0.018*	0.015	0.015
	(0.004)	(0.004)	(0.004)	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.011)
Constant	7.407***	7.398***	7.554***	-7.047***	-7.015***	-6.877***	-8.324***	-8.358***	-8.246***
	(2.000)	(1.988)	(1.896)	(2.003)	(2.001)	(2.118)	(1.710)	(1.756)	(1.848)
Observations	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.762	0.762	0.762	0.586	0.586	0.586	0.087	0.084	0.084
Designation	Yes	Yes	Yes	No	No	No	No	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 5: Powerful CEO

This table provides regression results of managerial compensation, firm value, and residual firm value on group identity when a manager is powerful. *Log Compensation* is log of total compensation. *Firm value* is ratio of market value of equity to the book value of equity. *Residual Firm Value* is residual of regression of firm value on manager's total compensation of last year. Total compensation, market value of equity, and book value of equity are in million USD. *Homophily Language* is a dummy variable whose value equal to 1 if the board (at least one director) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same caste otherwise 0. *Powerful CEO* is a dummy variable which takes value 1 if a manager is chairperson, CEO, and promoter; otherwise 0. In Columns (1) - (3), the dependent variable is logarithm of total compensation (*Log Compensation*) of a manager in that firm-year. In Columns (4)-(6), the dependent variable is firm value (*Firm Value*). In Columns (7)-(9), the dependent variable is residual firm value (*Residual Firm Value*). All other variables are defined in variable definition sheet in Appendix. The standard errors are in parentheses. All variables are winsorized at 1 %. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

VARIABLES	(1) Log Compensation	(2) Log Compensation	(3) Log Compensation	(4) FirmValue	(5) FirmValue	(6) FirmValue	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual FirmValue
Homophily Language	0.078***			0.124***			0.095***		
	(0.021)			(0.040)			(0.034)		
Homophily Place		0.077***			0.059			0.056	
		(0.022)			(0.041)			(0.034)	
Homophily Caste			-0.137***			-0.122			-0.097
			(0.041)			(0.076)			(0.065)
Powerful CEO	0.032	0.031	0.031	-0.035	-0.036	-0.032	-0.027	-0.027	-0.028
	(0.021)	(0.021)	(0.021)	(0.039)	(0.039)	(0.040)	(0.033)	(0.033)	(0.033)
Size	0.093***	0.095***	0.092***	0.241***	0.241***	0.243***	0.226***	0.226***	0.225***
	(0.017)	(0.017)	(0.017)	(0.033)	(0.033)	(0.033)	(0.028)	(0.028)	(0.028)
Firm Age	0.032*	0.032*	0.032*	0.053*	0.053	0.051	0.044	0.044	0.044
	(0.017)	(0.017)	(0.017)	(0.032)	(0.032)	(0.032)	(0.027)	(0.027)	(0.027)
ROA	1.132***	1.129***	1.125***	4.244***	4.239***	4.233***	3.954***	3.949***	3.947***
	(0.099)	(0.099)	(0.099)	(0.185)	(0.185)	(0.185)	(0.157)	(0.157)	(0.157)
Institutional Ownership	0.015***	0.015***	0.015***	0.048***	0.047***	0.048***	0.045***	0.045***	0.045***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Promoters Ownership	0.007***	0.007***	0.007***	0.042***	0.042***	0.042***	0.037***	0.037***	0.037***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Leverage	-0.146**	-0.145**	-0.148**	0.135	0.136	0.136	0.152	0.153	0.151
	(0.068)	(0.068)	(0.068)	(0.128)	(0.128)	(0.128)	(0.108)	(0.108)	(0.108)
Volatility	-5.923***	-5.914***	-5.882***	-9.152***	-9.118***	-9.076***	-9.160***	-9.137***	-9.113***
	(0.960)	(0.960)	(0.961)	(1.803)	(1.803)	(1.803)	(1.526)	(1.527)	(1.527)
Big 5	0.097**	0.095**	0.097**	0.316***	0.314***	0.314***	0.262***	0.261***	0.262***
	(0.038)	(0.038)	(0.038)	(0.071)	(0.071)	(0.071)	(0.060)	(0.060)	(0.060)
Number of Meetings	0.015***	0.015***	0.015***	0.044***	0.044***	0.043***	0.043***	0.043***	0.043***
	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Board Size	-0.005	-0.005	-0.005	0.013*	0.013*	0.016**	0.011*	0.011*	0.011*
	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Constant	7.433***	7.416***	7.594***	-7.021***	-6.986***	-6.838***	-8.352***	-8.336***	-8.207***
	(0.667)	(0.668)	(0.668)	(1.253)	(1.253)	(1.254)	(1.060)	(1.061)	(1.061)
Observations	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.762	0.762	0.762	0.586	0.585	0.586	0.084	0.084	0.084
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 6: Group Identity and Ties Based on College and Past Experience

This table reports the main results controlling the induced identity variables such as college level ties and the ties based shared experiences. *Log Compensation* is log of total compensation. *Firm value* is ratio of market value of equity to the book value of equity. *Residual Firm Value* is residual of regression of firm value on manager's total compensation of last year. Total compensation, market value of equity, and book value of equity are in million USD. *Homophily Language* is a dummy variable whose value equal to 1 if the board (at least one director) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same caste otherwise 0. College ties (*College Ties*) is a dummy variable which takes value 1 if a manager and at least one of the board members went to the same college; otherwise 0. Past experience ties (*Past Experience Ties*) is a dummy variable which takes value 1 if a manager and at least one of the board members worked for the same firm in the past; otherwise 0. Panel A panel provides regression results of managerial compensation, firm value, and residual firm value on group identity controlling for college level ties between a manager and the board members. Panel B provides regression results of managerial compensation, firm value, and residual firm value on group identity controlling for past employment level ties between a manager and the board members. All other variables are defined in variable definition sheet in Appendix. The standard errors are in parentheses. All variables are winsorized at 1 %. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Panel A: College Ties

This panel provides regression results of managerial compensation, firm value, and residual firm value on group identity controlling for college level ties between a manager and the board members. In Columns (1) - (3), the dependent variable is logarithm of total compensation (*Log Compensation*) of a manager in that firm-year. In Columns (4)-(6), the dependent variable is firm value (*Firm Value*). In Columns (7)-(9), the dependent variable is residual firm value (*Residual Firm Value*).

VARIABLES	(1) Log Compensation	(2) Log Compensation	(3) Log Compensation	(4) Firm Value	(5) Firm Value	(6) Firm Value	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual Firm Value
Homophily Language	0.131**			0.305**			0.313**		
	(0.058)			(0.136)			(0.123)		
Homophily Place		0.126**			0.094			0.118	
		(0.060)			(0.148)			(0.134)	
Homophily Caste			-0.125			-0.162			-0.137
			(0.087)			(0.212)			(0.191)
College Ties	0.071	0.068	0.080	-0.055	-0.039	-0.037	-0.044	-0.028	-0.025
	(0.061)	(0.061)	(0.061)	(0.129)	(0.129)	(0.129)	(0.116)	(0.116)	(0.116)
Size	0.040*	0.040*	0.037	0.019	0.012	0.009	0.074	0.067	0.064
	(0.023)	(0.023)	(0.023)	(0.129)	(0.129)	(0.129)	(0.116)	(0.117)	(0.117)
Firm Age	0.005***	0.006***	0.005***	-0.052	-0.056	-0.066	-0.060	-0.063	-0.072
	(0.001)	(0.001)	(0.001)	(0.085)	(0.085)	(0.086)	(0.077)	(0.077)	(0.078)
ROA	1.940***	1.960***	1.972***	6.082***	6.148***	6.164***	5.801***	5.869***	5.882***
	(0.329)	(0.329)	(0.329)	(0.635)	(0.635)	(0.635)	(0.574)	(0.574)	(0.574)
Institutional Ownership	0.013***	0.013***	0.013***	0.053***	0.051***	0.051***	0.052***	0.050***	0.050***
	(0.003)	(0.003)	(0.003)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Promoters Ownership	-0.004*	-0.004*	-0.004*	0.127***	0.125***	0.125***	0.115***	0.113***	0.113***
	(0.002)	(0.002)	(0.002)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)
Leverage	0.798***	0.830***	0.846***	2.221***	2.269***	2.278***	2.004***	2.052***	2.061***
	(0.203)	(0.202)	(0.203)	(0.548)	(0.548)	(0.548)	(0.495)	(0.495)	(0.495)
Volatility	-27.132***	-27.014***	-26.525***	15.765	15.476	15.421	17.476*	17.167*	17.144*
	(5.352)	(5.351)	(5.345)	(11.419)	(11.428)	(11.428)	(10.321)	(10.331)	(10.332)
Big 5	0.482***	0.485***	0.484***	-0.234	-0.231	-0.228	-0.180	-0.178	-0.174
	(0.058)	(0.058)	(0.058)	(0.194)	(0.194)	(0.194)	(0.175)	(0.175)	(0.175)
Number of Meetings	-0.011	-0.011	-0.010	-0.001	-0.002	-0.001	-0.010	-0.010	-0.009
	(0.014)	(0.014)	(0.014)	(0.024)	(0.024)	(0.024)	(0.022)	(0.022)	(0.022)
Board Size	-0.010	-0.010	-0.009	-0.059***	-0.060***	-0.060***	-0.051***	-0.052***	-0.052***
	(0.008)	(0.008)	(0.008)	(0.020)	(0.020)	(0.020)	(0.018)	(0.018)	(0.018)
Constant	10.976***	10.981***	11.190***	-2.470	-1.975	-1.346	-6.659*	-6.186	-5.594
	(0.435)	(0.436)	(0.435)	(4.273)	(4.273)	(4.312)	(3.861)	(3.862)	(3.898)
Observations	3,299	3,299	3,299	3,299	3,299	3,299	3,299	3,299	3,299
R-squared	0.412	0.412	0.411	0.789	0.789	0.789	0.377	0.375	0.375
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Past Employment Ties

This panel provides regression results of managerial compensation, firm value, and residual firm value on group identity controlling for past employment level ties between a manager and the board members. In Columns (1) - (3), the dependent variable is logarithm of total compensation (*Log Compensation*) of a manager in that firm-year. In Columns (4)-(6), the dependent variable is firm value (*Firm Value*). In Columns (7)-(9), the dependent variable is residual firm value (*Residual Firm Value*).

VARIABLES	(1) Log Compensation	(2) Log Compensation	(3) Log Compensation	(4) Firm Value	(5) Firm Value	(6) Firm Value	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual Firm Value
Homophily Language	0.145**			0.403**			0.413***		
	(0.064)			(0.164)			(0.139)		
Homophily Place		0.183***			0.445**			0.323**	
		(0.067)			(0.182)			(0.155)	
Homophily Caste			-0.229**			0.191			0.215
			(0.096)			(0.281)			(0.238)
Past Employment Ties	-0.065	-0.066	-0.070	-0.144	-0.151	-0.156	-0.143	-0.151	-0.155
	(0.068)	(0.068)	(0.068)	(0.123)	(0.123)	(0.123)	(0.104)	(0.104)	(0.105)
Size	0.009	0.011	0.008	0.081	0.080	0.070	0.142	0.137	0.131
	(0.024)	(0.024)	(0.024)	(0.146)	(0.146)	(0.146)	(0.124)	(0.124)	(0.124)
Firm Age	0.006***	0.006***	0.005***	-0.015	-0.001	0.001	-0.027	-0.014	-0.009
	(0.001)	(0.001)	(0.001)	(0.138)	(0.138)	(0.139)	(0.117)	(0.117)	(0.117)
ROA	3.043***	3.029***	3.092***	8.293***	8.232***	8.337***	7.042***	7.017***	7.085***
	(0.387)	(0.387)	(0.386)	(0.857)	(0.858)	(0.858)	(0.726)	(0.727)	(0.727)
Institutional Ownership	0.018***	0.018***	0.017***	0.058***	0.058***	0.055***	0.054***	0.053***	0.052***
	(0.003)	(0.003)	(0.003)	(0.011)	(0.011)	(0.011)	(0.009)	(0.009)	(0.009)
Promoters Ownership	-0.003	-0.003	-0.004	0.135***	0.134***	0.134***	0.120***	0.119***	0.119***
	(0.002)	(0.002)	(0.002)	(0.013)	(0.013)	(0.013)	(0.011)	(0.011)	(0.011)
Leverage	0.800***	0.828***	0.864***	-3.298***	-3.243***	-3.202***	-2.292***	-2.219***	-2.196***
	(0.216)	(0.215)	(0.215)	(0.709)	(0.708)	(0.709)	(0.601)	(0.600)	(0.601)
Volatility	-20.705***	-20.855***	-20.719***	-2.473	-1.970	-1.776	5.176	5.712	5.901
	(5.557)	(5.556)	(5.557)	(13.066)	(13.065)	(13.080)	(11.069)	(11.077)	(11.086)
Big 5	0.586***	0.588***	0.585***	-0.356	-0.394	-0.369	-0.410*	-0.441**	-0.424**
	(0.063)	(0.063)	(0.063)	(0.253)	(0.253)	(0.253)	(0.214)	(0.214)	(0.214)
Number of Meetings	0.007	0.008	0.009	-0.015	-0.016	-0.016	-0.018	-0.018	-0.019
	(0.014)	(0.014)	(0.014)	(0.028)	(0.028)	(0.028)	(0.024)	(0.024)	(0.024)
Board Size	0.012	0.012	0.013	0.007	0.005	0.005	-0.001	-0.002	-0.003
	(0.009)	(0.009)	(0.009)	(0.025)	(0.025)	(0.025)	(0.021)	(0.021)	(0.021)
Constant	10.663***	10.604***	10.938***	-5.299	-5.770	-5.647	-9.094*	-9.365*	-9.509*
	(0.468)	(0.469)	(0.467)	(6.147)	(6.153)	(6.213)	(5.207)	(5.217)	(5.265)
Observations	2,921	2,921	2,921	2,921	2,921	2,921	2,921	2,921	2,921
R-squared	0.425	0.425	0.425	0.750	0.750	0.749	0.413	0.412	0.411
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 7: Group Identity and Promoter (Founder) as Manager

This table provides regression results of executive compensation, firm value, and residual firm value on group identity when a manager is also the promoter (founder) of the firm. *Log Compensation* is log of total compensation. *Firm value* is ratio of market value of equity to the book value of equity. *Residual Firm Value* is residual of regression of firm value on manager's total compensation of last year. Total compensation, market value of equity, and book value of equity are in million USD. *Homophily Language* is a dummy variable whose value equal to 1 if the board (at least one director) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same caste otherwise 0. In Columns (1) - (3), the dependent variable is logarithm of total compensation (*Log Compensation*) of an executive in that firm-year. In Columns (4)-(6), the dependent variable is firm value (*Firm Value*). In Columns (7)-(9), the dependent variable is residual firm value (*Residual Firm Value*). All other variables are defined in variable definition sheet in Appendix. The standard errors are in parentheses. All variables are winsorized at 1 %. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

VARIABLES	(1) Log Compensation	(2) Log Compensation	(3) Log Compensation	(4) Firm Value	(5) Firm Value	(6) Firm Value	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual Firm Value
Homophily Language	0.056			0.201***			0.160***		
	(0.038)			(0.067)			(0.057)		
Homophily Place		0.017			0.156**			0.108*	
		(0.039)			(0.069)			(0.058)	
Homophily Caste			-0.189**			-0.197			-0.170
			(0.082)			(0.145)			(0.122)
Size	0.075**	0.076**	0.076**	0.296***	0.298***	0.301***	0.296***	0.298***	0.296***
	(0.032)	(0.032)	(0.032)	(0.056)	(0.056)	(0.056)	(0.047)	(0.047)	(0.047)
Firm Age	0.063**	0.063**	0.063**	0.113**	0.118***	0.108**	0.086**	0.087**	0.085**
	(0.025)	(0.025)	(0.025)	(0.045)	(0.044)	(0.045)	(0.038)	(0.038)	(0.038)
ROA	1.558***	1.554***	1.546***	4.194***	5.135***	4.185***	4.030***	4.015***	4.010***
	(0.180)	(0.180)	(0.180)	(0.319)	(0.329)	(0.320)	(0.268)	(0.268)	(0.268)
Institutional Ownership	0.015***	0.015***	0.015***	0.035***	0.036***	0.035***	0.035***	0.035***	0.035***
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Promoters Ownership	0.006***	0.006***	0.006***	0.045***	0.044***	0.045***	0.039***	0.039***	0.039***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Leverage	-0.210*	-0.212*	-0.213*	0.363*	0.396*	0.374*	0.271	0.264	0.264
	(0.123)	(0.123)	(0.123)	(0.218)	(0.217)	(0.218)	(0.183)	(0.183)	(0.183)
Volatility	-3.582**	-3.587**	-3.548**	-8.854***	-9.043***	-8.697***	-9.128***	-9.139***	-9.104***
	(1.635)	(1.635)	(1.635)	(2.906)	(2.890)	(2.910)	(2.442)	(2.442)	(2.443)
Big 5	0.101	0.101	0.100	0.165	0.144	0.159	0.108	0.106	0.109
	(0.065)	(0.065)	(0.065)	(0.116)	(0.116)	(0.116)	(0.098)	(0.098)	(0.098)
Number of Meetings	0.015**	0.015**	0.015**	0.042***	0.040***	0.041***	0.040***	0.040***	0.040***
	(0.006)	(0.006)	(0.006)	(0.011)	(0.010)	(0.011)	(0.009)	(0.009)	(0.009)
Board Size	-0.005	-0.006	-0.005	0.014	0.011	0.018	0.014	0.013	0.014
	(0.007)	(0.007)	(0.007)	(0.013)	(0.013)	(0.013)	(0.011)	(0.011)	(0.011)
Constant	6.829***	6.852***	7.036***	-10.214***	-10.280***	-9.896***	-10.968***	-10.970***	-10.697***
	(1.031)	(1.032)	(1.033)	(1.831)	(1.821)	(1.839)	(1.538)	(1.540)	(1.542)
Observations	10,083	10,083	10,083	10,083	10,071	10,083	10,083	10,083	10,083
R-squared	0.773	0.773	0.773	0.576	0.582	0.576	0.193	0.193	0.193
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 8: Group Identity and Business Group Firms

This table provides regression results of managerial compensation and residual firm value on group identity for business group firms compare to non-business group firms. *Log Compensation* is log of total compensation. *Firm value* is ratio of market value of equity to the book value of equity. *Residual Firm Value* is residual of regression of firm value on manager's total compensation of last year. Total compensation, market value of equity, and book value of equity are in million USD. *Homophily Language* is a dummy variable whose value equal to 1 if the board (at least one director) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same caste otherwise 0. Business group (*BG*) is a dummy variable which takes value 1 if a firm belongs to a business group; otherwise 0. In Columns (1) - (3), the dependent variable is logarithm of total compensation (*Log Compensation*) of a manager in that firm-year. In Columns (4)-(6), the dependent variable is residual firm value (*Residual Firm Value*). Please refer to Appendix A for variable definitions. The standard errors are in parentheses. All variables are winsorized at 1 %. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

VARIABLES	(1) Log Compensation	(2) Log Compensation	(3) Log Compensation	(4) Residual Firm Value	(5) Residual Firm Value	(6) Residual Firm Value
BG	0.287***	0.290***	0.144***	-0.069**	-0.057**	0.016
	(0.024)	(0.024)	(0.054)	(0.029)	(0.028)	(0.064)
Homophily Language	0.075***			0.012		
	(0.024)			(0.029)		
BG x Homophily Language	-0.035			0.073*		
	(0.036)			(0.042)		
Homophily Place		0.093***			0.006	
		(0.025)			(0.029)	
BG x Homophily Place		-0.043			0.049	
		(0.036)			(0.043)	
Homophily Caste			-0.099**			0.014
			(0.042)			(0.049)
BG x Homophily Caste			0.141**			-0.063
			(0.057)			(0.067)
Size	0.066***	0.066***	0.065***	-0.005	-0.005	-0.005
	(0.007)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)
Firm Age	0.005***	0.005***	0.005***	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
ROA	2.310***	2.310***	2.310***	2.311***	2.308***	2.312***
	(0.100)	(0.100)	(0.100)	(0.119)	(0.119)	(0.119)
Institutional Ownership	0.029***	0.030***	0.029***	0.005***	0.005***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Promoters Ownership	0.008***	0.008***	0.008***	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Leverage	0.115**	0.119**	0.116**	0.134**	0.135**	0.136**
	(0.054)	(0.054)	(0.054)	(0.064)	(0.064)	(0.064)
Volatility	-16.513***	-16.538***	-16.438***	-1.983	-1.977	-1.934
	(1.027)	(1.027)	(1.027)	(1.220)	(1.221)	(1.220)
Big 5	0.464***	0.465***	0.457***	-0.041	-0.042	-0.044*
	(0.022)	(0.022)	(0.022)	(0.026)	(0.026)	(0.026)
Number of Meetings	0.059***	0.059***	0.059***	0.013***	0.013***	0.013***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Board Size	0.030***	0.030***	0.031***	-0.007*	-0.007*	-0.006*
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Constant	7.905***	7.899***	8.042***	-0.182	-0.174	-0.184
	(0.129)	(0.129)	(0.133)	(0.153)	(0.152)	(0.157)
Observations	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.545	0.545	0.545	0.020	0.020	0.020
Year	Yes	Yes	Yes	Yes	Yes	Yes

Table 9: Group Identity with Independent Directors

This table reports the regression results of executive compensation, firm value, and residual firm value on group identity with independent directors. Log Compensation (*Log Compensation*) is logarithm of total managerial compensation which is in million USD. *HIL* is a dummy variable whose value equal to 1 if the independent directors (at least one) and the manager speaks same Indian language otherwise 0. *HIP* is dummy variable whose value equal to 1 if the independent directors (at least one) and the manager belong to same Indian state otherwise 0. *HIC* is dummy variable whose value equal to 1 if the independent directors (at least one) and the manager belong to same caste otherwise 0. All other variables are defined in variable definition sheet in Appendix. The standard errors are in parentheses. All variables are winsorized at 1 %. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

VARIABLES	(1) Log Compensation	(2) Log Compensation	(3) Log Compensation	(4) Firm Value	(5) Firm Value	(6) Firm Value	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual Firm Value
HIL	0.021			-0.105**			-0.109***		
	(0.026)			(0.050)			(0.042)		
HIP		0.032			-0.112**			-0.074	
		(0.029)			(0.054)			(0.045)	
HIC			-0.140***			-0.203***			-0.190***
			(0.041)			(0.076)			(0.064)
Size	0.106***	0.106***	0.106***	0.239***	0.238***	0.243***	0.224***	0.224***	0.225***
	(0.017)	(0.017)	(0.017)	(0.033)	(0.033)	(0.033)	(0.028)	(0.028)	(0.028)
Firm Age	0.030*	0.030*	0.029*	0.052	0.053	0.050	0.042	0.044	0.043
	(0.017)	(0.017)	(0.017)	(0.033)	(0.033)	(0.033)	(0.027)	(0.027)	(0.027)
Institutional Ownership	0.016***	0.016***	0.016***	0.047***	0.047***	0.048***	0.045***	0.045***	0.045***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Promoter Ownership	0.007***	0.007***	0.007***	0.042***	0.042***	0.042***	0.037***	0.037***	0.037***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
ROA	1.201***	1.204***	1.201***	4.238***	4.230***	4.237***	3.948***	3.943***	3.948***
	(0.099)	(0.099)	(0.099)	(0.186)	(0.186)	(0.186)	(0.157)	(0.157)	(0.157)
Leverage	-0.150**	-0.149**	-0.149**	0.136	0.135	0.141	0.150	0.150	0.153
	(0.068)	(0.068)	(0.068)	(0.128)	(0.128)	(0.128)	(0.108)	(0.108)	(0.108)
Volatility	-6.094***	-6.086***	-6.100***	-9.040***	-9.082***	-9.099***	-9.066***	-9.112***	-9.139***
	(0.964)	(0.964)	(0.963)	(1.814)	(1.814)	(1.814)	(1.526)	(1.526)	(1.526)
Big5	0.104***	0.104***	0.104***	0.319***	0.319***	0.318***	0.262***	0.263***	0.263***
	(0.038)	(0.038)	(0.038)	(0.072)	(0.072)	(0.072)	(0.060)	(0.060)	(0.060)
Number of Meetings	0.014***	0.014***	0.014***	0.044***	0.044***	0.043***	0.044***	0.044***	0.043***
	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Board Size	-0.004	-0.004	-0.004	0.013*	0.013*	0.016**	0.011*	0.011*	0.011*
	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Constant	7.319***	7.326***	7.468***	-6.872***	-6.916***	-6.757***	-8.199***	-8.258***	-8.101***
	(0.670)	(0.670)	(0.671)	(1.261)	(1.260)	(1.262)	(1.061)	(1.060)	(1.062)
Observations	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.766	0.766	0.766	0.585	0.585	0.585	0.084	0.084	0.084
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 10: Group Identity in Complex Firms

This table reports the regression results of executive compensation, firm value, and residual firm value on group identity in complex firms. Firm Complexity (*Firm Complexity*) is a dummy variable which takes value 1 if complexity of a firm is above median; otherwise 0. The complexity of a firm is measured as square root of number of subsidiaries. Log Compensation (*Log Compensation*) is logarithm of total managerial compensation which is in million USD. *Homophily Language* is a dummy variable whose value equal to 1 if the board (at least one director) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same caste otherwise 0. All other variables are defined in variable definition sheet in Appendix. The standard errors are in parentheses. All variables are winsorized at 1 %. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Log Compensation	Log Compensation	Log Compensation	Firm Value	Firm Value	Firm Value	Residual Firm Value	Residual Firm Value	Residual Firm Value
Firm Complexity	0.465***	0.432***	0.388***	-0.273***	-0.231***	-0.166	-0.127***	-0.124***	-0.003
	(0.025)	(0.025)	(0.055)	(0.048)	(0.047)	(0.104)	(0.030)	(0.030)	(0.065)
Homophily Language	0.111***			-0.066			0.011		
	(0.023)			(0.044)			(0.028)		
Firm Complexity x Homophily Language	-0.129***			0.128*			0.084*		
	(0.036)			(0.069)			(0.043)		
Homophily Place		0.088***			-0.060			-0.003	
		(0.023)			(0.044)			(0.028)	
Firm Complexity x Homophily Place		-0.053			0.022			0.079*	
		(0.036)			(0.069)			(0.043)	
Homophily Caste			-0.053			-0.011			0.026
			(0.037)			(0.070)			(0.044)
Firm Complexity x Homophily Caste			0.026			-0.062			-0.103
			(0.057)			(0.108)			(0.068)
Size	0.053***	0.054***	0.053***	0.242***	0.241***	0.242***	-0.000	-0.001	-0.001
	(0.007)	(0.007)	(0.007)	(0.014)	(0.014)	(0.014)	(0.009)	(0.009)	(0.009)
Firm Age	0.006***	0.006***	0.006***	-0.005***	-0.005***	-0.005***	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Institutional Ownership	0.027***	0.027***	0.026***	0.031***	0.031***	0.031***	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Promoter Ownership	0.008***	0.008***	0.008***	0.020***	0.020***	0.020***	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ROA	2.319***	2.316***	2.320***	5.867***	5.869***	5.869***	2.306***	2.307***	2.309***
	(0.100)	(0.100)	(0.100)	(0.190)	(0.190)	(0.190)	(0.119)	(0.119)	(0.119)
Leverage	0.141***	0.144***	0.141***	-0.873***	-0.874***	-0.872***	0.134**	0.136**	0.137**
	(0.054)	(0.054)	(0.054)	(0.102)	(0.102)	(0.102)	(0.064)	(0.064)	(0.064)
Volatility	-16.744***	-16.712***	-16.586***	-27.011***	-27.024***	-27.118***	-1.927	-1.924	-1.934
	(1.023)	(1.024)	(1.023)	(1.945)	(1.945)	(1.944)	(1.220)	(1.220)	(1.219)
Big5	0.432***	0.432***	0.427***	0.707***	0.704***	0.708***	-0.032	-0.033	-0.036
	(0.022)	(0.022)	(0.022)	(0.042)	(0.042)	(0.042)	(0.026)	(0.026)	(0.026)
Number of Meetings	0.051***	0.050***	0.050***	-0.018***	-0.018***	-0.018***	0.015***	0.015***	0.015***
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Board Size	0.031***	0.031***	0.031***	-0.042***	-0.041***	-0.041***	-0.007*	-0.007*	-0.006*
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Constant	8.125***	8.136***	8.232***	-2.002***	-1.996***	-2.024***	-0.261*	-0.251	-0.270*
	(0.129)	(0.129)	(0.133)	(0.245)	(0.244)	(0.252)	(0.154)	(0.153)	(0.158)
Observations	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.548	0.548	0.548	0.193	0.193	0.193	0.020	0.020	0.020
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 11: Group Identity in Public Sector Undertakings (PSUs)

This table provides regression results of managerial compensation and residual firm value on group identity for public sector undertakings (PSUs). *Log Compensation* is log of total compensation. *Firm value* is ratio of market value of equity to the book value of equity. *Residual Firm Value* is residual of regression of firm value on manager's total compensation of last year. Total compensation, market value of equity, and book value of equity are in million USD. *Homophily Language* is a dummy variable whose value equal to 1 if the board (at least one director) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same caste otherwise 0. PSU (*PSU*) is a dummy variable which takes value 1 if a firm is central or state government's commercial enterprise (also known as public sector undertakings (PSUs)); otherwise 0. In Columns (1) - (3), the dependent variable is logarithm of total compensation (*Log Compensation*) of a manager in that firm-year. In Columns (4)-(6), the dependent variable is residual firm value (*Residual Firm Value*). All other variables are defined in variable definition sheet in Appendix. The standard errors are in parentheses. All variables are winsorized at 1 %. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

VARIABLES	(1) Log Compensation	(2) Log Compensation	(3) Log Compensation	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual Firm Value
PSU	-2.258***	-2.248***	-1.736***	-0.160**	-0.146*	-0.132
	(0.067)	(0.065)	(0.338)	(0.081)	(0.078)	(0.413)
Homophily Language	0.006			0.044**		
	(0.018)			(0.022)		
PSU x Homophily Language	0.205			0.148		
	(0.133)			(0.163)		
Homophily Place		0.006			0.029	
		(0.018)			(0.022)	
PSU x Homophily Place		0.231			0.112	
		(0.153)			(0.187)	
Homophily Caste			0.007			-0.009
			(0.028)			(0.034)
PSU x Homophily Caste			-0.490			-0.001
			(0.342)			(0.419)
Size	0.117***	0.118***	0.117***	-0.015	-0.015*	-0.015*
	(0.007)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)
Firm Age	0.007***	0.007***	0.007***	-0.001*	-0.001*	-0.001*
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Institutional Ownership	0.033***	0.033***	0.033***	0.008***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Promoter Ownership	0.011***	0.011***	0.011***	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ROA	2.120***	2.121***	2.119***			
	(0.098)	(0.098)	(0.098)			
Leverage	0.012	0.012	0.013	-0.301***	-0.299***	-0.300***
	(0.053)	(0.053)	(0.053)	(0.061)	(0.061)	(0.061)
Volatility	-18.165***	-18.174***	-18.158***	-5.570***	-5.563***	-5.515***
	(1.006)	(1.006)	(1.006)	(1.214)	(1.214)	(1.214)
Big5	0.354***	0.354***	0.354***	-0.046*	-0.047*	-0.050*
	(0.022)	(0.022)	(0.022)	(0.027)	(0.027)	(0.026)
Number of Meetings	0.060***	0.060***	0.060***	0.014***	0.014***	0.014***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Board Size	0.050***	0.050***	0.050***	-0.005	-0.005	-0.005
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Constant	6.857***	6.857***	6.860***	0.043	0.056	0.085
	(0.127)	(0.127)	(0.129)	(0.154)	(0.154)	(0.156)
Observations	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.563	0.563	0.563	0.007	0.007	0.007
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Manager Designation	Yes	Yes	Yes	No	No	No
Year	Yes	Yes	Yes	Yes	Yes	Yes

Appendix A

Variable Definition Sheet

Variable	Description	Source
TotalCompensation	Log of manager's total compensation. Total compensation is in USD.	ProwessDx
Salary	Log of manager's salary. Salary is in USD.	ProwessDx
Firm Value (Market to Book Ratio)	Ratio of market value of equity to book value of equity. The market value of equity and book value of equity are in USD.	ProwessDx
Market Value	Sum of book value of debt and market value of equity. The market value of equity and book value of debt are in USD.	ProwessDx
Sales Growth	Growth in sales compare to last year	ProwessDx
Residual Firm Value	Residual from regression of market to Book ratio on last year managerial compensation	ProwessDx
Homophily Language	Dummy variable which takes value 1 if a manager and majority of board members have the same native language; otherwise 0.	Linguistic Survey of India (LSI)
Homophily Place	Dummy variable which takes value 1 if a manager and majority of board members have the same native place; otherwise 0.	Socio-Economic Caste Census (SECC)
Homophily Caste	Dummy variable which takes value 1 if a manager and majority of board members have the same caste; otherwise 0.	Socio-Economic Caste Census (SECC)
Size	Log of market capitalization. Market capitalization is in USD.	ProwessDx
ROA	Ratio of net income to total assets.	ProwessDx
ROE	Ratio of net income to book value of equity.	ProwessDx
Promoters Ownership	Perentage of equity ownership by promoters.	ProwessDx
Institutional Ownership	Perentage of equity ownership by institutional investers.	ProwessDx
Leverage	Ratio of total debt to total assets.	ProwessDx
South	Dummy variable which takes value 1 if the state is in south part of India; otherwise 0.	Socio-Economic Caste Census (SECC)
Firm Age	Age of the firm from incorporation	ProwessDx
Board Size	Number of board members	BoardEx
Number of Meetings	Number of board meetings in a year.	BoardEx
Board Diversity	Percentage of women in board	BoardEx
BG	Dummy variable which takes value 1 if a firm belongs to a business group; otherwise 0.	ProwessDx
Firm Complexity	Dummy variable which takes value 1 if complexity of a firm is above median; otherwise 0. The complexity of a firm is measured as square root of number of subsidiaries.	ProwessDx
PSU	Dummy variable which takes value 1 if a firm is central or state government's commercial enterprise (also known as public sector undertakings (PSUs)); otherwise 0.	ProwessDx
Volatility	Volatility of stock return.	Global Compustat
HIL	Dummy variable whose value equal to 1 if the independent directors (at least one) and the manager speaks same Indian language; otherwise 0.	Linguistic Survey of India (LSI)
HIP	Dummy variable whose value equal to 1 if the independent directors (at least one) and the manager belongs to same Indian state; otherwise 0.	Socio-Economic Caste Census (SECC)
HIC	Dummy variable whose value equal to 1 if the independent directors (at least one) and the manager have the same caste; otherwise 0.	Socio-Economic Caste Census (SECC)
Shock	Dummy variable which takes value 1 if a director died or non-voluntary retired in that year; otherwise 0.	ProwessDx
Age	Managerial age in years.	BoardEx
Gender	Dummy variable which takes value 1 if a manager is male and 0 if a manger is female.	BoardEx
Qualification(Education)	Number of educational degrees by a manager.	BoardEx

Appendix B

The following model setup is from Holmstrom (1979),

$maxE(V-c)$ (Principal's Problem)
s.t. $E(u(c-g(a_t))) \geq E(u(w))$ (Participatory Constraint)
$c = \phi + \theta V$ (Managerial Compensation)
$a_t^* \epsilon$ argmax $E(c-g(a_t))$ (Agent's Problem)
$V = S + b(S)a_t + \epsilon$ (Firm Value)

(1)

a_t: Managerial effort in the traditional models such as Holmstrom.
$g(a_t)$: The cost function of the manager's efforts
c: Managerial compensation
V:Firm Value
S: Firm Size
ϕ : Fixed component of the manager's compensation (Salary)
θ: Variable component of the manager's compensation
w: Reservation wage
ϵ: Random variable with a normal distribution. $\epsilon \sim N(0, \sigma^2)$

Agent utility function from Akerlof and Kranton (2005)[35]

$$U(c, a, m) = c\text{-}a + I_m - t_m|\hat{a} - a| \tag{2}$$

$\hat{a}$: Ideal effort
I_m : Additional utility due to group identity (or homophily)

For simiplicity, $I_m = 0$

Managerial efforts with identity $(a_n) = a + t_m|\hat{a} - a|$, where $\hat{a} \in a_{in-\hat{}group}, a_{out-\hat{}group}$

Here, $a_{in-\hat{}group}$ and $a_{out-\hat{}group}$ are working as upper bound and lower bound on the managerial ideal efforts.

$|\cdot|$ captures the deviation of manager's efforts from ideal in-group efforts (i.e., $|\cdot| > 0$) or ideal out-group efforts (i.e., $|\cdot| < 0$).

Complete Homophily (or in-group): When a manager and board have the social identity
In-complete Homophily (or out-group): When a manager and board have different social identities.
m: Variable to categorize different social identities such as native language, native place, or caste.

$$a_n = \begin{cases} a + t_m(\hat{a} - a), & \text{Where } a_{in-\hat{}group} > a \\ a - t_m(\hat{a} - a), & \text{Where } a_{out-\hat{}group} < a \end{cases} \tag{3}$$

Managerial cost function, $g(a_n) = \frac{1}{2} g a_n^2$

[35] I_m is identity utility, which is considered as 0 for simplicity, and $t_m|\hat{a} - a|$ is disutility for diverging from the ideal effort level for in-group member, denoted by $\hat{a}$.

Here, g is a constant.

$$\text{Managerial cost function, } g(a_n)= \begin{cases} \frac{1}{2}g(a + t_m(\hat{a} - a))^2 & \text{[In-group]} \\ \frac{1}{2}g(a - t_m(\hat{a} - a))^2 & \text{[Out-group]} \end{cases} \tag{4}$$

$$\text{Firm value, } V = \begin{cases} S + b(S)(a + t_m(\hat{a} - a)) + \epsilon & \text{[In-group]} \\ S + b(S)(a - t_m(\hat{a} - a)) + \epsilon & \text{[Out-group]} \end{cases} \tag{5}$$

$$\text{Managerial Compensation, c= } \begin{cases} \phi + \theta(S + b(S)(a + t_m(\hat{a} - a))) & \text{[In-group]} \\ \phi + \theta(S + b(S)(a - t_m(\hat{a} - a))) & \text{[Out-group]} \end{cases} \tag{6}$$

Maximizing agent (manager's) problem (or agent's net compensation) w.r.t its efforts (a),

I. When there is a homophily between a manager and the board,

$\frac{\partial}{\partial a}(\phi + \theta(S + b(S)(a + t_m(\hat{a} - a))) - \frac{1}{2}g(a + t_m(\hat{a} - a))^2) = 0$

$\theta b(S)(1 + t_m) - g(a + t_m(\hat{a} - a))(1 + t_m) = 0$

$\frac{\theta b(S)}{g} = a + t_m(\hat{a} - a)$

$$a^*_{in-group} = \frac{1}{1 - t_m}(\frac{\theta b(S)}{g} - t_m\hat{a})$$

II. When there is a no homophily between a manager and the board,

$$\frac{\partial}{\partial a}(\phi + \theta(S + b(S)(a - t_m(\hat{a} - a))) - \frac{1}{2}g(a - t_m(\hat{a} - a))^2) = 0$$

$\theta b(S)(1 - t_m) - g(a - t_m(\hat{a} - a))(1 - t_m) = 0$

$$\frac{\theta b(S)}{g} = a - t_m(\hat{a} - a)$$

$$a^*_{out-group} = \frac{1}{1 + t_m}(\frac{\theta b(S)}{g} + t_m\hat{a})$$

$a^*_{in-group}$ and $a^*_{out-group}$ depends on the managerial effort which depends on three factors (- sensitivity of effort to identity (t_m), sensitivity of the firm value to effort ($b(S)$), and sensitivity of the managerial compensation (variable component) to firm value(θ)) and their dynamics. On combining the above two questions of efforts and rearranging,

$(1 - t_m)a^*_{in-group} + (1 + t_m)a^*_{out-group} = \frac{2}{k}(b\theta)$

As b is sensitivity of the firm value to effort and θ is sensitivity of compensation to firm value. Following the chain rule in derivatives, above equation can be seen as,

$(1 - t_m)a^*_{in-group} + (1 + t_m)a^*_{out-group} \cong \frac{dc}{da}$

This shows that impact of a manager's effort on compensation depends on identity overlap between a

manager and the board and the sensitivity of manager's utility to the identity i.e. (t_m). In other terms, the impact of a manager's effort on compensation is weighted average of manager's in-group efforts and out-group efforts weighted by sensitivity of manager's utility to the identity.

The $\frac{dc}{da}$ has upper bound of 1 (complete effort convert into compensation) and lower bound of 0 (effort does not convert into compensation at all).

$0 \leq (1 - t_m)a^*_{in-group} + (1 + t_m)a^*_{out-group} \leq 1$

Lower bound does provide the real solution as in complete homophily and for incomplete homophily, both $a^*_{in-group}$ and $a^*_{out-group}$ would be 0 and it cannot be the solution. Therefore, from the upper bound,

$(1 - t_m)a^*_{in-group} + (1 + t_m)a^*_{out-group} = 1$

For complete homophily, $a^*_{in-group}$ is $\frac{1}{1-t_m}$ and for incomplete homophily, $a^*_{out-group}$ is $\frac{1}{1+t_m}$. As t_m is positive and lies between 0 and 1, this shows,

$a^*_{in-group} > a^*_{out-group}$

Proposition. *The managerial efforts are higher when he is in-group (complete homophily) with board members than when he is out-group with board members.*

This implies that in the case of complete homophily, the manager's effort at equilibrium is higher than it is for incomplete homophily.

Lemma 1. *Higher managerial efforts lead to higher firm value at equilibrium for complete homophily compare to incomplete homophily case.*

Proof. As,

$a^*_{in-group} > a^*_{out-group}$

Therefore, at equilibrium (for same b(S)),

$S + b(S)(a^*_{in-group} + t_m(\hat{a} - a^*_{in-group})) > S + b(S)(a^*_{out-group} - t_m(\hat{a} - a^*_{out-group}))$

$V^*_{in-group} > V^*_{out-group}$

This implies that a firm in which there is complete homophily has a higher value than a firm with incomplete homophily, this derives from the difference in managerial efforts in both cases.

Lemma 2. *Higher managerial efforts and firm value for complete homophily lead to higher managerial compensations at equilibrium compared to incomplete homophily.*
Proof. From the participation constraint,

$$E[c] = \begin{cases} E[c]_{in-group} = w + \frac{1}{2}k(a^*_{in-group})^2 - \frac{\nu}{2}\theta^2_{in-group}\sigma^2, \text{ For complete homophily} \\ E[c]_{out-group} = w + \frac{1}{2}k(a^*_{out-group})^2 - \frac{\nu}{2}\theta^2_{out-group}\sigma^2, \text{ For in-complete homophily} \end{cases} \quad (7)$$

$$\text{Principal's utility} = E[V - c] = \begin{cases} S + b(S)a^*_{in-group} - E[c]_{in-group} \\ S + b(S)a^*_{out-group} - E[c]_{out-group} \end{cases}$$

First Order Condition (FOC) with respect to θ,

$$\frac{\partial}{\partial\theta}(E[V-c])=0$$

I. When there is a homophily between a manager and the board,

$$\frac{b(S)^2}{(1-t_m)g})-\frac{kb(S)}{(1-t_m)^2g^2}(\theta b(S)-t_m\hat{a}g)-\nu\sigma^2\theta=0$$

$$\theta^*_{in-group}=\frac{1}{b(S)}[\frac{1+t_mk\hat{a}}{1-\frac{\nu k(1-t_m)}{(\frac{b}{\sigma})^2}}]$$

II. When there is a no homophily between a manager and the board,

$$\frac{b(S)^2}{(1+t_m)g})-\frac{kb(S)}{(1+t_m)^2g^2}(\theta b(S)+t_m\hat{a}g)-\nu\sigma^2\theta=0$$

$$\theta^*_{out-group}=\frac{1}{b(S)}[\frac{1-t_mk\hat{a}}{1-\frac{\nu k(1+t_m)}{(\frac{b}{\sigma})^2}}]$$

if $t_m=\theta=b(S)=k=\sigma=\nu=0.5$ and $\hat{a}\leq 0.667$,

$\theta^*_{out-group}>\theta^*_{in-group}$

As, $a^*_{in-group}>a^*_{out-group}$

Therefore, from eq(3), $E[c]^*_{in-group}>E[c]^*_{out-group}$

This implies that the manager's total compensation is higher for a firm with group identity. This is the cost of the in-group favoritism to the firm.

There is one limitation of this model, such that, the cost of managerial effort can be pecuniary and non-pecuniary. Following the work of Holmstrom and Milgrom (1987), only the pecuniary cost of effort is considered.

In sum, the firm value and the manager's total compensation (and salary) are higher if the manager and the board share the same identity (i.e., group identity or homophily). Although the additional managerial compensation is a cost to the firm due to the in-group favoritism, this cost is compensated for by the manager's extra efforts to increase the firm value. In the next section, I verify whether these theoretical predictions sustain the empirical analysis.

Internet Appendix (IA)

Group Identity and Agency Frictions

Table IA1: Distribution of Firms and Observations

This table provides yearly distribution of firms and observations. The sample is from 2004-2018 with 28,026 observations.

Year	Unique Firms	Observations
2004	237	360
2005	365	554
2006	959	1,386
2007	1120	1,621
2008	1258	1,797
2009	1351	1,920
2010	1391	1,943
2011	1458	2,045
2012	1546	2,172
2013	1558	2,184
2014	1539	2,128
2015	1562	2,244
2016	1760	2,564
2017	1767	2,627
2018	1552	2,481
Total		28,026

Panel B : Caste-wide Distribution of Managers and Directors

This panel reports the distribution of managers and directors based on their caste. The caste classification is based on administrative data (Socio-Economic Caste Census (SECC) database) and categorize as Schedule Caste (SC), Schedule Tribe (ST), and other than SC and ST ($Other$).

	Caste			
	Other	SC	ST	Total
Directors	94,280	9,407	5,922	109,609
Manager (& Equivalent)	23,865	2,427	1,734	28,026
Total	118,145	11,834	7,656	137,635

Table IA2: SECC Data Preview and Surname Extraction
Panel A: Preview of Socio-Economic Caste Census (SECC) Data

head_of_hh	gender	age
AJAY KUMAR SAH\अजय कुमार साह	M	34
AJAY KUMAR SINGH\अजय कुमार सिंह	M	47
AJIT KUMAR SINGH\अजित कुमार सिंह	M	33
AKHILESH SINGH\अखिलेश सिंह	M	47
AMERIKA BAITHA\अमेरिका बैठा	M	32

state	district	tehsil	panchayat
BIHAR	Purba Champaran	Phenhara	Bishunpur Basant(Village)
BIHAR	Purba Champaran	Phenhara	Bishunpur Basant(Village)
BIHAR	Purba Champaran	Phenhara	Bishunpur Basant(Village)
BIHAR	Purba Champaran	Phenhara	Bishunpur Basant(Village)
BIHAR	Purba Champaran	Phenhara	Bishunpur Basant(Village)

social_cat	fathers_and_mothers_name	total_members	village
Other	गोपाल जी\सीया देवी	6	Bishunpur Basant
Other	राजमंगल सहि\मंजू कुँवर	7	Bishunpur Basant
Other	लालबाबू सहि\श्रीकांत देवी	5	Bishunpur Basant
Other	रामश्रय सहि\फुलमत देवी	3	Bishunpur Basant
SC	भखिारी बैठा\धनेशरी कुँवर	2	Bishunpu

Panel B: Cleaning of Socio-Economic Caste Census (SECC) Data

The SECC 2011 data contains information on an individual name, parent's name, gender, age, caste, address (state and district), and others. This information on 474 million Indians is available in 48 CSV files. The surname data developed from SECC provide a surname, state, district, and caste information. To scrape surnames from SECCC data, the cloud computing on Microsoft Azure is used with an instance of 32 GB RAM and 1 TB hard disk.

The steps to scrape surnames from Indian names are as follows:

- The observations with missing names are omitted.
- All the special characters and names in the native language are replaced by white space.
- The slicing of the words is done till all surnames are obtained.
- If the surname length is 2 or 1, it might be initial rather than surname.

In SECC data, south Indian names and long names are recorded differently. For example, names like "P.K. Chandrashekhar Rao" is penned as "Chandrashekhar Rao P.K." in the SECC data. Therefore, when surnames are either 1 or 2 length, those are not surnames but initials. These initials are replaced by second name if the name has three words such as "Rao" from "Chandrashekhar Rao P.K." or with first name if a name has two words such as "Gupta" from "Gupta S.K." Fifth, even after all these steps, if there are surnames with length 1 or 2, those surnames are removed to curb noise in the final dataset. Otherwise, the final dataset would contain certain discrepancies, such as the surname "r" which is part of "Sahil Kuma r", but, in this name, the surname is "Kumar" and not "r". All the designations such as Shri, Sh., Mr., Mr, Dr., Dr, Col., Maj., and others are removed in the beginning. Also, Indian names are written differently in northern India and southern India. To mitigate these concerns, various books and websites are referred. The Python code to scrape Indian surnames and mapped identities from SECC data is available on request.

Panel C: Common Indian Surnames among Managers

This panel shows the common surnames among the managers and their ranking. The ranking is based on the frequency of the surnames in the sample.

Surname	Rank	Surname	Rank	Surname	Rank	Surname	Rank
Gupta	1	Bangur	51	Mathur	101	Ramachandran	151
Jain	2	Kabra	52	Parikh	102	Tamboli	152
Agarwal	3	Kalyani	53	Sawhney	103	Agarwala	153
Shah	4	Bhartia	54	Shahra	104	Mahajan	154
Patel	5	Choudhary	55	Bhansali	105	Malik	155
Mehta	6	Rathi	56	Choksi	106	Basu	156
Reddy	7	Thakkar	57	Mehra	107	Iyer	157
Kumar	8	Kapur	58	Pillai	108	Khatau	158
Rao	9	Modi	59	Roy	109	Murthy	159
Goenka	10	Bansal	60	Muthiah	110	Babu	160
Poddar	11	Daga	61	Ravi	111	Baid	161
Sharma	12	Godrej	62	Ambani	112	Chhabria	162
Khaitan	13	Jalan	63	Gaur	113	Ganesh	163
Agrawal	14	Maheshwari	64	Shriram	114	Kapadia	164
Garg	15	Sethi	65	Wadia	115	Adani	165
Birla	16	Nair	66	Das	116	Godha	166
Mittal	17	Chaturvedi	67	Patil	117	Mahindra	167
Singhania	18	Kanoria	68	Garware	118	Rai	168
Jindal	19	Sheth	69	Bhat	119	Adhikari	169
Aggarwal	20	Soni	70	Krishnan	120	Bhattacharya	170
Doshi	21	Dhanuka	71	Advani	121	Chatterjee	171
Kothari	22	Mallya	72	Jha	122	Ghosh	172
Patodia	23	Patni	73	Puri	123	Kaul	173
Jhunjhunwala	24	Taparia	74	Ansal	124	Mansukhani	174
Arora	25	Bhargava	75	Chowdhary	125	Mirchandani	175
Kapoor	26	Lodha	76	Gogri	126	Misra	176
Parekh	27	Sanghvi	77	Khurana	127	Mukherjee	177
Goyal	28	Krishna	78	Mohan	128	Parakh	178
Oswal	29	Mistry	79	Munot	129	Patwari	179
Gandhi	30	Ruia	80	Punj	130	Raja	180
Munjal	31	Seth	81	Uppal	131	Subramaniam	181
Surana	32	Jatia	82	Chopra	132	Tripathi	182
Goel	33	Hinduja	83	Dhoot	133	Gopalakrishnan	183
Saraf	34	Narayanan	84	Mafatlal	134	Kanodia	184
Desai	35	Piramal	85	Rajan	135	Labroo	185
Somany	36	Shetty	86	Ramakrishnan	136	Mundhra	186
Chandra	37	Khanna	87	Tandon	137	Murugappan	187
Mody	38	Sinha	88	Wadhwa	138	Tuteja	188
Srinivasan	39	Jaipuria	89	Kejriwal	139	Vellayan	189
Bajaj	40	Banerjee	90	Minda	140	Chowgule	190
Raju	41	Datta	91	Oberoi	141	Ganguly	191
Shroff	42	Ram	92	Saxena	142	Relan	192
Amin	43	Bhandari	93	Trivedi	143	Srivastava	193
Malhotra	44	Bhatia	94	Handa	144	Taneja	194
Prasad	45	Dalmia	95	Joshi	145	Chordia	195
Kulkarni	46	Nanda	96	Kanwar	146	Dasgupta	196
Menon	47	Jhawar	97	Saboo	147	Kashyap	197
Somani	48	Mishra	98	Bagrodia	148	Mammen	198
Tata	49	Kirloskar	99	Chandran	149	Sardana	199
Thapar	50	Lalbhai	100	Dandekar	150	Shankar	200

Table IA3: Homophily in Different Industries

This table provides industry distribution based on homophily measures. The homophily measures are based on native language, native place, and caste. Industry classification uses ProwessDx database definition. Please refer to Appendix A for variable definitions.

Industry	Homophily Language	Homophily Place	Homophily Caste
Abrasives	0.273	0.182	0.896
Agricultural machinery	0.364	0.424	1.000
Air transport services	0.059	0.088	0.824
Air-conditioners & refrigerators	0.120	0.120	0.978
Aluminium & aluminium products	0.216	0.279	1.000
Animation content provider	0.294	0.235	1.000
Bakery products	0.133	0.000	0.733
Beer & alcohol	0.432	0.358	0.802
Boilers & turbines	0.207	0.276	0.931
Books & cards	0.355	0.355	0.645
Business services & consultancy	0.448	0.440	0.890
Castings & forgings	0.456	0.479	0.927
Caustic soda	0.000	0.000	1.000
Cement	0.330	0.360	0.891
Ceramic products	0.545	0.625	0.670
Cloth	0.416	0.393	0.848
Coal & lignite	0.160	0.080	1.000
Cocoa products & confectionery	0.667	0.500	1.000
Coffee	0.133	0.133	1.000
Commercial complexes	0.188	0.125	0.750
Commercial vehicles	0.130	0.014	0.812
Communication equipment	0.364	0.182	0.818
Computer software	0.313	0.314	0.920
Computers, peripherals & storage devices	0.483	0.483	0.897
Consumer electronics	0.590	0.872	1.000
Conventional electricity	0.343	0.146	0.961
Copper & copper products	0.219	0.438	0.943
Cosmetics, toiletries, soaps & detergents	0.391	0.370	0.929
Cotton & blended yarn	0.423	0.440	0.936
Courier services	0.333	0.333	1.000
Crude oil & natural gas	0.265	0.193	0.976
Dairy products	0.418	0.304	0.968
Diversified	0.474	0.433	0.883
Diversified automobile	0.000	0.000	1.000
Diversified cotton textile	0.307	0.281	0.784
Diversified machinery	0.370	0.348	0.957
Diversified metal & metal products	0.720	0.560	0.980
Drugs & pharmaceuticals	0.424	0.398	0.907
Dry cells	0.433	0.500	0.417
Dyes & pigments	0.632	0.555	0.907
Education	0.365	0.394	0.865
Electricity distribution	1.000	0.364	1.000
Electricity transmission	0.091	0.091	1.000
Engines	0.360	0.260	0.780
Exhibition of films	0.629	0.571	0.657
Ferro alloys	0.571	0.578	1.000
Fertilisers	0.307	0.259	0.820
Floriculture	1.000	1.000	1.000
Footwear	0.794	0.706	0.941
Gems & jewellery	0.502	0.515	0.783
General purpose machinery	0.431	0.434	0.912
Generators, transformers & switchgears	0.353	0.335	0.909
Glass & glassware	0.198	0.189	0.468
Granite	0.589	0.589	1.000
Health services	0.642	0.572	0.994
Hotels & restaurants	0.469	0.406	0.901
Housing construction	0.561	0.585	1.000
ITES	0.403	0.368	0.924
Industrial construction	0.286	0.292	0.842
Industrial machinery	0.530	0.561	0.944
Infrastructural construction	0.458	0.470	0.863
Inorganic chemicals	0.484	0.339	0.890
Lubricants, etc.	0.467	0.417	0.983
Machine tools	0.287	0.230	0.955
Man-made filaments & fibres	0.434	0.338	0.844
Marine foods	0.000	0.000	1.000
Media-broadcasting	0.450	0.424	0.900
Media-content	0.383	0.353	0.985
Media-print	0.644	0.637	0.778

Table IA3: Homophily in Different Industries (Contd.)

This table provides industry distribution based on homophily measures. The homophily measures are based on native language, native place, and caste. Industry classification uses ProwessDx database definition. Please refer to Appendix A for variable definitions.

Industry	Homophily Language	Homophily Place	Homophily Caste
Milling products	1.000	1.000	1.000
Minerals	0.455	0.455	0.791
Mining & construction equipment	0.629	0.477	0.917
Miscellaneous electrical machinery	0.311	0.311	0.977
Miscellaneous manufactured articles	0.632	0.472	0.962
Natural gas trading & distribution	0.393	0.115	0.984
Online marketplaces	0.125	0.125	0.625
Organic chemicals	0.365	0.403	0.962
Other agricultural products	0.518	0.447	0.961
Other automobile ancillaries	0.382	0.344	0.911
Other chemical products	0.431	0.447	0.941
Other construction & allied activities	0.491	0.448	0.827
Other construction materials	0.280	0.326	0.970
Other consumer goods	0.602	0.637	0.832
Other domestic appliances	0.524	0.444	0.841
Other electronics	0.389	0.436	0.940
Other ferrous metal products	0.420	0.510	0.874
Other industrial machinery	0.480	0.600	0.640
Other leather & related products	0.174	0.000	0.783
Other miscellaneous services	0.538	0.487	0.952
Other non-ferrous metals & metal products	0.224	0.168	0.825
Other recreational & allied services	0.390	0.378	0.976
Other textiles	0.476	0.363	0.855
Other transport equipment & ancillaries	0.440	0.368	0.800
Paints & varnishes	0.336	0.252	0.969
Paper & newsprint	0.334	0.268	0.863
Paper products	0.600	0.600	0.900
Passenger vehicles	0.067	0.000	0.733
Pesticides	0.395	0.328	0.993
Pig iron	1.000	1.000	1.000
Plastic films & flexible packaging	0.235	0.204	0.864
Plastic furniture, floorings & miscellaneous items	0.561	0.606	0.947
Plastic packaging goods	0.486	0.646	0.913
Plastic tubes, pipes, fittings & sheets	0.390	0.444	0.912
Polymers	0.432	0.432	0.821
Poultry & meat products	0.250	0.250	0.250
Processed foods	0.483	0.510	0.867
Production & distribution of films	0.293	0.293	0.683
Railway transport services	0.095	0.143	0.857
Readymade garments	0.409	0.440	0.789
Refinery	0.388	0.321	0.993
Refractories	0.360	0.027	1.000
Renewable electricity	0.318	0.136	1.000
Retail trading	0.332	0.234	0.889
Road transport infrastructure services	0.600	0.600	1.000
Road transport services	0.769	0.808	0.885
Rubber products	0.440	0.493	0.853
Shipping transport infrastructure services	0.246	0.000	1.000
Shipping transport services	0.374	0.325	0.894
Soda ash	0.550	0.450	1.000
Sponge iron	0.404	0.372	0.862
Starches	0.091	0.000	1.000
Steel	0.391	0.351	0.899
Steel pipes & tubes	0.544	0.424	0.921
Storage & distribution	0.667	0.383	0.433
Storage batteries	0.238	0.032	0.635
Sugar	0.350	0.355	0.928
Tea	0.520	0.376	0.769
Telecommunication services	0.305	0.374	0.920
Textile processing	0.421	0.381	0.802
Tobacco products	0.316	0.316	0.908
Tourism	0.083	0.111	0.833
Transport logistics services	0.426	0.414	0.975
Two & three wheelers	0.557	0.585	0.802
Tyres & tubes	0.541	0.557	0.907
Vegetable oils & products	0.592	0.598	0.858
Wholesale trading	0.401	0.378	0.849
Wires & cables	0.593	0.586	0.947
Wood & wood products	0.548	0.444	0.976

Table IA4: Univariate Analysis

This table shows univariate analysis of managerial compensation, firm value and residual firm value with respect to homophily variables. The homophily variables are based on native language (*Homophily Language*), native place (*Homophily Place*), and caste (*Homophily Caste*).Log Compensation (*Log Compensation*) is logarithm of total managerial compensation which is in million USD. Firm Value (*Firm Value*) is price to book ratio. Residual Firm Value (*Residual Firm Value*) is the residual of regressing price to book ratio on last year's managerial total compensation. The detailed variable definitions are provided in Appendix A.

		Log Compensation	Firm Value	Residual Firm Value
Homophily Language	1	10.632	2.375	0.001
	0	10.595	2.049	-0.024
	Diff	0.037	0.325	0.025
	t-value	1.435	8.767	1.178
Homophily Place	1	10.645	2.376	-0.007
	0	10.587	2.028	-0.019
	Diff	0.058	0.349	-0.012
	t-value	2.203	9.313	-0.564
Homophily Caste	1	10.607	2.315	0.003
	0	10.635	2.229	-0.016
	Diff	-0.027	0.085	0.019
	t-value	-0.67	1.452	0.584

Table IA5: Covariate Balance

This table shows co-variate balance between treated group (*Homophily* equals to 1) and control group (*Homophily* equals to 0). The homophily variables are based on native language (*Homophily Language*), native place (*Homophily Place*), and caste (*Homophily Caste*). P-value suggests whether the difference between treated and control value is statistically significant or not. The P-value of 0.001, 0.05, and 0.1 shows the significance at 1 %, 5%, and 10%. The variable definitions are provided in Appendix A.

	Homophily Language			Homophily Place			Homophily Caste		
	Treated	Control	P-value	Treated	Control	P-value	Treated	Control	P-value
Size	19.057	19.005	0.335	19.042	19.006	0.531	19.218	19.162	0.084
Firm Age	37.918	39.730	0.069	36.874	35.922	0.339	38.621	35.979	0.000
ROA	0.063	0.061	0.631	0.066	0.069	0.465	0.067	0.069	0.260
Institutional Ownership	23.502	23.937	0.428	24.002	23.743	0.643	23.877	25.812	0.000
Promoters Ownership	52.675	52.290	0.586	52.473	53.873	0.053	53.958	51.055	0.000
Leverage	0.176	0.182	0.443	0.160	0.158	0.820	0.165	0.167	0.584
Volatility	0.023	0.023	0.386	0.023	0.023	0.755	0.022	0.022	0.282
Big 5	0.540	0.564	0.242	0.541	0.544	0.895	0.587	0.621	0.006
Number of Meetings	9.385	9.451	0.477	9.342	9.379	0.697	9.252	9.377	0.015
Board Size	12.608	12.551	0.697	12.515	12.294	0.131	12.689	12.489	0.026
Age	64.737	64.417	0.499	64.888	64.881	0.989	64.247	64.227	0.944
Gender	1.985	1.988	0.477	1.979	1.982	0.636	1.981	1.984	0.336
Education	1.982	2.016	0.303	2.048	2.077	0.410	2.047	1.951	0.000

Table IA6: Group Identity and Managerial Compensation (with all controls)

This table provides regression results of executive compensation on three measures of group identity - homophily language, homophily place, and homophily caste. *Log Compensation* is log of total compensation. The total compensation is in million USD. *Homophily Language* is a dummy variable whose value equal to 1 if the board (at least one director) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same caste otherwise 0. All other variables are defined in variable definition sheet in Appendix. The standard errors are in parentheses. All variables are winsorized at 1 %. In all panels, *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)
VARIABLES	Log Compensation	Log Compensation	Log Compensation
Homophily Language	0.178***		
	(0.068)		
Homophily Place		0.033	
		(0.076)	
Homophily Caste			-0.718***
			(0.112)
Size	-0.032	-0.035	-0.037
	(0.064)	(0.064)	(0.063)
Firm Age	0.201***	0.200***	0.169***
	(0.043)	(0.043)	(0.043)
Instiutional Ownership	0.012***	0.011***	0.011**
	(0.004)	(0.004)	(0.004)
Promoters Ownership	0.016***	0.015***	0.014***
	(0.005)	(0.005)	(0.005)
ROA	0.602*	0.614*	0.671**
	(0.336)	(0.336)	(0.334)
Leverage	-0.040	-0.025	-0.012
	(0.276)	(0.277)	(0.275)
Volatility	-5.788	-5.747	-6.563
	(5.267)	(5.273)	(5.240)
BIG5	0.061	0.059	0.072
	(0.102)	(0.102)	(0.101)
Number of Meetings	0.011	0.011	0.011
	(0.012)	(0.012)	(0.012)
Board Size	-0.017	-0.017	-0.014
	(0.010)	(0.010)	(0.010)
Age	0.002	0.002	0.002
	(0.003)	(0.003)	(0.003)
Gender	0.585***	0.588***	0.603***
	(0.219)	(0.219)	(0.218)
Education	0.014	0.012	0.021
	(0.039)	(0.039)	(0.039)
Constant	1.932	2.153	4.065*
	(2.168)	(2.170)	(2.174)
Observations	3,542	3,542	3,542
R-squared	0.750	0.750	0.753
Manager Designation	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Table IA7: External Validity Test

This table shows the regression results for external validity tests. *Homophily Language* is a dummy variable whose value equals to 1 if a manager and at least one board member speaks same Indian language, otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if a manager and at least one board member belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if a manager and at least one board member belong to same caste, otherwise 0. Panel A - Panel C report the regression results of external validity tests for compensation variables - compensation relative to industry average (*Ind_Average Relative*), log Salary (*Log Salary*), and log of salary plus bonus plus sitting fees (*Log SBF*). Panel D reports the regression results of external validity test for homophily construct. Panel E and Panel F reports the regression results of external validity of firm value using sales growth (*Sales Growth*) and sum of market value of equity and book value of debt (*Log Market Value*) . The standard errors are in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Panel A: Compensation Relative to Industry Peers

Panel A reports the regression results of external validity tests for compensation variable using the compensation relative to industry average (*Ind_Average Relative*).

	(1)	(2)	(3)
VARIABLES	Ind_Average Relative	Ind_Average Relative	Ind_Average Relative
Homophily Language	0.046**		
	(0.021)		
Homophily Place		0.024	
		(0.022)	
Homophily Caste			-0.117***
			(0.041)
Size	0.083***	0.083***	0.082***
	(0.018)	(0.018)	(0.018)
Firm Age	0.003	0.003	0.003
	(0.017)	(0.017)	(0.017)
Instiutional Ownership	0.012***	0.012***	0.012***
	(0.002)	(0.002)	(0.002)
Promoters Ownership	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)
ROA	0.857***	0.855***	0.852***
	(0.099)	(0.099)	(0.099)
Leverage	-0.115*	-0.114*	-0.116*
	(0.068)	(0.068)	(0.068)
Volatility	-3.661***	-3.649***	-3.634***
	(0.968)	(0.968)	(0.968)
BIG5	0.093**	0.092**	0.093**
	(0.038)	(0.038)	(0.038)
Number of Meetings	0.012***	0.012***	0.012***
	(0.004)	(0.004)	(0.004)
Board Size	-0.004	-0.005	-0.004
	(0.004)	(0.004)	(0.004)
Constant	-1.957***	-1.946***	-1.833***
	(0.673)	(0.673)	(0.673)
Observations	28,026	28,026	28,026
R-squared	0.737	0.737	0.737
Manager Designation	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Panel B: Managerial Salary and Group Identity

Panel B reports the regression result of external validity test for compensation variable using log Salary (*Log Salary*).

VARIABLES	(1) Log Salary	(2) Log Salary	(3) Log Salary
Homophily Language	0.031*		
	(0.019)		
Homophily Place		0.057***	
		(0.019)	
Homophily Caste			-0.128***
			(0.030)
Size	0.119***	0.120***	0.120***
	(0.007)	(0.007)	(0.007)
Leverage	-0.020	-0.019	-0.019
	(0.041)	(0.041)	(0.041)
Firm Age	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
ROA	0.048***	0.047***	0.049***
	(0.016)	(0.016)	(0.016)
Promoters Ownership	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)
Instituional Ownership	0.032***	0.032***	0.032***
	(0.001)	(0.001)	(0.001)
Constant	8.400***	8.379***	8.530***
	(0.124)	(0.124)	(0.126)
Observations	12,990	12,990	12,990
R-squared	0.256	0.256	0.257
Managerial Designation	Yes	Yes	Yes
Year	Yes	Yes	Yes

Panel C : Salary, Bonus, and Sitting Fees

Panel C report the regression results of external validity tests for compensation variables using log of salary plus bonus plus sitting fees (*Log SBF*)

VARIABLES	(1) SBF	(2) SBF	(3) SBF
Homophily Language	0.079***		
	(0.021)		
Homophily Place		0.072***	
		(0.022)	
Homophily Caste			-0.133***
			(0.040)
Size	0.110***	0.111***	0.109***
	(0.017)	(0.017)	(0.017)
Firm Age	0.038**	0.038**	0.039**
	(0.017)	(0.017)	(0.017)
Institutional Ownership	0.016***	0.016***	0.016***
	(0.001)	(0.001)	(0.001)
Promoters Ownership	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)
ROA	1.239***	1.236***	1.232***
	(0.098)	(0.098)	(0.098)
Leverage	-0.132*	-0.130*	-0.133**
	(0.068)	(0.068)	(0.068)
Volatility	-5.968***	-5.956***	-5.930***
	(0.957)	(0.957)	(0.957)
Big5	0.097**	0.095**	0.097**
	(0.038)	(0.038)	(0.038)
Number of Meetings	0.014***	0.014***	0.014***
	(0.004)	(0.004)	(0.004)
Board Size	-0.005	-0.005	-0.004
	(0.004)	(0.004)	(0.004)
Constant	6.791***	6.780***	6.951***
	(0.664)	(0.664)	(0.664)
Observations	27,972	27,972	27,972
R-squared	0.761	0.761	0.761
Manager Designation	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Panel D : Alternative Homophily Construct

Panel D reports the regression results of managerial compensation on homophily variables constructed using the alternative dataset. To check the external validity of the homophily variables, rather than using the SECC 2011 and LSI data, data from Indianchildnames.com is used. *Homophily* is a dummy variable whose value equals to 1 if the manager speaks the same Indian language as (at least one) of the board members, otherwise 0. Please refer to Appendix A for variable definitions. The standard errors are in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

VARIABLES	(1) Log Compensation	(2) Log Salary
Homophily	0.138***	0.102***
	(0.032)	(0.032)
Size	0.368***	0.337***
	(0.029)	(0.024)
ROA	1.364***	0.854***
	(0.136)	(0.145)
Leverage	-0.308***	-0.228**
	(0.091)	(0.089)
Board Size	0.004	-0.002
	(0.007)	(0.005)
Constant	7.935***	8.016***
	(0.226)	(0.189)
Observations	14,523	14,229
R-squared	0.805	0.804
Firm Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes

Panel E: Sales Growth and Group Identity

This panel reports regression results of sales growth on homophily (group identity) measures. In Columns (1)-(3), the dependent variable is sales growth (*Sales Growth*). Please refer to Appendix A for variable definitions.

VARIABLES	(1) Sales Growth	(2) Sales Growth	(3) Sales Growth
Homophily Language	0.031***		
	(0.009)		
Homophily Place		0.028***	
		(0.009)	
Homophily Caste			0.009
			(0.014)
Size	0.003	0.003	0.002
	(0.003)	(0.003)	(0.003)
Leverage	-0.107***	-0.106***	-0.106***
	(0.019)	(0.019)	(0.019)
Firm Age	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)
ROA	0.048***	0.047***	0.049***
	(0.016)	(0.016)	(0.016)
Promoters Ownership	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Instituional Ownership	0.001**	0.001**	0.001**
	(0.000)	(0.000)	(0.000)
Constant	0.117**	0.119**	0.131**
	(0.058)	(0.058)	(0.059)
Observations	15,612	15,612	15,612
R-squared	0.076	0.076	0.076
Managerial Designation	No	No	No
Year	Yes	Yes	Yes

Panel F: Market Value of Firm and Group Identity

This panel reports regression results of log of market value of equity and book value of debt on homophily (group identity) measures. Please refer to Appendix A for variable definitions.

VARIABLES	(1) Log Market Value	(2) Log Market Value	(3) Log Market Value
Homophily Language	0.030***		
	(0.011)		
Homophily Place		0.026**	
		(0.012)	
Homophily Caste			-0.003
			(0.022)
Size	0.262***	0.262***	0.262***
	(0.009)	(0.009)	(0.009)
Firm Age	0.011	0.011	0.011
	(0.009)	(0.009)	(0.009)
Institutional Ownership	0.025***	0.025***	0.025***
	(0.001)	(0.001)	(0.001)
Promoters Ownership	0.015***	0.015***	0.015***
	(0.001)	(0.001)	(0.001)
Leverage	-0.062*	-0.061*	-0.062*
	(0.035)	(0.035)	(0.035)
Volatility	-5.404***	-5.399***	-5.331***
	(0.511)	(0.511)	(0.510)
Big5	0.174***	0.173***	0.173***
	(0.020)	(0.020)	(0.020)
Number of Meetings	0.026***	0.026***	0.026***
	(0.002)	(0.002)	(0.002)
Board Size	0.019***	0.019***	0.022***
	(0.002)	(0.002)	(0.002)
Constant	-2.102***	-2.105***	-2.097***
	(0.356)	(0.356)	(0.356)
Observations	28,026	28,026	28,026
R-squared	0.934	0.934	0.934
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Table IA8: Identities Index, Managerial Compensation and Firm Value

This table provides regression results of managerial compensation, firm value, and residual firm value on homophily index. Homophily index (*Homophily*) is a dummy variable which takes value 1 if all the three identities match between a manager and at least one of the board members; otherwise 0. Please refer to Appendix A for variable definitions. Panel A uses the homophily variable which takes value 1 if all the three identities match between a manager and at least one of the board members; otherwise 0. Panel B uses the continuous variable on homophily which is defined as a ratio of number of identity match between a manager and the maximum board members to the total number of board members. The standard errors are in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Panel A: All Identities Match

Panel A uses the homophily variable which takes value 1 if all the three identities match between a manager and at least one of the board members; otherwise 0. In Column (1), the dependent variable is logarithm of total managerial compensation (*Log Compensation*) of an executive in that firm-year. In Column (2), the dependent variable is firm value (*Firm Value*). In Column (3), the dependent variable is residual firm value (*Residual Firm Value*).

VARIABLES	(1) Log Compensation	(2) Firm Value	(3) Residual Firm Value
Homophily	0.029***	0.033**	0.027**
	(0.008)	(0.016)	(0.013)
Size	0.095***	0.245***	0.230***
	(0.017)	(0.033)	(0.028)
Firm Age	0.031*	0.049	0.040
	(0.017)	(0.033)	(0.028)
ROA	1.132***		
	(0.099)		
Institutional Ownership	0.015***	0.053***	0.050***
	(0.001)	(0.003)	(0.002)
Promoters Ownership	0.007***	0.047***	0.042***
	(0.001)	(0.002)	(0.002)
Leverage	-0.146**	-0.606***	-0.539***
	(0.068)	(0.125)	(0.106)
Volatility	-5.922***	-11.838***	-11.668***
	(0.961)	(1.818)	(1.542)
Big 5	0.095**	0.321***	0.267***
	(0.038)	(0.072)	(0.061)
Number of Meetings	0.015***	0.050***	0.049***
	(0.004)	(0.007)	(0.006)
Board Size	-0.005	0.009	0.007
	(0.004)	(0.007)	(0.006)
Constant	7.455***	-6.867***	-8.219***
	(0.667)	(1.265)	(1.073)
Observations	28,026	28,026	28,026
R-squared	0.762	0.577	0.062
Managerial Designation	Yes	No	No
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Panel B: Homophily Index

Panel B uses the continuous variable on homophily which is defined as a ratio of number of identity match between a manager and the maximum board members to the total number of board members. In Column (1)-(3), the dependent variable is logarithm of total managerial compensation (*Log Compensation*) in that firm-year. In Column (4)-(6), the dependent variable is residual firm value (*Residual Firm Value*).

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Log Compensation	Log Compensation	Log Compensation	Residual Firm Value	Residual Firm Value	Residual Firm Value
Homophily Index Language	0.151**			0.187*		
	(0.066)			(0.104)		
Homophily Index Place		0.166**			0.262**	
		(0.070)			(0.110)	
Homophily Index Caste			-0.023			-0.200**
			(0.059)			(0.093)
Size	0.095***	0.095***	0.094***	0.226***	0.227***	0.224***
	(0.017)	(0.017)	(0.017)	(0.028)	(0.028)	(0.028)
Firm Age	0.032*	0.031*	0.031*	0.044	0.044	0.044
	(0.017)	(0.017)	(0.017)	(0.027)	(0.027)	(0.027)
Institutional Ownership	0.015***	0.015***	0.015***	0.045***	0.045***	0.045***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Promoter Ownership	0.007***	0.007***	0.007***	0.037***	0.037***	0.037***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
ROA	1.130***	1.129***	1.130***	3.949***	3.947***	3.949***
	(0.099)	(0.099)	(0.099)	(0.157)	(0.157)	(0.157)
Leverage	-0.146**	-0.145**	-0.148**	0.155	0.157	0.148
	(0.068)	(0.068)	(0.068)	(0.108)	(0.108)	(0.108)
Volatility	-5.903***	-5.869***	-5.893***	-9.133***	-9.083***	-9.128***
	(0.961)	(0.961)	(0.961)	(1.526)	(1.526)	(1.526)
Big5	0.096**	0.095**	0.096**	0.263***	0.261***	0.263***
	(0.038)	(0.038)	(0.038)	(0.060)	(0.060)	(0.060)
Number of Meetings	0.015***	0.015***	0.015***	0.043***	0.043***	0.043***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)
Board Size	-0.005	-0.005	-0.005	0.011*	0.011*	0.012*
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)
Constant	7.421***	7.432***	7.500***	-8.366***	-8.373***	-8.169***
	(0.668)	(0.668)	(0.668)	(1.061)	(1.061)	(1.062)
Observations	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.762	0.762	0.762	0.084	0.084	0.084
Manager Designation	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Table IA9: Return on Equity (ROE) and Group Identity

This table provides regression results of return on equity and residual return on equity on group identity measures. In Columns (1) - (3), the dependent variable is return on equity (*ROE*) of an executive in that firm-year. In Columns (4)-(6), the dependent variable is residual return on equity (*Residual ROE*). Please refer to Appendix A for variable definitions. The standard errors are in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ROE	ROE	ROE	Residual ROE	Residual ROE	Residual ROE
Homophily Language	0.027**			0.026**		
	(0.013)			(0.012)		
Homophily Place		0.044***			0.036***	
		(0.015)			(0.013)	
Homophily Caste			-0.003			-0.001
			(0.022)			(0.019)
Size	0.011	0.012	0.010	0.009	0.009	0.008
	(0.013)	(0.013)	(0.013)	(0.011)	(0.011)	(0.011)
Firm Age	0.012	0.012	0.011	0.011	0.011	0.011
	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)
ROA	1.267***	1.266***	1.269***	1.180***	1.179***	1.182***
	(0.067)	(0.067)	(0.067)	(0.058)	(0.058)	(0.058)
Institutional Ownership	0.001	0.001	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Promoters Ownership	-0.000	-0.000	-0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Leverage	0.309***	0.311***	0.312***	0.219***	0.220***	0.221***
	(0.055)	(0.055)	(0.055)	(0.048)	(0.048)	(0.048)
Volatility	0.162	0.192	0.159	0.045	0.070	0.045
	(1.047)	(1.046)	(1.048)	(0.911)	(0.910)	(0.912)
Big 5	0.078***	0.075***	0.079***	0.065***	0.062***	0.065***
	(0.020)	(0.020)	(0.020)	(0.018)	(0.018)	(0.018)
Number of Meetings	-0.002	-0.002	-0.002	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Board Size	-0.001	-0.001	-0.001	-0.002	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age	0.001	0.001	0.001	0.001	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Gender	0.036	0.035	0.037	0.032	0.032	0.033
	(0.042)	(0.042)	(0.042)	(0.036)	(0.036)	(0.036)
Education	-0.008	-0.008	-0.008	-0.006	-0.006	-0.006
	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Constant	-0.823*	-0.842**	-0.774*	-0.850**	-0.860**	-0.809**
	(0.430)	(0.429)	(0.434)	(0.374)	(0.374)	(0.377)
Observations	3,541	3,541	3,541	3,536	3,536	3,536
R-squared	0.421	0.422	0.421	0.215	0.216	0.214
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Table IA10: Excluding Chairperson

This table provides regression results of total managerial compensation, firm value, and residual firm value on group identity measures after excluding the observations with chairperson designation. In Columns (1) - (3), results are for homophily language (*Homophily Language*). In Columns (4)-(6), the results are for homophily place(*Homophily Place*). In Columns (7) - (9), results are for homophily caste (*Homophily Caste*). Please refer to Appendix A for variable definitions. The standard errors are in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

	Homophily Language			Homophily Place			Homophily Caste		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Log Compensation	Firm Value	Residual Firm Value	Log Compensation	Firm Value	Residual Firm Value	Log Compensation	Firm Value	Residual Firm Value
Homophily Language	0.134***	0.134**	0.108**						
	(0.028)	(0.054)	(0.046)						
Homophily Place				0.109***	0.132**	0.119**			
				(0.029)	(0.056)	(0.048)			
Homophily Caste							-0.177***	-0.224**	-0.174*
							(0.056)	(0.107)	(0.092)
Size	0.091***	0.224***	0.225***	0.092***	0.226***	0.226***	0.089***	0.223***	0.223***
	(0.023)	(0.043)	(0.037)	(0.023)	(0.043)	(0.037)	(0.023)	(0.043)	(0.037)
Firm Age	0.126***	0.013	0.006	0.125***	0.013	0.006	0.124***	0.012	0.005
	(0.021)	(0.041)	(0.035)	(0.021)	(0.041)	(0.035)	(0.021)	(0.041)	(0.035)
ROA	1.210***			1.193***			1.186***		
	(0.132)			(0.132)			(0.132)		
Institutional Ownership	0.013***	0.046***	0.043***	0.013***	0.046***	0.043***	0.013***	0.046***	0.043***
	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)
Promoters Ownership	0.008***	0.037***	0.034***	0.008***	0.037***	0.033***	0.008***	0.037***	0.034***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Leverage	-0.299***	-0.620***	-0.620***	-0.295***	-0.612***	-0.613***	-0.299***	-0.614***	-0.616***
	(0.089)	(0.164)	(0.141)	(0.089)	(0.164)	(0.141)	(0.089)	(0.164)	(0.141)
Volatility	-5.699***	-9.909***	-10.145***	-5.701***	-9.908***	-10.149***	-5.621***	-9.801***	-10.060***
	(1.245)	(2.375)	(2.032)	(1.245)	(2.375)	(2.031)	(1.245)	(2.375)	(2.032)
Big5	0.137***	0.362***	0.291***	0.133***	0.358***	0.288***	0.134***	0.359***	0.289***
	(0.050)	(0.096)	(0.082)	(0.050)	(0.096)	(0.082)	(0.050)	(0.096)	(0.082)
Number of Meetings	0.012***	0.055***	0.053***	0.012**	0.055***	0.053***	0.012***	0.055***	0.053***
	(0.005)	(0.009)	(0.008)	(0.005)	(0.009)	(0.008)	(0.005)	(0.009)	(0.008)
Board Size	-0.001	0.013	0.009	-0.001	0.013	0.009	-0.001	0.014	0.009
	(0.005)	(0.010)	(0.008)	(0.005)	(0.010)	(0.008)	(0.005)	(0.010)	(0.008)
Constant	3.507***	-4.956***	-6.585***	3.520***	-4.969***	-6.611***	3.796***	-4.622***	-6.321***
	(0.838)	(1.602)	(1.370)	(0.839)	(1.602)	(1.370)	(0.839)	(1.603)	(1.371)
Observations	15,741	15,741	15,741	15,741	15,741	15,741	15,741	15,741	15,741
R-squared	0.833	0.598	0.099	0.833	0.598	0.099	0.833	0.598	0.099
Managerial Designation	Yes	No	No	Yes	No	No	Yes	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table IA11: Profitability and Homophily

This table provides regression results of firm performance on homophily language and homophily place. Firm performance is measured as return on assets and future return on assets. In Columns (1) - (2), results are for return on assets (*ROA*). In Columns (3)-(4), the results are for future return on assets(*Future ROA*). Please refer to Appendix A for variable definitions. The standard errors are in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

VARIABLES	(1) ROA	(2) ROA	(3) Future ROA	(4) Future ROA
Homophily Language	0.006**		0.006**	
	(0.003)		(0.003)	
Homophily Place		0.008**		0.008**
		(0.004)		(0.004)
Size	0.002	0.002	0.002	0.002
	(0.003)	(0.003)	(0.003)	(0.003)
Firm Age	0.000	0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Institutional Ownership	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Promoters Ownership	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Leverage	-0.078***	-0.078***	-0.078***	-0.078***
	(0.014)	(0.014)	(0.014)	(0.014)
Volatility	-1.678***	-1.666***	-1.678***	-1.666***
	(0.267)	(0.267)	(0.267)	(0.267)
Big 5	-0.001	-0.002	-0.001	-0.002
	(0.005)	(0.005)	(0.005)	(0.005)
Number of Meetings	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Board Size	-0.001*	-0.001*	-0.001*	-0.001*
	(0.001)	(0.001)	(0.001)	(0.001)
Age	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Gender	0.017	0.017	0.017	0.017
	(0.011)	(0.011)	(0.011)	(0.011)
Education	0.002	0.001	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.016	0.015	0.016	0.015
	(0.110)	(0.109)	(0.110)	(0.109)
Observations	3,388	3,388	3,388	3,388
R-squared	0.718	0.718	0.718	0.718
Firm	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes

Table IA12 : Endogeneity Test

This table reports Durbin score and Wu-Hausman statistics conducted post two-stage least square regression (2SLS). The p-value lower than critical value suggests to reject the null-hypothesis i.e. log compensation and firm value are endogenous. Please refer to Appendix A for variable definitions.

Homophily Language		
	Statistic	p-value
Durbin (score)	23.167	0
Wu-Hausman	23.163	0

Homophily Place		
	Statistic	p-value
Durbin (score)	22.429	0
Wu-Hausman	22.424	0

Homophily Caste		
	Statistic	p-value
Durbin (score)	23.191	0
Wu-Hausman	23.187	0

Table IA13: Alternative Measure of Firm Size

This table reports the regression results of managerial compensation, firm value, and residual firm value on group identity measures using alternate firm size construct. Log Compensation (*Log Compensation*) is logarithm of total managerial compensation which is in million USD. *Firm value* is ratio of market price of equity to the book value of equity. *Residual Firm Value* is residual of regression of firm value on manager's total compensation of last year. *Homophily Language* is a dummy variable whose value equal to 1 if the board (at least one director) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same Indian state otherwise 0.*Homophily Caste* is dummy variable whose value equal to 1 if the board (at least one director) and the manager belong to same caste otherwise 0. Panel A reports the regression results of managerial compensation on group identity measures using alternate firm size construct. Panel B This panel reports the regression results of firm value on group identity measures using alternate firm size construct. This panel reports the regression results of residual firm value on group identity measures using alternate firm size construct. All other variables are defined in variable definition sheet in Appendix. The standard errors are in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Panel A: Executive Compensation, Group Identity, and Alternative Measure of Firm Size

This panel reports the regression results of managerial compensation on group identity measures using alternate firm size construct.

VARIABLES	(1) Log Total Compensation	(2) Log Total Compensation	(3) Log Total Compensation
Homophily Language	0.075***		
	(0.021)		
Homophily Place		0.078***	
		(0.022)	
Homophily Caste			-0.144***
			(0.040)
Size	0.331***	0.331***	0.331***
	(0.019)	(0.019)	(0.019)
Firm Age	0.029*	0.029*	0.030*
	(0.017)	(0.017)	(0.017)
Institutional Ownership	0.012***	0.011***	0.011***
	(0.001)	(0.001)	(0.001)
Promoters Ownership	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)
ROA	0.995***	0.991***	0.988***
	(0.099)	(0.099)	(0.099)
Leverage	-0.228***	-0.226***	-0.229***
	(0.068)	(0.068)	(0.068)
Volatility	-4.815***	-4.805***	-4.774***
	(0.958)	(0.958)	(0.958)
Big5	0.053	0.052	0.053
	(0.038)	(0.038)	(0.038)
Number of Meetings	0.013***	0.013***	0.013***
	(0.004)	(0.004)	(0.004)
Board Size	-0.011***	-0.011***	-0.011***
	(0.004)	(0.004)	(0.004)
Constant	6.429***	6.427***	6.573***
	(0.608)	(0.608)	(0.608)
Observations	27,928	27,928	27,928
R-squared	0.765	0.765	0.765
Manager Designation	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Panel B: Firm Value, Group Identity, and Alternative Measure of Firm Size

This panel reports the regression results of firm value on group identity measures using alternate firm size construct.

VARIABLES	(1) Firm Value	(2) Firm Value	(3) Firm Value
Homophily Language	0.116***		
	(0.040)		
Homophily Place		0.051	
		(0.041)	
Homophily Caste			-0.136*
			(0.077)
Size	-0.102***	-0.102***	-0.110***
	(0.037)	(0.037)	(0.037)
Firm Age	0.042	0.042	0.040
	(0.033)	(0.033)	(0.033)
Institutional Ownership	0.056***	0.056***	0.056***
	(0.003)	(0.003)	(0.003)
Promoters Ownership	0.047***	0.047***	0.047***
	(0.002)	(0.002)	(0.002)
Leverage	-0.591***	-0.589***	-0.585***
	(0.125)	(0.125)	(0.125)
Volatility	-11.736***	-11.697***	-11.676***
	(1.824)	(1.824)	(1.823)
Big5	0.351***	0.350***	0.352***
	(0.072)	(0.072)	(0.072)
Number of Meetings	0.053***	0.053***	0.052***
	(0.007)	(0.007)	(0.007)
Board Size	0.016**	0.016**	0.020***
	(0.007)	(0.007)	(0.008)
Constant	-1.710	-1.670	-1.424
	(1.158)	(1.158)	(1.160)
Observations	27,928	27,928	27,928
R-squared	0.577	0.577	0.578
Manager Designation	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Panel C: Residual Firm Value, Group Identity, and Alternative Measure of Firm Size

This panel reports the regression results of residual firm value on group identity measures using alternate firm size construct.

VARIABLES	(1) Residual Firm Value	(2) Residual Firm Value	(3) Residual Firm Value
Homophily Language	0.087**		
	(0.034)		
Homophily Place		0.049	
		(0.035)	
Homophily Caste			-0.111*
			(0.065)
Size	-0.100***	-0.100***	-0.100***
	(0.031)	(0.031)	(0.031)
Firm Age	0.034	0.034	0.034
	(0.028)	(0.028)	(0.028)
Institutional Ownership	0.053***	0.053***	0.053***
	(0.002)	(0.002)	(0.002)
Promoters Ownership	0.042***	0.041***	0.041***
	(0.002)	(0.002)	(0.002)
Leverage	-0.523***	-0.521***	-0.523***
	(0.106)	(0.106)	(0.106)
Volatility	-11.603***	-11.576***	-11.553***
	(1.546)	(1.546)	(1.546)
Big5	0.296***	0.295***	0.296***
	(0.061)	(0.061)	(0.061)
Number of Meetings	0.052***	0.052***	0.052***
	(0.006)	(0.006)	(0.006)
Board Size	0.014**	0.014**	0.014**
	(0.006)	(0.006)	(0.006)
Constant	-3.332***	-3.309***	-3.203***
	(0.982)	(0.982)	(0.983)
Observations	27,928	27,928	27,928
R-squared	0.061	0.060	0.060
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Table IA14: Geographical Variation in Group Identity, Managerial Compensation, and Firm Value

This table shows the regression results of impact of geographical variation in group identity on managerial compensation and firm value. *Total Compensation* is log of total compensation. *Firm Value* is the ratio of market value to the book value of an equity. *Residual Firm Value* is the firm value condition on managerial compensation of last year. *Homophily Language* is a dummy variable whose value equals to 1 if a manager and at least one board member speaks same Indian language, otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if a manager and at least one board member belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if a manager and at least one board member belong to same caste, otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. Panel A shows the regression results excluding the surnames for those communities which migrated from one part of India to another after independence. Panel B shows the regression results of homophily difference in southern Indian states compared to northern Indian states on executive compensation, firm value, and residual firm value. For definition of other variables, refer Appendix A. To control for unobservables, firm, year, and managerial designation fixed effects are used. All regressions include a constant. Standard errors are in parentheses. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Panel A: Community Migration and Group Identity

This panel shows the regression results excluding the surnames for those communities which migrated from part of India to another after independence. Columns (1)-(3) shows the results for total compensation. Columns (4)-(6) shows the results for firm value. Columns (7)-(9) shows the results for residual firm value.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Total Compensation	Total Compensation	Total Compensation	Firm Value	Firm Value	Firm Value	Residual Firm Value	Residual Firm Value	Residual Firm Value
Homophily Language	0.079***			0.112***			0.090***		
	(0.021)			(0.040)			(0.034)		
Homophily Place		0.077***			0.058			0.060*	
		(0.022)			(0.041)			(0.035)	
Homophily Caste			-0.127***			-0.107			-0.081
			(0.041)			(0.078)			(0.066)
Constant	7.431***	7.415***	7.589***	-7.070***	-7.043***	-6.903***	-8.339***	-8.330***	-8.207***
	(0.668)	(0.668)	(0.669)	(1.263)	(1.264)	(1.265)	(1.071)	(1.071)	(1.072)
Observations	27,718	27,718	27,718	27,718	27,718	27,718	27,718	27,718	27,718
R-squared	0.762	0.762	0.762	0.575	0.575	0.575	0.064	0.064	0.063
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Geographical Differences and Group Identity

This panel shows the regression results of homophily difference in southern Indian states compared to northern Indian states on managerial compensation, firm value, and residual firm value. Columns (1)-(3) shows the results for total compensation. Columns (4)-(6) shows the results for firm value. Columns (7)-(9) shows the results for residual firm value.

VARIABLES	(1) Total Compensation	(2) Total Compensation	(3) Total Compensation	(4) Firm Value	(5) Firm Value	(6) Firm Value	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual Firm Value
Homophily Place	0.037*				-0.019			0.027	
	(0.020)				(0.039)			(0.024)	
South	-0.116***	-0.122***	-0.159*	0.226***	0.245***	0.169	-0.048	-0.034	0.057
	(0.028)	(0.028)	(0.085)	(0.053)	(0.054)	(0.164)	(0.033)	(0.033)	(0.102)
Homophily Place x South	0.131***				-0.105			0.032	
	(0.044)				(0.085)			(0.053)	
Homophily Language		0.020		0.014			0.034		
		(0.020)		(0.038)			(0.024)		
Homophily Language x South		0.156***		-0.056			0.073		
		(0.045)		(0.085)			(0.053)		
Homophily Caste			-0.045			-0.051			0.000
			(0.030)			(0.058)			(0.036)
Homophily Caste x South			0.103			0.041			-0.084
			(0.088)			(0.169)			(0.105)
Constant	7.730***	7.737***	7.799***	-1.248***	-1.221***	-1.201***	0.113	0.124	0.146
	(0.128)	(0.129)	(0.130)	(0.245)	(0.245)	(0.248)	(0.152)	(0.152)	(0.154)
Observations	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.542	0.542	0.542	0.165	0.165	0.165	0.007	0.007	0.007
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes