Traditional Institutions Meet the Modern World: Caste, Gender and Schooling Choice in a Globalizing Economy *

Kaivan Munshi† Mark Rosenzweig‡

September 2004

Abstract

This paper addresses the question of how traditional institutions interact with the forces of globalization to shape the economic mobility and welfare of particular groups of individuals in the new economy. We explore the role of one such traditional institution - the caste system - in shaping career choices by gender in Bombay using new survey data on school enrollment and income over the past 20 years. Bombay's labor market was historically organized along rigid caste lines; such restrictions on mobility can be welfare enhancing when network externalities are present. But there was a large change in the returns to different occupations in the 1990s. We find that male working class - lower caste - networks continue to channel boys into local language schools that lead to the traditional occupation, despite the fact that returns to non-traditional white collar occupations rose substantially, suggesting the possibility of a dynamic inefficiency. In contrast, lower caste girls, who historically had low labor market participation rates and so did not benefit from the network, are taking full advantage of the opportunities that became available in the new economy by switching rapidly to English schools. Thus, caste continues to play a particular (gender-specific) role in shaping schooling choices in the new economy of the 1990s. But the overall increase in English schooling in recent years, and the growing mismatch in education choices and hence occupational outcomes between boys and girls in the same caste, suggest that the remarkably resilient caste system might finally be starting to disintegrate.

*Leena Abraham and Padma Velaskar of the Tata Institute of Social Sciences collaborated in the design of the survey instrument, carried out the survey, and provided many valuable insights. We are also very grateful to Suma Chitnis for her support and advice at every stage of the project. We received helpful comments from two referees, David Card, Abhijit Banerjee, Jan Eeckhout, Lakshmi Iyer, Duncan Thomas, and seminar participants at Columbia, Harvard-MIT, IZA, Princeton, and the World Bank. Research support from the Mellon Foundation at the University of Pennsylvania and the National Science Foundation is gratefully acknowledged. We are responsible for any errors that may remain.

†Brown University and NBER

‡Harvard University

1 Introduction

The collapse of the former Soviet Union, followed by the economic and financial liberalization of the 1990s, has restructured and "globalized" many economies throughout the world. One consequence of this restructuring, which has been widely observed, is that some groups have taken advantage of the new benefits afforded by globalization, while others appear to have been left behind. This paper addresses the question of whether and how old institutions clash with the forces of globalization in shaping the response of particular groups of individuals to the new economy. Informal institutions, such as community networks, are generally believed to play an important role in low-income countries by facilitating economic activity when markets function imperfectly (Townsend 1994, Fafchamps and Lund 2000). Less well understood is how such institutions affect the transformation of economies undergoing change, affecting in turn the distribution of benefits from macroeconomic structural reform. Although traditional institutions may have served a useful purpose when they were first put in place, it is possible that they may constrain the response to new opportunities when the structure of the economy changes unexpectedly.[1]

One example of such a traditional institution is the caste system in India. We study the role of the caste system in shaping career choices by gender in a dynamic urban context, using new data on schooling choices and income covering the past 20 years in Bombay city, the industrial and financial center of the Indian economy. Bombay is a useful and important setting in which to study the role of institutional rigidities in a dynamic context as the Bombay labor market was historically organized along caste lines, with individual sub-castes or *jatis* controlling particular occupational niches over the course of many generations. Chandavarkar (1994: 122,223), for instance, describes how "[caste] clusters formed within particular trades and occupations ... [this] occupational distribution reflected neither [traditional rural] caste vocation nor the inheritance of special skills. It was produced partly by exclusionary practices by which social groups, once they obtained a foothold in a particular occupation, would not admit an outsider."[2] A particularly important feature of these caste networks is that they

[1] While these traditional institutions might have enhanced welfare in a second-best world, they were very likely associated with static inefficiencies of their own prior to the restructuring. For example, Fafchamps (2002) and Banerjee and Munshi (2003) show how network segmentation along community lines could impose significant efficiency costs with respect to investment by traders in Africa and small manufacturing units in South India respectively.

[2] In prior work on caste in the economics literature, caste is treated as an individual characteristic like ethnicity or race that is one of many determinants of human capital investment or income in India (e.g., Behrman 1988). In contrast, we study how the caste system operates as a community network, providing employment assistance to its members, which could in turn result in the persistence of occupational distributions within castes over many generations.

were most active in working class occupations dominated by *lower caste men*. Women historically did not participate in Bombay's labor market and so did not benefit from the caste networks, but both men and women scrupulously adhered to the social rule of endogamous marriage within the *jati*.

Although Bombay was a predominantly industrial city for a hundred years, starting from the last quarter of the nineteenth century, the liberalization of the Indian economy in the 1990s saw a shift in the city's economy towards the corporate and financial sectors. We study how members of different *jatis*, by gender, responded to these changes in the returns to different occupations, and we will show that the historical pattern of networking within the *jati* continues to shape gender-specific, individual responses to these new opportunities in ways that will importantly affect the future distributions of incomes, independent of pre-school human capital effects or liquidity constraints.[3]

Our strategy in this paper is to assess how schooling choice, measured by the language of instruction, varied across *jatis*, across boys and girls within *jatis*, and over time. We focus on schooling choice because most adults were already locked in to their occupations when the unexpected economic changes occurred. Schooling choice is an important determinant of future occupational outcomes in the Bombay economy and thus reflects the contemporaneous perceptions of expected occupational returns. University education in Bombay is entirely in English, but children choose between English and Marathi (the local language) as the language of instruction in primary and secondary school at the time they enter primary school. Schooling in Marathi channels the child into working class jobs, while more expensive English education significantly increases the likelihood of obtaining a coveted white-collar job. If the economic liberalization of the 1990s effectively increased white-collar incomes, and by extension the returns to English education, then (future) occupational mobility can be identified from changes in the choice of the language of instruction made by parents of school-age children.[4] Examination of the changing patterns of schooling choice by *jati* and gender thus permits an assessment of the interactions between traditional institutions and the new realities of globalization.

Our empirical analysis is based on a survey of 4,900 households belonging to the Maharashtrian

[3]A recent literature has shown that historical institutions have long run consequences for growth in low-income countries (Acemoglu, Johnson and Robinson 2001; Banerjee and Iyer, 2002). These empirical findings, however, do not provide insight into the mechanisms underlying such persistence. In our analysis, we show how caste-based job networks can preserve existing occupational distributions even in the face of the increasing returns to individual occupational mobility that accompany globalization.

[4]Prior work on schooling enrollment in India has ignored this important dimension of schooling, which affects importantly schooling attainment and economic mobility (Jacoby and Skoufias 1997; Kochar, 2002). These studies examined rural households. In 1990, approximately 28% of rural village schools provided instruction in a local language (other than English or Hindi).

community and residing in Bombay's Dadar area and a survey of the schools in the locale that we conducted in 2001-2002. Secondary schools in Bombay run from grade one to grade 10. The household survey was based on a stratified random sample of students who entered 28 of the 29 schools in Dadar (in the first grade), over a twenty year period, 1982-2001.[5] English is the language of instruction in 10 schools in Dadar, while Marathi is the language of instruction in the remaining 18 schools.

The parents of the randomly selected students in the sample were interviewed at their homes. The survey data suggest that the returns to English education, for given years of schooling, increased in the 1990s. Based on retrospective information on the annual earnings of the parents of the sampled children, we estimated the returns to English and the returns to years of schooling at five points in time from 1980 through 2000 for working adults between the age of 30 and 55.[6] Figures 1 and 2 provide the estimated returns to schooling attainment and schooling language, for men and women, respectively, in each time period. As can be seen, the returns to years of schooling increased only mildly over time for both men and women. In contrast, the English premium increased sharply from the 1980s to the 1990s for both sexes, rising from 15% in 1980 to 24% in 2000 for men and from approximately 0% in 1980 to 27% in 2000 for women. The returns to English for men increase from the mid-1980s, which is most likely due to the decline around that time in manufacturing jobs in Bombay that we will discuss in the next section, but continue to rise through the 1990s.

The survey collected information on schooling choice for 20 cohorts of students who entered the 28 neighborhood schools (in the first grade) over the 1982-2001 period. The times-series data on enrollments in English- and Marathi-medium schools suggest that the changes in the returns to English significantly affected schooling choice for both boys and girls in the sample, across castes and over time. Figure 3 and Figure 4 display the changing proportions of students enrolled in English schools for the 20 entering cohorts from 1982 (cohort=1) to 2001 (cohort=20) for three caste groupings - low, medium and high - and by gender.[7] The figures were constructed using the Epanechnikov kernel function to

[5]One school refused to provide us with information on its students and will be ignored in all the discussion that follows. Details of the survey and the sampling design are provided in Section 4.1.

[6]The details of the estimation procedure are provided in the Appendix. The estimates of the returns to English and the returns to schooling (with standard errors) are provided in Table A1.

[7]Children enter first grade at the age of six and complete tenth grade at the age of 15, so the current age of the students in our sample, with only a few exceptions, ranges from six to 25. Students in Bombay typically do not change the language of instruction midstream or switch schools after they enter first grade. High castes include all the Brahmin *jatis*, as well as a few other elite *jatis* (CKP and Pathare Prabhus). Low castes include formerly untouchable and backward castes (Scheduled Castes, Scheduled Tribes, and Other Backward Castes, as defined by the government of India). Medium castes are drawn mostly from the cultivator *jatis*, such as the Marathas and the Kunbis, as well as other traditional vocations that were not considered to be ritually impure.

nonparametrically regress schooling choice (1=English medium; 0=Marathi medium) on the cohort variable for each caste group, taking into account the strong intergenerational state-dependence with respect to the language of instruction within the family.[8] Although *jatis* define the relevant boundary for the labor-market networks and form the relevant social unit in our analysis,[9] we aggregate the 59 sub-castes in our data for expositional convenience in these figures.

Figures 3 and 4 show that enrollment rates in English-medium schools have grown substantially over time for both boys and girls and for all castes. However, the trajectory is much steeper for the 10 most recent cohorts, who would have entered school in the post-reform 1990s. Thus, the increase in the returns to English observed in Figure 1 and Figure 2 appears to have shifted schooling choice towards English education. The figures also indicate substantial differences in English schooling between castes at the beginning of the sample period, reflecting in part the circumstances of the colonial regime. The high castes gained access to clerical and administrative jobs under the British, while the lower castes were confined for the most part to working class jobs. Consistent with the view that Marathi education channels students into working class jobs, and that English education increases the likelihood of obtaining a white-collar job, we see in Figure 3 and Figure 4 that high-caste boys and girls currently 25 years old (the oldest cohort) were much more likely to have been schooled in English, and that this caste difference in schooling persists over the next 10 cohorts. But although the caste-gap narrows dramatically for the girls in the 1990s, there is no convergence for the boys. Thus it appears that caste continues to play a role in shaping schooling choices in the new economy of the 1990s, but only for boys. The key question is why the lower-caste boys seemingly fail to take advantage of the new economic opportunities.

There are many potential explanations for the absence of convergence in Figure 3. For example, the distribution of pre-school human capital could vary across castes, in which case it might be optimal for the lower caste boys to continue to select into the working class jobs, and hence Marathi schooling, even as the returns to English grow. Alternatively, liquidity constraints could prevent the poorer lower

[8] There is a very high degree of state dependence in schooling choice across generations within the family; if both parents have been schooled in English, it is very unlikely that the child will be sent to a Marathi school. All the parametric regressions that we later report will also account for such state dependence at the level of the family. Nonparametric estimates of schooling choice over time that do not take account of parents' language of schooling are similar to Figure 3 and Figure 4, except that the convergence for the girls is less sharp. Details of the nonparametric estimation procedure are provided in the Appendix and parametric estimates of the schooling regression (with standard errors) are provided in Table A2.

[9] As Morris (1965: 76) emphasizes in his historical account of the Bombay labor market, "for any analysis of labor recruitment [in Bombay] ... it is entirely inappropriate to lump into larger groups because of similarity of name, function, social status, or region-of-origin subcastes that are not endogamous."

caste boys from switching to more expensive English schooling. These explanations do not, however, explain the differences for boys and girls that we observed above. The gender-specific explanation for the observed pattern in Figure 3 that we pursue in this paper is based on network externalities. It is commonly observed that working class occupations, which tend to be dominated by men, are associated with higher levels of networking, possibly because the information and enforcement problems that give rise to networks are more acute in the working class jobs. For example, Rees (1966) found that informal sources accounted for 80% of all hires in eight blue-collar occupations versus 50% of all hires in four white-collar occupations in an early study set in Chicago. Similarly, 68% of blue-collar workers and 38% of white-collar workers reported having received help finding a job in Gore's (1970) study of migrants in Bombay. These studies focus on men, the primary occupants of blue-collar jobs. And among the household heads in Dadar, 68% of the men in working class jobs found employment through a relative or a member of the community, while the corresponding statistic for white-collar workers was 44%. Once the (working class) network is in place, there is a positive externality associated with participation in the network, and hence with the traditional occupational choice in the *jati*. This externality could give rise to intergenerational occupational persistence *at the level of the* jati, with labor market networks channelling boys into particular (traditionally male) occupations and hence towards particular schooling choices.

Once the returns to the white-collar occupation grow, however, schooling choice must ultimately converge across castes. The explanation for the absence of convergence in Figure 3 that we put forward in this paper is based on the idea that the caste networks might place tacit restrictions on the occupational mobility of their members to preserve the integrity of the network.[10] We will show that although these restrictions might have been welfare-enhancing and indeed equalizing when they were first put in place, such restrictions could result in dynamic inefficiencies when the structure of the economy changes.

The paper is organized in six sections. The next section provides a brief description of the institutional setting. In Section 3 we briefly outline a heuristic model of schooling choice for boys when caste-based networks are active, which incorporates within-caste heterogeneity in ability. The model is used to provide an explanation of intergenerational persistence in occupational choice and hence

[10]Given the historical connection between caste and occupation in the Bombay labor market, caste and class are closely related in Bombay. The restrictions on occupational mobility among the lower castes may then take the form of a conformist working class culture, observed in many economies, that tends to be resistant to change (Kohn, 1977, is the classic reference in the large sociological literature on class, values, and conformity).

schooling choice within castes, and why networks might attempt to constrain mobility in the face of economic change. The framework also provides predictions for which castes restrict mobility and which individuals within castes will first exit the traditional occupations as the economy changes, which has implications for the selectivity of school choice. Section 4 describes the data and presents preliminary evidence that occupational patterns persist across generations by caste, for men but not women, consistent with the presence of male caste-based labor market networks. The basic assumptions of the model, relating the type of occupation (working class versus white-collar) to the level of referrals and the type of schooling (English versus Marathi), are also verified in this section, using information collected from the parents of the sampled students. Section 5 tests the implications of the framework, assessing the role of caste networks in determining school choice differentially for boys and girls, and the selectivity in terms of pre-school human capital of those who exit the traditional occupations. Intergenerational state dependence at the level of the *jati* in school choice, and hence career choice, is found to remain strong through the 1990s, providing us with a simple explanation for the absence of convergence across castes for the boys that we observed in Figure 3. Alternative explanations for the absence of convergence are considered in Section 6, including liquidity constraints, caste differences in pre-school human capital, preferences for the parental occupation, and discrimination in the labor market. Section 7 concludes.

The results in this paper provide empirical support for the view that historical occupation patterns kept in place by caste-based networks continue to shape occupational choice, and hence schooling choice, for the boys in the new economy. Although our results indicate that differences in family income and in pre-school human capital help explain differences in schooling choice across castes, these factors cannot by themselves explain why groups that have traditionally been occupied in working class jobs remain in those jobs despite rising individual returns to non-traditional occupations. In contrast, the lower-caste girls who historically kept away from the labor market, and so have no network ties to constrain them, take full advantage of the opportunities that become available in the new economy. The growing disparities in school choices between boys and girls within the traditional *jatis* not only suggest a new balance of economic opportunities by gender, but also could threaten the long-run stability of the caste system, which is based on endogamous marriages within the sub-caste. A complete understanding of the development process must thus not only take account of the initial conditions and the role of pre-existing institutions in shaping the response to modernization and globalization, but must also consider how these traditional institutions are shaped in turn by the

forces of change.

2 The Institutional Setting

2.1 A Brief History of Bombay City

Bombay's economic history begins in the last quarter of the eighteenth century when it replaced Surat as the largest trading port and the center of commerce on the West Coast of India (Chandavarkar 1994, Katzenstein 1979). Bombay's first cotton-textile mill was established in 1851, and from the late nineteenth century onwards this industry dominated the city's economy. The 1921 census indicates that this one industry employed as much as 16.2% of the male population of the city (Chandavarkar 1994). But other industries started to expand from the 1930s onwards, and by the 1961 census, manufacturing accounted for 41% of Bombay's employment (D'Monte 2002).

Although manufacturing continues to play an important role in the city's economy, the past two decades have seen a steady decline in the importance of this sector. The textile industry, which had dominated the economy from the last quarter of the nineteenth century up till the middle of the twentieth century, began to decline after a crippling strike shut down the mills for eighteen months in 1982-83. More industries began to leave Bombay in the 1990s, as real estate prices rose and the industrial wage gap between the city and the rest of the country widened (D'Monte 2002). But apart from its manufacturing base, Bombay has also historically dominated India's commercial and service sector. The city is the center of trade, banking, insurance and financial services, publishing, and advertising in India. The decline in manufacturing and the removal of controls on industrial licensing and foreign investment in 1991 gave a further boost to this sector, sharply increasing wages in corporate jobs, and commercial occupations in general. This exogenous shock to the local economy serves as the point of departure for our analysis.

Although Bombay is located in the modern Indian state of Maharashtra, the city was largely established by outsiders. It was only with industrialization in the mid-nineteenth century that the mills and manufacturing companies started to recruit Maharashtrian labor from the Konkan coast and the interior Deccan region on a massive scale (Morris 1965). The Pathare Prabhus, who had lived in Bombay from early times, and later several Maharashtrian Brahmin castes, found administrative and clerical jobs under the British. With the formation of the linguistic state of Maharashtra with Bombay as its capital in 1960, the Maharashtrian community also quickly gained control of local political power

and state government jobs (Katzenstein 1979). But the Maharashtrian community was historically conspicuously absent from trade and commerce, the major source of recent economic opportunities. The question is how did different groups within this community respond to these new opportunities, as observed in the schooling choices of their children?

2.2 Bombay's Labor Market

Bombay's industrial economy in the late nineteenth century and through the first half of the twentieth century was characterized by wide fluctuations in the demand for labor (Chandavarkar 1994). It is well known that such frequent job turnover can give rise to labor market networks, particularly when the quality of a freshly hired worker is difficult to assess and performance-contingent wage contracts cannot be implemented. One scheme to organize these networks (Munshi 2003) uses incumbent workers to search for fresh hires; incumbent workers will typically have built up a reputation within the firm, and so they have the incentive to bring back high quality recruits to maintain that reputation (and the economic rents that go with it). The presence of such recruitment networks has indeed been documented by numerous historians studying Bombay's economy prior to independence in 1947 (Chandavarkar 1994, Morris 1965, Burnett-Hurst 1925). These networks appear to have been organized around the jobber, a foreman who was in charge of a work gang in the mill, factory, dockyard, or construction site, and more importantly also in charge of labor recruitment.[11] The jobber is almost ubiquitous in the Indian industrial system, but the fluctuations in labor demand gave his role particular prominence in Bombay.

Given the information and enforcement problems that are associated with the recruitment of short-term labor, it is not surprising that the "jobber had to lean on social connections outside his workplace such as his kinship and neighborhood connections" (Chandavarkar 1994: 107). Here the endogamous subcaste or *jati* served as a natural social unit from which to recruit labor, because marriage ties strengthen information flows and improve enforcement. This widespread use of caste-based networks led to a fragmentation of the Bombay labor market along social lines. The presence of caste clusters in the textile mills, for example, has been well documented. Gokhale's (1957) survey of textile workers in the 1950s showed that Maratha and Kunbi (both middle caste) men were evenly distributed throughout the mill, while Harijan (low caste) men were employed for the most part in the

[11] The jobber was known as the *mukadam* in the mills and the factories, *maistry* in the construction industry, and the *serang* in the shipping industry.

spinning section. Consistent with the presence of local (jobber-specific) networks, the caste clusters that were observed in particular mills often differed from the general pattern for the industry as a whole. The same sort of caste-based clustering has been documented among Bombay's dock workers (Cholia 1941), construction workers, and in the railway workshops (Burnett-Hurst 1925), in the leather and dyeing industries, and in the Bombay Municipal Corporation and the Bombay Electric Supply and Transportation Company (Chandavarkar 1994).

Most historical accounts of caste-based networking in Bombay's labor market are situated prior to independence in 1947. But a few studies conducted over the subsequent decades suggest that these patterns tended to persist over many generations. Patel (1963) surveyed 500 mill workers in the Parel area, close to the site of our study, in 1961-62 and found that 81% of the workers had relatives or members of their *jati* in the textile industry. 50% of the workers got jobs in mills through the influence of their relatives and 16% through their friends, many of whom would have belonged to the same *jati*. Similarly Dandekar (1986) traced the evolution of a network of Jadhavs (a particular subcaste) from Sugao village in Satara district to one jobber who went to Bombay in the 1930s, working first on the docks and later in the textile mills. In 1942, 67% of the Jadhav migrants from that village were working in the textile mills and 4% in other factories. By 1977, 58% were in textile mills and 10% were in other manufacturing industries, suggesting little change in occupational patterns over a 35 year period.

A noticeable feature of historical descriptions of caste-based networks in Bombay is that they were restricted to working class jobs. This is not surprising, because the information and enforcement problems that give rise to such networks tend to be more acute in those occupations. Further, most studies of caste-based networks in Bombay focus on male workers. Women were conspicuously absent from Bombay's labor force, particularly in the working class jobs (Morris 1965). We will later see that female labor force participation is low in our sample of households as well, and that when women do enter the workforce they tend to find clerical and professional jobs, which are less networked. These historical patterns of labor force participation by gender will later help explain the schooling choice dynamics, for boys and girls, that we saw in Figure 3 and Figure 4.

2.3 The School System

The British were not committed to providing mass education when they first began to colonize India on a large scale in the eighteenth century. The few schools and colleges that were set up were based in

the major cities and taught entirely in English; the objective at this stage was to create an Indian elite that would assist the British in administering the country (Nurullah and Naik 1947, Kamat 1985). But by 1850, missionary societies and some princely states not directly under the control of the British had begun to establish a network of rural schools that taught in vernacular (local) languages. And with Sir Charles Wood's Despatch of 1854, the British government finally assumed responsibility for the education of the entire population (Dakin, Tiffen and Widdowson 1968).

Over the next century, education which had previously been restricted to a few business communities and the Brahmin castes spread to the middle castes, with an accompanying shift in the language of instruction away from English. The Education Commission accepted Indian languages as a legitimate medium of instruction for all grades in 1902 (Dakin, Tiffen and Widdowson 1968), and by independence in 1947, a large fraction of Indian schools taught in the local language (Kamat 1985). But university education, at least in the major metropolitan areas, was always in English, and schooling in the local language was sometimes seen to put the prospective college student at a disadvantage. Thus, English and local language schools coexisted in the major cities.

Dadar was historically a suburb of Bombay - today with the expansion of the city, it lies at its center - and schools were established relatively late in the area. The average establishment year for the 18 Marathi schools in Dadar is 1947. The corresponding year for the 10 English schools is not much later, 1959, and all the schools in the area have now been operating for many decades. Our analysis highlights the medium of instruction as the salient feature of schooling choice. It is possible that the choice of the language of instruction merely proxies for school quality. Marathi schools can be private or municipal (government), while the English schools can be missionary or secular. These differences in the management lead to fairly wide variation in the availability of resources within each type of school. However, on average English-medium and Marathi-medium schools look very similar in terms of the facilities, attention and resources each student receives, and teacher qualifications.

In parallel with the household survey, we carried out a survey of schools, based on a questionnaire filled out by school principals. This questionnaire elicited information on a variety of school characteristics, as well as recent student performance on the standardized school leaving examination (common to both Marathi and English schools), which allows us to compare the two types of schools as well as the students across the schools. Table 1, Panel A, describes school infrastructure and faculty qualifications in the English and Marathi schools. The average student-teacher ratio, class size, number of students per desk, computers per student, and the proportion of teachers with B.Ed. degrees and

higher (postgraduate) degrees, are each very similar and statistically indistinguishable for the two types of schools.[12]

Despite the increase in the demand for English education in the last 10 years, as seen in Figures 3 and 4, no new schools were added in this period in Dadar. The English language schools accommodated this increased demand by adding divisions in each grade, increasing the number of desks in each classroom, and by doubling students on each desk. Because the supply of schools was effectively fixed, we would expect the English schools to extract some economic rents from this increased demand through higher fees and schooling costs in general. In contrast, fees in the Marathi schools are subsidized by the state government. Our household survey collected information on school fees and other expenses (transportation, coaching classes, textbooks, uniforms, and stationary) in the last year. Table 1, Panel B shows that school fees (in 1980 Rupees) are currently significantly higher in the English schools (480 versus 200 Rupees), as are other expenses (1,100 versus 710 Rupees).

One other difference between the schools is in the performance of the students on the Secondary School Certificate (S.S.C.) school-leaving examination. Table 1, Panel C reports student performance on this exam over a five year period, 1997-2001. Students in the English schools perform much better on this standardized test in terms of the percentage that pass, receive a first class, and a distinction.[13]

The test score differences across the schools suggest that the conventional measures of school quality seen in Panel A, which do not vary by language of instruction, may not capture all dimensions of school quality. We will show below, however, that these substantial differences in test performance can be explained by differential selection by ability into English and Marathi schools, an implication of our network model of school choice. Moreover, our survey provides direct evidence that the medium of instruction and its implications for children's future role in society, rather than differences in school quality, dominate the schooling choice of parents. The survey elicited from parents the reasons for their choice of school for their child. The percentage of parents reporting that the "quality of education" was a factor in choosing schools did not differ substantially across parents choosing English-medium schools or Marathi schools - 43.7% versus 35.2%, respectively. However, although almost 87% of parents who chose English as the medium of instruction for their child reported that career opportunities was a factor in choosing schools, less than 2% of parents choosing a Marathi-language school listed this

[12] A regression of school language medium on the set of school characteristics in Table 1, Panel A, indicates that the joint set of characteristics is not significantly different across the school types.

[13] Scores above 35% are required to pass, scores above 60% are required for a first class, and above 75% for a distinction. The same test is administered to all schools, with the questions translated into English and Marathi.

factor. In contrast, over 62% of parents who chose a Marathi-language school listed closer community ties as a reason; almost no parents choosing English-medium schools reported community ties as a factor in school choice.

3 A Simple Model of Schooling Choice

Our first objective in this section is to show how exogenous, historically-determined occupational differences across otherwise identical *jatis* can persist when network externalities are present. Because occupational choice translates into schooling choice, this explains the initial caste-gap that we observe for the boys in Figure 3 (the model that we lay out in this section applies to the boys, as we will show later that labor market networks are most active among the men). However, we will show that *jatis* should start to converge once the returns to English grow sufficiently large, which is inconsistent with what we observed in that Figure. Our second objective in this section will consequently be to show how network externalities could give rise to endogenous social restrictions on occupational mobility, and by extension schooling choice, preventing convergence across social groups in a changing economic environment.

3.1 Population, Community Structure, and Market Structure

Consider a population with a continuum of individuals. Each individual i is endowed with a level of ability $\omega_i \in \left\{0, \frac{1}{2}, 1\right\}$. Note that ability in this section, and throughout the paper, refers to pre-school human capital rather than genetic ability. He lives for three periods, studying in the first period and working in the remaining periods. Schooling choice is restricted to instruction in English or Marathi, the local language. Occupational choice is restricted to white-collar and working class jobs. Education in English is required to obtain a white-collar job, but is more expensive than Marathi education, which is assumed for simplicity to be costless. Occupational choice is based on the wage that the individual will receive in the white-collar and the working class job, net of the pecuniary cost of schooling. Each individual then makes his schooling decision based on the type of job that he (correctly) anticipates he will occupy in the subsequent period. If he prefers to hold a white-collar job then he will study in English, if not he will study in Marathi which is less costly.

Each individual is born into a community or *jati*. There is a large number of communities in this economy, and we normalize so that the measure of individuals in each generation of a *jati* is equal to one. To simplify and highlight the role of network externalities in intergenerational occupational

persistence we assume that the distribution of pre-school human capital within each generation does not vary over time or across *jatis*.[14] Within each *jati*-generation there is a measure P_L of low types (with ability $\omega = 0$) and a measure P_M of medium types (with ability $\omega = 1/2$).

On the demand side of this labor market, firms operate competitively in both the working class and the white-collar sectors. We noted earlier that working class jobs generally tend to be more heavily networked. For the purpose of this simple model we assume that the white-collar worker's ability, and hence his productivity, can be observed perfectly and so the white-collar wage (net of schooling costs) is specified to be $\theta\omega_i$. Here θ represents the returns to ability in the white-collar job, which in our set up also reflects the returns to English education. In contrast, the nature of the production technology prevents working class firms from directly observing their employees' ability before they commence work. We take it that the firm is unable to specify a performance-contingent wage contract, and so will use referrals from its incumbent workers to hire new employees, generating a role for the network in the working class jobs alone. Munshi (2003) provides evidence that experienced workers contribute disproportionately to labor market networks in the United States, and we would expect this pattern to hold up in other economies as well. In our model, the expected working class wage over the individual's working life is thus specified to be P, the proportion of the previous generation (three-year olds) in the *jati* that will be employed in the working class job when he enters the labor force.

3.2 The Schooling Equilibrium

We now proceed to derive the different occupational distributions, and hence schooling equilibria, that can be sustained across *jatis* with the same ability distribution in this set up. Each individual chooses the occupation, and hence the language of instruction, that maximizes his net return. This return depends on his own ability, as well as the proportion of his *jati* in the previous generation employed in the working class occupation, as described above.

Under conditions that we specify below, with three levels of ability, three distinct schooling equilibria can be sustained within *jatis*: (1) only low types choose Marathi education, (2) low and medium types choose Marathi education, (3) everyone in the *jati* chooses Marathi education.[15]

Condition 1: $P_L < \frac{\theta}{2}$

[14] We will relax this assumption in the empirical work by allowing for heterogeneity in ability across *jatis*.

[15] A previous working paper version of this paper considered the general case with N types and N equilibria, without altering the results that we present below.

Condition 2: $\frac{\theta}{2} < P_L + P_M < \theta$

Condition 3: $\theta < 1$.

It is easy to verify that once a *jati* is exogenously assigned a particular occupational distribution, this distribution will persist unchanged over many generations when the conditions above are satisfied.[16] This intergenerational state dependence is a consequence of the network externality associated with the working class occupation. It implies, in turn, that the probability that any individual i drawn randomly from *jati* j will be schooled in English ($E_{ij} = 1$) is related to the proportion of men in the previous generation employed in the working class job, P_j:

$$Pr(E_{ij} = 1) = 1 - P_j. \tag{1}$$

This expression will serve as the starting point for the empirical analysis described in Section 5, where we will examine the relationship between schooling choice in the current generation and the occupational distribution in the previous generation, to identify the presence of an underlying network organized around the *jati*.

3.3 Schooling Choice as the Returns to English Grow

The state dependence at the level of the *jati* derived above is obtained under the assumption that the parameters of the model, P_L, P_M, θ remain stable across generations. To explore the effect of the increase in the returns to English (θ) in the 1990s, we now allow for multiple cohorts within each generation.

If θ remains constant within a generation, then the results derived above follow through without modification for all cohorts. However, if θ increases across successive cohorts, then schooling choice within a *jati* could change over the course of a single generation. When θ just crosses one, high ability boys belonging to *jatis* that were traditionally in equilibrium 3 switch to English. When θ subsequently reaches $2(P_L + P_M)$, medium ability boys in *jatis* that were traditionally in equilibrium 2 or equilibrium 3 switch to English, at which point schooling choice across all *jatis* will converge.

Although the network externality described above can explain the persistence of traditional occupational patterns within the *jati* over many generations, and hence the initial caste-gap observed in

[16] It is merely necessary to show that no individual wishes to deviate from the occupation, and hence schooling choice, assigned to his type in his *jati* in the previous generation, for each of the schooling equilibria.

Figure 3, it cannot by itself explain the absence of convergence over the 1990s as the returns to English grew. To explain this absence of convergence, we consider the possibility that heavily networked (working class) *jatis* might have put restrictions on occupational mobility, and hence schooling choice, in place to preserve the viability of the community network.

To understand why restrictions on mobility might emerge, define a social welfare function that places equal weight on all members of the *jati*. Now the welfare in a *jati* situated in equilibrium 3, in which everyone studies Marathi, is simply the unweighted average of all the payoffs from the working-class occupation, $W = 1$. When θ just crosses one, in a given generation, all high types in the *jati* can expect to earn more in the white-collar sector than in the *jati's* "traditional" working-class jobs and will thus switch to English schooling. Welfare from the subsequent generation onwards is then $W = (P_L + P_M)^2 + (1 - P_L - P_M)$. The new welfare level is a weighted average of $P_L + P_M < 1$ and 1, and so *jati* -level welfare must unambiguously decline when schooling choice, and hence the occupational distribution, shifts. As noted in Section 2, there was intense competition for scarce working-class jobs historically in Bombay. Because larger numbers improve the *jati*'s competitiveness, and increase the working class wage in general, it is easy to see why social restrictions on occupational mobility could emerge endogenously. Moreover, the fact that the lower-caste girls in our sample do not display a similar resistance to change can be attributed to the gender-specific nature of these job networks.

Social restrictions on occupational mobility can be welfare-enhancing for small and medium changes in θ, because they preserve the externalities that the network provides. But they could give rise to substantial inefficiencies if they continue to persist when θ grows large. For example, it is easy to verify that the social restrictions described above for equilibrium 3 will be inefficient once θ reaches $1+(P_L+P_M)$, although a welfare calculation that identifies the presence of such a dynamic inefficiency is beyond the scope of this paper.

Although we conjecture that restrictions on occupational mobility might be in place in the heavily networked *jatis*, no direct evidence of their presence in Bombay is available. We can, however, test one important implication that is consistent with the presence of these restrictions; the relationship between schooling choice E_{ij} and the occupational distribution within the *jati* in the previous generation P_j must not weaken over successive cohorts in the current generation, even as the returns to English grow. This stability in intergenerational state dependence would then explain the wedge between high caste and lower caste schooling choices for boys that was observed through the 1990s in Figure 3.

3.4 Selection into Schools

The model of schooling choice as laid out in this section also has implications for selection, by ability, into English and Marathi schools. Within any *jati*, the average pre-school human capital of the English students must be greater than that of the Marathi students. Taking the average across all *jatis*, this implies that average ability must be greater among the English students at any point in time. This observation is consistent with the significantly higher test scores obtained by students in the English schools in Table 1, despite the fact that English and Marathi schools appear to be similar in terms of the resources available per student and the qualifications of the teachers. But how does the ability distribution *within* the English and Marathi schools change across successive cohorts in the current generation as the returns to English grow? Without social restrictions, deviation to English education is ordered by ability, so as θ grows there is a steadily worsening pool of Marathi students. *Jatis* that begin with a greater proportion of their members in working-class jobs have higher ability among the Marathi students, but their shift into English, and hence the decline in ability, must also be more rapid, because all *jatis* ultimately converge. With social restrictions, heavily networked *jatis* continue to have a superior ability distribution within Marathi schools, but now there could be no convergence in ability among Marathi students across *jatis*.

Although the model implies that the quality of the Marathi students unambiguously declines over time as the returns to English increase, the change in the quality of the pool of English students is ambiguous. Average ability among the English students is greater than average ability among the Marathi students at any point in time, but among the Marathi students it is those with the highest ability that deviate as θ grows. For example, in the three-type case with no social restrictions, some *jatis* (in equilibrium 2) have only high ability children in English schools, while other *jatis* (in equilibrium 1) have both medium and high ability children in English schools to begin with. This implies that the quality of the English pool must improve when θ reaches one, because only the high types from *jatis* in equilibrium 3 deviate at that point. But average ability drops below its initial level when θ reaches $2(P_L + P_M)$, because medium and high types in all *jatis* will have switched into English schools by that point. With social restrictions, the change in ability within the English schools becomes even more difficult to predict.

4 The Household Data

4.1 The Survey

To examine empirically the role of caste networks in shaping mobility during a period of change we carried out a household survey based on a stratified (by caste) random sample of students who entered the 28 secondary schools in Dadar (in the first grade) over a 20 year period, 1982-2001. This design provides information for the periods before and after the major Indian economic reforms. As noted, secondary schools in Bombay run from grade one to grade 10. We obtained a complete list of all students enrolled in grades one to 10 in 2001 (the year of the survey), as well as a list of students who were enrolled in grade 10 from 1991 to 2000. Ignoring dropouts, this leaves us with 20 cohorts of students who entered school over the 1982-2001 period. A total of 101,567 students were enrolled in the schools in 2001 or studied in grade 10 over the previous 10 years. We drew the roll numbers of 20,596 students randomly from these 20 cohorts, and recovered their names and addresses from the school records. Restricting attention to Maharashtrians residing in Dadar and the immediately adjacent neighborhoods of Mahim, Matunga, Wadala, Prabhadevi, and Parel, we were left with 8,092 eligible students to serve as the sampling frame for the survey. The student's name is typically a good indicator of the caste, and we wanted close to one thousand upper castes in the sample, so all 1,082 students from this population who appeared to be upper castes were selected for the survey. We drew randomly from the remaining students in the sampling frame until the target sample size was reached. The upper castes account for 17.5% of the final sample of 4,945 observations, which is slightly higher than the 13.4% that we began with in the sampling frame.

The research team interviewed the parents of the selected students at their residences. The survey instrument elicited detailed sub-caste information from the respondents and included sections on grandparents' education and occupation, parents' education and occupational and income histories (at five year intervals from 1980 to 2000), as well as the student's and siblings' subsequent education, occupation, income and marriage outcomes (where relevant). Information on transfers, assistance in finding jobs, and ties to the community was also collected.

82.5% of eligible households provided completed schedules. This is a relatively high response rate, especially given that some of our addresses were 20 years old. But we might still have obtained a selective sample of households, for a number of different reasons. First, households residing in Dadar who sent their children to study outside the area would be missing from the sample. Second,

households who moved out of the area would be among the 17.5% of the respondents who did not complete the survey. And third, students from the first 10 cohorts who did not reach the tenth grade, and current students who have dropped out, would be missing from the sample. In Section 6 we will discuss how our identification strategy is unlikely to be undermined by these potential sources of bias.

4.2 Descriptive Statistics: Caste, Occupational Networks and Schooling

The data provide empirical support for three features of the model of schooling choice laid out in Section 3. First, the occupational distribution, a product of historical circumstances, varies by caste, and persists across generations, particularly among the men. Second, working class jobs are associated with a higher level of referrals (networking). And third, working class jobs are associated with lower levels of English schooling.

The survey elicited information on parental occupations at five year intervals from 1980 to 2000. For the grandparents, we simply asked for the main occupation over the individual's working life. The 90 occupations in the data were divided by roughly increasing levels of human capital into seven aggregate categories: unskilled manual, skilled manual, organized blue-collar, petty trade, clerical, business, and professional. We further classified unskilled manual, skilled manual, and organized blue-collar as working class occupations. Clerical, business, and professional were classified as white-collar occupations. Petty trade is treated as an intermediate unclassified occupation.

Table 2 describes the occupational distribution across broad caste categories (low, medium, high), separately for the employed fathers, based on information in 1995, and the paternal grandfathers of the students in the sample. Columns 1-3 of Table 2 indicate that lower-caste fathers are much more likely to be employed in working class occupations (53% and 43%) as compared with high-caste fathers (18%).[17] The same cross-caste pattern is obtained for individual occupations within the working class and white-collar classifications, with the exception of clerical jobs. Notice also that the cross-caste patterns for the two major occupations, organized blue-collar and professional, are particularly striking for the fathers.

The comparison of the fathers in Columns 1-3 with the grandfathers in Columns 4-6 suggests that there is little change in the basic occupational distribution, measured by the percentage of working class jobs, across the generations within broad caste categories. The major difference is that farming is listed as the primary occupation for a large proportion of lower caste grandfathers, which suggests

[17]Note that we only use working class and white-collar occupations when computing this statistic.

that many of the lower caste fathers are first-generation migrants.[18] There appears to have been a switch from farming to manual jobs, clerical jobs, and professional jobs. But the percentage of blue-collar jobs, and working class jobs more generally, remains stable across the generations. For the high castes, the major change is the decline in blue-collar and clerical jobs, and the increase in professional occupations, but once again the proportion of working class jobs remains stable across generations.

Although most men are employed, we see that labor force participation for the women in Table 3 is relatively low but is growing. Only 15% of high caste grandmothers worked; whereas just over half of high caste mothers entered the labor force (based on their employment status in 1995). Among the lower castes, the percentage employed remains stable at 20% across the generations, but notice that farming is listed as the primary occupation for a large number of working grandmothers. This suggests that urban employment must have increased sharply for the lower-caste women as well.

The occupational distribution across castes for the mothers in Table 3, Columns 1-3, displays a pattern similar to that for the fathers. Lower-caste women are much more likely to be employed in working class occupations (44% and 32%) as compared with high caste women (9%). However, there is an important difference between men and women - although the large difference within the working class occupations for the men was in access to blue-collar jobs for the lower castes, for women the major difference is in access to unskilled manual jobs; many of the lower caste women work as sweepers and domestic servants.

Columns 1-3 and Columns 4-6 suggest that there has been, in contrast to the men, significant intergenerational change for women within the castes. Among the high-caste women there has been an increase in the proportion of clerical jobs and a decline in the proportion of professional jobs. This is most likely because only the highest ability women of the older generation (grandmothers) would have entered the labor force. For the lower castes, women have evidently shifted out of farming into unskilled manual jobs (particularly among the medium castes), skilled manual jobs, clerical jobs, and professional jobs. The decline in the percentage of working class jobs among the lower caste women, across a single generation, is particularly dramatic. This contrasts with the stability of the occupational distribution for the men, for all castes, that we noted earlier, consistent with the view that labor networks are weak among the women.

[18] A parent or grandparent will only list farming as their primary occupation if they were residing outside Bombay, in their home village. Not surprisingly, farming is rarely reported as the father's primary occupation. But it is for a substantial fraction of the grandfathers (19% for the low caste and 27% for the medium castes), which tells us that roughly one-quarter of the lower caste fathers are first-generation migrants. Migrants are by definition newcomers in the labor market, and so will be more susceptible to the information problems that generate a need for the caste networks.

The model assumes that working class jobs are associated with higher levels of networking (referrals). The parents of the selected students were asked how they learned about their first job; through a childhood friend, through a college friend, through a relative, through a member of the community (*jati*), or by some other means (which was left open-ended in the questionnaire). This open-ended category ultimately included the many cases in which no help was received, or in which the job was found through newspaper advertisements, campus interviews, and other impersonal information channels. A binary referral variable was then constructed, taking the value one if the parent learned about the first job from a relative or member of the community, zero otherwise.

Table 4 indicates that, as assumed in the model, working class occupations are associated with a higher level of job referrals in our sample. Column 1 shows that 68% of the working class men received help from a relative or member of the community in finding their first job (or starting their first business if self-employed), which is significantly higher than the 44% of men in the white-collar jobs who received a referral. The corresponding statistics for the women in Column 3 reveal essentially the same pattern, although the level of referrals for the women is generally lower than that for the men, perhaps because the networks for female jobs are less developed.[19]

The model also assumes that Marathi schooling channels the student into a working class job, while English schooling leads to the white-collar occupation. The survey elicited information on the language of instruction (English versus Marathi) for fathers and mothers, in secondary school. Table 4, Columns 2 and 4, show that there is a clear distinction between working class and white-collar jobs with respect to the language of instruction in secondary school. The percentage of men in working class jobs that attended secondary school in English is just over 1%, compared with the 6% of men in white-collar jobs. A similar pattern is obtained for the women in Column 4.

We have described the relationship between the broad occupational categories (working class versus white-collar), the level of referrals, and English schooling. But inspection of Table 4 indicates that the level of referrals and English schooling vary systematically within these categories as well. For example, among the white-collar occupation, the professional jobs have lower referrals and higher English schooling, as we would expect. Similarly, blue-collar jobs, the dominant occupation in the working class category, have extremely high referrals and correspondingly low English schooling. We will take advantage of this finer relationship between particular occupations, the level of referrals, and

[19]The variation in networking across occupational categories remains unchanged when assistance from friends (not belonging to the *jati*) is included in our measure of referrals. 75% of working class men versus 53% of white-collar men receive a referral with this alternative measure. The corresponding statistics for women are 66% and 34%.

English schooling, to later characterize the occupational distribution in the *jati*.

Table 2 and Table 3 suggest that lower-caste men and women are much more likely to hold working class jobs. Table 5 combines these results with the results in Table 4 to describe referrals and English schooling across castes. Not surprisingly the table indicates that a much higher proportion of lower caste men received referrals, and that such men are much less likely to have been schooled in English. In contrast, although lower caste women are also much less likely to be schooled in English, the level of referrals is statistically indistinguishable across castes. Although we noted earlier that lower caste women who work are more likely to hold working class jobs, which are associated with more referrals, we also saw that lower caste women are less likely to enter the labor force. These two opposing effects appear to cancel each other, leaving little variation in the level of referrals across castes for the women. The level of referrals is low in any case, especially when compared with the corresponding level for the men, and we will later establish that labor market networks are effectively available for the men only.

Tables 2 through 4 provide support for the basic features and assumptions of the model with respect to the variation in the occupational distribution across castes and its persistence, the relationship between the type of occupation and the level of referrals, and the relationship between the type of occupation and English schooling. The data also suggest, however, that the modelling assumption that the distribution of pre-school human capital is the same across castes (or *jatis*) does not appear to be supported empirically. Children in a wealthy, educated *jati* that have had access to white-collar jobs for many generations will be nurtured very differently from children in a *jati* that was historically confined to manual jobs. Thus, the occupation that a particular *jati* gets locked into, as a consequence of historical circumstances, could affect future occupational choice and hence schooling via human capital effects. Table 6 reports the mean years of schooling and monthly income separately by caste for men and women.[20] As expected, high caste mothers and fathers have significantly more years of schooling and significantly higher incomes. This suggests that pre-school human capital could vary in practice across broad caste categories, and across *jatis*, as well. When estimating the effect of the historical occupational distribution on the child's schooling choice we will consequently take account of the possibility that the occupational distribution could be correlated with the ability distribution in the *jati*.

[20]Recall that income information was collected from each parent at five equal points in time from 1980 to 2000. We use the income (in 1980 Rupees) that coincides as closely as possible with the year in which the child entered school. Thus, the income in 2000 is used for students aged 6-10, the income in 1995 for students 11-15, the income in 1990 for students 16-20, and the income in 1985 for students 21-25. The same income statistic is used later in the schooling regressions.

5 Empirical Analysis

5.1 Specification and Identification

The first implication of the model incorporating network externalities is that the occupational distribution in the *jati* should persist across generations when networks are active. Because schooling choice maps into occupational choice, equation (1) depicts this implication in terms of schooling choice in the current generation and the occupational distribution in the previous generation:

$$Pr(E_{ij} = 1) = 1 - P_j.$$

Recall that $E_{ij} = 1$ if individual i belonging to *jati* j is schooled in English, $E_{ij} = 0$ if he is schooled in Marathi. P_j is the proportion of men in the *jati* in the previous generation who are employed in working class jobs and so in a position to provide referrals.

The particular relationship between E_{ij} and P_j in the equation above is, of course, a consequence of the modelling assumption that schooling choice maps one-for-once into future occupational outcomes. More generally we would expect to see a negative coefficient, but not necessarily with magnitude one, on P_j. The model laid out in Section 3 also does not allow for intergenerational state dependence in schooling choice at the level of the *household*. Moreover, we noted above that pre-school human capital and family incomes appeared to vary systematically across castes with different occupational backgrounds. The schooling regression that we estimate is consequently specified as,

$$Pr(E_{ij} = 1) = \alpha P_j + X_{ij}\beta + \omega_j, \tag{2}$$

where X_{ij} includes the parents' language of schooling to reflect household-level state dependence in schooling choice, as well as a cohort variable to capture the increase in the returns to English over successive cohorts in the current generation. ω_j measures pre-school human capital and family income in the *jati*, which could independently determine schooling choice.

P_j measures the proportion of men in the previous generation employed in working class jobs. Although the model assumes that only two types of jobs - working class and white-collar - are available, as many as 90 occupations are listed in the data. A relatively strong relationship between the level of referrals and the type of occupation was observed earlier in Table 5, and so one convenient statistic that accurately and parsimoniously describes the occupational distribution in the *jati* would be the proportion of fathers (the previous generation) that received a job referral. Working class jobs were

also associated with lower levels of English schooling in Table 5. An alternative measure of the occupational distribution in the previous generation would compute the proportion of fathers that attended English secondary schools. Most of the regressions reported in this paper will use the referrals statistic to measure P_j; English schooling levels were generally low in the previous generation and so there is substantially more variation in the referrals statistic across *jatis*. We will, however, verify that the results hold up with the English schooling statistic as well.

Following the discussion above, we expect to find $\alpha < 0$ when networks are active and the occupational distribution persists across generations. Recall that α must also remain stable across cohorts in the current generation to explain the absence of convergence in Figure 3. Although much of the analysis treats α as constant, we will later verify that α does indeed remain stable across cohorts.

An identification problem arises when P_j and ω_j are correlated in equation (2). Although *jatis* might have been the same to begin with, we noted in the previous section that their members now have very different characteristics (income and education), depending on the type of occupation that the *jati* has historically been engaged in. These differences could result in different levels of pre-school human capital in the current generation, which would independently determine schooling choices. In addition, given imperfections in credit markets, liquidity constraints could prevent students from attending more expensive English schools and this determinant of schooling choice could vary across *jatis* with different levels of income. A traditionally working-class *jati* could thus be associated with high P_j and low ω_j, in which case a family effect would be erroneously interpreted as a network effect because individuals with lower family resources independently select into Marathi schools.

Our solution to this identification problem exploits the fact, documented in Table 4, that networks are concentrated in working class jobs dominated by men. Recall from Table 5 that the levels of referrals for women were relatively low, consistent with the significant change in the occupational distribution across generations for women indicated in Table 3. Thus, although the networks might affect schooling choice for the boys, they should have had little or no impact on the girls. The model in Section 3, then applies to boys only. Instead of using variation in the level of referrals across *jatis* to identify the presence of networks, as in equation (2), we proceed instead to exploit this gender difference in the access to job networks by pooling both sexes in the schooling regression to identify the presence of the network *within* the *jati*:

$$Pr(E_{ij} = 1) = (\alpha - \tilde{\alpha})P_j \cdot B_{ij} + X_{ij}\tilde{\beta} + X_{ij} \cdot B_{ij}(\beta - \tilde{\beta}) + \gamma B_{ij} + f_j \tag{3}$$

where $\tilde{\alpha}$, $\tilde{\beta}$ represent the effect of the network and parents' language of schooling on the girls. B_{ij} is a dummy variable that takes a value of one for boys and zero for girls. The advantage of pooling the boys and girls is that the schooling regression can be estimated with *jati* fixed effects, $f_j \equiv \tilde{\alpha} P_j + \omega_j$. Although we can no longer identify α directly, we can obtain a consistent estimate of $\alpha - \tilde{\alpha}$, the coefficient on the $P_j \cdot B_{ij}$ interaction term. For the special case with exclusively male networks, $\tilde{\alpha} = 0$ and the coefficient on the interaction term identifies network-based occupational persistence for the boys directly. More generally, the coefficient on the interaction term provides a conservative estimate of the effect of caste-based networks on schooling choices for the boys.

The identifying assumption in this estimation strategy is that no variable $\phi_j \cdot B_{ij}$ appears in the residual of equation (4), where ϕ_j is correlated with P_j. A sufficient condition for this identifying assumption to be satisfied is that the unobserved determinant of schooling choice should not vary by gender or have a differential effect on schooling choice by gender, *within* the *jati*. Later in Section 6 we will discuss alternative explanations for the negative and significant $\alpha - \tilde{\alpha}$ coefficient that we obtain in the schooling regression. These explanations either relax the assumptions of the model, made earlier in Section 3, or build on the failure of the identifying assumption. We will argue that none of these explanations fit the data quite as well as the male labor market network explanation that we put forward in this paper.

5.2 Caste-based Networks and Schooling Choice

Table 6, Column 1 reports the estimates of the schooling choice regression, equation (2), for the boys. As noted, the sample covers 20 cohorts of students aged six to 25, who entered school between 1982 (cohort=1) and 2001 (cohort=20). The student's cohort (1 to 20), the proportion of fathers in his *jati* that received a referral, and the father's and the mother's language of instruction in secondary school, are included as regressors.

The cohort term is included in this regression to account for the increase in the returns to English over time. While the linear cohort effect that we specify in Table 6 is clearly restrictive, we verify below that the estimated referral coefficient is unchanged when we allow for more flexible cohort effects. The referrals coefficient is also specified to be constant over time in Table 6, and we will subsequently relax this restriction as well. For now, we see that the referral coefficient is negative and significant; children belonging to (historically) working class and more heavily networked *jatis* are less likely to be schooled in English, consistent with the first implication of the model. The cohort effect is positive

and significant, implying a shift into English over time, which is consistent with the increase in the returns to English that we saw in Figure 1. Finally, the results imply that a boy is much more likely to be schooled in English if his parents were educated in that language, indicating significant state dependence in schooling choice at the level of the household.

Table 6, Column 2 reports estimates from a specification that includes variables that measure the determinants of the student's pre-school human capital as well as the household budget constraint, which could independently determine schooling choices. The parents' years of education, conditional on their language of instruction in secondary school and the level of referrals in the *jati*, are likely significant determinants of children's pre-school human capital. The family's access to own resources is measured by the total income of the father and the mother at the time when the child entered school.[21] Inclusion of these variables results in a substantial decline in the referrals coefficient, suggesting that the level of referrals was previously proxying to some extent for unobserved family-specific determinants of schooling choice, but it remains negative and significant. The coefficient on the cohort variable is quite stable. And the coefficients on the additional regressors all have sensible signs; the boy is more likely to be schooled in a more-expensive English-medium school if his father or mother are more educated, or if the family is wealthier. Occupational persistence in Bombay thus in part arises due to human capital effects.

The estimates for girls are reported in Columns 3 and 4 in Table 6. Column 3 reports the estimates based on equation (2); Column 4 reports the estimates from the augmented specification that adds the parents' years of schooling and family income as additional regressors. The estimated cohort effects, and the coefficients on parents' language of instruction, parents' education, and family income are similar to those for boys in Columns 1 and 2. However, the referral coefficient becomes negligible for the girls in Column 4 once the observed determinants of pre-school human capital and access to own family resources are included. One explanation for this result is that girls receive help from the women, not the men, in their *jati*. But we saw in Table 6 that the level of referrals for the women is very low, across all castes. Although not reported, we also found no correlation between referrals and schooling choice, for both boys and girls, when we replaced the level of referrals for the fathers with the level of referrals for the mothers.

The results that we have just described are consistent with the view that caste-based networks, net

[21] We use the income in 2000 for students current aged 6-10, the income in 1995 for students 11-15, the income in 1990 for students 16-20, and the income in 1985 for students 21-25. All incomes are computed in 1980 Rupees.

of individual and family characteristics, affect schooling decisions for the boys, but not for the girls. But up to this point, we have only controlled for unobserved ability with a limited number of family characteristics. A more robust identification strategy estimates the schooling regression with *jati* fixed effects, as in equation (4). These estimates are reported in Column 5 of Table 6. As noted, only the referral-boy interaction coefficient, and not the linear referral coefficient, can now be identified. The coefficient on this term is negative and significant, and very similar to the referrals coefficient for the boys in Column 2. Recall from equation (4) that the coefficient on the referral-boy interaction term provides us with a direct estimate of the referral coefficient for the boys if the referral coefficient for the girls is zero. The result that we obtained earlier for the girls in Column 4 suggests that this might well be the case.

The coefficient on the cohort variable in Column 5, which now measures the cohort effect for the girls, is similar to what we obtained previously for both boys and girls in Columns 1-4. The coefficient on the cohort-boy interaction term, which measures the difference in the cohort effect for boys and girls, is consequently small and statistically insignificant. The boy dummy itself is positive and significant, reflecting the historically higher levels of English schooling among the men. This is most likely because the very large proportion of women who stayed outside the workforce in previous generations would have been schooled in Marathi, which is less expensive.

We have argued throughout this paper that the persistence of occupational patterns across generations is due to the presence of networks organized at the level of the *jati*. But an alternative explanation is that occupations get passed down at the family level, from father to son. We account for such occupational persistence at the level of the family in Table 7, Column 6, by including a full set of (90) dummies for the student's father's occupation, in addition to the regressors listed above. The set of paternal occupational dummies is statistically significant, suggesting that persistence in occupational choice additionally operates at the family level. However, comparing Column 5 and Column 6, we see little change in any of the estimated coefficients. There is still an independent, *jati*-level source of occupational persistence.

Although the labor market in Bombay was historically organized by caste, the region of origin (within Maharashtra) also has been seen as determining the scope of the network (Chandavarkar 1994). To allow for this finer partition of the network, we divide Maharashtra into 4 regions: Bombay, Konkan, Deccan, and Other. Our survey elicited information on where the grandparents grew up. We use this response, for the paternal grandfather, to assign an origin region for each student. The

Deccan (in interior Maharashtra) and the Konkan (along the coast), were the main source regions for the supply of labor to Bombay under the British. We find that Bombay is listed as the origin region for 27% of the grandfathers, the Konkan for 42% of the grandfathers, the Deccan for 20% of the grandfathers, and that the remaining 11% of the grandfathers came from other parts of the state. Replacing the *jati* by the *jati*-region as the boundary of the network in Table 6, Column 7, the point estimates are almost identical to what we obtained earlier in Column 5. There are 59 *jatis* in our sample, and with four regions this leaves us with a very large number of *jati*-region units in Column 7. Not surprisingly, there is wide variation in the size of these units. To check the robustness of our results to the size of the network, we dropped very large *jati*-regions (more than 250 observations) and very small *jati*-regions (less than 10 observations). The estimated referral coefficient based on this substantially reduced sample, reported in Column 8 of Table 6, is close to what we obtained with the full sample.[22]

The regression specifications with *jati* fixed effects in Table 6, Columns 5-8, did not include family characteristics. Table 7, Column 1, includes parents' language of schooling, both uninteracted and interacted with the boy dummy, as additional regressors. Subsequently, Table 7, Column 2 adds parents' years of schooling and family income, also uninteracted and interacted with the boy dummy. Including uninteracted household characteristics in the schooling regression has no effect on the estimated referrals-boy coefficient by construction, once *jati* fixed effects are included. But we see in Table 7, comparing columns 1 and 2, that the inclusion of parents' years of schooling and family income, interacted with the boy dummy, has no effect on the estimated referrals-boy coefficient as well. Indeed, this coefficient is no smaller than the corresponding coefficient estimated earlier in Table 6, Columns 5-8, without *any* household characteristics. This stability contrasts with the decline in the referrals coefficient in Table 6, Columns 1-4, when household characteristics were included, providing some support for the view that the *jati* fixed effects absorb much of the unobserved heterogeneity in this environment.

Notice also that parents' years of schooling and family income, which had a strong influence on schooling choice for both boys and girls in Table 6, do not differentially affect schooling choice by gender in Table 7. It is only the *jati*-level referrals variable that has such a differential effect on schooling choice, as measured by the negative and significant referrals-boy coefficient. This observation will be

[22]It is possible to trim the sample even further - restricting attention to networks between 50 and 150 observations - without affecting the estimated referrals coefficient.

useful later in Section 6 when we consider alternative explanations for the results presented in this paper.

5.3 Schooling Choice over Time

The second implication of the model laid out in Section 3 is that the relationship between schooling choice and the occupational distribution in the previous generation will weaken across successive cohorts in the current generation as the returns to English grow, unless restrictions on occupational mobility are in place. To empirically assess the stability of the referrals coefficient we create four cohort categories that evenly divide the 20 cohorts, and then estimate the referral coefficient separately for each category.

We begin with a benchmark *jati*-fixed effects regression, which maintains a constant referral coefficient but relaxes the restriction imposed thus far that cohort effects are linear, by including the cohort categories in Table 8, Column 1. The estimated negative referral-boy coefficient is unaffected by the inclusion of the flexible cohort effect and remains very similar to the results shown in Tables 6 and 7. Inclusion of the family background variables, uninteracted and interacted with the boy dummy, as additional regressors again has no effect on the estimated referral coefficient (Column 2).

Table 8, Column 3, allows for the changes in the referral coefficient across cohort categories. All the referral-boy-cohort coefficients are negative and significant except for the coefficient on the first cohort category which is slightly less precisely estimated. The referral coefficient is actually increasing for the later cohorts, and we can easily reject the convergence hypothesis which implies a decline in the referral effect over time. Once more, the estimated referral coefficients are robust to the inclusion of the family background variables as regressors (Column 4). Although not reported here, the referral coefficient remained stable when the schooling regression was estimated with boys only, including parents' years of education and family income as additional regressors. It is this *jati*-level effect that presumably sustains the gap in schooling choice between broad caste categories observed for the boys in Figure 3.

5.4 Robustness Tests and Validation: Alternative Measures of the Occupational Distribution and Schooling

We noted that the occupational distribution in the *jati* could be measured either by the proportion of fathers that received a referral for their first job or by the proportion of fathers that were schooled

in English. The regression results reported thus far used referrals to measure the occupational distribution, and we now proceed to verify that our results are robust to this alternative measure of the distribution. Table 9, Columns 1-4, repeat the schooling regressions for boys, girls, and the pooled sample, reported earlier in Table 6, Columns 2, 4, 5 and Table 7, Column 2, with the proportion of fathers schooled in English as the measure of the occupational distribution.

The coefficients on the cohort variable and the household characteristics in Table 9, Columns 1-2, are very similar to the estimates reported in Table 6. The coefficient on the English proportion, which can be interpreted as state dependence in schooling choice at the *jati* level, is positive and significant as expected. The coefficient on the English-boy interaction term in Table 9, Columns 3-4, which include *jati* fixed effects, is very similar to the English coefficient for the boys in Column 1, matching the results reported earlier with the referrals variable. Although not reported, parents' education and family income also do not have a differential effect on schooling choice by gender once again; it is only English schooling among the fathers at the level of the *jati* that appears to have this gender-specific effect.[23]

In our framework, schooling choice reflects occupational choice and the relationship between the *jati*-level parental occupational measure, whether referral- or language-based, and a household's choice of the medium of instruction for its children measures intergenerational occupational persistence (for boys). This intergenerational relationship between the *jati* occupational distribution and school choice should therefore not be observed for measures of schooling that do not strongly map into future occupational choices. School quality could also affect occupational choice, and one set of measures of school quality, for given student abilities, are student test scores. We have seen, however, that the English-medium and Marathi-medium schools do not appear to differ by quality, and thus the quality dimensions of the schools are not important determinants of careers. Test scores may vary across *jatis* because of differences in parental resources and thus in student pre-school human capital. However, under the assumption that boys and girls within households and *jatis* do not differ in pre-school human capital, there should be no differential effect by gender of the occupational distribution of the *jati* on test scores even though within *jatis* boys and girls attend different schools.

Table 9, Columns 5-7, replace medium of instruction choice with performance on the school-

[23]We repeated this robustness test separately with average income, average years of education, and the proportion of working class jobs, among the fathers in the *jati*, as measures of the occupational distribution. The pattern reported in Columns 1-4 was replicated in each case. Our measure of working class jobs is a binary variable, and not surprisingly the coefficients with this coarse measure of the occupational distribution are less precisely estimated (although they continue to remain significant at the 5 percent level).

leaving S.S.C. examination as the dependent variable. Referrals are once more used to measure the occupational distribution in the *jati*, to be consistent with the specifications used elsewhere in the paper. We restrict attention to the first 10 cohorts (age 16-25), which have already attained school-leaving age, in these regressions. Just 17% of the students aged 16-25 in the sample never passed the S.S.C. examination, so we focus on the test score conditional on having passed the exam in these regressions.[24]

Table 9, Column 5 restricts attention to boys, and includes the cohort, the parental background variables, and the level of referrals in the *jati* as regressors. The cohort effect is negative and significant, suggesting, consistent with the model, that there is a decline in the quality of students over time. The referral coefficient is also negative and precisely estimated as well, consistent with students from high-referral *jatis* having lower levels of pre-school human capital. Consistent with this interpretation, family characteristics, particularly parents' years of education, have a very large positive effect on test performance. Subsequently we repeat the exercise just described for the girls (Table 9, Column 6). The cohort effect is now absent, but the coefficient on referrals remains negative and statistically significant - both boys and girls from high-referral *jatis* do less well on exams. However, the fixed-effects estimates, reported in Table 9, Column 7, of the cohort effect, the cohort-boy interaction, and the boy dummy are not statistically significantly different from zero. More importantly, the coefficient on the referral-boy interaction term is small and statistically insignificant, in contrast to the specifications in which choice of language of instruction is the dependent variable. The caste system affects the language but not the quality of instruction of its members.

5.5 Selection into Marathi Schools over Time

The framework laid out in Section 3 also has implications for the compositional change in the students who attend Marathi schools over time by *jati*: First, the pre-school human capital of boys entering Marathi schools should decline on average as the returns to English grow. Second, when there are no restrictions on mobility put in place to exploit network externalities, the distribution of pre-school ability among the boys entering Marathi schools will converge across all *jatis* over time. It is possible that such convergence across *jatis* will be absent when restrictions are in place. Note that the model has no prediction for selection by ability into English schools.

[24]Previous versions of the paper studied the effect of referrals on the probability of success in the S.S.C. exam and obtained results that are qualitatively the same as what we report below with the test score, conditional on success, as the dependent variable.

We do not have a direct measure of pre-school human capital. However, the results in Tables 9 suggest that, net of income, parental schooling has a positive and significant effect on school performance. In particular father's schooling has a significant positive effect on test scores for boys and girls, and the effects do not differ significantly by the gender of the child. We thus use the father's schooling level as a proxy for pre-school ability.[25] The question we address is whether boys with more educated fathers increasingly exit Marathi schools and whether and how the rate of decline in the pre-school ability of boys entering Marathi schools varies by *jati*.

To test the implications for school selectivity described above, we estimate regressions on the sub-sample of boys entering Marathi schools of the form:

$$E(S_{ij} \mid E_{ij} = 0) = \kappa + \lambda R_j + \mu C_{ij} + \nu R_j \cdot C_{ij} + \psi \omega_j \tag{4}$$

where S_{ij} is boy i in *jati* j's father's years of schooling and C_{ij} is the boy's cohort. R_j measures the level of referrals in the *jati* and ω_j measures pre-school ability in the entire *jati*; these terms reflect the fact that pre-school ability conditional on selection into Marathi is in general a function of ability in the *jati* and the level of referrals. The cohort terms reflect the change in this selection process over time. The model, which ignores the variation in ability ω_j across *jatis*, predicts that $\lambda > 0$, $\mu < 0$, $\nu < 0$ without restrictions and (possibly) $\nu > 0$ with restrictions.

Because the major shift into English schooling occurred in the 1990s, we first estimated equation (4) for the boys in cohorts 11-20. The estimates are reported in Table 10, Column 1. The cohort coefficient is negative and significant as predicted, which implies that the pre-school human capital of the boys who entered Marathi schools was declining substantially in the 1990s. The coefficient on the referrals-cohort interaction term is positive, consistent with the results in Table 9 showing that restrictions on mobility in the high-referral networks were still in place during this period - the more-able boys in high-referral *jatis* were shifting to English-medium schools at lower rates.

The referral coefficient is negative and significant, consistent with the results in Table 9, which indicate that *jatis* with higher referrals R_j have lower ability ω_j; the negative $R_j - \omega_j$ correlation appears to dominate the positive selection $\lambda > 0$ effect in this case. But this tells us that the positive referrals-cohort coefficient that we reported above might also be spurious. To assess the robustness of the results in Column 1, we add *jati* fixed effects, which subsume $\kappa + \lambda R_j + \psi \omega_j$, in Table 10, Column

[25] The results reported below are essentially the same if we replace father's schooling by mother's schooling.

2. The referral coefficient λ is no longer identified, but the estimated cohort and referral-cohort coefficients are very similar to the results in Column 1.

The within-*jati* estimates allow ability to vary across *jatis* but assume that ability is constant over time (both within and across generations). The level of parental schooling could, however, also depend on the access to education, which might have changed over time. If there was convergence in the access to education across *jatis* in the parent generation, then that could explain the positive referrals-cohort coefficient in Columns 1-2 without requiring networks to be active. One test to rule out this alternative interpretation of our result would be to estimate the school selectivity regression for girls rather than boys; we have already seen that the network has no effect on schooling choice for the girls, and so both the cohort and the referrals-cohort effect should be absent. In contrast, if the referrals-cohort term is picking up convergence in (fathers') schooling levels across *jatis*, then this coefficient should be positive and significant for the girls as well.

Table 10, Column 3 reports the basic selectivity regression for the girls attending Marathi schools with cohort, referrals-cohort, and referrals included as determinants of father's schooling, while Table 10, Column 4 repeats this regression with *jati* fixed effects. The referrals coefficient in Column 3 is again negative (but insignificant), consistent with the lower levels of ability in high-referral *jatis*. The cohort coefficient is positive but insignificant. More importantly, the referrals-cohort coefficient is small in magnitude and statistically insignificant - the point estimate is actually negative - consistent with the results obtained earlier that girls in families belonging to high-referral *jatis* are not restricted in their mobility.

An alternative strategy to control for the confounding effect of changes in access to schooling among the fathers across *jatis* and over time pools boys and girls in the selectivity regression, which can then be estimated with a full set of *jati* dummies interacted with the cohort variable:

$$E(S_{ij} \mid E_{ij} = 0) = (\mu - \tilde{\mu})C_{ij} \cdot B_{ij} + (\nu - \tilde{\nu})R_j \cdot C_{ij} \cdot B_{ij} + f_j + g_j \cdot B_{ij} + h_j \cdot C_{ij} \tag{5}$$

where $\tilde{\mu}$, $\tilde{\nu}$ are the coefficients on the cohort variable and the referrals-cohort interaction for the girls, and B_{ij} is a boy dummy as before. The fixed effects, f_j which subsume $\tilde{\kappa} + \tilde{\lambda}R_j + \psi\omega_j$, allow for the possibility that ability varies across *jatis*. The fixed effects interacted with the boy dummy $g_j \cdot B_{ij}$, which subsume $(\kappa - \tilde{\kappa})B_{ij} + (\lambda - \tilde{\lambda})R_j \cdot B_{ij}$, also allow ability to vary by gender across *jatis*. Finally, the fixed effects interacted with the cohort variable $h_j \cdot C_{ij}$, subsume $\tilde{\mu}C_{ij} + \tilde{\nu}R_j \cdot C_{ij}$ and

control for changes in access to schooling for the fathers both across *jatis* and over time.

For the special case with $\tilde{\mu} = 0$, $\tilde{\nu} = 0$, as is consistent with the model, the estimated coefficients in equation (5) should match the cohort coefficient and the referrals-cohort coefficient when equation (4) is estimated with *jati* fixed effects for boys only. Table 10, Column 5 suggests that this is indeed the case; the cohort-boy coefficient is negative and significant and the referrals-cohort-boy coefficient is positive and significant, and the point estimates are very similar to the corresponding coefficients in Columns 1-2. These results confirm that in the most heavily-networked *jatis* high-ability girls were exiting to English-medium schools at significantly faster rates than were boys.[26] The 0.1 quantile - 0.9 quantile of the referrals distribution ranges from 0.2 to 0.7. The point estimates in Column 5 thus suggest that over the period of the 1990's the gap in father's schooling between boys and girls schooled in Marathi grew by 2.3 years in the highest-referral *jatis* (at the 0.9 quantile level). In contrast, the ability-differential measured by the difference in the father's schooling between boys and girls, declined by as much as 5 years in the low-referral *jatis* (at the 0.1 quantile level) over the same period. This increasing mismatch in ability levels between the sexes within *jatis* and school types could have important implications for the future stability of the caste system, which relies on endogamous marriage, as discussed below.

Table 10, Columns 6-10 reports the estimates of the selectivity equations for the first 10 cohorts of students, who entered school in the 1980s. Schooling choices were stable over this period and thus we do not expect to find changing selectivity effects for the boys or the girls. As before, the referrals coefficient, in Column 6 and Column 8, is negative and significant, reflecting the persistent differences in ability across *jatis*. However, as expected and in contrast to the cohorts making schooling choices in the post-1990's new economy, the cohort effect and the referrals-cohort effect, both uninteracted and interacted with the boy dummy, are insignificant in the pre-reform period. Disparities in ability across gender and caste groups within schools were evidently stable in the pre-reform period.

5.6 Alternative Interpretations of the Empirical Results

The simple model of schooling choice laid out in Section 3 generates two predictions when *jati*-based labor market networks are active. First, the occupational distribution at the level of the *jati* should persist across generations. Second, this intergenerational state dependence could persist even as the

[26]The negative cohort-boy coefficient implies that the boy-girl pre-school human capital differential is declining over time, independent of the influence of the male job network. This may be due to differences in labor force participation or changes in the returns to English by gender (as in Figures 1 and 2).

returns to alternative occupations grow, if restrictions on occupational mobility are in place. The analysis above provided empirical support for both these predictions.

The discussion that follows considers alternative explanations for these results. The identifying assumption in the fixed effects schooling choice regression is that unobserved determinants of schooling choice that are correlated with the occupational distribution should not vary by gender, or have a differential effect by gender on schooling choice, within the *jati*. Some of the alternative explanations that we pursue are associated with the failure of this identifying assumption. Other explanations are generated by relaxing the assumptions of the model. We will argue that none of these alternative explanations match all the results as well as our preferred explanation, based on underlying male labor market networks.

1. **Liquidity constraints**: The model assumes that schooling choices are based entirely on the individual's ability and the historical occupational distribution in his *jati*, which determines the labor market network that he inherits. When credit markets function imperfectly, liquidity constraints could in addition prevent individuals from choosing more expensive English schooling. For example, households from working class *jatis* with lower incomes might lack the resources to send their children to English schools. When credit networks are active, access to credit will depend on (average) income in the *jati*, generating an alternative link between the occupational distribution in the *jati* and schooling choice.

The schooling regression accounts for liquidity constraints by including family income at the time the child entered school. Schooling expenses for the children in our sample are relatively low (6.3 % of family income in English schools and 6.0% in Marathi schools), and not surprisingly family income has a relatively weak effect on schooling choice for both boys and girls. However, because liquidity does not vary by the gender of the child *within* the *jati*, the *jati* fixed- effects regression would appear to be robust to one channel through which the identifying assumption could be violated.

Boys and girls have different labor market opportunities. It is thus conceivable that liquidity could have a differential effect on schooling choice by gender. The estimates of family income income effects on school choice, however, do not differ from each other in Table 6 and Table 7 at conventional levels of significance. If statistical significance is ignored, then the household income estimates imply that the liquidity effect is smaller for boys than for girls. If so, then, given that high-referral *jatis* have lower average incomes in the data, the referrals-boy coefficient will be biased upward when *jati*-level

credit is omitted from the schooling regression. This implies, that unobserved liquidity constraints cannot spuriously generate the negative referrals-boy coefficient in Table 6 and Table 7.

2. **Differences in ability or preferences**: The model assumes that pre-school human capital does not vary across *jatis*. But the schooling regressions, which include parents' years of education to control for the child's pre-school human capital, indicate that this modelling assumption might not be appropriate. Again, the *jati*-level fixed effects absorb all variation in the *jati* that is not gender specific. In an economy where men and women historically performed very different roles, however, the parental and societal inputs that boys and girls received in childhood might have been very different, and such differences might vary by household or *jati* environment.

The estimates do not provide evidence of gender distinctions in pre-school human capital within households or *jatis*. The estimates reported in Table X do not reject the hypothesis that the effects of parental human capital characteristics on school choice are equal for boys and girls. The results reported in Table 9 with test performance as the dependent variable are also consistent with the assumption in the fixed effects schooling regression that pre-school human capital does not vary by gender within the *jati*. Recall that higher referrals were associated with lower test scores for both boys and girls, but that referrals did not have a differential effect on test scores by gender *within* the *jati*. Assume, as before, that test scores depend only on individual pre-school human capital and school quality. We have seen, however, based on the school survey statistics in Table 1 and the household survey evidence on motives for school choice that English and Marathi schools are of similar quality. These results together imply that high referral (working class) *jatis* have lower pre-school human capital, but that this pre-school human capital does not vary by gender within the *jati*. The fact that referrals have a differential effect on schooling choice is irrelevant for test scores, if school quality does not vary by the language of instruction.

Students who expect to enter working class occupations in the future could conceivably put less effort in their studies. If we relax the assumption that test scores depend on pre-school human capital and school quality alone, and allow effort to play a role as well, then both pre-school human capital and effort could vary by gender within the *jati*, as long as these competing effects just cancel each other. But knife-edge results of this sort are generally difficult to justify.[27]

[27]We could relax the assumption that school quality is the same in English and Marathi schools as well, in which case school quality, effort, and pre-school human capital could all vary by gender within the *jati*, and just cancel out. This knife-edge result, involving variation along three dimensions, is even more difficult to justify.

The discussion this far has assumed that the individual is endowed with a particular level of pre-school human capital, which might vary across *jatis* and which independently determines schooling choice. Although the model rules this out, preferences for schooling or future career choices could also be determined endogenously within the *jati*. For example, an individual belonging to a *jati* that has traditionally channelled its members into the civil service might well place symbolic value on that profession and end up becoming a bureaucrat himself. Similarly, social interactions within the *jati* could lead individuals to make similar schooling choices, and hence career choices, across generations.

An instrumental interpretation of the symbolic decisions and social interactions described above is that they preserve the integrity of the network by effectively restricting occupational mobility. Indeed, these preferences could persist long after the network that gave rise to them has ceased to serve its originally intended purpose. We will not attempt to distinguish between preferences that are complementary to the network, and the network itself, in this paper. Any preferences that emerge for other reasons and that do not vary by gender will be subsumed in the *jati* fixed effects.

3. **Discrimination**: The model assumes that individuals with the same ability whatever their caste or gender are face the same school and labor market parameters determining their schooling and earnings. In practice, schools and employers could discriminate against individuals from particular socioeconomic backgrounds. Unless there is a gender-based component to this discrimination, it will be subsumed entirely by the *jati* fixed effects. But it is possible that firms or schools discriminate against boys from working class backgrounds, perhaps because they are difficult to discipline, while treating girls from different backgrounds more equally. The referrals-boy coefficient would proxy for underlying discrimination in that case.

The question is whether this discrimination is based on the individual's family background, which will of course be correlated across members of the *jati*, or whether it is based on the individual's *jati* alone. Recall that household characteristics, such as parental education and family income, had the *same* effect on schooling choice for boys and girls within the *jati*. It was only the *jati*-level referrals statistic that had a gender-specific effect on schooling. If these results are attributed to discrimination, then it implies that firms or schools do not discriminate by family background *within* the *jati*, but by *jati* alone.

It is not obvious why we would expect to see discrimination purely along caste lines. Family characteristics, such as parental education and income, were seen to be correlated with pre-school

human capital and are at least as easy to observe as caste identity. Historically there does not appear to have been a policy of caste discrimination by employers in any industry in Bombay. The jobber was most likely responsible for the specific caste concentrations that emerged in individual manufacturing units in the city (Morris 1965).

4. **Restrictions on high-caste women**: We focused in Figure 3 and Figure 4 on the absence of convergence for the boys, which was attributed to restrictions on occupational mobility among the lower castes. An alternative interpretation of these Figures is that the ability distribution varies across the population such that it remains optimal for individuals to sort by caste into different careers even as the returns to English grow. The convergence among the girls with this alternative interpretation is attributed to restrictions on the *high-caste* girls.

There is no evidence that such restrictions are in place, or have been in place historically. High-caste women have always had higher labor-force participation rates and more English schooling than lower-caste women in Bombay, as observed in Table 3 and Table 5. Within the high castes, boys are substantially more likely to be schooled in English than girls, and so the girls could easily switch into English schools without creating a mismatch on the marriage market.

Moreover, although the pooled schooling choice regression with fixed effects cannot distinguish between the alternative explanation, based on female restrictions, and our view that male networks shape schooling choice for the boys alone, recall that we also ran regressions separately for boys and girls. The *jati*-level statistic, measured either by the proportion of referrals or the proportion of fathers with English schooling, affects schooling choice for the boys but not for the girls in these regressions. We thus appear to be picking up restrictions on mobility that are specific to the boys.

5. **Sampling bias**: We noted three potential sources of sampling bias in Section 4.1. First, particular households might school their children outside the Dadar area. Second, particular households might have moved from Dadar over the past 20 years. Third, children from particular households might have dropped out of school.

The first two sources of sampling bias are easily accommodated in the fixed effects regression framework. Although school locations and out-migration might vary by *jati*, there is no reason to expect these decisions to vary by the gender of the child *within* the *jati*. The third source of sampling bias is potentially more problematic, because drop-outs could vary by gender within the *jati*. The decision to drop out would depend on the child's pre-school human capital and future employment opportu-

nities, both of which determine schooling choice. Selective drop outs, by gender across *jatis*, could consequently violate the identifying assumption underlying the fixed-effects estimation procedure.

However, the sex-ratio of students in the most recent eight cohorts (grades one through eight) in which there would be relatively few drop-outs is statistically indistinguishable from the sex-ratio in the older 12 cohorts. Regressions not reported here also reveal that the sex-ratio is uncorrelated with the level of referrals in the *jati*, both in the first 12 cohorts and in the eight most recent cohorts.

6 Conclusion

As modernization proceeds around the world, there is a perception that indigenous pre-existing institutions may importantly shape the course of the development process across different countries. Yet little is known about how such institutions actually affect the transformation of economies undergoing change or their impact on the economic mobility of particular groups of individuals. This paper examines the role of one long-standing traditional institution - the Indian caste system - in shaping career choices by gender in a rapidly globalizing economy.

We have found that male working class networks, organized at the level of the sub-caste or *jati*, continue to channel boys into traditional occupations despite the fact that returns to non-traditional (white collar) occupations have risen substantially during the post-1990s reform period. In contrast, girls, who have had historically low labor-market participation rates and few network ties to constrain them, appear to be taking full advantage of the opportunities that have become available in the new economy. It is generally believed that the benefits of globalization have accrued disproportionately to the elites in developing countries. In this setting we find instead that a previously disadvantaged group (girls) might surpass boys in educational attainment and employment outcomes in the future in the most heavily networked *jatis*.

Although we have focused on how traditional institutions shape the responses of particular groups of individuals to the new opportunities that accompany globalization, our findings suggest that these institutions are likely to be affected in turn by the forces of change. In our framework, an individual schooled in English no longer needs the traditional caste network; indeed, it has been remarked that "the English educated form a caste by themselves" (M.P. Desai, 1952, quoted in Dakin, Tiffin and Widowson, 1968: 24). Simple statistics on marriage and migration that we computed for the elder siblings of the students in our sample would appear to support the view that English education will

ultimately undermine the caste network. Among the 825 married siblings in our sample, 11.9% married outside their *jati*. This contrasts with the parent generation in which only 3.6% of the partners were not members of the same *jati*. Schooling in English appears to be contributing to this increase in inter-caste marriage, as 31.6% of the English-educated married outside their *jati* versus only 9.7% of the Marathi-educated. And among the 1073 siblings who are currently employed, 13.9% of the English-educated work outside Maharashtra versus only 2.1% of the Marathi-educated (these differences between the Marathi educated and the English educated are statistically significant at the 5% level). Both marriage outside the *jati* and out-migration weaken caste ties and the caste network. Increasing exposure to the modern economy through English education, and the mismatch in educational choices and future occupational outcomes between boys and girls in the same *jati* that we have documented, suggest that the forces of modernization could ultimately lead to the disintegration of a system that has remained firmly in place for thousands of years.

7 Appendix

To estimate the returns to schooling attainment and medium of instruction, separately for men and women in each of the five time periods, we used a specification of the wage regression given by:

$$\ln(W_{it}) = \alpha_t E_i + \beta_t S_i + \gamma A_{it} + \delta_t + \epsilon_{it}$$

where W_{it} is individual i's real income in period t (nominal income normalized by the consumer price index in that year), E_i, S_i measure the language of instruction and the years of schooling for that individual, A_{it} measures his or her age in that period, δ_t is a set of time period dummies, and ϵ_{it} is a mean-zero disturbance term. α_t, β_t measure the returns to English and the returns to education in each period, and δ_t measures secular changes in earnings over time. Table A1 reports the estimated coefficients that are displayed in Figures 1 and 2. As can be seen, the coefficient on the years of schooling is very precisely estimated in each time period for both men and women. The coefficient on English schooling is also significant at the 5 percent level in each period except 1980 for the men and 1980 and 1985 for the women. An F-test rejects the hypothesis that the English coefficients are the same across the times periods. The time period dummies, which reflect the returns to Marathi schooling once the English schooling dummies are included, are small and insignificant in the 1980s, but significant and increasingly negative in the 1990s. Because Marathi schooling channels the individual into working class jobs, this result suggests that working class wages declined over the 1990s in real terms.

To describe the dynamics of schooling choice nonparametrically in Figures 3 and 4, while taking into account intergenerational state dependence within the family, we ran a first stage parametric regression of schooling choice on the cohort variable, interacted with the broad caste categories, as well as the father's and the mother's language of education in secondary school. Less than 3% of parents in the sample had gone to an English medium school, although this figure was above 12% in the high castes. We constructed the cohort variable based on the year in which the child entered school; for example, cohort=1 for individuals who are currently 25 years old and entered school in 1982, cohort=20 for six year olds who entered school in 2001. This first stage regression is reported in Table A2. The dependent variable in the second stage nonparametric regression is schooling choice net of the estimated parental education effect from the first stage (this two-step procedure is based on Porter 1996). Figure 3 and Figure 4 indicate that the cohort effects might be nonlinear. We experimented

with a more flexible first-stage regression that included cohort-squared, without changing the patterns reported in the Figures. The advantage of the parsimonious specification reported in Table A2 is that we can verify statistically that schooling choice is converging for the girls but not for the boys.

References

[1] Acemoglu, Daron, Simon Johnson and James Robinson. 2001. The Colonial Origins of Comaprative Development: An Empirical Investigation, *American Economic Review*, December, Vol. 91, p. 1369-1401.

[2] Banerjee, Abhijit and Laksmi Iyer (2002). History, Institutions and Economic Performance: The Legacy of Colonial land Tenure Systems in India, mimeo, MIT.

[3] Banerjee, Abhijit and Kaivan Munshi (forthcoming). How Efficiently is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur, *Review of Economic Studies*.

[4] Behrman, Jere R.(1997). Intrahousehold Distribution and the Family. *Handbook of Population and Family Economics*, Mark R. Rosenzweig and Oded Stark, eds. Amsterdam: North-Holland.

[5] Burnett-Hurst, A.R. 1925. *Labour and Housing in Bombay: A Study in the Economic Conditions of the Wage-Earning Classes of Bombay*. London: P.S. King and Son.

[6] Chandavarkar, Rajnarayan. 1994. *The Origins of Industrial Capitalism in India: Business strategies and the working classes in Bombay, 1900-1940*. Cambridge: Cambridge University Press.

[7] Cholia, R.P. 1941. *Dock Labourers in Bombay*. Bombay.

[8] Dakin, Julian, Brian Tiffen, and H.G. Widdowson. 1968. *Language in Education: The Problem in Commonwealth Africa and the Indo-Pakistan Sub-continent*. Oxford: Oxford University Press.

[9] Dandekar, H.C. 1986. *Men to Bombay, Women at Home: Urban Influence on Sugao Village, Deccan Maharashtra, India, 1942-1982*. Ann Arbor, Michigan.

[10] D'Monte, Darryl. 2002. *Ripping the Fabric: The Decline of Mumbai and its Mills.* New Delhi: Oxford University Press.

[11] Dobbin, Christine. 1972. *Urban Leadership in Western India: Politics and Communities in Bombay City, 1840-1885*. London: Oxford University Press.

[12] Fafchamps, Marcel and Susan Lund (2000). Risk-Sharing Networks in Rural Philippines, mimeo, Oxford University.

[13] Fafchamps, Marcel (2001). Networks, Communities and Markets in Sub-Saharan Africa: Implications for Firm Growth and Investment. *Journal of African Economies*. Vol. 10 (0). p 109-42. Supplement 2.

[14] Gokhale, R.G. 1957. *The Bombay Cotton Mill Worker*. Bombay.

[15] Gore, M.S. 1970. *Immigrants and Neighborhoods: Two Aspects of Life in a Metropolitan City*. Tata Institute of Social Sciences, Bombay.

[16] Jacoby, Hanan G and Emmanuel Skoufias (1997). Risk, Financial Markets, and Human Capital in a Developing Country. *The Review of Economic Studies*, Vol. 64, No. 3.

[17] Kamat, A.R. 1985. *Education and Social Change in India*. Bombay: Somaiya Publications Private Limited.

[18] Katzenstein, Mary Fainsod. 1979. *Ethnicity and Equality: The Shiv Sena Party and Preferential Policies in Bombay*. Ithaca: Cornell University Press.

[19] Kochar, Anjini. 2002. Schooling, Wages and Profits: Some Negative Pecuniary Externalities from the Spread of Schooling in Rural India and Their Consequences for Schooling Investments, mimeo, Stanford University.

[20] Kohn, Melvin L. 1977. *Class and Conformity*. Chicago: The University of Chicago Press.

[21] Komarovsky, Mirra. 1967. *Blue-Collar Marriage*. New York: Random House, Inc.

[22] Morris, Morris David. 1965. *The Emergence of an Industrial Labor Force in India: A Study of the Bombay Cotton Mills, 1854-1947*. Berkeley: University of California Press.

[23] Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *Quarterly Journal of Economics*, Vol. 118(2), pp. 549-597.

[24] Nurullah, Syed and J.P. Naik. 1949. *A Student's History of Education in India, 1800-1947*. Bombay: Macmillan and Company Limited.

[25] Patel, Kunj. 1963. *Rural Labor in Industrial Bombay*. Bombay: Popular Prakashan.

[26] Porter, Jack. 1996. *Essays in Econometrics*, unpublished M.I.T. Ph.D. dissertation.

[27] Rees, Albert. 1966. "Information Networks in Labor Markets." *American Economic Review (Papers and Proceedings).* 56, pp. 559-66.

[28] Townsend, Robert (1994). "Risk and Insurance in Village India," *Econometrica*, Vol. 62 (3), pp. 171-84.

Figure 1: Returns to English and Schooling by Year, 1980-2000 - Men Aged 30-55

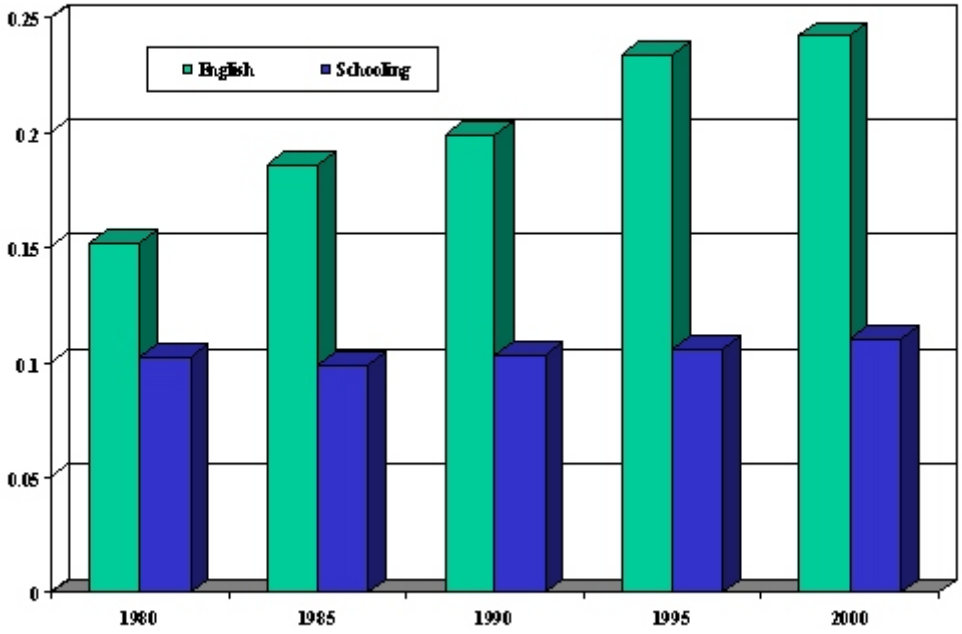

Figure 2: Returns to English and Schooling by Year, 1980-2000 - Women Aged 30-55

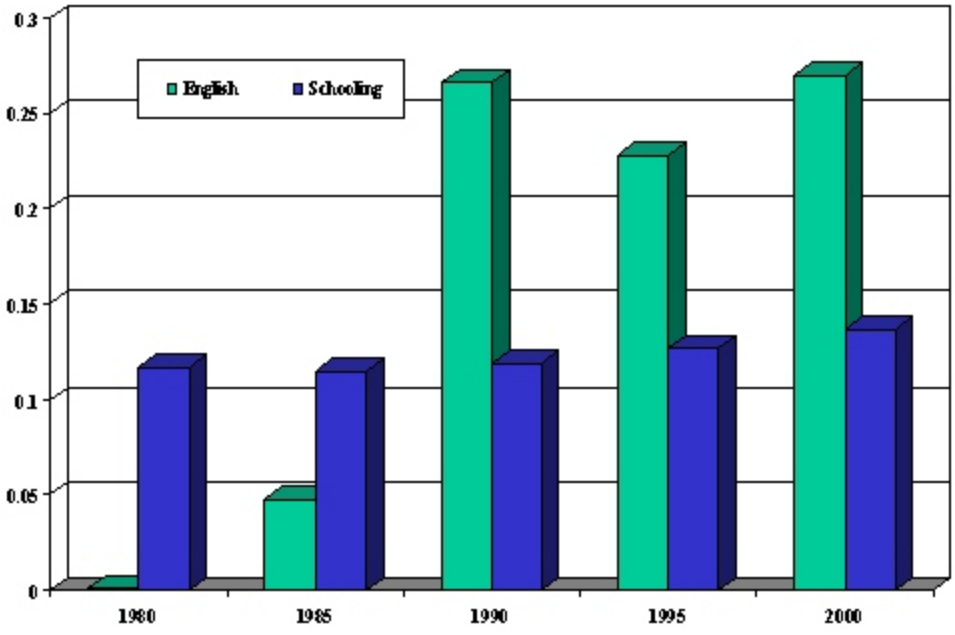

Figure 3: English Schooling - net parental education effect - Boys

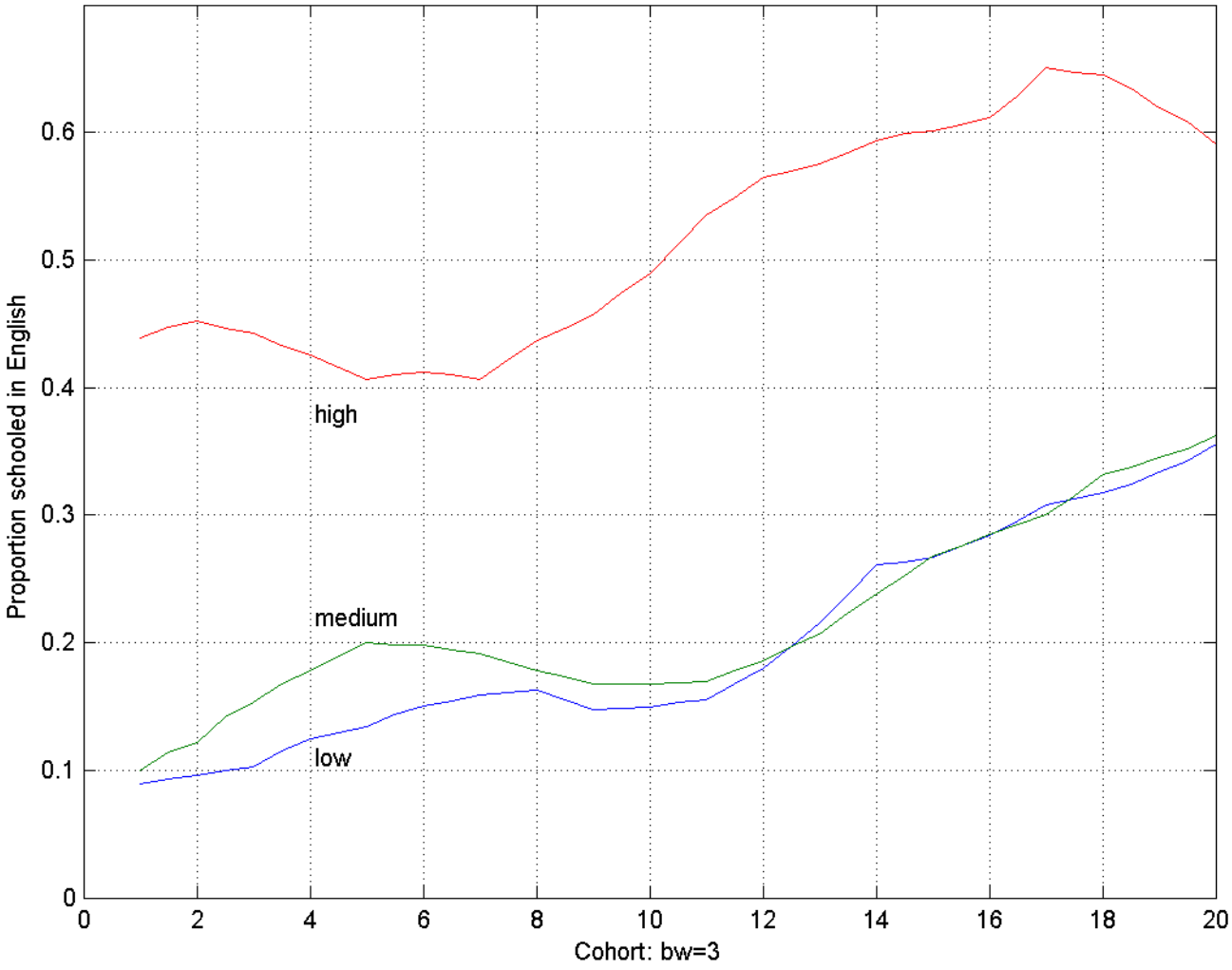

Figure 4: English Schooling - net parental education effect - Girls

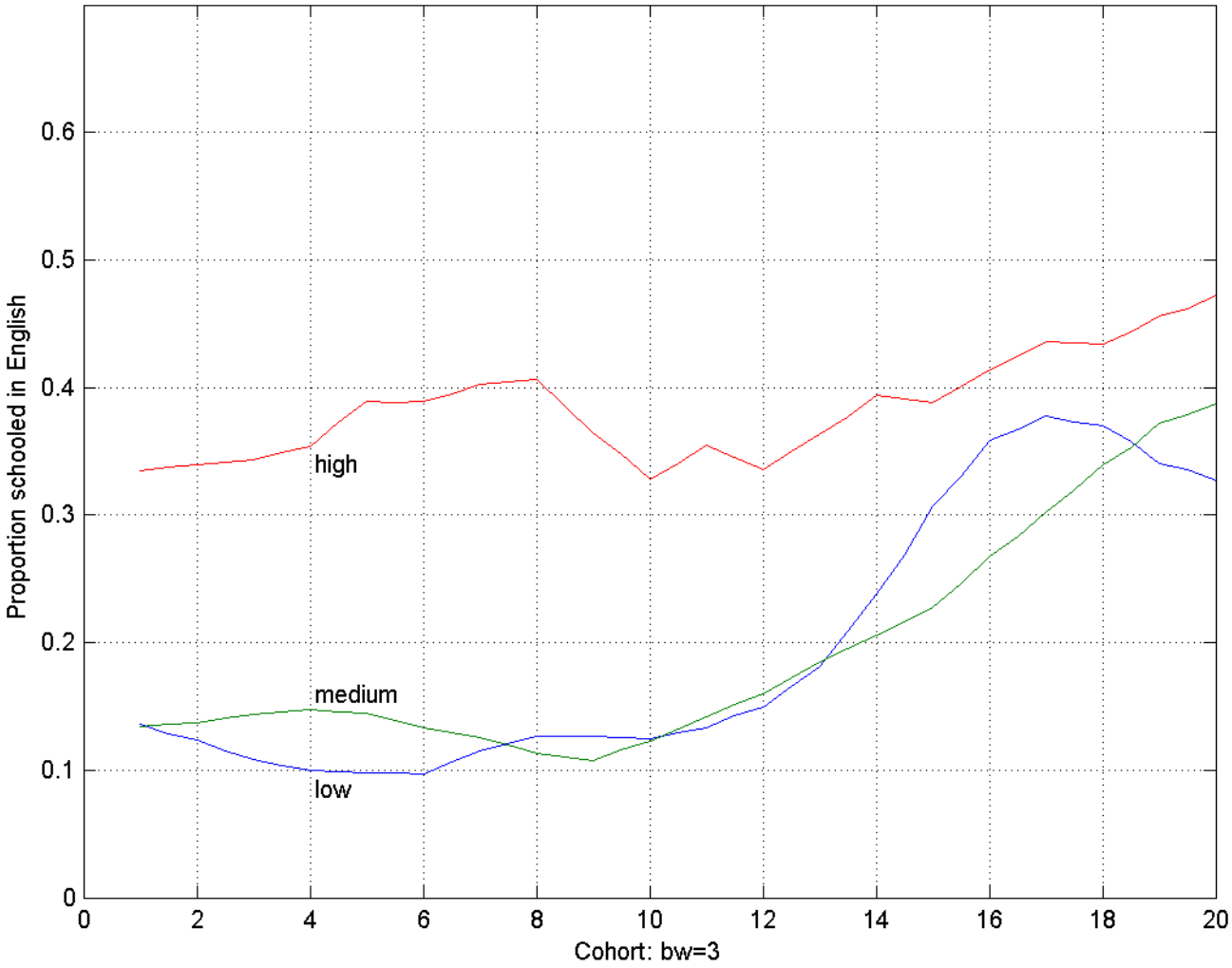

Table 1: School Characteristics and Student Performance

School type:	English medium	Marathi medium
	(1)	(2)
Panel A: School characteristics (secondary section)		
Student-teacher ratio	36.71	35.76
	(2.40)	(2.17)
Class size	61.90	62.28
	(3.69)	(3.15)
Students per desk	2.40	2.36
	(0.10)	(0.11)
Proportion of teachers with B.Ed.	0.72	0.70
	(0.07)	(0.05)
Proportion of teachers with higher degree	0.08	0.10
	(0.03)	(0.03)
Computers per student	0.02	0.02
	(0.004)	(0.005)
Student enrollment in secondary section	1528.40	1059.00
	(360.64)	(175.73)
Panel B: School expenses		
Fees	0.48*	0.20*
	(0.01)	(0.01)
Other expenses	1.10*	0.71*
	(0.04)	(0.01)
Panel C: S.S.C. school leaving exam results (1997-2001)		
Percentage passed	92.59*	51.62*
	(2.04)	(5.95)
Percentage first class among passed	36.2*	24.23*
	(1.69)	(3.35)
Percentage distinction among passed	23.94*	6.90*
	(3.92)	(1.87)
Number of schools	10	18

Note: standard errors are in parentheses.

* denotes rejection of the equality of means for the two types of schools with greater than 95% confidence.

Panels A and C use data from the school survey, Panel B uses data from the household survey.

School expenses are measured for the past year (2000-2001) in thousands of 1980 Rupees.

To convert to 2000 Rupees, multiply by 4.44.

Other school expenses include transportation, coaching classes, textbooks, uniforms, and stationary.

Scores above 35% are required to pass S.S.C., scores above 60% are required for first class, and above 75% for distinction.

Table 2: Occupational Choice by Caste - Men

Relationship to student:	fathers			grandfathers		
Caste:	low	medium	high	low	medium	high
	(1)	(2)	(3)	(4)	(5)	(6)
Employment (%)	97.37	97.31	99.06	98.87	98.86	99.28
Occupational Distribution (%)						
Unskilled manual	11.09	7.84	4.41	9.00	3.63	2.10
Skilled manual	17.35	13.70	10.21	11.67	6.72	8.42
Organized blue-collar	22.87	19.22	2.90	22.89	24.23	7.67
Petty trade	4.00	4.51	2.52	3.11	3.20	3.34
Clerical	28.09	36.64	20.81	22.22	23.79	28.84
Business	7.95	8.79	15.51	6.11	4.72	13.00
Professional	8.30	8.79	43.51	5.56	6.18	33.66
Farming	0.35	0.51	0.13	19.44	27.53	2.97
Percent working class	53.64	42.91	18.01	56.24	49.92	19.42
(standard error)	(1.18)	(1.17)	(1.73)	(1.29)	(1.34)	(1.75)
Total number of observations	1,860	1,774	793	1,866	1,934	839

Note: occupational distribution is computed using employed individuals only.
Employment for fathers is computed as of 1995. Statistics in Columns 4-6 are reported for paternal grandfathers.
Working class = 1 if unskilled manual, skilled manual, organized blue-collar, 0 if clerical, business, professional.

Occupational categories

Unskilled manual : daily wage labor, deliveryman, servant, hotel worker, helper, cleaner/sweeper, porter, assistant, watchman, fisherman, gardener, barber, cobbler (chambhar), unskilled laborer, seaman.

Skilled manual : machine operator, plumber, welder, technician, electrician, mechanic, carpenter, fitter/turner, tailor, painter, film developer, goldsmith, artist, priest, lab assistant, skilled worker, traditional healer (vaidhya), computer operator.

Organized blue collar : mill worker, factory worker, peon, Bombay Port Trust (BPT) worker, Bombay Electric Supply and Transportation (BEST) worker, Bombay Municipal Corporation (BMC) worker.

Petty trade : hawker, storeman (storekeeper), salesman, agent, shopkeeper.

Clerical : supervisor, driver, police, clerk, conductor, stenographer, postmaster, receptionist, foreman/draftsman, secretary.

Business : self business, medical representative, transporter, marketing, consultant, employer, contractor, politician (social worker/leader), merchant.

Professional : tuitions, teacher, programmer, engineer, officer, manager, doctor, lawyer, nurse, lecturer, vice-chancellor, librarian, superintendant, director, principal, architect, salaried employee (service), chartered accountant, big business

Farming : farmer, agricultural laborer.

Table 3: Occupational Choice by Caste - Women

Relationship to student:	mothers			grandmothers		
Caste:	low	medium	high	low	medium	high
	(1)	(2)	(3)	(4)	(5)	(6)
Employment (%)	20.56	20.01	51.23	19.31	18.59	15.57
Occupational Distribution (%)						
Unskilled manual	29.95	16.94	2.36	24.65	7.18	3.13
Skilled manual	8.82	8.47	6.15	1.70	1.44	3.13
Organized blue-collar	4.01	4.92	0.47	8.50	4.31	0.78
Petty trade	3.74	3.83	1.18	1.13	0.57	0.00
Clerical	31.55	40.71	46.34	4.25	2.30	19.53
Business	4.55	2.46	3.78	2.27	1.44	3.91
Professional	17.38	22.68	39.72	5.10	8.62	67.97
Farming	0.00	0.00	0.00	52.41	74.14	1.56
Percent working class	44.44	31.53	9.09	75.00	51.14	7.14
(standard error)	(2.22)	(2.25)	(2.06)	(3.19)	(4.36)	(3.64)
Total number of observations	1,887	1,954	857	1,885	1,953	854

Note: occupational distribution is computed using employed individuals only.

Employment for mothers is computed as of 1995. Statistics in Columns 4-6 are reported for maternal grandmothers.

Working class = 1 if unskilled manual, skilled manual, organized blue-collar, 0 if clerical, business, professional.

Occupational categories were defined in Table 2.

Table 4: Referrals and Schooling by Occupation

Relationship to student:	father		mother	
Outcomes and choices:	percentage that received referrals	percentage that studied in English	percentage that received referrals	percentage that studied in English
	(1)	(2)	(3)	(4)
Occupation				
Unskilled manual	65.95	0.80	61.29	0.00
Skilled manual	60.13	2.24	45.56	5.56
Organized blue-collar	76.43	0.91	69.44	5.56
All working class	68.44	1.36	57.69	2.24
(standard error)	(1.11)	(0.28)	(2.80)	(0.84)
Petty trade	57.89	1.75	61.76	2.94
Clerical	47.41	2.89	30.56	7.26
Business	49.29	8.53	41.86	9.30
Professional	32.77	11.38	29.25	14.47
All white-collar	43.76	6.20	30.64	10.13
(standard error)	(1.02)	(0.49)	(1.60)	(1.05)
Number of observations	4,515	4,513	1,215	1,215

Note: statistics are computed using employed individuals only. Farmers are also excluded.

A parent is said to have received a referral if a relative or member of the community found him/her a job.

A parent is said to have studied in English if he/she studied in that language in secondary school.

Occupational categories were defined in Table 2.

Table 5: Referrals, English Schooling and Parental Characteristics by Caste

Relationship to student:	father			mother		
Caste:	low	medium	high	low	medium	high
	(1)	(2)	(3)	(4)	(5)	(6)
Percentage that received referrals	60.58	58.02	37.25	13.50	12.13	19.16
	(1.14)	(1.13)	(1.70)	(0.80)	(0.79)	(1.19)
Percentage that studied in English	2.16	1.90	12.47	1.35	1.42	12.93
	(0.44)	(0.44)	(0.66)	(0.42)	(0.41)	(0.62)
Years of schooling	9.63	10.22	13.82	8.03	8.73	13.49
	(0.08)	(0.08)	(0.11)	(0.09)	(0.09)	(0.13)
Monthly income	1.92	1.99	4.61	0.23	0.30	1.37
	(0.08)	(0.08)	(0.12)	(0.03)	(0.03)	(0.04)
Number of observations	1,852	1,896	835	1,852	1,896	835

Note: Standard errors are in parentheses.

Statistics are computed using all parents in the sample, regardless of whether they are employed or not.

Monthly income is measured in thousands of 1980 Rupees, in the year that most closely corresponds to the year in which the child entered school.

A parent is said to have received a referral if a relative or member of the community found him/her a job.

A parent is said to have studied in English if he/she studied in that language in secondary school.

Table 6: Caste-based Networks and Schooling Choice

Dependent variable:	English schooling							
Network boundary:	jati						jati-region	
Sample:	boys only		girls only		boys and girls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Referrals	-1.060	-0.377	-0.646	0.124	--		--	--
	(0.164)	(0.148)	(0.160)	(0.167)				
Referral - boy	--	--	--	--	-0.398	-0.486	-0.389	-0.458
					(0.091)	(0.104)	(0.079)	(0.083)
Cohort	0.013	0.009	0.013	0.009	0.017	0.019	0.016	0.018
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Cohort-boy	--	--	--	--	-0.002	-0.003	-0.002	-0.002
					(0.002)	(0.002)	(0.002)	(0.003)
Boy	--	--	--	--	0.270	0.331	0.265	0.327
					(0.049)	(0.059)	(0.049)	(0.059)
Father studied in English	0.320	0.236	0.388	0.309	--	--	--	--
	(0.037)	(0.033)	(0.037)	(0.026)				
Mother studied in English	0.351	0.220	0.441	0.269	--	--	--	--
	(0.041)	(0.028)	(0.071)	(0.045)				
Father's years of education	--	0.023	--	0.020	--	--	--	--
		(0.004)		(0.003)				
Mother's years of education	--	0.023	--	0.026	--	--	--	--
		(0.003)		(0.003)				
Family income	--	0.005	--	0.009	--	--	--	--
		(0.005)		(0.003)				
Occupation dummies	No	No	No	No	No	Yes	No	No
R^2	0.173	0.274	0.146	0.272	0.163	0.265	0.200	0.228
Number of observations	2,405	2,286	2,228	2,093	4,635	4,414	4,635	2,622

Standard errors are in parentheses.

Standard errors are robust to heteroscedasticity and clustered residuals within each jati or jati-region.

English schooling = 1 if the child is sent to an English school, 0 if he/she is sent to a Marathi school.

Referrals are measured at the level of the jati in Columns 1-6, by jati and region in Columns 7-8. There are 59 jatis and 4 regions (Bombay, Konkan, Deccan, Other).

Referrals measures the proportion of fathers in the social unit who received a referral.

Boy = 1 if the student is a boy, 0 if girl.

Family income is measured in thousands of 1980 Rs. in the year that most closely corresponds to the year in which the child entered school.

Column 6 includes a full set of 90 father's occupation dummies.

Column 1-2: schooling choice for boys.

Columns 3-4: schooling choice for girls.

Columns 5-8: schooling choice for both boys and girls. Use both jati and jati-region as the social unit. Social unit dummies are included in all these regressions.

Column 8 excludes very small (less than 10 observations) and very large (more than 250 observations) social units.

Table 7: Within-Jati Schooling Choice, Including Parental Characteristics

Dependent variable:	English schooling	
Additional regressors:	parents' language of schooling	parents' language of schooling + measures of ability
	(1)	(2)
Referral - boy	-0.348	-0.464
	(0.089)	(0.105)
Cohort	0.014	0.010
	(0.002)	(0.002)
Father studied in English	0.361	0.301
	(0.034)	(0.026)
Mother studied in English	0.389	0.259
	(0.059)	(0.043)
Father's years of education	--	0.021
		(0.003)
Mother's years of education	--	0.024
		(0.003)
Family income	--	0.007
		(0.003)
Boy	0.232	0.297
	(0.051)	(0.077)
Cohort-boy	-0.0005	-0.001
	(0.002)	(0.002)
Father studied in English - boy	-0.080	-0.091
	(0.039)	(0.044)
Mother studied in English - boy	-0.069	-0.044
	(0.053)	(0.042)
Father's years of education - boy	--	0.002
		(0.005)
Mother's years of education - boy	--	-0.001
		(0.004)
Family income - boy	--	-0.003
		(0.005)
R^2	0.207	0.299
Number of observations	4,633	4,379

Standard errors are in parentheses.

Standard errors are robust to heteroscedasticity and clustered residuals within each jati.

English schooling = 1 if the child is sent to an English school, 0 if he/she is sent to a Marathi school.

Referrals measures the proportion of fathers in the jati who received a referral.

Boy = 1 if the student is a boy, 0 if girl.

Family income is measured in thousands of 1980 Rs.

Sample includes boys and girls.

Table 8: Schooling Choice over Time

Dependent variable:	English schooling			
Additional regressors:	without family characteristics	with family characteristics	without family characteristics	with family characteristics
	(1)	(2)	(3)	(4)
Referral - boy	-0.426	-0.478	--	--
	(0.090)	(0.106)		
Referral-boy-cohort1	--	--	-0.269	-0.416
			(0.168)	(0.167)
Referral-boy-cohort2	--	--	-0.352	-0.333
			(0.100)	(0.112)
Referral-boy-cohort3	--	--	-0.523	-0.540
			(0.145)	(0.143)
Referral-boy-cohort4	--	--	-0.607	-0.665
			(0.256)	(0.238)
Cohort 1	-0.261	-0.161	-0.261	-0.161
	(0.031)	(0.032)	(0.030)	(0.032)
Cohort 2	-0.231	-0.146	-0.231	-0.146
	(0.031)	(0.028)	(0.031)	(0.028)
Cohort 3	-0.161	-0.121	-0.161	-0.121
	(0.030)	(0.023)	(0.030)	(0.023)
Boy	0.236	0.261	0.338	0.364
	(0.065)	(0.091)	(0.156)	(0.149)
Cohort 1 - boy	0.033	0.031	-0.152	-0.106
	(0.038)	(0.037)	(0.209)	(0.169)
Cohort 2 - boy	0.052	0.031	-0.090	-0.153
	(0.042)	(0.035)	(0.174)	(0.151)
Cohort 3 - boy	0.041	0.041	-0.007	-0.030
	(0.032)	(0.024)	(0.117)	(0.114)
R^2	0.164	0.301	0.164	0.301
Number of observations	4,635	4,379	4,635	4,379

Standard errors are in parentheses.
Standard errors are robust to heteroscedasticity and clustered residuals within each jati.
English schooling = 1 if the child is sent to an English school, 0 if he/she is sent to a Marathi school.
Referrals measures the proportion of fathers in the jati who received a referral.
Boy = 1 if the student is a boy, 0 if girl.
Cohort 1: age 21-25, Cohort 2: age 16-20, Cohort 3: age 11-15, Cohort 4: age 6-10.
Column 2 and Column 4 include family characteristics, separately and interacted with the boy dummy.
Family characteristics include parents' language of schooling and years of education, and total family income.
Jati dummies are included in all regressions. Sample includes boys and girls.

Table 9: Alternative Measures of the Occupational Distribution and Schooling

Dependent variable:	English schooling				test scores		
Occupational distribution measure:	proportion of fathers schooled in English				proportion of fathers that received a referral		
Sample:	boys only	girls only	boys and girls		boys only	girls only	boys and girls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Occupational distribution	0.847	0.083	--	--	-23.151	-23.650	--
	(0.262)	(0.427)			(5.045)	(4.080)	
Occupational distribution - boy	--	--	0.701	0.869	--	--	-0.734
			(0.224)	(0.221)			(5.761)
Cohort	0.008	0.010	0.017	0.010	-0.505	-0.180	-0.190
	(0.002)	(0.002)	(0.002)	(0.002)	(0.134)	(0.204)	(0.223)
Boy	--	--	0.026	-0.006	--	--	3.794
			(0.028)	(0.031)			(4.357)
Father studied in English	0.217	0.301	--	0.313	4.901	2.323	1.847
	(0.032)	(0.027)		(0.026)	(1.397)	(3.711)	(4.028)
Mother studied in English	0.220	0.266	--	0.259	3.312	-2.596	-2.772
	(0.030)	(0.046)		(0.042)	(2.200)	(1.905)	(1.632)
Father's years of education	0.024	0.019	--	0.020	0.929	0.765	0.812
	(0.004)	(0.002)		(0.003)	(0.195)	(0.225)	(0.238)
Mother's years of education	0.023	0.025	--	0.023	0.617	1.074	0.984
	(0.003)	(0.003)		(0.003)	(0.221)	(0.199)	(0.200)
Family income	0.005	0.008	--	0.007	0.260	0.122	0.107
	(0.004)	(0.003)		(0.003)	(0.118)	(0.076)	(0.076)
R^2	0.275	0.272	0.162	0.298	0.322	0.334	0.354
Number of observations	2,286	2,093	4,635	4,379	849	775	1,624

Standard errors are in parentheses. Standard errors are robust to heteroscedasticity and clustered residuals within each jati.

The sample in Columns 5-7 is restricted to children in cohorts 1-10, past the school-leaving age, who passed the S.S.C. exam. Test scores range from 35 to 100.

Boy = 1 if the student is a boy, 0 if girl.

Family income is measured in thousands of 1980 Rs.

Column 3 also includes cohort interacted with boy.

Column 4 and Column 7 also include cohort, father/mother studied in English, father's/mother's years of education, and family income, interacted with boy.

Table 10: Selection into Marathi Schools over Time

Dependent variable:	father's years of education									
Cohort:	11-20					1-10				
Sample:	boys		girls		boys and girls	boys		girls		boys and girls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Cohort	-0.697	-0.577	0.423	0.219	--	0.153	0.240	-0.543	-0.087	--
	(0.230)	(0.155)	(0.230)	(0.189)		(0.191)	(0.146)	(0.326)	(0.156)	
Cohort-boy	--	--	--	--	-0.792	--	--	--	--	0.117
					(0.252)					(0.232)
Referral-cohort	1.430	1.204	-0.596	-0.280	--	-0.147	-0.258	1.054	0.310	--
	(0.400)	(0.260)	(0.376)	(0.311)		(0.323)	(0.250)	(0.554)	(0.270)	
Referral-cohort-boy	--	--	--	--	1.469	--	--	--	--	-0.231
					(0.414)					(0.415)
Referrals	-30.991	--	-3.651	--	--	-10.256	--	-18.624	--	--
	(6.894)		(6.866)			(2.035)		(2.793)		
Jati dummies	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
R^2	0.106	0.205	0.138	0.205	0.215	0.136	0.254	0.184	0.278	0.285
Number of observations	839	839	815	815	1,654	866	866	851	851	1,717

Standard errors are in parentheses.

Standard errors are robust to heteroscedasticity and clustered residuals within each jati.

Referrals measures the proportion of fathers in the jati who received a referral.

Column 2, Column 4, Column 7, and Column 9 include jati dummies.

Column 5 and Column 10 include jati dummies, jati-boy dummies, and jati-cohort dummies.

All regressions are restricted to students in Marathi schools.

Table A1: The Returns to Schooling Attainment and English

Dependent variable:	log(income)	
Parent:	father	mother
	(1)	(2)
Age	0.017	0.030
	(0.001)	(0.004)
Period 2	0.048	0.077
	(0.058)	(0.197)
Period 3	-0.003	0.018
	(0.070)	(0.215)
Period 4	-0.293	-0.372
	(0.076)	(0.225)
Period 5	-0.495	-0.700
	(0.081)	(0.229)
Education - period 1	0.102	0.117
	(0.007)	(0.017)
Education - period 2	0.099	0.115
	(0.005)	(0.009)
Education - period 3	0.103	0.119
	(0.004)	(0.006)
Education - period 4	0.106	0.127
	(0.003)	(0.006)
Education - period 5	0.110	0.137
	(0.003)	(0.005)
English - period 1	0.152	-0.054
	(0.157)	(0.287)
English - period 2	0.186	0.048
	(0.093)	(0.140)
English - period 3	0.199	0.266
	(0.060)	(0.095)
English - period 4	0.234	0.228
	(0.051)	(0.084)
English - period 5	0.242	0.270
	(0.051)	(0.071)
R^2	0.285	0.419
Number of observations	13,638	3,068

Standard errors are in parentheses.
Standard errors are robust to heteroscedasticity and correlated residuals for each individual parent.
Income is measured in 1980 Rupes per year.
Period 1: 1980, Period 2: 1985, Period 3: 1990, Period 4: 1995, Period 5: 2000.
The sample is restricted to working individuals between the age of 30 and 55 in each period.
Education is measured by the number of years of schooling.
English = 1 if English was the medium of instruction in the parent's secondary school, 0 otherwise.

Table A2: The Dynamics of Schooling Choice by Caste

Dependent variable:	English schooling	
Sample:	boys	girls
	(1)	(2)
Cohort	0.014	0.005
	(0.004)	(0.003)
Cohort - low caste	0.0007	0.012
	(0.004)	(0.004)
Cohort - medium caste	-0.002	0.009
	(0.004)	(0.004)
Low caste	-0.306	-0.314
	(0.051)	(0.049)
Medium caste	-0.262	-0.283
	(0.051)	(0.050)
Father studied in English	0.315	0.396
	(0.045)	(0.048)
Mother studied in English	0.337	0.426
	(0.048)	(0.049)
R^2	0.170	0.151
Number of observations	2,405	2,228

Standard errors are in parentheses.
The broad caste categories are low, medium and high.