Title
Three Applied Essays in Development Economics

Author
Kapoor, Shruti

Publication Date
2009

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Three Applied Essays in Development Economics

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Shruti Kapoor

March 2010

Dissertation Committee:

Professor Anil Deolalikar, Co-Chairperson
Professor Aman Ullah, Co-Chairperson
Professor Jorge Agüero
Professor Mindy Marks

The Dissertation of Shruti Kapoor is approved:

Co-Chairperson

Co-Chairperson

University of California, Riverside

Acknowledgments

Somewhere ages and ages hence

Two roads diverged in a wood, and I

I took the one less traveled by

And that has made all the difference

(By Robert Frost)

Maybe not ten years from now, but as I pen down this acknowledgement section, my PhD. has been the single biggest challenge and achievement in my life so far. These five years have been a mixed bag of memories and emotions. A strong sense of achievement with every step I took forward, coupled with frustration for the times I did not make progress. But in retrospect, moments like these have made me emerge stronger, work harder and teach me the Art of Economics.

I am very grateful to my chairs Dr. Anil Deolalikar and Dr. Aman Ullah, for being a constant source of support and inspiration. They have not only been excellent teachers but great mentors guiding me all along. I would like to acknowledge the senior faculty members including Prof. Russell, Prof. Pattanaik, Prof. Guo, Prof. Gonzalez and Prof. Cullenberg whose classes I have thoroughly enjoyed and have learnt much from. I thank Professors, Jorge Agüero, Mindy Marks and Todd Sorensen for their positive attitude and unending support and guidance. They are an inspiration to be around and work with.

I thank my parents and family for giving me the opportunity to finish what I started. For their long-lasting patience and support all throughout. Thank you mom for the constant pushing and nagging, without it maybe I would still be writing my first chapter.

I would also like to thank my friends and peers at UCR, without whom life in Riverside would have been really boring and dull. Thank You to the wonderful staff in the Economics department at UCR (past and present).

I hope this achievement makes my grandfather very proud. He always wanted me to be a Doctor, so what if I cheated and turned out to be the other kind. Finally this acknowledgement section would not be complete if I did not thank Siddharth, a BIG thank you for your support all through out and for paving the way for a beautiful future. I promise to work even harder.

ABSTRACT OF THE DISSERTATION

Three Applied Essays in Development Economics

by

Shruti Kapoor

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, March 2010
Professor Anil Deolalikar, Co-Chairperson
Professor Aman Ullah, Co-Chairperson

My dissertation is comprised of three independent empirical chapters. Below is a brief description of each.

The first Chapter is titled "Infant Mortality Rates in India: District-Level Variations and Correlations". This paper examines the correlates of infant mortality in India using district-level data from the 1991 and 2001 Census of India. While infant mortality rates have dropped across districts over this ten year period, there still remains a lot of heterogeneity across districts and hence across the states. Using a panel dataset of 666 districts, the analysis seeks to determine which of socio and or economic factors play an important role in reducing infant mortality rates. In our empirical work, the explanatory variables used are male and female literacy, male and female labor force participation, the level of poverty, urbanization and other socio-economic variables. We use quantile regression analysis to determine which of these factors impact infant mortality. Quantile regression is preferred over OLS because it allows us to estimate models for the conditional median

function, and the full range of other conditional quintile functions and therefore provides a more complete statistical analysis of the stochastic relationship among random variables. The analysis brings out the powerful influence of women's characteristics on infant mortality, especially literacy and labor force participation. Increases in both of these variables significantly reduce child mortality at the district level. Improvements in male laborers in non-agricultural work and reductions in poverty also reduce child mortality, but their quantitative impact is weak in comparison. Further the non-parametric analysis reinforces the results found in the parametric section. They indicate that the action or the impact of the covariates is strongest in the districts which lie in the center of the conditional distribution, rather than those at the extreme. This analysis allows us to determine in which districts the impact of additional target policies would yield the greatest reduction in infant mortality.

The second paper is titled "Same-sex siblings and their affect on mothers labor supply in South Africa". This study aims to look at the labor-supply consequences of childbearing for women in South Africa. However due to the endogeneity of fertility, the research question becomes complicated. Using parental preferences for a mixed sibling-sex composition I construct instrumental variables (IV) estimates of the effect of childbearing on labor supply of all women aged 15-35 years having more than two children. The data used for the study is the 10% household sample from the 2001 census. The covariate of interest in the labor supply model is the indictor *More than two children*. Demographic variables include mother's age, age of the mother at first birth, years of schooling, indicators for race, and an urban dummy. Labor-supply variables include hours worked per week, worked for pay and total income. Unlike previous studies which restrict their sample to include only female household heads or spouse of male household heads, this study expands the sample size to include all childbearing women in the household aged 15-35. The IV estimates that exploit the fertility consequences of sibling sex do not confirm the OLS estimates showing that more children lead to lowering of female labor supply. While OLS estimates exaggerate the causal effect of children, children seem to have a smaller effect on the labor supply of college-educated women. I find that labor market outcomes of childbearing are more severe for married women in South Africa.

The third paper is on "The effect of Immigration on Ethnic Composition and Occupational Reallocation". Over the last 30 years, the U.S. labor market has been transformed by the 'second great migration'. Much of this immigration has been among the lower skilled; the share of High School Dropout (HSD) workers who are foreign born increased from 12% in 1980 to 44% in 2007. At the same time, native born HSD workers grew more slowly than any other educational category, falling by nearly 6%. These two outcomes have inevitably lead to much speculation that immigrants depress the wages of similarly skilled natives. The labor economics literature, however, has found little empirical evidence to support this claim. We aim to assess whether the impact of immigration is mitigated by occupational transition of natives. Being over represented among HSDs, we focus on the labor market outcomes for Black workers. We use data from the 5% public use sample of the census (1980, 1990 and 2000) as well as the 1% sample of the population from the American Community Survey (2005, 2006 and 2007) to estimate the effect of occupational reallocation on the wages of Black workers as well as the effect of immigration on reallocation. A shift-share analysis reveals that occupational transitions caused wages for Blacks to rise by 46% more than they would have with a static occupational distribution. However, we find that these occupational shifts were due to crowding out effect of Hispanics on Black occupations: a 10 percentage point increase in the share of workers in an occupation who are Hispanics leads to a 5 percentage point decrease in the share of Black workers in that occupation. This is significantly large to explain substantially, occupations that declined in importance for Blacks during the period of study. We find a strong correlation between importance of occupations to Hispanic and Blacks, suggesting that most occupational transition for these two groups has not only been driven by outside factors such as trade and technological change, but that these shocks are affecting the two groups similarly.

Contents

List of Figures **xii**

List of Tables **xiii**

1 Infant Mortality in India: District Level Variations and Correlations **1**

1.1 Introduction . . . 1

1.1.1 Cross-State Variation . . . 3

1.2 Objective . . . 5

1.3 Issues and Hypotheses . . . 7

1.3.1 Mother's Education . . . 8

1.3.2 Female Labor Force Participation . . . 9

1.3.3 Other Influences . . . 11

1.4 Empirical Model and Methodology . . . 12

1.5 Data Source . . . 13

1.6 Results and Discussion . . . 16

1.6.1 Ordinary Least Squares Regression . . . 16

1.6.2 Fixed Effects . . . 18

1.6.3 Simultaneous Quantile Regression . . . 19

1.6.4 Non Parametric Analysis . . . 22

1.7 Conclusion . . . 26

2 Same sex siblings and their affect on mother's labor supply in South Africa **41**

2.1 Introduction . . . 41

2.2 Lit Review . . . 45

2.3 Data and Summary Statistics . . . 48

2.3.1 Fertility and Sex Mix . . . 51

2.4 Empirical Model . . . 54

2.4.1 Wald Estimates . . . 54

2.4.2 Method of Two-Stage Least Squares Estimation . . . 55

2.5 Results . . . 58

2.5.1 2SLS Results 58
2.5.2 Other Issues 62
2.6 Conclusions and Future Work 64

3 The Effect Of Immigration On Ethnic Composition And Occupational Reallocation **73**
3.1 Introduction 73
3.2 Lit Review 76
3.3 Data and Summary Statistics 79
3.4 Empirical Model 83
3.5 Results 87
3.5.1 Reallocation and Wages 88
3.5.2 Evidence on Causes of Reallocation 89
3.6 Conclusion 90

4 Conclusion **106**

Bibliography **113**

List of Figures

1.1 Infant Mortality Rate World Map, 2006 28
1.2 Infant Mortality rate, by residence 1971-2006 28
1.3 Mortality Rate Under-5 (per 1,000), 1970-2005, selected countries in Asia . 29
1.4 Infant Mortality Rates across Indian states, 1990 - 2006 29
1.5 Percentage Annual reduction in Infant Mortality Rates across Indian states, 1990 - 2006 30
1.6 Contribution of the 21 larger states to national infant deaths, 2000 30
1.7 IMR, Indian Districts, 1991 and 2001 31
1.8 Distribution of District-level changes in IMR, 1991-2001 31
1.9 The Distribution of Infant Mortality Rate, 1991 and 2001 32
1.10 Median of Betas over the quantiles of the covariates 38
1.11 Median of Betas over the quantiles of the covariates 39

3.1 Increase in Group sizes 102
3.2 Change Consistent with Crowding Out 103
3.3 Actual and Counterfactual Black wages 104
3.4 Actual and Counterfactual hisp wages 105

List of Tables

1.1 Effect of individual, family, community and services on infant mortality . . 32
1.2 Definition of District Level Variables used in the analysis, India 1991 & 2001 33
1.3 Descriptive Statistics of District Level Variables, India 1991 & 2001 34
1.4 Results from the OLS Estimation of IMR, 666 Districts. 35
1.5 Results from the OLS Estimation of IMR with state and district dummies, 666 Districts. 36
1.6 Results from the Simultaneous Quantile Regression 37
1.7 Comparison of Non-parametric results in 1991, 2001 and entire panel 40

2.1 Fertility And Labor Supply Measures 67
2.2 Descriptive Statistics, Women aged 15-35 with 2 or more children 68
2.3 Fraction of Mothers that had another child by Parity and Sex of Children . 69
2.4 Differences in Means for Demographic Variables by Same Sex 70
2.5 OLS estimates of More than 2 children equation 71
2.6 OLS and 2SLSL estimates of labor Supply Model using 2001 Census Data . 72

3.1 Total Share of Immigrant Population in each Skill Category 93
3.2 Mean wages for natives in the working age by education (2007 dollars) . . . 93
3.3 Prior Elasticity Estimates 94
3.4 Summary Statistics 94
3.5 Most Common Occupations 95
3.6 Largest Black Occupations in 1980 95
3.7 Largest Black Occupations in 2007 96
3.8 Largest Decrease in Employment from 1980-2007 96
3.9 Wage Changes for Largest Decrease in Employment from 1980-2007 for Blacks 97
3.10 Largest Decrease in Employment from 1980-2007 for Black Men 97
3.11 Wage Changes for Largest Decrease in Employment from 1980-2007 for Black Men 98
3.12 Largest Decrease in Employment from 1980-2007 for Blacks aged 18 - 35 . . 98
3.13 Average wage of Black workers weighted by actual and counterfactual occupational distribution 99

3.14 OLS and Fixed Effect Regression Analysis 100
3.15 Largest Decrease in Employment from 1980-2007 for Blacks 101

Chapter 1

Infant Mortality in India: District Level Variations and Correlations

1.1 Introduction

Children are important assets of a nation, therefore reduction in infant and child mortality is likely the most important objective of the Millennium Development Goals (MDG). Infant and child mortality rates reflect a country's level of socio-economic development and quality of life and are used for monitoring and evaluating population, health programs and policies. It is an outcome rather than a cause and hence directly measures results of the distribution and use of resources, Haines(1995). The world map Figure 1.1 shows the level of Infant Mortality Rate (IMR) across various countries in 2006. While countries like Australia and Canada have IMRs well below 10 per 1,000 live birth, most of the African countries are struggling with mortality levels over 50 and in some cases 100 deaths per 1,000 live births. According to the United Nations estimates, 10 million infant deaths occur annually in the world. India accounts for a quarter of those. Thus any study of Indian infant mortality has global significance.

India has experienced an impressive decline in infant mortality since the 1970s. From 130-140 deaths per 1,000 live births in the early 1970's, mortality levels have declined to as low as 60 deaths per 1,000 live births in 2000. This represents an annual rate of decline of around 2.6 percent 1.2. However the absolute levels of infant and child mortality are still too high (about 68 infant and 95 child deaths per 1,000 live births in 1998-99).

According to the Registrar General of India, the IMR for the country as a whole is 57 infant deaths for every 1,000 live births in the year 2006. The National Population Policy 2000, aims at achieving IMR of 30 by the year 2010, Government of India (2000). The MDG is to reduce infant and child mortality by two-thirds between 1990 and 2015. In the case of India this would imply a reduction of the IMR to 27 and of the under-five mortality rate to 32 by 2015, The World Bank (2004). Figure 1.2 also shows that IMR has declined both in the urban and rural areas. The latest data from the sample registration system shows that the infant mortality rates for the urban and rural regions are 40 and 64, respectively, for every 1,000 live births in the year 2004, Government of India (2006).

While there has been a significant decline in IMR in India over the last three decades, it's performance with respect to other countries in Southeast Asia is not that impressive. Countries such as Indonesia, Sri Lanka, and Bangladesh have managed to reduce their IMR levels by between 3-5% annually. Figure 1.3 shows child mortality rates for select Asian countries over the period 1970-2005. As is seen in the figure, countries like

Thailand and Sri Lanka have had stellar performance.

Using a sample of 666 districts from the 1991 and 2001 census, this paper attempts to study the relevant relationships of demographic and socio-economic variables with IMR. The choices of independent variables are partly guided by previous literature on IMR and partly by the availability of data. Special attention has been paid to female literacy, female labor force participation rates, urbanization and some socio-economic variables. The prime intention of this analysis is to see which variables, economics or social, have a greater impact in reducing infant mortality levels and in which quintile the impact is the strongest.

1.1.1 Cross-State Variation

India is demographically a very diverse country. India is administratively divided into 28 states and 7 union territories. The states are further divided into 593 districts for political and judicial purposes. There are variations in basic demographic indicators not only across states but districts also. At one end of the spectrum, Kerala has demographic features which are similar to those of middle income countries like Bulgaria, Russia and Ukraine: life expectancy at birth is 72 years, infant mortality rate is 12 per thousand live births, total fertility rate is 1.8 births per women and ratio of females to males in the population is well above unity (1.04). At the other end we have the large north Indian states which find themselves in the same league as some of the least developed countries for the same indicators. In Uttar Pradesh[1], infant mortality rate is 72, life expectancy at birth is 61, total fertility rate is 5.1 and female-male ratio is (0.8), lower than that of any

[1] According to the 2001 census, Uttar Pradesh accounts for 16% of India's population.

country in the world Murthi & Guio (1995). In the state of Arunachal Pradesh, there exist a districts such as East Kameng which has an IMR of 158, as well as districts like East Siang where the IMR is only 64.

Due to wide inter-state and intra-state variations, it is therefore almost meaningless to talk about an average infant mortality rate for India. Figure 1.4 shows IMR across selected states in India. As far as the performance of individual states goes in terms of reducing their IMRs, the results are intriguing (Refer to Figure 1.5). Kerala was the slowest; it reduced its infant mortality at an annual rate of 12% between 1990 and 2006. On the other hand Bihar and U.P. which had the highest level of infant mortality achieved significant reductions (almost 30%) during the same period. States like Andhra Pradesh and Karnataka which are usually perceived to be good human development performers had the highest rate of decline in the above mentioned periods. Rajasthan, U.P., M.P. Bihar and Orissa are the high mortality states. These states also have high fertility and population growth rates. This implies that by the year 2015 a majority of India's population will be concentrated in these states. Thus if the MDG of 27 is to be achieved, these states will have to work even harder to reduce their IMRs to lower levels.

Due to heterogeneity, a look at the contribution of individual states to the number of infant deaths nationally is important from a policy perspective. In order to achieve the MDGs it is critical to target populous states with high IMRs. Figure 1.6 below shows the individual as well as the cumulative contribution of the 21 larger states to the total number of infant deaths nationally in 2000. About a quarter of the deaths occurred in the state of Uttar Pradesh. The four largest states of Uttar Pradesh, Madhya Pradesh, Bihar and

Rajasthan account for a little more than half of the nation's infant deaths.

Kernel density plot, Figure 1.7 using district-level data for the years 1991 and 2001 shows that the entire distribution of IMRs has shrunk in magnitude. A greater proportion of district have lower mortality levels in 2001 than in 1991. The mean IMR level in 1991 was around 79. This decreased to 72 by the year 2001. The maximum level of IMR across districts also dropped from 166 to 128. Figure 1.8 shows the kernel density plot for changes in the district-level IMR between 1991 and 2001. This graph confirms that overall decline in IMR during the decade of the 1990s. The distribution has a mean value of -6 with 59% of the districts in the country experiencing a decline between 1991 and 2001.

1.2 Objective

India is not only described by heterogeneity across states but also by large demographic variations within states (across districts). While a number of studies have looked at the correlates of infant mortality, most of them have exclusively concerned themselves with estimating the mean effect on infant mortality of variables such as mother's education, child's sex, urbanization level and birth order etc. Such estimates miss an important point for policy makers: exogenous variables and policy interventions may affect infant mortality differentially at different points in the conditional distribution. For example, while the effect of electrification may not influence infant mortality "on average", it might play an important role in the case of infants at the bottom of the conditional mortality distribution (i.e. infants at highest risk of mortality).

This paper seeks to address this shortcoming in the existing literature. Using a unique dataset (at the district level) from India, we estimate quantile regressions to analyze the affect of demographic, socio-economic and other correlates on infant mortality rates at different points of the conditional distribution of infant mortality. This technique will allow us to answer not only the question, ”can policy influence infant mortality rate?” but more importantly, the question, ”for whom do policy interventions matter the most?”

The unit of analysis is the district level for the following reasons. First, most of the studies on IMR have been done using data at either the household level or aggregated to the state level. To the best of my knowledge, no previous study has used the 2001 census data at the district level data to answer the above question. Thus this study will be a analyzing data across districts, which we saw in the earlier section can vary quite a bit within a state. Second, since infant mortality rates have changed a lot in the last two decades, it is important for policy purposes to use the most current data at the most disaggregated level. A study done by The World Bank (2004) analyzed the National Family Health Survey (NFHS-2) data and found that infant deaths in India were heavily concentrated in a relatively small number of districts and villages. During the period 1994 - 99, 20% of the villages and 22.5% of the districts accounted for half of all infant deaths in the country. Since the survey covered only a fraction of the villages in the country and the sample of households from each village was small, one could infer that these numbers are suggestive of patterns at the district level. It would be thus worthwhile to explore the data at the district level and identify factors most correlated with IMR.

Finally, using panel data at the district level (compared to state level) gives me larger number of observations and more variation in the covariates. It also gives me the ability to control for state*year fixed effects. This study will be unique because it will be the first study to analyze the 1991 and 2001 census data for infant mortality at the district level. The study will make use of quintile regressions and some basic non-parametric techniques to see whether it is mostly the economic factors or the social ones that play an important role in bringing down mortality levels at the district level.

1.3 Issues and Hypotheses

Vast bodies of empirical studies have focused on analyzing the determinants of infant mortality in India, and elsewhere in the world. Household incomes, female education, access to health services and immunizations have been the key determinants of infant mortality. This indicates that public policies which promote access to schools, better health facilities and encourage economic growth can help reduce infant mortality. Other environmental factors like electricity, access to safe water and sanitation and use of cooking fuels all have been found to have important health implications on infants. A recent World Bank study on the role of public policy and service delivery to help India achieve it MDGs found a strong association between infant mortality and various other factors like government health spending, access to roads, supply of electricity and immunization.

There are multiple factors that go into determining whether a young infant will survive in his/her early years. During pregnancy and after, external factors like the family, the situation and practices of the household, the norms of the community can influence both the mother and the infant. A brief explanation of how the individual, family, community and services might affect infant mortality is explained in Table 1.1.

This section is an attempt to take cues from various studies and hypothesis made to the extent that these help in understanding the resultant IMRs. Surveying the diverse connections emphasized in the literature to explain India's high IMRs would help create a framework for my analysis in the subsequent sections.

1.3.1 Mother's Education

Basic female education is considered as one of the most powerful factors that influence infant mortality and fertility. Education of the mother has often been treated as a proxy for socio-economic status. Caldwell (1979) has argued that it has a more direct effect on child mortality through improved child-care. Mothers who are more educated tend to get married upon adulthood, this in turn delays child bearing. An educated mother is likely to be more knowledgeable about nutrition, health care and hygiene of the infant (for example washing and feeding practice, taking better care of the sick child, immunization etc). This aspect of maternal education is particularly significant since large parts of rural India still practice childcare which is deficient in nature. For example, it is still quite common in villages to cut the umbilical cord with unsterilized sickles, keep the cooked food uncovered and exposed, leave the child un-immunized or follow orthodox methods to cure common childhood diseases like tetanus and diarrhea. If the mother is educated she can take advantage of public health services and request other members of the family to tend to the child's need. There can also be an effect due to income levels as more educated mothers are likely to have higher income. In Asia, the mortality among children of uneducated Nepalese women is almost 15 times greater than it is among those of Malaysian women with seven more years of schooling."[2] Caldwell (1979) examines education as a factor in mortality decline using the Nigerian data. He stresses on the role of parental education, particularly, that of the mother, in reducing infant and child mortality. He argues that a well educated mother can change the range of feeding and child care practices without imposing significant extra cost on the household, she is more capable of handling the modern world and that the education of women greatly changes the traditional balance of familial relationships with profound effects on child care. Thus, a priori, holding other factors constant, one would expect female education to help in lowering IMR.

1.3.2 Female Labor Force Participation

The status of a woman in society can be measured by her ability to participate in economically gainful work outside the household. Gainful employment outside the house not only gives the women increased bargaining power within the household Blumberg (1991), it also increases the potential economic "worth" of female children due to higher discounted parental value for them. Various studies in the literature have reiterated the connection between female work participation and female worth (See for example Dasgupta, Palmer-Jones & Parikh (2000), Miller (1982), Rosenzweig & Schultz (1982), Sen (1987), Desai & Jain (1994), Murthi & Guio (1995)).

Work status of the mother can have a two way affect on mortality. The need to work outside the house, may affect child survival rates simply by preventing the mother from caring for the infant. The dual burden of employment and household work can reduce

[2] John C. Caldwell, 'Cultural and Social Factors Influencing Mortality Levels in Developing Countries', Annals of the American Academy of Political and Social Sciences, Vol. 510, World Population: Approaching the Year 2000 (1990), pp. 44-59.

the time available for childcare activities. This could lead to substantial effect through a lack of feeding, especially breast feeding early in life. On the other hand, working outside the home leads to higher family income and gives the mother a modern outlook, both of which could increase the probability of survival. Kishor & Parasuraman (1998) using data from the 1992-93 National Family Health Survey, found that mother's income translates into greater control over the expending of resources, increased exposure and access to relevant information about childbearing and childrearing practices, and an enhanced ability to engage the world outside the home to better meet the nutritive, medical and survival needs of infants. Their study showed that mothers who are employed have a 10 percent higher infant-mortality rate and a 36 percent higher child mortality rate than mothers who are not employed. The study also found that male mortality increases more than female mortality if mothers work. In the case of a girl child higher levels of female labor force participation may increase the importance attached to the survival of the girl child.

In India, female work outside the house is largely influenced by local customs and traditions. Female Work Participation Rate (FWPR) is generally higher in the southern states than in the northern states. Miller (1982) found a clear relationship between female labor-force participation rates and regional pattern of female seclusion. Female seclusion was especially prevalent in the northern and north-western parts of the country. As a result female did little rural work. On the other hand women in the south were found to be very active as workers. Some studies also attribute the differential participation in economic activities between women in the northern and southern states to the type of production in that region (See Bardhan (1974, 1982, 1988), Gupta & Attari (1994), Miller

(1982), Rosenzweig & Schultz (1982)). According to Boserup, Kanji, Tan & Toulmin (2007), seclusion of women will be found in areas where plough agriculture is practiced. Dry-field system of cultivation does not require much female labor. Thus wheat cultivation which is more prevalent in the north and requires less tedious but more physical power is primarily dominated by males. On the other hand, wet rice cultivation (staple for southern and eastern parts of the country); involve a lot of female intensive labor.

1.3.3 Other Influences

Apart from the educational and economic impact discussed earlier, several other variables' impact on mortality can be investigated using the district-level data on India.

To begin with it is worth investigating whether infant mortality rates vary across social groups. Scheduled Castes (SCs) and Scheduled Tribes (STs) are Indian population groupings that are explicitly recognized by the Constitution of India as previously "depressed". According to the 2001 Census, SCs/STs together comprise over 24% of India's population, with SC at over 16% and ST above 8%. The proportion of SC/STs in the population of India has steadily risen since independence in 1947. Scheduled Caste/Scheduled Tribe along with the female-headed households is among the poorest in the country. The vulnerability of SC/ST households to acute poverty is evident in the arrangements for job reservations made for these groups in India. Other backward class households are also relatively poor compared to forward class households in most states. While one would argue that given their scarcity of resources on account of poverty, probability of male and female infant deaths could be expected to be greater in these households that might be just one side of the coin. Studies done by Liddle & Joshi (1989), Basu (1990), Miller (1982) have found that SC/ST households have a higher value for female children than their non-SC/ST counterparts. This is because originally bride price was more prevalent among lower castes and tribes. This would gives less reasons of discrimination between male and female children.

Another issue of interest is whether urbanization plays an important role in influencing mortality levels. Greater urbanization should lead to lower mortality levels. Finally, the relationship between poverty and mortality is worth noticing. Does poverty have a strong effect on mortality rates after controlling for the other explanatory variables?

1.4 Empirical Model and Methodology

Due to wide inter-state variations in mortality levels it would be more meaningful to investigate the effect of income and other interventions on infant mortality rates at different quantiles of the conditional distribution of IMR. In this study I estimate regression equations at different points of the dependent variable's conditional distribution using the quantile regression technique. This technique was initially developed as a "robust" regression technique that would allow for estimation where the usual assumption of normality of the error term might not be strictly satisfied, Koenker(1978). This technique is commonly used to understand the relationship between the dependent and independent variables over the entire distribution of the dependent variable and not just at the conditional mean.

The following basic model will be used for the analysis [3]: $IMR_{it} = \alpha + X_{it}\beta +$

[3] The author acknowledges the fact the model might change over the 10 year period. The functional form may not be the same between 1991 and 2001. However in this paper we assume that the functional form is linear and same for both the census years.

$D_i\gamma + e_it$, where $i = 1$ to n, $t = 2$ time periods, 1991 and 2001

IMR_{it} is the Infant Mortality Rate in district i for time 1991 and 2001 respectively.

X_{it} is the vector of demographic, socio-economic and health care, exogenous variables like female literacy, poverty, urbanization, etc.

D_i is a vector of dummies (n=332 for districts, n= 16 for states).

e_i is independent and identically distributed error term for each district.

Factors associated with IMR i.e. variables comprising the X matrix, may be broadly grouped into three categories: the population's socio-cultural characteristics, availability and opportunities of female employment, material prosperity and economic development of a region. I begin to analyze IMR by first doing a simple Ordinary Least Square (OLS) regression. I then add some district/state dummies to capture the heterogeneity that might have not been addressed by the covariates. The use of Simultaneous Quintile Regressions helps me analyze the affect of the covariates at various points of the IMR's distribution. I finally use some simple non-parametric and compare my results from the parametric analysis.

1.5 Data Source

The analysis uses a sample of 666 districts for the years, 1991 and 2001. This represents 71% of the 938 districts from 17 states in India. The Census of India[4] is the primary

[4]The Census is conducted by the Government of India and is the largest source of data on population characteristics. Data is available at the state, district and at the village level for rural and urban areas.

source of data for population related variables such as male and female work participation, literacy, place or residency and the scheduled caste and tribes. Mortality is measured by IMR, which represents the number of deaths before the first birthday for every 1,000 live births. With respect to the exogenous variables, the indicator of female (male) literacy is the crude female (male) literacy rate, defined as the proportion of literate females (males) above the age of six in the total population[5]. Female labor force participation is defined as the percentage of female "main workers" in the total female population above the age of 15 years[6]. Urbanization is measured by the percentage of total population living in urban areas[7]

It is difficult to obtain reliable district level data on income and poverty in India as there is no nationwide comparable source for such data. The National Sample Survey [8] is the largest nation wide survey on per capita household expenditure levels and over the years, the NSS unit data has been used to calculate estimates of district level income and poverty. Due to inherent lacunae in such calculations[9], this paper uses data on poverty from Amaresh Dubey's calculations in Debroy & Bhandari (2003)[10]. The reference years for the

[5]Literacy is defined in the Census of India as the ability to both read and write with understanding in any language for a person aged 7 years and above.

[6]Female work participation rate is the ratio of female "main workers" (women engaged in economically productive work for at least six months) to the total female population. Household duties were not counted as economically productive work.

[7]According to the 1991 and 2001 Census, "urban areas" were all statutory places with a municipality, corporation, cantonment board or notified town area committee, etc, and a place satisfying the following three criteria's: A population of over 5,000; those with at least 75% of the male labor force in the non-agricultural sector and those with a population density greater than 1,000 per square mile.

[8]The National Sample Survey is conducted by the NSSO. Data on per capita consumption expenditure is collected quinquennially at the household level, using stratified random sampling.

[9]The NSS data is based on households and then aggregated at the state level. Strictly, the data is not collected at the district level and is not a census but a large survey.

[10]For the purpose of identifying poor persons, a poverty line specified by the Planning Commission, Government of India in 1979 at 1973-74 prices was used. This poverty line is the cost of a bundle of commodities (share of expenditure on food items is over 80 percent and the remaining for other essential items such as clothing etc) that could provide a little over 2400 kilo calories to an average Indian living in

poverty variables are 1993-94 for the 91 census and 1999-00 for the 2001 census. Although I do acknowledge that the reference years are not exact and there might be some limitations of using this data for the analysis, to my knowledge it is the best possible information on poverty at the district level.

While a comprehensive set of explanatory variables affecting IMR across districts of India is desirable, I am bound in my selection of variables by the limitation of the data set. A gamut of factors such as income levels, wage rates, medical facilities, which could potentially be good indicators for understanding IMR are simply not available at the district level. This definitely puts a limitation on my analysis and on the interpretation of my results.

Tables 1.2 and 1.3 give the definitions and summarizes the means and the coefficient of variation of all the variables used for 666 districts in India. Note in Table 1.3 while the total number of observations for the panel data set should be 1,186 (593 districts for each year), I have data only 666 districts over the two years. This discrepancy in observations is due to the creation of new districts after the 1991 census. The new states (districts) created after the 1991 census have data only for the 2001 census. In some other cases, the data was simply missing or did not exist for the two time periods.

---

the rural areas. For the urban areas, the commodity bundle used was similar, except that the average food energy requirements were fewer that 2100 kilo calories on the average. The cost of these bundles was worked out to be Rs 49.09 per person per day for the rural population. For the urban population it was Rs 56.64. The National Sample Survey Organization through periodic representative surveys from roughly 120,000 households collects information on household expenditure. The information on expenditure is collected at prevailing market prices. Consequently the Poverty Line has to be updated for price changes and this is done using the State and Sector specific price deflator.

1.6 Results and Discussion

1.6.1 Ordinary Least Squares Regression

Table 1.4 below shows the results of the OLS regression of IMR on a set of explanatory variables[11]. Because the OLS estimated do not control for unobserved heterogeneity, they cannot be interpreted as causal effects. These results only tell us the signs and statistical significance of different coefficients. In arriving at the estimates in Table 1.4, I began with a fairly general specification (1) and went on to add a quadratic term (3) and also a log-log specification (5). With respect to IMR, the following observations are particularly noteworthy.

Female work participation rate: Higher female work participation reduces the extent of IMR and this effect is statistically significant. This result is in accordance to earlier studies. The sign on the coefficient in all the 5 specifications (Table 1.4) is negative and significant at the 5% level. A mother who works outside the house has greater resources, better access and information to health care facilities and hence can take better care of the child, implying lower mortality rates in the household. A 1% increase in the female work participation rate is associated with 6% decrease in IMRs(refer to specification 5). However one must be careful when examining the effect of female work participation rate on infant mortality. It is important to control for both economic and social disadvantages that motivates the women to seek gainful employment outside the household. So for example the level of poverty might be an important factor that determines the female's participation rate, and hence its effect on mortality rate.

[11]For a detailed description of the variables refer to the Data section.

Female Literacy: Female literacy has a negative and statistically significant effect on infant mortality. A 1% increase in female literacy is associated with a 23% drop in IMRs on average (specification 5). These results further enforce the claims made in the existing literature, that there indeed is a strong association between female's education and infant mortality rates.

Percentage of female agricultural workers: Apart from the role of female work participation rate in reducing IMRs, percentage of women as agricultural laborers also seem to have a significant (at 5%) and negative influence of IMR. On average a 1% increase in female agricultural laborers is associated with IMRs by almost 2%. The percentage of males working in non-agricultural areas has a significant(at 5%) and negative effect on IMR.

Poverty (as measured by the headcount ratio): Was found to be significant (at 5%) and positively associated with IMR. Higher levels of poverty are associated with higher levels of infant mortality. A 1% increase in the poverty level leads to an increase in IMR by 4% on average. Poor people are less informed about health care issues, have fewer resources and access to medical facilities and are not able to take proper care of their child, resulting in higher mortality rates on average.

Scheduled Caste & Scheduled tribes: A higher proportion of Scheduled caste and Scheduled tribe in the population increases the extent of infant mortality. Increasing the percentage of scheduled class and scheduled tribes in the population is associated with an increase in IMRs by 13% and 3% on average respectively. The coefficients on both the variables are significant.

Other variables: The effect of urbanization on IMR is positive and significant.

There could be various other factors that might influence IMR in urban areas like greater level of pollution leading to higher mortalities etc. These omitted factors might cause urbanization to have a positive influence on IMR. Male literacy was not found to be significant and that could be due to the strong correlation between male and female literacy.

I also test the hypothesis, that for given values of explanatory variables included in the analysis, child mortality is higher in areas of higher gender inequality. Areas with high levels of gender inequality tend to suppress the role of women in society, this could lead to higher infant mortality rates (as the health of the child depends a lot on the women). In order to conduct this test, I used the juvenile sex ratio (number of females, per 1,000 males in the age group 0-10 years) as my exogenous variable in the equation of infant mortality. The validity of this procedure is based on the implicit assumption that juvenile sex ratio is not affected by IMR. Juvenile sex ratio here serves as a proxy for gender inequality. I found that holding other factors constant, higher juvenile sex ratio in a district is associated with lower IMR. The effect was statistically significant. This supports the hypothesis that in districts with lower gender inequality, infant mortality rates might be higher.

1.6.2 Fixed Effects

Due to diversity of cultural norms in India, each state is strikingly different from the other. It has its own norms and cultures which might influence and govern the attitudes towards female children, female work participation rates and other factors important to IMR. I tried to capture these unknown and unquantifiable unique factors by introducing state and district dummies respectively. While there is no denying that there might exist vast cultural, socio-economic differences within a state (especially the larger ones), I will assume that they are acting as rough cultural boundaries capturing mostly non quantifiable factors. Thus the state dummy would attempt to capture any state specific impact that might not have been captured by the existing variables in use. With the introduction of these state dummies most of the significant variables were rendered insignificant. With reference to Table 1.5, specification 1 shows the simple OLS regression results. In Specification 2, the states dummies have been introduced. Female literacy and percentage of male non-agricultural laborers were the only two significant variables. The state dummies were found to be jointly significant[12]. This indicates that there exists some (unobserved)heterogeneity across states which might affect IMR and is not being captured by these existing covariants. Specification 3 is with just the district dummies. These again were found to be jointly significant[13]. Finally in specification 4, I introduced state*time interaction terms which were found to be jointly significant, indicating that there are some time varying state effects affecting IMR and that is introducing heterogeneity across states.

1.6.3 Simultaneous Quantile Regression

In the previous section I estimated regression equations at the average (i.e. at the conditional mean level of infant mortality rate). However for any policy to be effective, it would make more sense to evaluate the effect of the above mentioned covariates at different levels of the infant mortality distribution. Policy makers could be more effective in terms of total lives saved if they allocated spending to the places where it would be most effective). In

[12]The F statistics was F(16,640) with Probability $> F = 0.000$.
[13]The F statistics was F(331, 325) with Probability $> F = 0.000$.

this section I estimate the reduced-form IMR equations at different points of its conditional distribution, using the quantile regression technique,Koenkar and Hallock (2001). This also helps me address the issue of heterogeneity in my district level data. The word "quantile" refers to one of the class of values of a variate that divides the total frequency of a sample or population into a given number of equal proportions[14]. So for example the median is the 50th quantile (percentile) and divides the population (sample) into two halves: one would have values greater than the median and the other half lesser than the median. The OLS estimator minimizes the sum of least squares and estimates a conditional mean function. In contrast, the method of quantile regression is based on minimizing the sum of asymmetrically weighted absolute residuals and provides estimates for the conditional median function, and the full range of other conditional quantile functions. Because of the potential non-independence of the error term, the errors in the decile may be heteroscedastic and the quantile regression variances may be biased. To resolve this problem, bootstrap estimates of the asymptotic variances of the quantile coefficients are calculated with 100 repetitions and are used in the reported asymptotic t-rations. This method provides a greater insight into the various quantiles of the distribution of the IMR, which is not possible with the OLS alone. Figure 1.9 shows the conditional distributional graph of IMR for the entire panel[15].

Table 1.6 below shows the simultaneous quantile regression results for each selected quantiles. For comparison purposes I also report the OLS estimates. To capture any possible time affect, I introduced a time variable:

[14] Source: Dictionary.com
[15] Source: Census of India, 1991 and 2001

$$T = \begin{cases} 0 & \text{if year is 1991} \\ 1 & \text{if year is 2001} \end{cases} \tag{1.1}$$

The quantile regression results suggest important differences at different points of the distribution. The effect of female work participation rate is highest (in magnitude) and significant in the lowest 25% of the IMRs conditional distribution. As the distribution of IMR improves the affect is still significant but lesser. This indicates that increasing female work participation rate in districts with worst IMRs will be less effective than doing so in districts which are comparatively better off. Female literacy rates on the other hand had a strong and increasing negative affect on IMR as the quantiles increase. So for the districts which lie in the middle and the higher quintiles providing one more year of education to a female had a much stronger impact on reducing IMRs than for the lowest 10% of the districts. The affect of the percentage of female in agricultural labor was similar to that of female literacy in terms of reducing IMRs. The higher the quantiles the stronger was the negative affect. The districts in the 50th and 75th quintile showed a stronger impact of female workers in agriculture on IMR.

Poverty, percentage SC and percentage ST all had a significant and positive affect on IMR. These increased with each quantile. Thus increasing poverty in districts which were already worst off to begin with, had a greater negative effect on increasing IMR compared to those that fell in the lowest 10% of the IMR distribution. Percent of urbanization was effective only in the higher levels of the distribution, though the impact on IMR was still positive. The results of the 50th quantile looks similar to the results obtained from the

OLS estimation. Female work participation, female literacy, female participation as an agricultural laborer, greater male movement to non agricultural work all help in reducing IMRs and do so more in districts which are worst off in terms of IMRs. On the other hand, poverty and percentage of SC/ST all associate positively with higher IMRs. These results do point out to an encouraging conclusion: Socio-economic variables though important have a lesser impact on IMRs compared to economic variables. This might have significant policy implications in terms of targeting specific districts to reduce IMRs.

1.6.4 Non Parametric Analysis

In the last section I discussed the results of quantile regression analysis which happens to be a more specific form of non-parametric. It takes into consideration heterogeneity but only across low and high mortality districts. Non-parametric analysis is a more general form with no restrictions on the above. Let's consider a simple equation like $y_i = \alpha + \beta x_i + u_i$, parametric methods specify a form of $m(x_i)$. But typically the exact functional form connecting $m(x_i)$ with x is not known. If we force a linear or a quadratic function it may affect the accuracy of estimation of $m(x_i)$, hence we deploy the nonparametric method of estimation of the unknown function.

In this paper I have used the method of local linear regression estimator. It minimizes the following function: $\sum_i (y_i - m - (x_i - x)\beta)^2 K[(x_i - x)/h]$, $i = 1$ to n, with respect to m and β. The estimated is found by performing a weighted least squares regression of y_i against $z_i^{'} = (1, (x_i - x))$ with weights $K_i^{1/2}$. Thus the local linear approximation fits a straight line. The advantage of using local linear estimation is that it can be analyzed with standard regression techniques. It also has the same first-order statistical properties irrespective of whether the x_i is stochastic or non stochastic. Pagan and Ullah (1999).

In order to obtain the β for each of the covariates, a non parametric regression was run of IMR on the relevant covariates. This produced 666 βs for each covariate. The median value of the βs was then taken over the respective quantiles of the covariates. The graphs (Figures 1.10 and 3.2) below show the quantiles of each covariate (X-axis) plotted against the median of the βs (Y-axis).

Higher female work participation rate in districts is associated with greater reduction in IMRs. This is expected; working mothers have better resources and are more informed regarding the care and upbringing of the child. So as the percentage of women working outside the house increases, infant mortality rates in those districts should fall. Thus in districts where work participation rate is higher its impact on reducing IMR is stronger.

According to the non-parametric estimates male literacy does not seem to have the desired affect on IMR. Only in districts where the levels of literacy are low, the affect is negative and significant. At higher quintiles of literacy the sign is positive. Educating more males in a district is associated with higher infant mortality rates in most of the quintiles. Once again as discussed earlier, this opposite affect could be due to the high correlation between male and female literacy. The introduction of both in the analysis could reduce the significance and the magnitude of the affect of male literacy on IMR.

Female literacy on the other hand has a strong and negative affect on IMR across almost all quintiles. Increased female literacy levels lead to greater reductions in IMRs.

Educating more females leads to significant reductions in IMRs across those districts. The impact is strongest in the middle quintiles. By the time you reach the upper tails of the distribution the impact is still negative but weaker.

Again higher the percentage of female laborers in agriculture, higher is their association in reducing IMR. Districts with low percentage of female agricultural laborers are faced with lower reduction in IMR. As discussed this could be because the mother who do not work are less aware of medical care facilities, have access to fewer resources and probably have less income to take better care of the child. On average the affect of female laborers in agriculture on IMR is negative.

While the affect of the percentage of male laborers in non agricultural work on IMR is negative, surprisingly it is strongest in the lower quintiles (Refer to Figure 3.2). One would expect that increasing the percentage of male laborers in non-agricultural work would lead to reduction in IMR but my results show that the impact is strongest in districts with the lowest percentage of male workers.

According to the graph below, lower levels of poverty lead to increased IMR. Though the graph does show that the in the lower quintiles the impact is the strongest, but positive, by the time we reach the upper tail of the distribution reducing poverty leads to lowering IMRs. With regards to the SC, ST population, the impact on IMR is positive in all the quintiles. This is in agreement with the parametric results. As the percentage of SC/STs are increased in the population, infant mortality increases.

Increasing the percent of urban population should help in reducing mortality levels, however both my parametric and non-parametric results show a positive effect on IMR.

Though the impact is small it becomes stronger with subsequent quintiles. While theoretically one would believe the urbanization leads to lower IMR, this data suggests otherwise. One of the reasons for this could be that there are a lot other factors in the urban areas which might influence IMR and have not been captured here in this analysis. Maybe there is greater poverty in urban areas or parents spend less time on child care etc. A lot of unexplained and omitted factors might be the reason for this positive association between IMR and urbanization.

We do realize that the panel data set is not continuous over the 10 year period as a result of which the results might be different for the two census years individually. In order to ensure the same, we repeated the same exercise for each of the years separately (graphs not shown here). We find that for both the years, the non-parametric graphs are similar in shape to the ones presented here. In terms of magnitudes, the coefficients are not identical but similar for most of the variables regardless of the census years. Table 1.7 compares the mean effect of each covariate on IMR for the periods 1991, 2001 and the entire panel including both the census years. The above non-parametric analysis lends support to my parametric results. Economic variables like female work participation rate, female literacy and the percentage of female laborers in agriculture all had a strong and negative impact on IMR. The affect was stronger in districts were IMRs was higher. Some of the male variables and urbanization has a strong affect though in the opposite direction.

1.7 Conclusion

In this paper, we have used district-level data from the 1991 and 2001 Censuses to analyze the determinants of IMR in India. In particular we have focused on male/female literacy, male/female work participation rates and other variables which can be changed using policy. We estimated the impact of these variables on IMRs using OLS, Fixed effects, Quantile regressions and Non-parametric methods. We control for State-fixed heterogeneity by including state fixed effects in the estimating equations of IMRs. Based on all the empirical analysis above, the results can be summarized as follows:

First, female work participation rate, female literacy and the percentage of female laborers in agricultural work seemed to have the strongest affect on IMR. Improving the quality of female human capitals does seem to have a significant and positive association with reducing IMRs.

Second, male variables like male literacy do play a small but not a very significant role in reducing infant mortality. The increase in percentage of male workers outside of agriculture is associated with IMRs through factors discussed earlier.

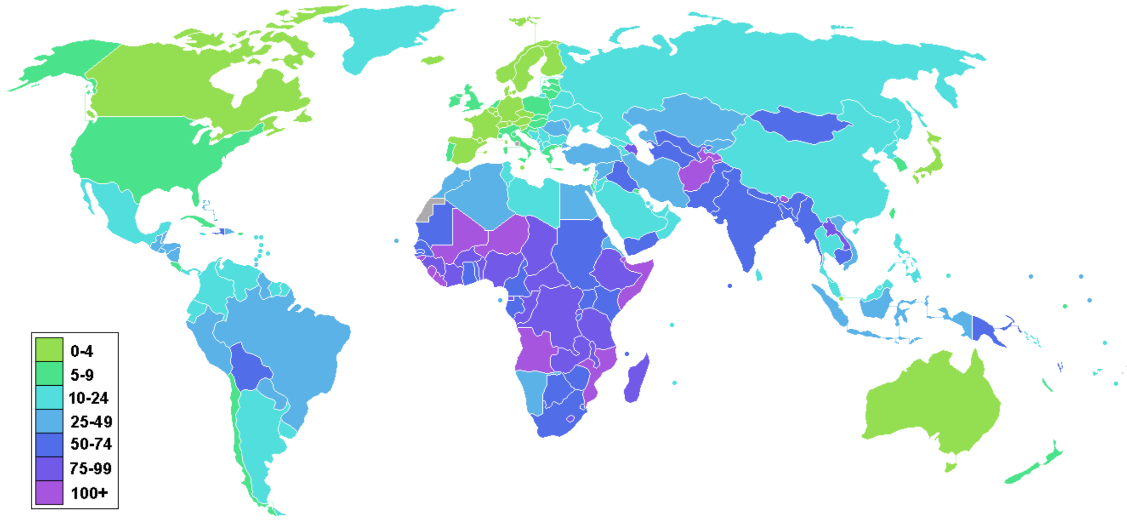
Third, other economic and cultural factors like poverty, percentage of SC/STs surely contribute to reducing infant mortality. Thus any improvements in these variables will have direct or indirect affects on the child.

Finally, the impact of the covariates discussed above is strongest in districts with high levels of IMRs. So from a policy perspective programs and aid should be devised keeping in mind the districts which fall in the middle of the IMR distribution and not at the tails.

The findings of this study clearly demonstrate the role of woman's agency and empowerment in reducing infant mortality. Furthermore the quintile analysis helps us identify that the results are strongest in districts with high levels of IMRs. This is helpful in targeting policies in the districts that lie in the middle of the distribution and not the tails. Economic variables like work participation rate and literacy levels were found to be negatively associated with IMRs. A typical response thus might be to increase female literacy levels in these districts. Other policies that have been initiated to prevent prenatal sex-determination and sex-selective abortions (which lead to higher IMRs) include the 1994 Pre-conception and Prenatal Diagnostic Techniques (Prohibition of Sex Selection) Act, the Hindu Succession (Amendment) Bill [16] etc.. In this paper I have attempted to analyze the district level IMRs in India using the 1991 and 2001 Census data. High IMRs across districts of India is an issue of concern. This paper serves the task of comprehensively discussing the issues in the recent Indian context. Using parametric and non-parametric techniques various important factors have been identified which could be targeted to improve the Infant mortality rates across districts and help in achieving the Millennium Development Goals of target IMR of 27.

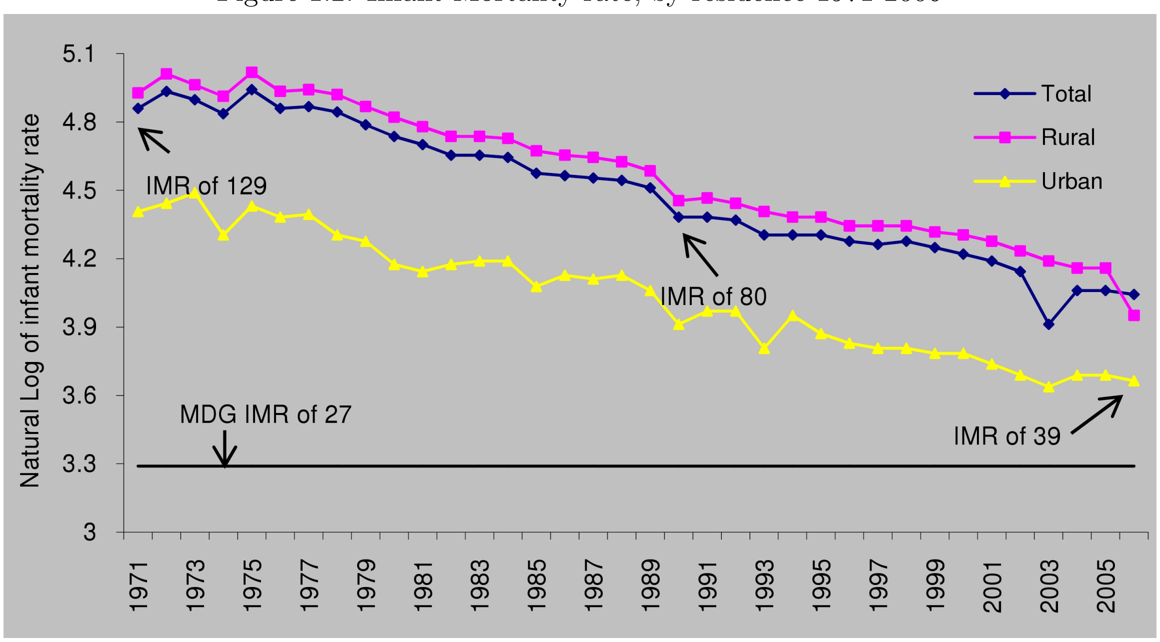
[16] This Bill provided women the right to inherit ancestral property. In the Bill enacted in 1956 according to Hindu inheritance customs, ancestral property could only be inherited by sons. The 2004 bill removed discriminatory provisions of the 1956 Act and allowed parents to bequeath their property to their daughters.

Figure 1.1: Infant Mortality Rate World Map, 2006



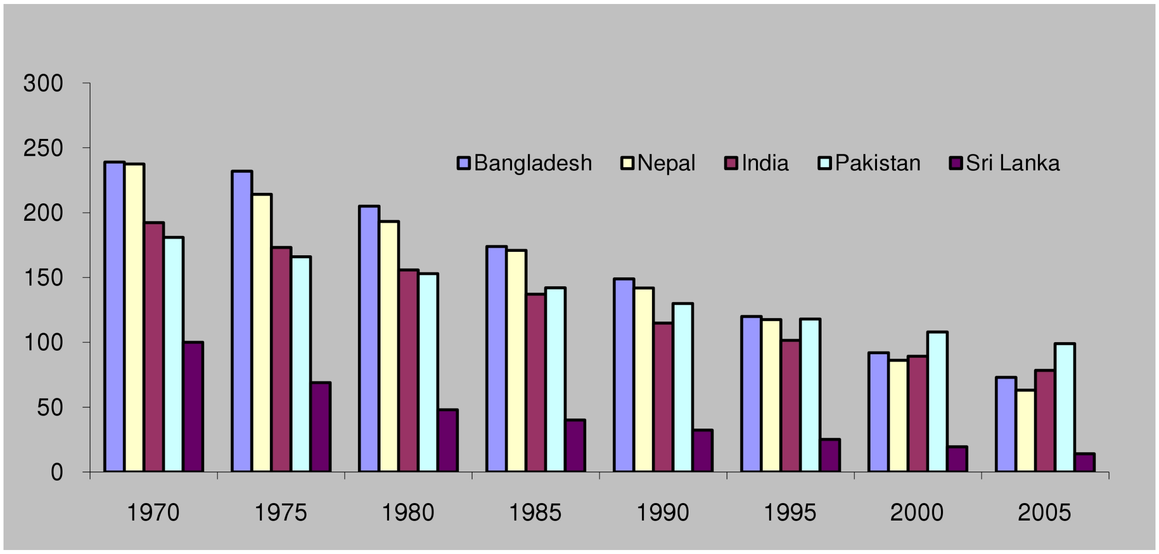
Source: CIA Factbook, 2006

Figure 1.2: Infant Mortality rate, by residence 1971-2006



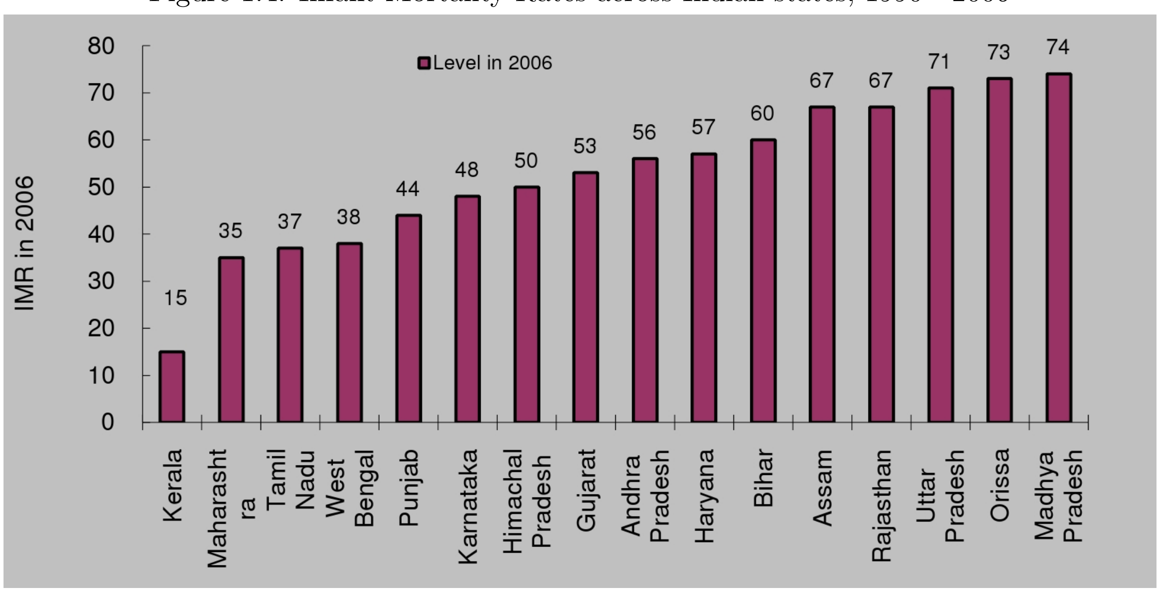
Source: Compendium of India's Fertility and Mortality Indicators 1971-1997, Registrar General & Family Welfare Program in India 2001, Department of family Welfare, Ministry of Health and Family Welfare, Govt. of India

Figure 1.3: Mortality Rate Under-5 (per 1,000), 1970-2005, selected countries in Asia



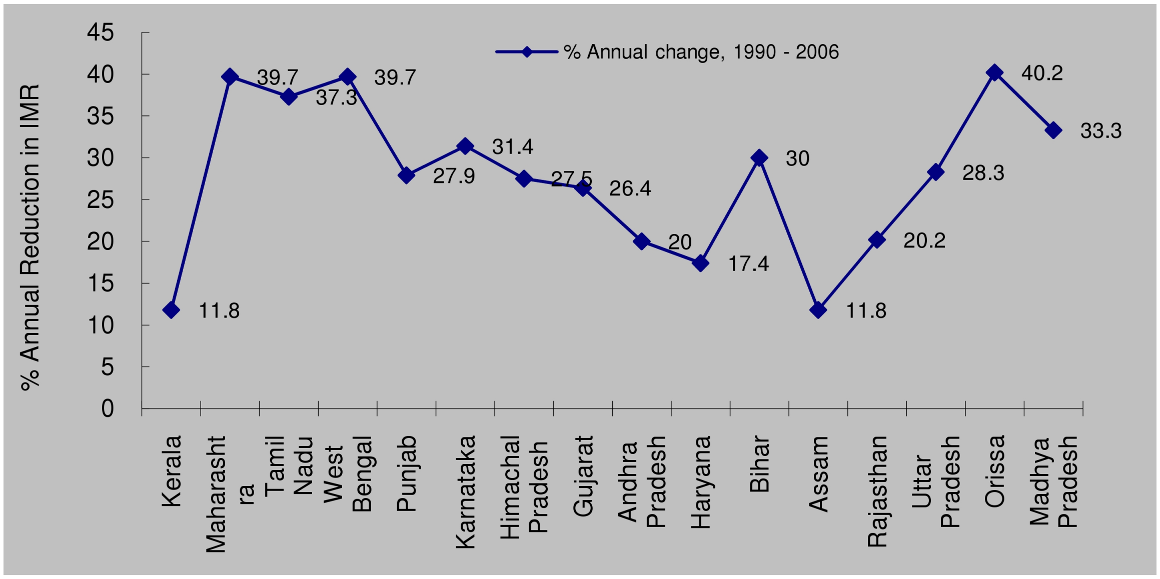
Source: World Development Indicators, World Bank

Figure 1.4: Infant Mortality Rates across Indian states, 1990 - 2006



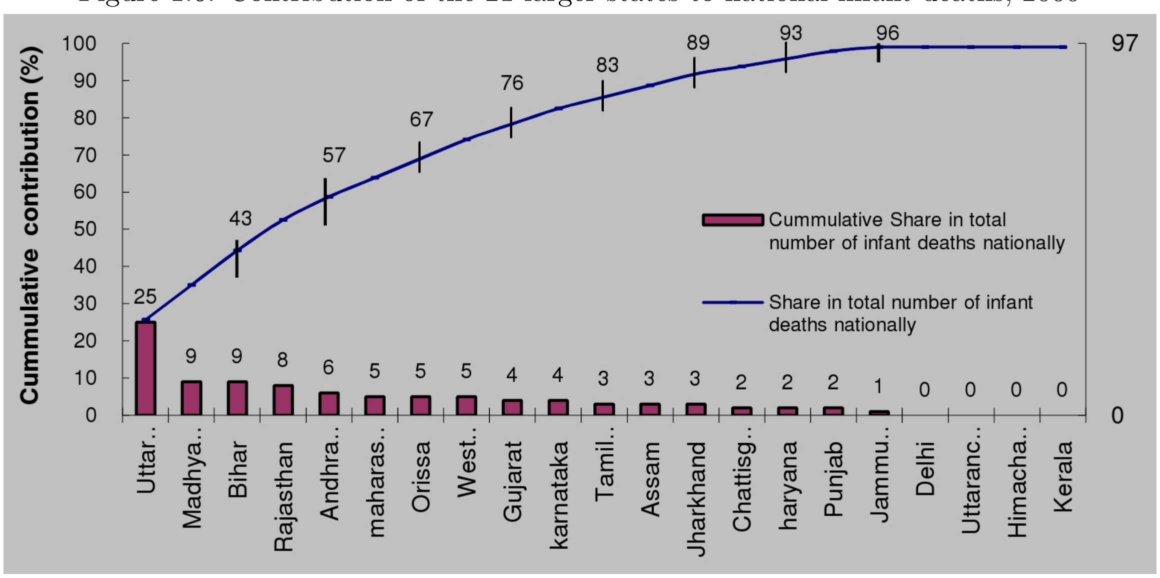
Source: Indiastat.com

Figure 1.5: Percentage Annual reduction in Infant Mortality Rates across Indian states, 1990 - 2006



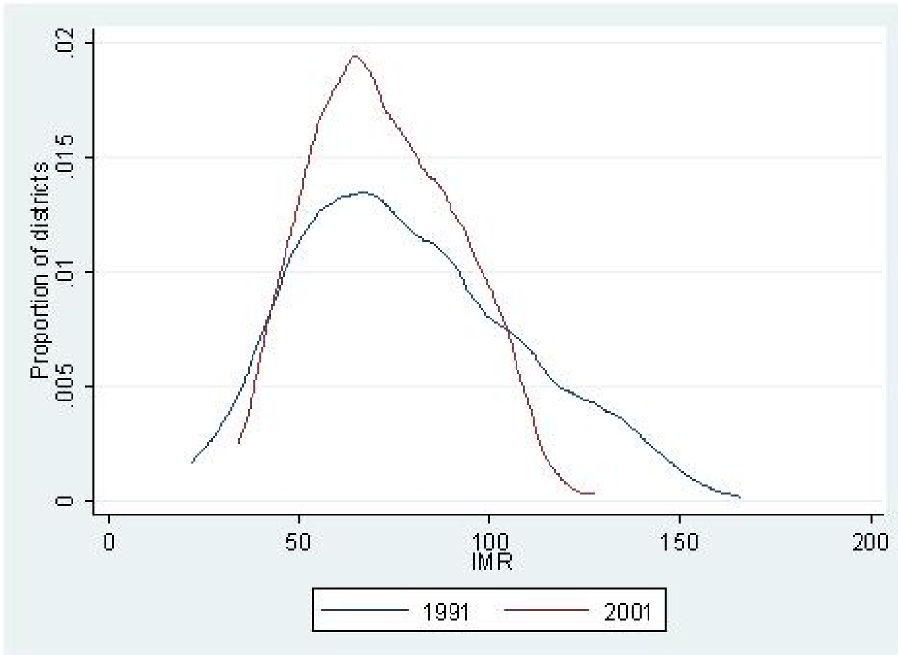
Source: Indiastat.com

Figure 1.6: Contribution of the 21 larger states to national infant deaths, 2000



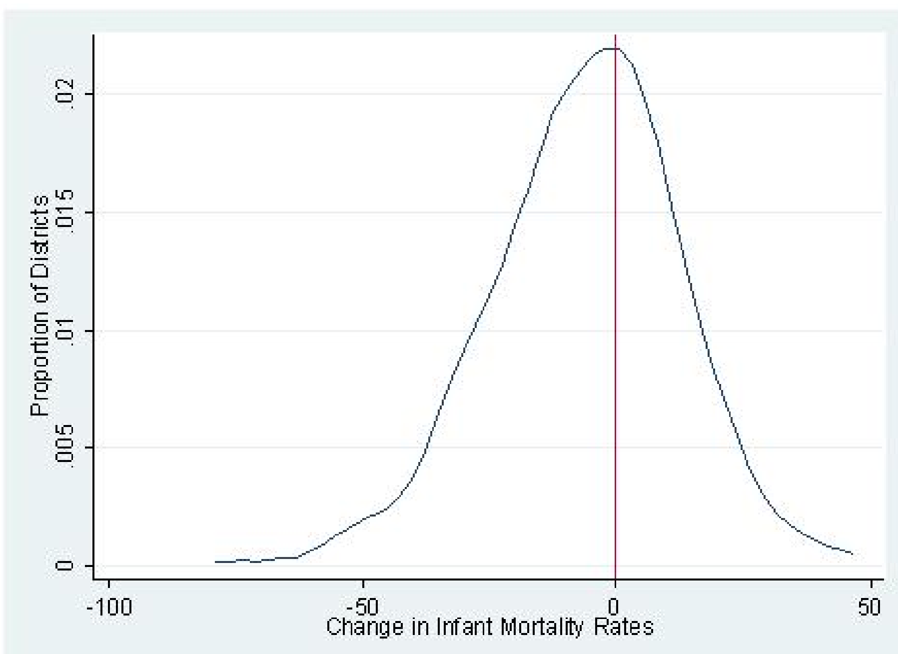
Source: World Bank (2004)

Figure 1.7: IMR, Indian Districts, 1991 and 2001



Source: World Bank (2004)

Figure 1.8: Distribution of District-level changes in IMR, 1991-2001



Source: 1991 and 2001 Indian census data from *census.india.gov.in*

Table 1.1: Effect of individual, family, community and services on infant mortality

Health Outcome	Lower child Mortality; Healthier babies and young children
Individual Woman/Mother	Education (esp. port-primary) Nutrition (during pregnancy and breastfeeding) Age at first birth and spacing between births Hygiene practice (esp. hand washing)
Household	Income and Wealth Intra-household dynamics Water (piped into households) Cooking fuels Hygiene practice Use of insecticide nets (malaria)
Community	Environmental health practices (water, sanitation, solid waste) Beliefs and practices (e.g. at birth) Women's self help group
Service Provision	Basic health/nutrition services in village/outreach to households Access to health facilities for emergency obstetric and sick child care

Source: World Bank (2007)

Figure 1.9: The Distribution of Infant Mortality Rate, 1991 and 2001

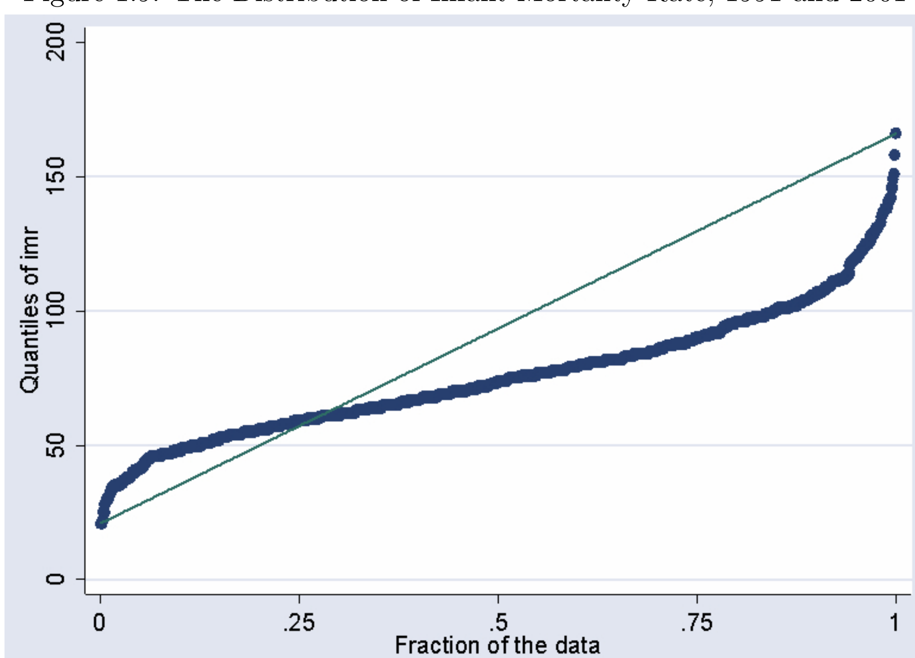


Table 1.2: Definition of District Level Variables used in the analysis, India 1991 & 2001

Name used	Name of Variable	Definition
IMR	Infant Mortality Rate	The probability of dying before the first birthday per 1,000 live births
MWPR	Male Work Participation Rate	Male workers among the male population above 15 years
FWPR	Female Work Participation Rate	Female workers among female population above 15 years
MALE_LIT_AB_6	Male Literacy above the age of 6	Percentage of male literates above the age of 6
FEM_LIT_AB_6	Female Literacy above the age of 6	Percentage of female literates above the age of 6
AGR_WRK_IN_POP	Agricultural workers in the population	Percentage of agriculture workers in population
PER_F_AG_LAB	Percentage of female workers who are agricultural laborers	Percentage of agricultural workers among all female workers
PER_M_NON_AG_LAB	Percentage of male workers who are non agricultural laborers	Percentage of non agricultural workers among all male workers
POVERTY	Poverty	Headcount of poor (1993-94 values used for the 1991 census and 1999-00 used for 2001 census)
PER_SC	Percentage of SC	Percentage of Scheduled caste in the population
PER_ST	Percentage of ST	Percentage of Scheduled tribe in the population
PER_LIT	Percentage of literates	Percentage of literates in the population
PER_URBAN	Percentage of Urban population	Percentage of urban population among total population

Source: 1991 and 2001 Indian census data from *census.india.gov.in*.

Table 1.3: Descriptive Statistics of District Level Variables, India 1991 & 2001

VARIABLE	OBS	MEAN	STD DEV	MAX	MIN
Infant Mortality Rate	666	75.64	24.75	22	166
Male Worker Participation Rate	666	51.57	5.75	22.98	99.52
Female Worker Participation Rate	666	24.91	12.56	1.75	89.42
Male Literacy	666	68.67	13.64	26.29	97.46
Female Literacy	666	43.89	18.32	7.68	94.35
% of Female AG LAB	666	39.59	18.57	0.14	79.85
% of Male NON AG LAB	666	40.13	18.18	9.54	99.89
Poverty	666	25.76	16.22	1.75	83.76
% of Scheduled Caste	666	16.89	7.10	1.44	51.76
% of Scheduled Tribe	666	7.86	13.45	0	86.85
% of Urbanization	666	23.50	16.77	2.74	100
Year Dummy	666	0.5	0.5	0	1

Note: OBS = Number of Observations, MAX = Maximum value of that variable in the entire sample, MIN = Minimum Value of that variable for the entire sample. Year Dummy = 1 if year 2001, 0 if year = 1991. Variable definitions same as in Table 1.2. Source: Calculated from the 1991 and 2001 census data by the author.

Table 1.4: Results from the OLS Estimation of IMR, 666 Districts.

	(1) IMR	(2) IMR	(3) IMR	(4) LIMR	(5) LIMR
Male Worker Participation Rate	0.137 (0.94)				
Female Worker Participation Rate	-0.236 (-2.28)†	-0.200 (-2.30)†	-0.184 (-2.09)†	-0.003 (-2.83)†	
Male Literacy above age 6	0.313 (1.52)	0.253 (1.35)	0.192 (1.04)	0.006 (2.51)	
Female literacy above age 6	-0.545 (-3.25)‡	-0.492 (-3.23)‡	-0.478 (-3.20)†	-0.009 (-4.74)‡	
Percentage of female Agricultural laborers	-0.339 (-6.16)‡	-0.338 (-6.11) ‡	0.148 (0.82)	-0.004 (-5.50)‡	
Percentage of male Non-agricultural laborers	-0.663 (-6.13)‡	-0.679 (-6.45)‡	-0.645 (-6.08)‡	-0.009 (-6.48)‡	
Poverty	0.296 (4.10)‡	0.291 (4.05)‡	0.293 (4.12)‡	0.003 (3.68)‡	
Percentage of Scheduled Caste	0.437 (3.82)‡	0.445 (3.89)‡	0.402 (3.52)‡	0.007 (4.95)‡	
Percentage of Scheduled Tribe	0.408 (4.83)‡	0.408 (4.80)‡	0.372 (4.39)‡	0.006 (5.51)‡	
Percentage of Urbanization	0.294 (3.93)‡	0.311 (4.64)‡	0.314 (4.91)‡	0.004 (4.36)‡	
Square of percentage of female Agricultural workers			-0.006 (-2.74)†		
Log of Female Work Participation Rate					-0.059 (-3.06)‡
Log of Female Literacy above age 6					-0.230 (-6.00)‡
Log of Percentage of Female Agricultural laborers					-0.038 (-2.58)‡
Log of Percentage of Male Non-agricultural laborers					-0.296 (-5.97)‡
Log of Poverty levels					0.041 (3.01)
Log of Percentage of Scheduled Caste					0.136 (6.09)‡
Log of Percentage of Scheduled Tribe					0.036 (5.85)‡
Log of Percentage of Urbanization					0.075 (3.26)‡
Year Dummy	2.505 (1.74)	2.450 (1.71)	2.658 (1.84)	0.088 (4.28)‡	0.120 (4.92)‡
Constant	90.443 (6.73)‡	98.698 (10.56)‡	93.746 (9.82)‡	4.477 (40.55)‡	5.707 (33.46)‡
Observations	666	666	666	666	666
R-Squared	0.44	0.44	0.45	0.48	0.41

Note: LIMR = log(IMR), OLS = Ordinary Least Squares. Columns 1 - 3 are regression of IMR on various covariates. All variables are in levels. Column 4 is log-linear specification Column 5 is log-log specification. Absolute value of t statistics in parentheses, † significant at 5%; ‡ significant at 1%. Source: Author's calculations.

Table 1.5: Results from the OLS Estimation of IMR with state and district dummies, 666 Districts.

	(1) IMR	(2) IMR	(3) IMR	(4) IMR
State Dummies	NO	YES	NO	NO
District Dummies	NO	NO	YES	NO
State*Time Interaction	NO	NO	NO	YES
Female Worker Participation	-0.163	-0.154	0.449	-0.102
Rate	(-2.04)†	(0.19)	(3.07)†	(-1.25)†
Female literacy above age 6	-0.320	-0.401	-0.471	-0.309
	(-3.99)‡	(-4.98)‡	(-3.15)‡	(-3.14)‡
Percentage of female Agricultural laborers	-0.353	-0.007	-0.318	0.009
	(-6.69)‡	(-0.14)	(-2.12)	(0.18)
Percentage of male Non-agricultural laborers	-0.680	-0.212	-0.138	-0.311
	(-6.42)‡	(-2.30)‡	(0.72)	(-3.41)‡
Poverty	0.302	0.067	0.054	0.014
	(4.18)‡	(1.17)	(0.34)	(1.35)
Percentage of Scheduled Caste	0.445	0.148	1.302	0.106
	(3.87)‡	(1.15)	(1.73)	(0.86)
Percentage of Scheduled Tribe	0.387	0.092	-0.069	0.110
	(4.44)‡	(1.41)	(-0.15)	(1.76)
Percentage of Urbanization	0.307	0.064	0.044	0.075
	(4.51)‡	(0.88)	(-0.20)	(1.05)
Year Dummy	2.604	0.855	-3.667	7.540
	(1.86)	(0.49)	(-1.27)	(1.69)
Constant	108.11	73.88	17.06	74.47
	(19.13)‡	(11.66)‡	(1.04)	(11.29)‡
Observations	666	666	666	666
R-Squared	0.44	0.62	0.74	0.66

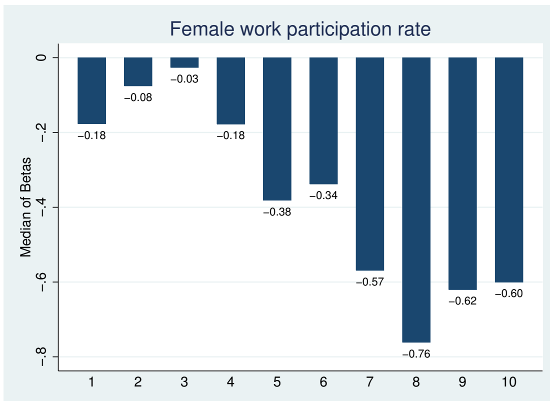
Note: IMR = Infant Mortality Rates, OLS = Ordinary Least Squares. Column 1 refers to the regression of IMR on the specified covariates. Column 2 includes the 16 state dummies for state-fixed effects. F-statistics for state dummies = F(16,640) with probability > F = 0.000. Column 3 includes 332 district dummies, F-statistics for district dummies = F(331,325) with probability > F = 0.000. Column 4 includes state∗time interactions. Absolute value of t statistics in parentheses. † significant at 5%; ‡ significant at 1%. Source: Author's calculations.

Table 1.6: Results from the Simultaneous Quantile Regression

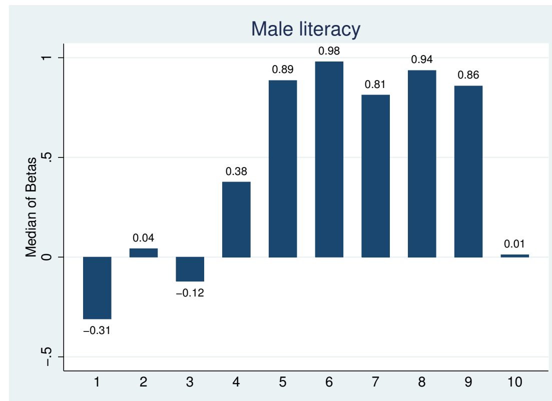
	Quantiles											
	0.10		0.25		0.50		0.75		0.90		OLS	
Independent Variable	Coeff.	T-ratio	Coeff.	T-ratio	Coeff.	T-ratio	Coeff.	T-ratio	Coeff.	T-ratio	Coeff.	T-ratio
Female work participation rate	-0.22	**-2.20**	-0.26	**-2.55**	-0.18	**-2.23**	-0.15	**-2.00**	-0.13	-1.14	-0.16	**-2.04**
Female literacy above the age of 6	-0.28	**-3.65**	-0.35	**-2.96**	-0.41	**-4.73**	-0.37	**-3.60**	-0.39	**-3.12**	-0.31	**-3.97**
% of female agricultural workers	-0.17	**-3.98**	-0.32	**-6.70**	-0.40	**-10.02**	-0.45	**5.85**	-0.31	**-4.23**	-0.35	**-6.69**
% of male non-agricultural workers	-0.29	**-2.48**	-0.41	**-3.95**	-0.59	**-5.32**	-0.77	**-6.34**	-0.58	**-4.41**	-0.68	**-6.42**
Poverty	0.18	**3.10**	0.25	**4.10**	0.33	**4.74**	0.26	**2.83**	0.45	**3.51**	0.30	**4.18**
% of scheduled caste	0.36	**3.16**	0.42	**3.95**	0.45	**4.27**	0.39	**2.19**	0.48	**2.21**	0.44	**3.87**
% of scheduled tribe	0.27	**3.46**	0.31	**5.09**	0.34	**4.56**	0.44	**3.32**	0.50	**2.30**	0.38	**4.44**
% of urbanization	0.04	0.60	0.15	**2.68**	0.26	**4.25**	0.38	**4.80**	0.35	**3.74**	0.30	**4.51**
Year	10.84	**12.97**	9.85	**4.97**	5.69	**3.55**	-0.005	-0.00	-15.39	**-3.43**	2.60	1.86
Constant	72.12	**12.97**	89.78	**12.64**	108.72	**18.09**	129.80	**17.80**	132.86	**18.8**	108.11	**19.13**
N	666										666	
R-Squared	0.27		0.27		0.29		0.29		0.29		0.44	

Note: IMR = Infant Mortality Rates, OLS = Ordinary Least Squares, Coeff. = Coefficients. Note: Asymptotic t-ratios are shown above (heteroscedasticity robust for OLS; bootstrapped for quantiles). Figures in bold indicate statistical significance of the estimated coefficient at the 10% or lower level. Source: Author's calculations.

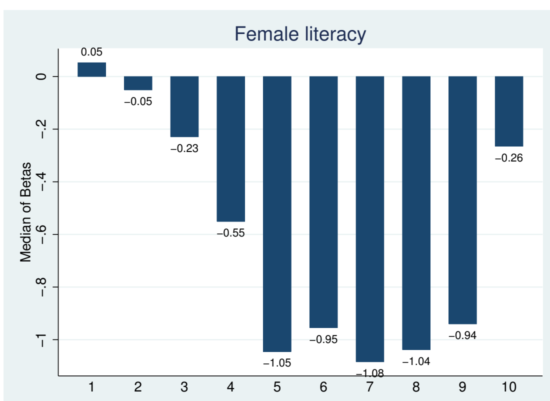
Figure 1.10: Median of Betas over the quantiles of the covariates



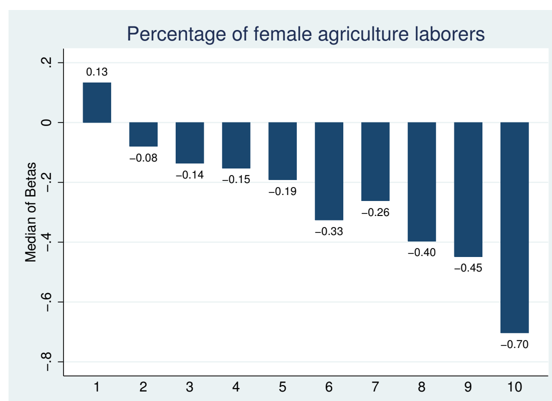
(a)



(b)



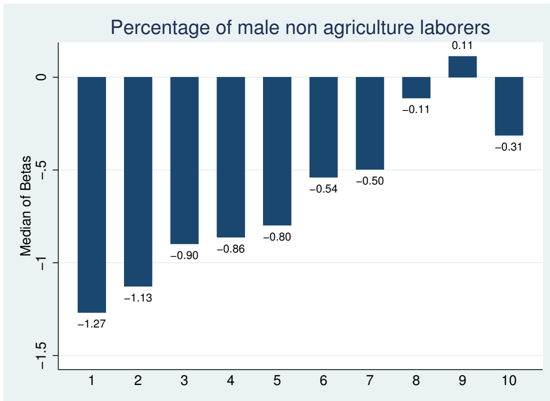
(c)



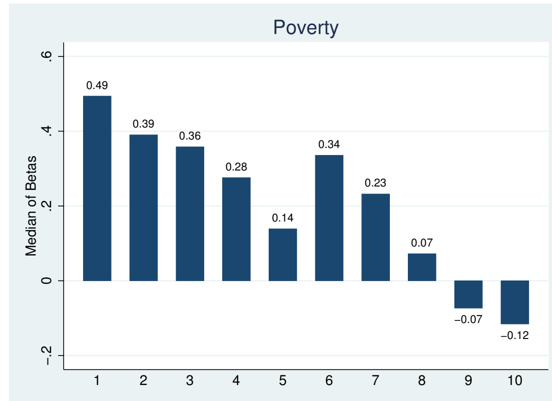
(d)

Note: The values plotted here are obtained by a nonparametric regression of IMR on covariates discussed in the summary statistics. On the X-axis we have the deciles of the respective covariate and on the Y-axis is the corresponding median value of the βs in that decile. Nonparametric estimation was done using the nonparametric software N by Jeff Racine. Through Least Square Cross Validation, Local Linear Fixed Bandwidth estimators were used. The Kernel used was a second order Epanechnikov kernel. The scale factors used were 2.06, 1.06, 1.06, 1.06, 1.06, 1.06, 1.06, 1.06, 1.06. Source: Author's calculations.

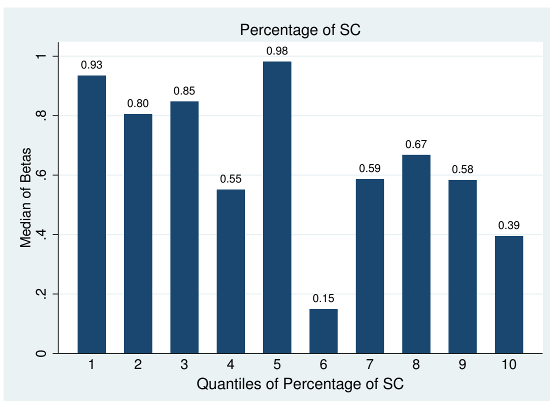
Figure 1.11: Median of Betas over the quantiles of the covariates



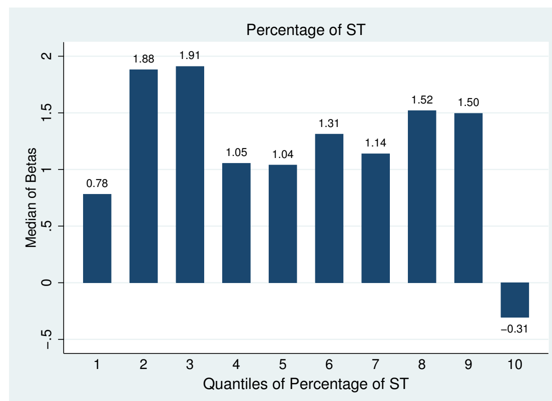
(a)



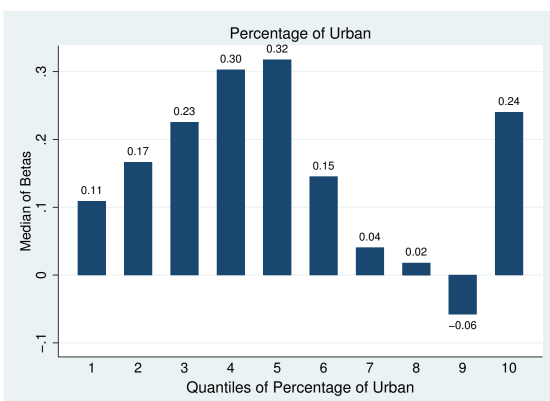
(b)



(c)



(d)



(e)

Note: The values are obtained by a nonparametric regression of IMR on relevant covariates. X-axis represents the deciles of the respective covariate and Y-axis represents the corresponding median value of the βs in that decile. Nonparametric software N by Jeff Racine was used for the estimation. Through Least Square Cross Validation, local linear fixed bandwidth estimators were used. Kernel used was a second order Epanechnikov kernel. The scale factors used were 2.06, 1.06, 1.06, 1.06, 1.06, 1.06, 1.06, 1.06, 1.06. Source: Author's calculations.

Table 1.7: Comparison of Non-parametric results in 1991, 2001 and entire panel

Variable Name	1991 only	2001 only	Both years
FWPR	-0.22	-0.18	-0.21
MLIT	0.55	0.58	0.57
FLIT	-0.89	-0.77	-0.83
PER_F_AG_LAB	-0.18	-0.16	-0.17
PER_M_NON_AG_LAB	-0.57	-0.77	0.67
POVERTY	0.34	0.14	0.24
PER_SC	0.90	0.26	0.59
PER_ST	0.73	1.15	0.94
PER_URBAN	0.28	0.39	0.33

Note: Variable definitions same as in Table 1.2. Column 2 gives the average effect of each covariate on IMR in 1991, Column 3 does the same for 2001 and Column 4 is for the entire panel data. Source: Author's calculations.

Chapter 2

Same sex siblings and their affect on mother's labor supply in South Africa

2.1 Introduction

While women constitute at least 50 % of the world's population and perform two-thirds of the world's work, they earn only 10 % of the world's total income and own only 1 % of the world's property, Rwomire (1992). In other words women throughout the world live in severe conditions of poverty. While women are generally disadvantaged in comparison to men, African women specifically, are poorer than women elsewhere. Olusi(1997) estimated that 80% of the poor in Africa were women compared to 70% in Asia and only 45% in Latin America.

Though Africa made impressive economic progress in the 1990s, introduction of economic programs like the structural adjustment programs, reversed many of Africa's accomplishments. Despite a decade of reforms in many African countries, economic growth remains fragile and poverty perpetuating. There has been little progress in reducing absolute poverty. In sub-Saharan Africa 52% of the people live on less that 1$ a day (in 1995 dollars adjusted for purchasing power parity) Economic Commission for Africa (2001). In 1998, the average monthly expenditure was only $14 a person by the rural poor and $27 a person by the urban poor. According to the 2001 government statistics 59% of rural and 43% of urban people live below the poverty line, Economic Commission for Africa (2001). While poverty affects men, women and children, studies have shown a stronger link between women and poverty. The situation worsens in the case of African women who are among the poorest of the poor. African women suffer from higher levels of both "human poverty", defined as the "deprivation in terms of short lives, illiteracy, and lack of basic services" and "income poverty",United Nations Development Program (1997).

One way to fight poverty is to earn higher income by working. In recent decades, women's labor force participation rates have increased dramatically. However, in most families the women continues to be responsible for a majority of the child care and household activities. Juster & Stafford (1991) find that American women spent over two times as many hours doing housework (30.5 hours/week) as American men (13.5 hours/week). If the women is married and has children under the age of five the average hours spent per week were 22.5 compared to the 9 hours spent by the husband Robinson (1988). Jacobsen (1994) finds that although both men and women spend time in child care activities, women spends four times as many hours than men. Conflicting demands on women's time have been one of the biggest factors in increasing the gender gap in earning and occupations. According to Fuchs (1989), women often drop out of the labor force to meet family obligation,while some others work on a part-time basis.

Understanding the relationship between the number of children born (fertility) and labor supply is of importance for a number of reasons. First, it would help shed some light on the increasing trend in women's labor force participation rates. Does having fewer children leads to an increase in labor-force attachment? Coleman & Pencavel (1993). Second it would also explain the phenomenon of fertility-induced withdrawals from the labor force and lower wages of women. Does childbearing keep women from developing their careers? Gronau (1988), Korenman & Neumark (1992). Third we might be able to observe how devoting more time to child care effects the quality of children, Stafford (1987), Blau & Grossberg (1992). Finally, the relationship might be of special importance to developing countries like South Africa, where women's labor force participation rates are really low and poverty level high.

This paper focuses on the causal link between fertility and labor supply of women in South Africa. We employ the well established instrumental variables strategy based on the sex-mix siblings in families with two or more children using the 2001 South African census data. While a number of studies have been done in this area of research including a recent paper by Guillermo & Galiani (2007)[1], our main contribution is the ability to use a larger sample of mothers. Unlike all other studies which restrict the sample of mothers to

[1]Cruces and Galiani, use the Angrist & Evans (1998) instrumental variable strategy to look at causal relationship between fertility and labor supply in two middle-income Latin American countries, Argentina and Mexico.

female household heads or spouse of male household heads, we include in our sample, all mothers present in a household. This is particularly significant for developing countries like South Africa, India, Mexico etc. where the culture of joint families is prevalent at large. Unlike the U.S. most developing countries have extended families or multiple families living in a household. Joint families can influence decisions of fertility [2] as in the case of South Africa where childbearing and child rearing take place within the larger social unit, often the extended family structure or clan. Unlike the western societies where childbearing decision are considered the exclusive decision of the couple or the woman, in Africa both fertility decision and child care are a collective approach. Large social units include multiple women of childbearing age, mothers-in-law, natal and affinal family and friends. This results in many types of female relationships like sisters-in-laws, cowives, mother-in-law, daughter-in-law, grandmother, granddaughter etc., all of whom have the potential to provide supportive environment for both the mother and the child. Such a family structure can lead to bias in a regular family size-labor supply model. The women in such a family structure can bear children with the expectation that they will receive child care support from other women in the family. She can also expect to receive support in child rearing activities which may lead to better child welfare. This is based on the assumption that all extended family relationships are cooperative. If the relationship is not cordial, women may bear more children to compete with other women and children residing under such conditional do not benefit from additional care.

All of the above will have implications on fertility and child survival and depending

[2] Fertility here refers to the number of children alive and living with the mother in a household.

on the components of the extended family units used in a study, may lead to different results. For example a women living in an extended family structure may decide to have more children while she is employed partly due to the fact that other family members can provide child care support. The implication of living in an extended family is that our instruments may not be able to truly capture the correct effect of family size on female's labor supply. They possible may overstate the effect of family size on labor related variables. It is for all these reasons that we do not restrict the sample to include only heads of a household (or their spouses). Moreover doing so gave me only 526,778 mothers in total. When i included all mothers in the household the sample size increased to 712,266 mothers.

The rest of the paper proceeds as follows. Section 3.2 reviews the relevant literature, section 3.3 describes the data, 3.4 discusses the empirical model used, section 3.5 discusses the OLS and 2 SLS results, and section 3.6 concludes.

2.2 Lit Review

There has been a wide and long standing interest in the connection between child-bearing and labor supply. It is there not surprising that a large number of studies have looked at this relationship (see e.g., Jacob (1962), Cain (1990), Gronau (1973), Heckman (1974), Rosenzweig & Wolpin (1980*a*), Schultz (1990), Angrist & Evans (1998), Aguero & Marks (2008)). Most studies done in this area find a negative relationship between fertility (or family size) and female labor supply. However there is much skepticism regarding the causal interpretation and association between fertility and labor supply. "...although we have a have a number of robust correlations, there are very few credible inferences that can be drawn from them", Browning (1992). Studies by Schultz (1981) and Goldin (1990), have shown that fertility and labor supply are jointly determined. Decisions about how much to work, when to work on one hand and decisions about the number of children and the timing of the birth, on the other hand are viewed as jointly endogenous consequences of the household's utility maximization problem, Jacobsen (1994). Thus the correlation between labor supply or earning and fertility may be misleading to a large extent.

Some empirical studies have tried to solve this endogeneity problem by using a simultaneous equation framework Schultz (1978), Dooley (1993), Moffit (1984), Rosenzweig & Schultz (1985). However implementing this procedure is marked with the difficulty in finding plausible indentifying restriction so that the underlying structural parameters can be recovered. In recent years three different instrumental variable techniques that have been employed.

First, is the use of Twins in first birth as an exogenous and unplanned event. Studies that have used this method include Rosenzweig & Wolpin (1980*a*), Bronars & Grogger (1994), and Gangdharan & Rosenbloom (1992). According to this method, occurrence of twins is randomly distributed with respect to other characteristics that may be related to labor force participation and earning. Therefore, it is possible to measure the effects of exogenous fertility variations using simple statistical techniques. The biggest challenge in using this approach is to find a sufficiently large dataset of mothers who have twins. Rosenzweig & Wolpin (1980*b*) did their study with a sample size of just 87 mothers, as a result most of their estimates were not precise. In the recent past the use of Public Use Microdata Samples (PUMS) have made it possible to have large sample sizes (see study done by

Angrist & Evans (1998)). Another drawback with the twinning methodology is that it does not allow us to measure the effect of the change from no children to one children.

This led to the introduction of a new instrument - "sex mix of the first two children" by Angrist & Evans (1998). Their instrument exploits the widely observed phenomenon of parental sex preference for a mixed-sibling sex composition. Parents who have same-sex sibling have a higher probability of having an additional child. Since sex mix is randomly assigned a dummy using the sex of the first two children of a mother serves as a plausible instrument for women who have at least two children and more.

Last year saw the introduction of an alternative instrument in a study done by Aguero & Marks (2008). Aguero & Marks (2008) propose the use of infertility as a source of exogenous variation in family size. Since infertility affects the number of children a women can have and is randomly assigned, an indicator variable for the infertility status of women of childbearing age is a plausible instrument for childbearing. An advantage to using this new instrument is that it enables one to investigate the differences in labor supply between women with children and without children. Thus one is able to identify the causal effect of having children on female labor force participation for a broader sample of women. Their results show that for women who do not actively control fertility having another child does not effect the women's participation in paid labor.

This study employs the method of Same-sex sibling for the case of South Africa. We prefer this method over the others because of the availability of a large enough sample of mothers who have same-sex sibling. We were restricted in terms of finding a sufficiently large sample of mothers who had first birth twins in South Africa. Due to high prevalence of HIV Aids [3] and other health related problems, the use of infertility as an instrument for the case of South Africa was not preferred.

2.3 Data and Summary Statistics

The sex-mix estimation strategy is implemented using the 2001 Census data on South Africa from the Integrated Public Microdata Series (IPUMS). We use information on labor supply, the sex of mother's first two children, and an indicator for multiple births in 2001. Table 2.1 shows the labor force participation rates and the probability of an additional child for mothers aged 15- 35 and aged 36 - 50 in 2001. Data for 2001 is from the census conducted on October 10, 2001 in South Africa. The data represents a 10% systematically stratified sample of household in South Africa. The sample size is 3,735,655 persons. It includes everyone who was present in the country on the night of 9-10 October 2001. People living in households across the country, as well as those in hostels, hotels, hospitals and all other types of communal living quarters, and even the homeless. The table shows higher fertility and labor supply for women aged 36-50. The mean number of children for women aged 15-35 was 1.97 and 50.9% of women in this age category had two or more children. The labor supply was extremely low for this age group and about a quarter were employed at the time of the census. While their married counterparts in this age group had higher fertility (2.87), employment levels were still low at 29.3% only.

Since there is no retrospective fertility information in the IPUMS data sets other than the number of children ever born, for each household, we match the children to their

[3] According to the South African Ministry of Health, at least 20% of those aged 15-49 are infected with HIV, while the prevalence among pregnant women attending antenatal clinics reached 24.5% in 2001.

mothers. In brief, we first identify the mothers in each household using the variable mother's locations in the household. We then attach all people in the household who had a valid mother's location with the person with the same mother's id. In household with multiple families, detailed relationship codes were used to match the children with the mothers. Unlike most studies done in this area which restrict their sample of mothers to either the head of a female headed household or the spouse of a male headed household, we use all mothers present in the household. This is important from the point of view of South Africa because of the existence of multiple families living in the same household. If we were to restrict our sample to only female headed households or spouse of male household head, we would loose out on a significant number of moms (185,488 mothers). Further we restricted the sample to mothers aged 15 - 35 . We chose the age of 15 compared to 18 as is done in the US related study because in developing countries girls start bearing children at a much younger age. 2.5 % of women aged 14 had given birth to atleast one child, increasing to 21.9 % by the age of 18 [4]. Unlike in western countries, another characteristic feature of developing countries is that children do not leave their house at the age of 18. It is common to see children older than 18 living with their parents This allows me to include in my sample mothers aged 15 - 35 years as the probability of her child over 18 being with her in the same household is higher. In fact a study done by the South African government using 2001 Census found that 41.7 % of youth (aged 14 - 19) live with both parents in a household. This drop to 33.4% of youth aged 20 - 24 living with both parents in the household.

Further we found that in the 2001 census data there are 713,816 mothers in the

[4] According to the 2001 Census published by Statistics South Africa, Stages in the life cycle of South Africans (2001).

sample who had given birth to a child (child was not born dead). Out of those who gave birth to a child, 657,276 had number of children surviving the same as number of children born. However only 346,792 mothers who gave birth to a child had equal number of children born, surviving and living in the household with them. In arriving at the number of mothers in our sample we use the variable number of children in the household. Thus our sample size of 698495 is higher because we do not restrict it to mothers for whom the number of children born = number of children surviving and living with them. Our sample does not take into account the number of children dead or not living with the mother and thus overestimates the sample of mothers.

The empirical analysis is conducted on two sub-samples from this data set. The first includes all women aged 15 - 35 with two or more children (Sample size is 144,643). The second includes only married women aged 15 - 35 years with two or more children. The 2001 married sample is restricted to women who were married or living in union at the time of the Census. There are 90,814 observations in the married sample. This latter sample is particularly interesting in the case of South Africa because marriage plays an important role in the lives of women. "Our parents were not keen to see us educated but very keen to see us married", Barrett, J., Dawber, A. and others (1985). Getting married was the first priority of all young women. It also seemed to create conflicts in time specially for working women. " ..they cook so that the husbands find food ready when they return home. They also clean the house. When they get home after a later night shift, their husbands are usually gone. They leave a mess, the bed unmade. He can't do anything, not even wash the dishes. So she cleans again...its a routine. This poor woman has no chance, at home or at work..” Barrett, J., Dawber, A. and others (1985).

Table 2.2 gives the descriptive statistics, variable definitions for covariates, instruments and the dependent variables. The most relevant covariate for our labor supply model is the variable *More than 2 children*. *Same sex* serves as our first instrumental variable for this indicator variable. Table 2.2 shows averages for the two components of *Same Sex*, the indicators *Two boys* and *Two girls*. 43.9% of all women with two children in 2001 had a third child. 49.9% of all mothers with two or more children had their first two children of the same sex. 48.9% of all women also had a boy as their first child. On the other hand 47.2% of all married women with two or more children had a third children. The average age of both married and unmarried mothers was 28 and 30 years respectively, while each group had their first child at the age of 20. The percentage of women in each group who worked for pay was as low as 27-28% . Average hours worked per week were 43 [5] Married mothers earned a higher annual income compared to their unmarried counterparts. Overall, this table shows similar characteristics between married and unmarried mothers in South Africa.

2.3.1 Fertility and Sex Mix

A number of studies have looked at parental preferences for mixed sibling-sex. Examples include Ben-Porath & Welch (1976). Using the 1970 U.S. Census data they find that 56% of families with either two boys or two girls has a third child. On the other hand

[5] The variable hours worked per week is a sub sample including only those who work and had a valid response for this variable. The entire analysis done has been repeated by creating a new hours worked per week variable which was coded as 0 for all mothers who did not work or did not respond to the questions. The results in either case were comparable. We feel coding this variable to 0 is not correct because a missing response does not imply 0 hours worked by the individual.

only 51% of families with one boy and one girl had a third child. The standard quantity-quality model of fertility, introduced by Becker & Lewis (1973) and Becker & Tomes (1977) can be used to explain the impact of sex mix on fertility. Rosenzweig & Wolpin (1980*b*) further extends the model and explains that, quantity and quality of children interact because parents prefer to have equal quality for each of their children. In these models, households maximize a utility function which includes number of children, a complementary good "child quality" and a composite commodity S. Purchase of inputs and the expenditure of parent's time in home production determine the level of child quantity. Ben-Porath & Welch (1980) establish that, there exists dependency of the tendency to have more children on the sex of the earlier children. Thus parents do care about the sex of their children. These preferences are more observable in less developed countries through length of birth intervals and sex ratios. Angrist & Evans (1998) discuss modeling sex-preferences using state-dependent utility. In brief, suppose a mother already has one or more than one children ($N_x \geq 1$) and she is trying to decide how many additional children to have (N_c).If parents prefer a mixed-sibling sex composition, then have two children of the same sex will reduce the utility from N_x. This in turn raises the marginal utility of N_c and increases the probability that parents will try and have an additional child.

Table 2.3 shows the impact of child mix and sex mix on fertility. In the top panel we show the fraction of mothers with at least one child who had a second child, conditional on the sex of the first child. These numbers shed light on the sex preference in families with one or more children. Results suggest that the probability of having your first child as a girl was almost the same as having a boy. 49.6% of all mothers in the sample have a girl first child. Of those mothers who had a first girl child, 49.7% went ahead and had another child. 50.3% of all mothers who had a first boy child went ahead and had another baby. The results are similar for the case of married women. Only 50.3% of all married women who had a first girl child went ahead with having another child.

The lower panel of the table shows the relationship between the sex of the first two children and the fraction of women who have a third child. Results suggest that women with two children of the same sex are much more likely to have a third child. For example 43% of women who had either two boy or two girls had another child. On the other hand only 41% of women with one boy and one girl had another child. In the married women sample the relationship between sex mix and the probability of having another child is even larger.

We further compare the demographic characteristics of mothers who have same-sex and mixed-sex sibling compositions. This is done to ensure that the regressions of fertility and labor-supply outcomes on the instruments have no casual interpretation. Table 2.4 shows the differences by Same sex for mother's age, age at first birth, race and years of education. Most of the means for both groups are not significantly different from zero at the 5% level for both married and the full sample.

2.4 Empirical Model

2.4.1 Wald Estimates

Before we illustrates how sex mix IV strategy can be used to identify the effect of fertility on mother's labor supply let's first discuss why do we need an instrument in the first place?

In order to use a linear regression the zero-conditional-mean assumption ($E[\epsilon \mid X] = 0$) must hold. Here ϵ is the error or disturbance term in the regression model and X is a vector of exogenous variables. Now in the light of simultaneous determination of the response variable and regressors (commonly known as the problem of endogeneity), this assumption may be violated in economic research. This problem of endogeneity is corrected using the instrumental variables (IV) estimator. A variable is endogenous if it is correlated with the disturbance terms. In other words, in a model:

$$y = \alpha + \beta_1 \ x_1 + \beta_2 \ x_2 + ...\beta_k \ x_k + \epsilon \tag{2.1}$$

x_j is exogenous if Cov $[x_j, \epsilon\ ] = 0$. If this condition is not satisfied the regressor is endogenous and the OLS estimator will not be consistent. So for example in our analysis family size or the number of children a mother has is endogenous with variables related to labor supply. They are simultaneously determined and hence our OLS estimators will be inconsistent. One way to get around this problem is to use Instrumental variables. In order to arrive at consistent estimates, the IV must satisfy two properties: the instrument z must be uncorrelated with ϵ but must be highly correlated with family size or the number of children a mother has.

Now that we have discussed the importance of using an IV let's start with a simple regression model:

$$y_i = \alpha + \beta\ x_i + \epsilon_i \tag{2.2}$$

where y_i measures labor supply of the mother and x_i is our measure of endogenous fertility. z_i represents the binary instrument, *Same Sex*. The IV estimate of β can then be represented by a reduced form relationship between y_i and z_i and between x_i and z_i.

$$\beta_{IV} = (\overline{y_i} - \overline{y_0})/(\overline{x_i} - \overline{x_0}) \tag{2.3}$$

where $\overline{y_i}$ is the mean of y_i for all observations with z_i =1 and a similar definition is used for all the other terms above.

As shown by Imbens & Angrist (1994), β_{IV} measures the average effect of x_i on y_i for all mothers whose fertility has been affected by their children's same sex mix. The β_{IV} can be interpreted as the local average treatment effect specific to the instrument z_i.

2.4.2 Method of Two-Stage Least Squares Estimation

In the last section we saw how the instruments identify the effect of children on labor supply. In this section we will discuss regression models relating labor market outcomes to fertility and a variety of exogenous covariates. In specific we will discuss two-stage least-squares (2SLS) and Ordinary least-squares (OLS) estimates of these regression models. 2SLS is preferred for the following reasons. First, it controls for exogenous covariates, thereby leading to more precise estimates if the treatment effects are roughly constant across groups. Second, we can control for any separate additive effects of child sex when using *Same sex* as an instrument. Since *Same sex* is an interaction term involving the sex of the first two children, it could be potentially correlated with the sex of either child. To illustrate, let f_1 and f_2 be indicators for boy firstborn and second-born children. The instrument can then be written as:

$$\textit{Same Sex} = f_1 f_2 + (1 - f_1)(1 - f_2) \tag{2.4}$$

If we assume that child sex is independent and identically distributed (i.i.d) over children, then the population regression of *Same sex* on either f_i produces a slope coefficient equal to 2E[f_i]-1, which reduces to zero only if E [f_i]=1/2 [6]. In our sample the probability of giving birth to a girl child is 0.50, therefore there is no positive association between *Same sex* and the sex of each child. The correlation would be a cause of concern if the sex of each child would effect the mother's labor supply for reasons other than family size. Studies like Morgan, Lye & Condran (1988) have shown that the sex of the child affects the father's commitment to the family. Sometimes parents change their treatment towards children depending on the sex of the child Butcher & Case (1994) Thomas (1994). A study done by Angrist, Imbens & Rubin (1996) shows that boys have a higher probability of being disabled than girls. Thus a parent who has a disabled child might alter his-her behavior towards the child. All these sources could lead to variable bias in the regressions. One way

[6] See Angrist & Evans (1998) for the proof.

to avoid it is by including f_1 and f_2 as regressors in the estimation equations.

Finally, 2SLS estimation allows us to decompose our instrument *Same sex* into two instruments, leading to an over identified model. We can use two separate indicators, *Two girls* [(1-f_1)(1-f_2)] and *Two boys*[f_1 f_2] as potential instruments.The advantage of having two potential instrument is that we can see the difference caused by any bias from secular effects of child sex on labor supply as labor supply consequences of childbearing are likely to be independent of whether Same sex equals two boys or two girls. We can test for over identification using the Hansen J test.

Labor-supply variables for moms are linked to the endogenous variable *More than 2 children* and other exogenous covariates including additive effects for the sex of each child, by the following regression models:

$$y_i = \theta_0\prime\ D_i + \theta_1\ f_{1i} + \theta_2\ f_{2i} + \gamma\ x_i + \epsilon_i \tag{2.5}$$

where D_i is a vector of demographic variables like age of the mother,age at first birth, indicators for race, indicator for whether or not living in an urban area and years of schooling. f_{1i} and f_{2i} are indicators for the sex of the first two children of mother i. When we only use one instrument like the *Same sex*, our model is just identified and the first-stage equation linking *More than 2 children* to sex mix is:

$$x_i = \pi_0\prime\ D_i + \pi_1\ f_{1i} + \gamma(Same\ sex_i) + \eta_i \tag{2.6}$$

γ is the first stage effect of the instrument.

If however we use as our instruments, the two components of *Same sex - Two boys* and *Two girls*, we must drop either f_{1i} or f_{2i} from the covariates list to avoid problems of linear dependence. We then estimate the following equation:

$$y_i = \theta_0 \prime \ D_i + \theta_1 \ f_{1i} + \gamma x_i + \epsilon_i \tag{2.7}$$

The first-stage relationship between x_i and sex mix is then given as:

$$x_i = \pi_0 \prime \ D_i + \pi_1 \ f_{1i} + \gamma_0 \ (Two \ boys_i) + \gamma_1 \ (Two \ girls_i) + \eta_i \tag{2.8}$$

here *Two boy* = (1-f_{1i})(1-f_{2i}) and *Two girls* = f_{1i} f_{2i}

2.5 Results

2.5.1 2SLS Results

Since the disturbance term ϵ cannot be observed, we cannot directly test the assumption of zero correlation between z and ϵ, known as the orthogonality assumption. However we can easily test the second assumption by regressing the included regressor *More than 2 children* on the instrument *samesex* and its components *Two Boys, Two Girls*. If γ from the above equation (2.7) is not statistically different from zero, then *Samesex* or its components are not a valid instrument.

Table 2.5 reports the first-stage results linking sex-mix and fertility. These estimates show that women in 2001 with same sex children are estimated to be 1.5 percentage points more likely to have a third child in a model with covariates. Married women have a higher estimate at 2.2 percent. Both estimates are statistically significant. The other covariates used in the model are age of the mother, age of the mother at the time of first birth, dummies for whether or not the mother was a black, white, asian or South African, years of schooling and whether or not she lived in an urban area. All coefficient are statistically significantly different from 0 at 10%. However rejecting the null of irrelevance is not sufficient to imply that the instrument is not weak.

The results also provide some evidence of an association between having a male child and reduced childbearing at higher parities. If a mother has a first male child then she is 0.5% less likely to have another child. In the case of married women this number is similar. However when we enter the two regressors *Two boys* and *Two girls* separately in the regression, the effect of *Boy 1st* is explained entirely by the difference in the two. In other words, there is no relationship between *Boy 1st* and fertility when the effects of sex mix are allowed to differ by sex. The *Boy 1st* effects remain significant for all specifications, but they are smaller in comparison to the effects of the sex mix.

The coefficient of 0.0109 for all mothers shows that if a mother has first two male children then she is 1.1% more likely to have another child. On the hand, if a mother has two girls, she is 1.9% more likely to have a third child. The numbers are all significant at the 5% level. In the case of married women the impact of same-sex children is stronger. If a married women has two girls then she is 2.9% points more likely to have a third child compared to a 1.7% chance if she has two boys. The F-statistics of all specifications are high of the OLS are high.

We further use the sex-mix to estimate the effect of *More than 2 children* on employment and earnings of mothers in 2001. Table 2.6 reports the results from the OLS estimates and two set of 2SLS estimates, one using *Same sex* and the other using *Two boys* and *Two girls* as instruments. The exogenous regressors are the same as in Table 2.5 (coefficients are not reported). In the model which uses *Two boys* and *Two girls* as instruments, we dropped the variable *Boy 2nd* from the list of covariates. The first three columns are for the entire sample of mothers between the age 15 and 35 and the last three columns are for married mothers of the same age.

Our OLS estimates for the full sample show that the presence of more than 2 children reduces the probability of working by about 7 percentage points and causes the total number of hours per week to drop by only 0.31 hours. Total income earned reduces by 2370 South African Rands. The presence of more than 2 children reduces the probability of working of a married women by 9 percentage points. Hours per week worked reduce by 0.45 hours and income earned by the mother drops by 2755 Rands. Total family income drops by 0.003 Rands only. All OLS coefficients are significant at the 5% level and are precisely estimated.

The first set of 2SLS estimators use the instrument *Same sex*. In the full sample, the coefficients on none of the labor supply variables are statistically significant. Our 2SLS results for both the sample indicate lower probability of working if the mother (aged 15 - 35 having same sex siblings) has more than two children. Hours worked per week decrease by 3 hours when more than 2 children are born to mothers who have the first two children of the same sex. Having more than two children is associated with having more total personal income for both the samples.

Results of 2SLS using *Two boys* and *Two girls* as instruments are reported in columns (3) and (6) of Table 2.6. For the full sample of mothers, having more than 2 children reduces the probability of working for pay by 5.8 percentage point. For married women, the drop in probability is lower 3.4 percentage points. Hours worked per week reduce by 3 for the entire sample and by 4.3 hours for married mothers.Unlike earlier results the two instruments only capture as expected the effect of more than two children on the income of married mothers. Having more than 2 children is associated with lower personal income for the married mothers. We find that the 2SLS results shown in Table 2.6 do not provide any additional benefit by separating the instrument *Same sex* into its two components. Overall our 2SLS estimates were not significant and did not provide better insight (compared to the OLS) into the effect of childbearing on women's labor supply.

Finally we also report the over-identification test statistics. The Hansen J statistics tests for over-identification of all instruments. (Values reported in square brackets in Table 2.6). The p-values for the two samples suggest that the instruments are appropriately uncorrelated with the distribution process. It also indicated that it does not matter which instrument is used. However, compared to using just one *Same sex* instrument, results from using two components of the Same sex instrument were slightly better.

In our regression model due to the suspected failure of the zero-conditional-mean assumption we turn to IV estimation for the sake of consistency. However turning to IV estimation for consistency leads to inevitable loss of efficiency. According to Wooldridge "[there is an] important cost of performing IV estimation when x and ϵ are uncorrelated; the asymptotic variance of the IV estimation is always larger, and sometimes much larger, than the asymptotic variance of the OLS estimator", Wooldridge (2006). In order to test the appropriateness of OLS and the necessity to resort to IV, we use the Durbin-Wu-Hausman (DWH) test. We want to see if the included endogenous regressor be appropriately treated as exogenous. This test involves fitting the model by both the OLS and IV approached and comparing the resulting coefficient vectors. The null hypothesis is that the OLS estimator is consistent and fully efficient. The test statistics is distributed as $\chi\hat{2}$ with k_1 degrees of freedom. It is best to interpret this test not as a test for the endogeneity or exogeneity of regressors, rather as a test of the consequence of using different estimation methods on the same equation. Under the null hypothesis that OLS is an appropriate estimation technique, if we turn to IV, only efficiency will be lost, however the point estimated should be qualitatively unaffected. The DWH test statistics reported in Table 2.6 (p-values) fail to reject the null in both the samples. The small test statistics indicate that the estimation equation with OLS does not yield inconsistent results. Thus the regressor *More than 2 children* cannot be considered endogenous in the fitted model.

2.5.2 Other Issues

We further check for the robustness of the results, the generality of the results and the validity of the instruments. Due to the random nature of the sex mix one might expect the results to be robust to any changes in basic set of covariates. Using the same data, we estimated models adding the following covariates to the vector of demographic variables: linear and quadratic term in mother's age, linear and quadratic terms in mother's education, a set of dummies for the different provinces and linear and quadratic terms in age of the mother at first birth. The estimates from the model including all of the above were: Worked for pay, -0.0136(0.1977); Hours/Week -3.259(5.952); Total Income 13230(29556.6).

The results are also robust to sample sizes of different ages. For example we expanded the sample to include all women in the age group of 18 to 45 years. Increasing the age limit would drop some women with two or more children from the sample because the oldest child might no longer be living with the mother. However while the sample size increased to 286,725 results were similar to what we found in table 2.6. The 2SLS estimates (standard errors) of *More than 2 children* using the instruments *Two boys* and *Two girls* for the following variables were: *Worked for pay*: -0.0437 (0.1185) ; *Hours/Week* : -3.320 (3,453); *Total Income*: 11510 (18949.98). All these values are within 5% of the corresponding values in Table 2.6.

Exclusion restriction can also be violated if the parents in South Africa have a strong son preference. This could affect the sex composition of children, either through stopping rules or selective sex abortion. However the evidence for our sample rules this concern out. In 2001 4.5 million infants were in the age group of 0-4. Both Male and Female constituted 50% each of the infant population. Statistics South Africa (2001). Basu & Das (2001) also point out that some societies exhibit a strong son preference because of the gap between son's and daughter's "ability to contribute to the physical, emotional and financial well being". In the case of South Africa, infant sex ratios are close to 1. The proportion of girls and boys aged 0-2 attending preschool is equal. In the age of three and four years, a slightly higher proportion of girls (19% at the age of three and 23.7% at the age of four) than boys (18.3% at the age of three and 22.8% at the age of four) were attending a pre-school Statistics South Africa (2001). Similar patterns were also observed for higher age groups (for example, among 13-years old, 55.3% of girls had completed grade 6 or higher compared with 54.2 % of boys). Further evidence shows that the proportion of male infants reported as disabled were slightly higher than the proportion of female infants for all age.

Finally, we also controlled for the sex of the child to see if the results were any different when used *Boy 2nd* compared to *Boy 1st*. It turns out that the 2SLS estimates were invariant to the inclusion of the regressors that controlled for the sex of each child. For example, the coefficient (standard error) on *Boy 2nd* in the Worked for pay model is -0.0342 (0.1926) for the full sample and -0.0345 (0.1289) in the married women sample.

2.6 Conclusions and Future Work

Compared to the 1950s where less that 20% of all women were workers, today, almost half the workforce comprise of women. Various studies have tried to explain this rise in labor force using the forces of demand and supply. For example Jacob (1962) concluded that 90% of the post-war rise in labor-force participation of women was due to increase in demand. This led to an increase in the total number of hours worked by women between 1850 to 1980. Goldin (1990) on the other hand explains the rise in labor-force participation of women during the period 1960 - 1980. In recent years the research question on causal link between fertility and labor-force participation of women has made some progress with the introduction of various types of instrumental variables.

While in most studies the OLS estimates confirm a negative relationship between fertility and labor supply, 2SLS and IV estimates give mixed results and sometimes with low predictive power. In this paper we have used one such instrument, same-sex siblings as introduced by Angrist & Evans (1998) on the 2001 census data for South Africa. We use a sample size of 144,643 women and 90,814 married women between the age of 15 and 35 and those who having two or more children. A priori, one would expect that if a mother has two children of the same sex she would want to have another child due to mixed-sex sibling preferences. This exogenous variation in fertility would would lead to a drop in the probability of working, the number of hours worked per week and total income for the mother.

Our OLS results are all significant and have the expected signs. Results for married mothers are stronger in general for all specifications. OLS results show that if a mother with same-sex siblings has more than two children it reduces the probability of working between 8-9% on average. Hours worked per week, total income all reduce by a statistically significant amount for both married and the full sample.

2SLS results using *Same sex* as instrument and the two components of *Same sex*: *Two boys* and *Two girls* seemed to have low predictive power for both the specifications. The Wald F-Statistics test confirms that the instruments used are not weak. Further the Hansen J Statistics rejects over-identification of instruments. For the 2SLS model with only Same sex as the instrument, we find mothers who with same-sex siblings reduce the probability of working by 4 percentage points when they have more than two children. For married mother the probability reduces by 5 percentage points. Hours worked reduce anywhere between and 3-4 hours per week. The instruments do not fair well in explaining the impact on total income earned by the mother. The estimates are insignificant. A final Hausman test fails to reject the null of consistent and fully efficient OLS estimators.

Further specifications test show that the results are robust even when the basic covariates are changed a little bit. The results also do not alter with the use of different indicators for sex like *Boy 1st* and *Boy 2nd*. In general our OLS estimates suggest that there is a negative impact of family size on the probability of working and hours worked per week for mothers who have their first two children of the same sex. The effect is stronger if the women is married. The DWH test supports the OLS specification over the IV specification. One of the reasons why *Same Sex* or its components do not play a significant role in explaining the impact of family size on labor related variables for a mother in South Africa are the presence of extended family support in a household. The presence of grandparents or other female relations in the household may influence the fertility decisions of a women in developing countries. Thus instrumental variables like *Same Sex* may be weak instruments in capturing this relationship. Other unobserved variables like career ambitions could play an important role in explaining both female employment and the number of children.

To summarize, the OLS estimates consistently show a negative relationship between the number of children a women has and labor supply variables in South Africa. The IV estimates show the expected effect but are statistically insignificant. The data in this paper lends support to the OLS estimates compared to the IV estimates. While the family size is negatively associated with labor supply decisions in South Africa, preference for mixed sex composition does not seem to effect the mother's labor related decisions.

Table 2.1: Fertility And Labor Supply Measures

Sample	2001 PUMS
Women aged 15-35	
Mean Children ever born	1.97
Percent with 2 or more children	50.99
Percent Employed	25.52
Observations	283,630
Women aged 36-50	
Mean Children ever born	2.80
Percent with 2 or more children	77.39
Percent Employed	39.81
Observations	248,273
Women aged 18-35 with 2 or more children	
Mean Children ever born	2.67
Percent with more than 2 children	42.19
Percent Employed	27.02
Observations	144,643
Married women aged 18-35 with 2 or more children	
Mean Children ever born	2.87
Percent with more than 2 children	45.62
Percent Employed	29.33
Observations	90,814

Notes: Married sample include women married or living in union at the time of the Census. Source: South Africa Census 2001, Integrated Public Use Microdata Series International: Version 5.0. Minneapolis: University of Minnesota, 2009.

Table 2.2: Descriptive Statistics, Women aged 15-35 with 2 or more children

	Means and Standard Deviations	
	2001 IPUMS	
Variable	All Women	Married Women
Children ever born	2.69 (1.269)	2.88 (1.285)
More than 2 children(=1 if Mother has more than 2 children, 0 otherwise)	43.97 (0.496)	47.21 (0.499)
Boy 1^{st}(s1)(=1 if first child was boy)	0.489 (0.499)	0.492 (0.499)
Boy 2^{nd} (s2) (=1 if second child was boy)	0.483 (0.499)	0.489 (0.499)
Two boys(=1 if first two children were boys)	0.246 (0.430)	0.248 (0.431)
Two girls(=1 if first two children were girls)	0.252 (0.434)	0.249 (0.432)
Same sex(=1 if first two children were the same sex)	0.499 (0.500)	0.500 (0.500)
Age	29.7 (3.980)	28.4 (4.298)
Age at first birth(Mother's age when the first child was born)	19.85 (3.628)	19.06 (3.294)
Worked for pay(=1 if worked for wages or was self employed)	0.268 (0.443)	0.287 (0.453)
Hour/week(average hours worked per week)	42.89* (14.546)	42.34** (14.12)*
Annual total income	10644.75 (60603.24)	13305.95 (66125.17)
Number of observations	144,643	90,814

Notes: The sample includes women aged 15-35 with two or more children. Married sample include women married or living in union at the time of the Census. *Sample size =39,078. **Sample size = 26,634. Source: South Africa Census 2001, Integrated Public Use Microdata Series International: Version 5.0. Minneapolis:University of Minnesota, 2009.

Table 2.3: Fraction of Mothers that had another child by Parity and Sex of Children

	All women 2001 IPUMS (144,643 obs.)		Married Women 2001 IPUMS (90,814 obs.)	
Sex of the first child For mothers with more than one children	Fraction of sample	Fraction that had another child	Fraction of sample	Fraction that had another child
(1) One Girl	0.496	0.497	0.505	0.503
		(0.001)		(0.006)
(2) One Boy	0.505	0.503	0.495	0.497
		(0.001)		(0.006)
Difference (2) - (1)	——	-0.019	——	-0.0118
		(0.002)		(0.003)

	All women 2001 IPUMS (144,643 obs.)		Married Women 2001 IPUMS (90,814 obs.)	
Sex of first two children For mothers with two or more children	Fraction of sample	Fraction that had another child	Fraction of sample	Fraction that had another child
One boy, one girl	0.479	0.420	0.491	0.445
		(0.005)		(0.007)
Two girls	0.253	0.439	0.249	0.480
		(0.007)		(0.009)
Two boys	0.246	0.429	0.248	0.469
		(0.007)		(0.009)
(1) One boy, one girl	0.479	0.421	0.492	0.445
		(0.005)		(0.007)
(2) Both same sex	0.499	0.434	0.498	0.467
		(0.005)		(0.006)
Difference (2) - (1)	——	0.012	——	0.011
		(0.009)		(0.009)

Notes: Sample includes women aged 15-35 with 2 or more children. Married sample includes women married or living in union at the time of the Census. One Girls refers to dummy = 1 if the mother's 1st child was a girl; One Boy refers to dummy = 1 if the mother's first child was a boy; One boy, one girl refers to the fraction of mothers whose first 2 children were a girl and boy. Two girls refers to the fraction of mothers whose first 2 children were girls. Two boys refers to the fraction of mothers whose first 2 children were boys. Both same sex refers to the fraction of mothers whose first 2 children were of the same sex. The samples are the same as in Table 2.2. Standard errors are reported in parentheses. Source: South Africa Census 2001, Integrated Public Use Microdata Series International: Version 5.0. Minneapolis: University of Minnesota, 2009.

Table 2.4: Differences in Means for Demographic Variables by Same Sex

	All Moms			Married Moms		
Variables	Same Sex	Same Sex Not	Difference	Same Sex	Same Sex Not	Difference
Age	29.714	29.721	0.0720	30.265	30.290	0.317
	(0.0144)	(0.0145)	(0.0206)	(0.0168)	(0.0169)	(0.0237)
Age at first birth	19.856	19.851	-0.0460	20.286	20.277	-0.009
	(0.0133)	(0.0136)	(0.0190)	(0.0174)	(0.0176)	(0.0248)
Black	0.784	0.782	-0.002	0.718	0.718	-0.000
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
White	0.074	0.075	0.001	0.110	0.110	0.000
	(0.0009)	(0.0010)	(0.0013)	(0.0014)	(0.0014)	(0.0020)
Asian	0.028	0.031	0.002	0.041	0.045	0.003
	(0.0006)	(0.0006)	(0.0008)	(0.0009)	(0.0098)	(0.0013)
South African	0.113	0.118	-0.001	0.128	0.126	-0.002
	(0.0011)	(0.0011)	(0.0016)	(0.0015)	(0.0015)	(0.0022)
Years of Education	8.201	8.201	0.117	8.285	8.301	0.022
	(0.0149)	(0.0154)	(0.2137)	(0.0190)	(0.0193)	(0.0270)

Notes: The sample includes women aged 15-35 with two or more children. Married sample include women married or living in union at the time of the Census. The samples are the same as in Table 2.2. Standard errors are reported in parentheses. Age refers to the age of the mothers; Age at first birth refers to age of the mother at first birth; Black is a dummy = 1 if the mother's race is black, 0 otherwise; White is a dummy = 1 if the mother's race is white, 0 otherwise; Asian is a dummy = 1 if the mother's race is Asian, 0 otherwise; South African is a dummy = 1 if the mother's race is South African, 0 otherwise; Years of Educations refers to the years of education of the mothers; Same sex is a dummy = 1 if the first two children of the mother are of the same sex. Same sex not refers to the dummy same sex = 0. Difference refers to the difference in the means for demographic variables by the dummy same sex. Source: South Africa Census 2001, Integrated Public Use Microdata Series International: Version 5.0. Minneapolis: University of Minnesota, 2009.

Table 2.5: OLS estimates of More than 2 children equation

2001 IPUMS

Independent variables	All Mothers			Married Mothers		
	(1)	(2)	(3)	(4)	(5)	(6)
Boy 1^{st}	—	-0.0055	-0.0013	—	-0.0055	0.0002
		(0.0023)	(0.0022)		(0.0028)	(0.0040)
Boy 2^{nd}	—	-0.0043	—	—	-0.0059	—
		(0.0023)			(0.0028)	
Same sex	0.0143	0.0150	—	0.0227	0.0235	—
	(0.0026)	(0.0023)		(0.0033)	(0.0028)	
Two Boys	—	—	0.0109	—	—	0.0176
			(0.0033)			(0.0040)
Two girls	—	—	0.0192	—	—	0.0292
			(0.0032)			(0.0040)
With other covariates	No	Yes	Yes	No	Yes	Yes
R^2	0.0002	0.2256	0.2256	0.0005	0.2615	0.2615
F-Stata	30.69	5722.88	5722.76	47.36	4562.99	4562.85
N	144643	144643	144643	90814	90814	90814

Notes: The sample includes women aged 15-35 with two or more children. Married sample include women married or living in union at the time of the Census. Other covariates in the model are indicators for age, age at first birth, Black, Asians, White, South African, years of Education, urban dummy. Boy 1st refer to a dummy = 1 if the mother's first child was a boy, 0 otherwise; Boy 2nd refers to a dummy = 1 if the mother's second child was a boy, 0 otherwise; Same sex refers to dummy = 1 if the first two children of the mother are of the same sex, 0 otherwide; Two Boys refers to dummy = 1 if the mother's first two children were boys, 0 otherwise; Two girls refer to dummy = 1 if the mother's first two children were girls, 0 otherwise. The variable Boy2nd is excluded from columns (3) and (6). Standard errors are reported in parentheses. OLS refers to Ordinary Least Squares. N refers to number of observations in the regression. F-stats refers to F statistics.

Source: South Africa Census 2001, Integrated Public Use Microdata Series International: Version 5.0. Minneapolis: University of Minnesota, 2009.

Table 2.6: OLS and 2SLSL estimates of labor Supply Model using 2001 Census Data

	All Mothers			Married Mothers		
	(1)	(2)	(3)	(4)	(5)	(6)
Estimation method	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Instrument for More than 2 children	—	Same Sex	Two Boy , Two Girls	—	Same Sex	Two Boy, Two Girls
Dependent variable:						
Worked for pay	-0.080	-0.043	-0.058	-0.092	-0.050	-0.034
	(0.002)	(0.136)	(0.134)	(0.003)	(0.114)	(0.113)
			[0.666]			[0.585]
Durbin-Wu Hausman Stats		0.670	0.753		0.678	0.552
N	144,643	144,643	144,643	90,814	90,814	90,814
Hours/Week	-0.301	-3.105	-3.070	-0.455	-3.523	-4.390
	(0.176)	(5.144)	(5.143)	(0.211)	(4.669)	(4.603)
			[0.598]			[0.253]
Durbin-Wu Hausman Stats*		0.664	0.469		0.668	0.344
N	39,078	39,078	39,078	26,634	26,634	26,634
Total Income in thousands of Rands	-2367.4	16992.97	8398.31	-2782.26	4791.30	-2066.27
	(333.33)	(20980.7)	(20141.39)	(452.14)	(18698.79)	(18154.1)
			[0.100]			[0.125]
Durbin-Wu Hausman Stats		0.244	0.462		0.443	0.717
N	144,643	144,643	144,643	90,814	90,814	90,814

Notes: The Table reports estimates of the coefficient on the *More than 2 children* variable in equations (4) & (6) in the text. The sample includes women aged 15-35 with two or more children. Married sample include women married or living in union at the time of the Census. Worked for pay is a dummy = 1 if the mother worked for pay, 0 otherwise; Hours/Week refers to the number of hours worked per week by the mother the week prior to the census day. Observations for this variable are fewer in both the sample because it includes only valid responses i.e. response from those who worked. The variable is not coded to 0 if the mother did not respond. Similar analysis has been done with the variable coded to 0 for all those who did not respond and the results are close to what we have here. Total income is annual personal income of the mother in rands from all sources for the twelve months prior to the census. In all samples the data are recoded to the midpoints of the broad intervals given in the original data. The top interval is coded to its lowest possible value (e.g, code 360001 for 360,001+). Other covariates in the models are age, age at first birth, plus indicators for Boy 1^{st}, Boy 2^{nd}, Blacks, Asians, Whites, years of education and urban dummy. The variable Boy 2^{nd} is excluded from equation (6). The p-value for the test of over identifying restriction associated with equation (6) is shown in square brackets. Standard errors are reported in parentheses. N refers to the number of observations in the regression.
Durbin-Wu Hausman Stats shows the p-value for the Hausman test associated with exogeneity of the instrument.

Source: South Africa Census 2001, Integrated Public Use Microdata Series International: Version 5.0. Minneapolis: University of Minnesota, 2009.

Chapter 3

The Effect Of Immigration On Ethnic Composition And Occupational Reallocation

3.1 Introduction

Over the last 30 years, increased immigration has been one of the key factors characterizing labor markets in the United States. The foreign born adult population has increased from 9% in 1980 to over 17% in 2007. This immigration has not been skill neutral. We see in Table 3.1 that the share of High School Dropout (HSD) workers who are foreign born has increased from 12% in 1980 to 44% by 2007. At the same time, we see in Table 3.2 that the wages of HSD workers have declined (both in absolute terms and when compared to other skill groups). Has this increase in low skilled immigration contributed to the decline in the wages of native low skilled workers?

The answer to these questions remain unclear, and the literature has typically found moderate to no effects of immigration on natives' wages (see Table 3.3). There are two potential factors that determine the impact on wages paid to native workers. The first is the degree of substitutability between a native-born and a foreign-born worker in performing particular productive tasks (i.e. will immigration affect occupational wages). The second is whether native workers respond to immigration by shifting occupations (i.e. does labor supply respond to these changes in occupational wages). Another possible response of Black workers could be to exit from employment. We exclude this option in our paper, thus estimating only the effect of increased Hispanic concentration on the occupational choice of Blacks, conditional upon choosing an occupation. Future work could consider the effect on Black exit from employment. However, an "unemployed' category could not be directly included in the present model as this would imply that Blacks could be "crowded out' of unemployment by an increasing Hispanic concentration among the unemployed. This does not necessarily bias [1] our results but that we are just finding a different effect, i.e. on occupational mobility, not on mobility in and out of the labor market. A higher concentration of Hispanics will typically mechanically create a smaller *concentration* of Blacks, however variation in the LHS variable of the regression due to overall growth or decline in the total number of jobs in the occupation will mitigate this mechanical correlation. A more detailed model would sequentially model the selection of Blacks into employment, and then after this decision, model their occupational choice.

[1] By not including the unemployment category our estimates might overstate the results of crowding out

Here, we shed some light on the second question by examining whether occupational transitions may affect the labor market outcomes of native workers, and to what extent these shifts can be attributed to immigration. We focus primarily on Black workers, as this group is over represented among HSD workers (14% of that group in 1980 while only 10% of the total population) and has often been posited as a group likely to be affected by low skilled immigration. We first perform a shift-share analysis to see how changes in Blacks' occupations have affected their wages. We then analyze correlations between the *occupational importance* (share among all Blacks who work in an occupation) and *occupational concentration* (share of workers in an occupation who are Black) and the *occupational concentration* of Hispanic workers. We find that occupational changes had a major impact on the wages of Black's, but that little of these changes in the occupational distribution of Black workers can be attributed to competition with Hispanic workers. Rather, our results are suggestive that most occupational shifts during this period were due to changes in labor demand that affected both groups similarly in magnitude. In other words, there were some occupations which saw a decrease in demand for workers in general. For example manufacturing might see a drop in demand for workers since a lot of manufacturing is being done outside the U.S.

The rest of the paper proceeds as follows. Section 3.2 reviews the relevant literature, section 3.3 describes the data, section 3.4 discusses the simple regression model used and the counterfactual wage analysis, section 3.5 discusses the results, and section 3.6 concludes and discusses scope for future work.

3.2 Lit Review

A seminal paper in the immigration literature was Card's (1990) paper. Using the impact of an exogenous shock (the Cuban government relaxing the restriction on out-migration) to identify the effect of immigration on wage and unemployment rates of less-skilled workers, he finds little evidence of adverse effects. The influx of Cubans in the early 1980s increased Miami's labor force by 7%. Surprisingly, the study found that the influx of Mariel immigrants had virtually no effect on the wages of less-skilled non-Cuban workers. Further, there was no increase in unemployment rates among the non-Cuban or Black workers. Rather it seemed that Miami's labor market was able to absorb the excess workers without changing either the wage rates or unemployment levels. This study spurred a number of other papers based on similar exogenous variations.

Kugler & Yuksel (2008) also found no wage effects on low skilled natives using a massive outmigration from Central America due to the Hurricane Mitch. Their OLS results showed that native wages are positively related to the recent influx of Latin Americans. However after controlling for out-migration, the native wage effect disappears and less-skilled employment of previous Latin American immigrants fall. Friedberg (2001) looks at the increased immigration to Israel after emigration restrictions were lifted in an unstable Soviet Union. Using least square estimates, she finds that occupations that employed more immigrants had lower native wage and employment growth compared to the others. However since this distribution of immigrants across occupations was not independent of relative labor market conditions, she re-estimates her results using an Instrumental Variable approach and finds that she cannot reject the hypothesis that mass migration of Russians to Israel did not affect the earnings or employment of native Israelis. Other papers using exogenous variations include those by Butcher & Card (1991) and Angrist & Kugler (2003).

More recently there has been a growing literature which looks at the mechanism through which immigration effects native wages. The paper by Peri & Sparber (2009) provides theoretical and empirical evidence that natives will change the occupations to which they supply their labor in response to immigration. They argue that production involves different labor skills. While less educated immigrants have a comparative advantage in manual and physical tasks, they have a disadvantage when it comes to communication and language intensive skills. Natives (with similar education levels) on the other hand have a comparative advantage in communication skills. Differences in skills lead to differences in specialization by immigrants and native-born workers. An increase in the supply of immigrants for manual tasks leads to a gradual shift by natives into language intensive jobs. This may either raise the return to communication skills, or partially offset a negative effect.

Peri & Sparber (2009) find evidence that less-educated immigrants are more likely to work in a manual job compared to a job which involved more communication skills. States which had a larger flow of less educated immigrants saw a higher transfer of natives from manual to communication tasks. This led to lower wages in manual task intensive occupation and a higher supply of natives in occupations demanding language skills. As a result, the wage loss (due to a shift in occupation by native workers) was much smaller than what was predicted by the model in which natives and immigrants were perfect substitutes. They also find that immigration of less-educated workers only reduces average real wages paid to less educated US-born workers by 0.2% between 1990 and 2000. Had there been no task specialization, the wage loss would have been 1.2%

In addition to the immigration literature, we also build on work done on occupational transition. A paper by Autor & Dorn (2008) describes a more than 50% rise in the share of hours worked in service occupations between 1980 and 2005 among workers with high school or lower education. Real hourly wages also increased by 20 log points in the same service occupations, exceeding wage growth in other low-skill occupations. They hypothesize that the rise in demand for service work is due to changes in task specialization induced partly by technical changes. They argue that primary job tasks of service occupations are difficult to automate or outsource since they require interpersonal and environmental adaptability as well as direct physical proximity. This in turn does not allow for substitution in outputs of service occupation, leading to rising wage and employment in service occupation.

Autor & Dorn (2008) study the determinants of employment and wages in services from 1950 to 2005 in 722 commuting zones in the US. They compare and contrast the period from 1980 to 2005 (during which rapid adoption of information technology took place) with a previous period from 1950 to 1980. They argue that if commuting zones differ initially in their share of employment in routine-intensive occupations, markets with higher routine share will see larger increases in service occupation employment and greater polarization of earnings between high and middle-skill worker as time progresses. Assuming goods and services are sufficiently complementary, their model implies that wages in service occupations will rise along with service employment. Their results show that there is reallocation of labor activity in response to exogenous shocks (technology, in their case). All of these papers point towards a process of employment and wage polarization within regional labor markets.

While a number of studies find moderate to no effect of immigration on native's wages, we add to the existing literature by examining whether native workers respond to immigration by shifting occupations. In other word we examine if the occupational distribution of Black workers is important in terms of their wages? And can "crowding out' types of effects explain some of the change in occupational distribution of Black workers.

3.3 Data and Summary Statistics

The primary sources of data for this study is the 5% public use sample of the decennial census (1980, 1990 and 2000) and the 1% sample of the population from the American Community Survey (2005, 2006 and 2007). Table 3.4 shows the summary statistics of the variables used in the analysis. Average age of all individual were 38 years. Hourly wage over all years for all occupations was $17 per hours. Table 3.5 shows the most common occupations in the sample. Our wage data trims the top and bottom 5% of the wages from the sample. Trimming took care of outliers and measurement error in usual hours and weeks worked. We restrict the data to include only workers in the age group of 18 to 64, both inclusive. While our data includes all Hispanics in the population, we did re-run our analysis using just Hispanic immigrants and got very similar results.

In Figure 3.1 we see an increase in the percentage of Hispanic workers both by male and female category over the period 1980 - 2007. Figure (c) shows the rise in the percentage of Hispanics compared to Black workers over the period 1980 to 2007. Further in Tables 3.6 through 3.9, we begin to look for evidence of competition between Blacks and Hispanics in the labor market. We begin by looking at the largest occupations (occupations with the largest percentage of Black workers) for Black workers in both 1980 and 2007. In Table 3.6, we see that these were (1) Freight and Stock workers, (2) Nursing Aides and Orderlies and (3) Janitors. These occupations employed 5.9%, 4.3% and 4.2% of Black workers, respectively. Looking forward to 2007, Table 3.7 shows that the importance of these occupations to the Black labor force changed significantly over time. Freight and Stock workers were now the 4th largest occupation and employed only 3.6% of Black workers. Nursing Aides and Orderlies were now the largest occupation, employing 6.1% of this group, and Janitors had fallen from 4.2% of employment to 2.6% of employment. A further look at wages for these occupations show that not only did the percentage of Blacks working as Freight and Material Handlers drop from 1980 to 2007, their corresponding wages also fell from \$13.46 an hour to \$12.84 an hour for Blacks. Nursing Aide and Orderlies saw an increase in the wage of Black workers while Black Janitor's wage dropped marginally over the same period.

We then turn to Figure 3.2 to see if there is any evidence that these changes in occupational importance may be negatively correlated with employment growth for Hispanics in these same occupations. Figure 3.2 (a) and .2(b) show the increase in concentration and importance of female Hispanic workers in the occupation "Housekeepers". A similar trend can be seen in Figures 3.2 (c) and (d), which show a large increase in the concentration and importance of Hispanic men in the occupation "Construction". We further turn our attention to some numbers in Table 3.8, we see that Freight and Stock workers was the occupation that experienced the second largest decline for Black workers over this period. However, the occupation declined in strikingly similar magnitudes for Hispanics also. Janitors faced the 5th largest decline for Blacks but grew slightly for Hispanics, consistent with a story of task re-specialization. Of these most declining occupations, we see that 8 of them declined for Hispanics as well. While the two occupations that moved in opposite directions for the two groups (Janitors and Housekeepers) are consistent with task re-specialization, this is the some preliminary evidence that the bulk of occupational transition is driven by outside factors that affect both Black and Hispanic workers, rather than competition between the two groups on the labor supply side. Nursing Aides and Orderlies do not show up in this table, as it is an occupation that grew during this period. But comparing the second columns of Tables 3.6 and 3.7, we can see that this occupation increased in importance for Hispanics as well as Blacks.

Table 3.9 shows us the wage change for Black occupations which decreased the most (as discussed above). Machine operators which saw the largest drop in Black and Hispanic workers also had a moderate decline in Black and Hispanic wages. Freight and Material Handlers saw the biggest drop in wages for both Blacks and Hispanics. What was interesting to see was that while Blacks workers as Janitors dropped in numbers (and Hispanic increased) the wage change was in the opposite direction. Wage of Black janitors increased while those of Hispanic janitors fell.

We further illustrate the occupation changes for each race by gender and age level. In the case of men, Table 3.10, occupations like machine operators, freight and material handlers and fabricators and assemblers, saw the largest drop Black men workers. These occupations also saw the largest drop in Hispanic men workers, thereby re-instating that these groups are similarly affected by common demand shocks Wages in the above three occupation dropped for both Black men and Hispanic men during the period 1980 to 2007.

We also look at the smallest occupation change for each of the four races by age. We have divided the population into two groups, those between the age of 18 and 35 and those above the age of 35 years. Age might be a determining factor in deciding whether or not one should change their occupation. For example a person who is younger (18 to 35) might have more flexibility and hence might be able to change occupations faster and easily compared to someone who is older (above 35 years). Table 3.12 shows the results for Blacks aged 18 to 35.The biggest drop in occupation were for Secretaries and Stenographers (3.13%), Machine operators (2.53 %) and Fabricators and Assemblers (2.51 %).

The final analysis done was for High School Dropouts (HSD)(Table not shown). A person counts as HSD if he has less than 10 years of education. About 9% of all Black HSD worked as freight, stock and material handlers in 1980 at an hourly wage of $13.01. During the period 1980 to 2007, those occupations which saw a fall in the number of Blacks, saw a comparable increase in the number of Hispanics. For examples while the number of Black HSD working as housekeepers, maids and butlers dropped from 1980 to 2007, it increased in numbers for Hispanics. As there are different changes for these subgroups we can in the future explore whether there are heterogenous effects of increase in Hispanic concentration on importance for these groups.

3.4 Empirical Model

In the last section, we looked to see if there was some preliminary evidence consistent with occupation reallocation. We define occupation reallocation as the shift (or reallocation) of occupations by a race. For example, if Blacks shift their occupation from being Janitors to Truck drivers due to either fall in wages in that occupation or due to higher competition from Hispanics, we would consider this as occupation reallocation. In this section we describe the empirical tools we will use to rigorously test for occupation reallocation. First, we motivate the importance of occupational reallocation through a shift share analysis. After finding how occupational re-allocation affects wages, we then turn our attention to correlating changes in occupational importance and concentration between Black and Hispanic workers to see if some of the Black occupational reallocation (and associated wage changes) can be attributed to immigration. In other works, we seek to answer two questions:

1. Has change in the occupational distribution for Black workers been important in terms of their wages?

2. Can we explain some of the change in the occupational distribution with "crowding out' types of effects?

To answer the first question, we analyze how wages have changed at the occupational level and how this change has affected overall wages. To do this, first find the average wage for Black workers in 1980. Essentially, this is finding a weighted average of occupational wages for this group, where weights are the share of the group employed in each occupation. Denote this by $\overline{Y80}$. The corresponding value for 2007 is denoted by $\overline{Y07}$. Finally we construct a counterfactual average wage for 2007, taking a weighted average of occupational wages using 1980 occupational weights. We denote this as $\overline{Y07^o}$. The difference between $\overline{Y07^o}$ and $\overline{Y80}$ shows us how much the average wage would have changed if the group's labor supply would not have changed at all between occupations. This change in wages is thus attributable only to changes in occupational wages and not to the changing occupational distribution of workers (the *within* estimate). Similarly by comparing $\overline{Y07}$ with $\overline{Y07^o}$, we obtain a *between* estimate, i.e. how much of the change in wages was due to shifts between occupations. Additionally, we use a kernel density estimation to perform a similar exercise across the entire distribution of wages.

To briefly summarize the results of our shift-share analysis, we find that wages for Blacks rose from \$14.37 in 1980 to \$16.45 by the mid-2000. This increase in wages for Blacks was 46% more due to occupational mobility. In the case of Hispanics,wages rose from \$13.98 to \$14.65.The between effect was marginal (\$0.19) compared to the Blacks.

To estimate whether there is a Hispanic crowding out effect, we begin by defining two shares:

$$S1_{ijt} = N_{ijt}/N_{jt} \tag{3.1}$$

$$S2_{ijt} = N_{ijt}/N_{it} \tag{3.2}$$

$S1$ captures the fraction of workers who are Black (or any other race) amongst all workers in occupation j, year t. We will refer to this as a measure of *occupational concentration*. $S2$ captures the fraction of Black(or any other race) workers in occupation j, year t amongst Black (other race) workers in all occupations in year t. We will refer to this as a measure of *occupational importance*.

We then estimate the following models:

$$S2_{ijt} = \alpha + \beta \; S1_{i'jt} + \epsilon_{ijt} \tag{3.3}$$

$$S2_{ijt} = \alpha + \beta \; S2_{i'jt} + \epsilon_{ijt} \tag{3.4}$$

$$\Delta S2_{ijt} = \alpha + \beta \Delta S1_{i'jt} + \epsilon_{ijt} \tag{3.5}$$

$$\Delta S2_{ijt} = \alpha + \beta \Delta S2_{i'jt} + \epsilon_{ijt} \tag{3.6}$$

Where:

$$
\begin{aligned}
i &= \text{Black, Hisp, White} \\
i' &= \text{Not i} \\
j &= \text{all occupations considered} \\
t &= 1980, 1990, 2000, 2005, 2006 \text{ and } 2007 \\
N_{ijt} &= \text{Number of Black/Hisp/White workers in occupation j in year t} \\
N_{jt} &= \text{Total number of workers in occupation j in year t} \\
N_{it} &= \text{Total number of Black/Hisp/White workers in year t} \\
\Delta S2_{ijt} &= S2_{ijt} - S2_{ijt-1} \\
\Delta S1_{ijt} &= S1_{ijt} - S1_{ijt-1}
\end{aligned}
$$

The first two equations (3.3 and 3.4) will be used to estimate the effect of Hispanic occupational concentration and Hispanic occupational importance, respectively, on Black occupational importance. These first two models are expressed in levels and will be estimated separately for all the years. Equations 3.5 and 3.6 regress changes in the LHS variable on changes in the RHS variable and are more arguably causal. In order to interpret the estimates from equation 3.5 as causal, the following assumption must hold: unobserved shocks to changes in importance of the occupation to Blacks are uncorrelated with changes in the occupational concentration of Hispanics. A higher concentration of Hispanics will typically mechanically create a smaller *concentration* of Blacks, however variation in the LHS variable due to overall growth or decline in the total number of jobs in the occupation will mitigate this mechanical correlation. A similar identifying assumption must also hold for equation 1.6. We hope to better identify these coefficients in future work, however we believe that estimating descriptive correlations in this case is informative as it describes trends in shifts in the occupational labor supply of these two groups.

Below we will focus primarily on equations 3.5 and 3.6. The coefficient estimate from equation 3.5 will be interpreted as a measure of the magnitude of crowding out between Hispanics and Blacks: what is the effect of a 10% point increase in Hispanic occupational concentration on Black occupational importance? The coefficient estimate from equation 3.6 will be interpreted as a measure of how correlated changes in occupational importance are for the two groups: when the share of the Hispanic labor force working in an occupation increases by 1% point, by how much does this increase for Blacks?

3.5 Results

To summarize our results, we find three things: (1) occupational reallocation has been important for Black wages, (2) there seems to be significant crowding out effect of Hispanics on Black occupational choice, and (3) there is a clear strong and positive correlation in the occupational importance for the two groups, suggestive that labor demand shocks affecting both groups explain most of the occupational reallocation that has taken place over the last 27 years. Below, we present these results in more detail.

3.5.1 Reallocation and Wages

Table 3.13 shows the results of our shift share analysis for Blacks, Hispanics, Whites and All others. Actual wages for Blacks rose from $14.37 to $16.45. This is a $2.08 increase in wages between 1980 to the middle of this decade. In 2007 the counterfactual wage per hour was $15.79, thereby implying that even if the share of Blacks remained constant in an occupation, wages would have risen by $1.42. This implies a between effect of 66 cents. Thus occupational mobility increased the wage growth by an additional 66 cents on what would have been only a $1.42 increase with a static occupational distribution. Thus the wage increases 46% higher than they would have been without the occupational mobility and 32% of the wage gain is attributable to the between effect. This is consistent with Blacks being able to re-specialize to avoid competition with Hispanics by shifting occupation. It is also consistent with re-specialization for any other number of reasons.

In the case of Hispanics, wages rose from $13.98 to $14.65. This is a $0.67 increase in wages between 1980 to the middle of this decade. In 2007 the counterfactual wage per hour was $14.46, thereby implying that even if the share of Hispanics remained constant, wages would have rise by $0.48. This is a between effect of 19 cents. Thus the wage increased 40% more than they would have been without the occupation mobility and 28% of the wage gain is attributable to the between effect. A further look at the kernel density figures (See Figures 3.3 and 3.4) allow us to see the actual and counterfactual wages for the entire wage distribution. They show us the between effect is more pronounced for the Blacks (than the Hispanics) and is prevalent across the entire distribution for Hispanics.

From these results it seems like occupational reallocation matters in terms of determining average wages. The regressions in the next section will tell us if occupational reallocation is because of cross ethnic labor market competition.

3.5.2 Evidence on Causes of Reallocation

In Table 3.14 we present our regression results. Column (1) show results of the regression of Black occupational importance on Hispanic occupational concentration for each of the 6 years. When doing the regression in levels, the coefficients are negative for 1980, 1990 and 2000 but positive for the subsequent years. They are significant only for 1980 and 2007. Similarly, column (2) shows results of the regression of Black occupational importance on Hispanic Occupational importance. The coefficient is negative and significant for all but the year 1980. While these results are suggestive of crowding *out*, they do not account for occupational fixed effects and are thus likely highly endogenous. Once we include the occupational fixed effects, we find we find estimates of substantial (or significant) crowding out. Column (3) show our point estimate of -.052 implying that an occupation where Hispanics grew from 20% of the occupation in 1980 to 30% of workers in 2007 would have declined in occupational importance to Blacks by approximately 5% points (for example, the occupation would have declined from employing 5% of Black workers to employing no Black workers). We will assess the magnitude of these effects in a moment, but first we turn our attention to the regressions that correlate occupational importance for the two groups.

We now look for more evidence of crowding by running "importance on importance" regressions, column (4). A negative coefficient would suggest that as a larger share of the Hispanic workforce moves into a particular occupation, Blacks move out of this occupation. This would be consistent with *occupational Balkanization*. A positive coefficient is evidence that the primary reason for changes in the occupational distribution are driven by labor demand shocks that affect both groups. A coefficient of +1 would indicate that these two groups are affected homogenously by labor demand shocks. Our regressions yield coefficients between .778 (column 2) and .770 (column 4), providing strong evidence that the dominant theme in changes to the occupational distributions of Black and Hispanic labor are labor demand shocks that effect each group similarly.

Turning our attention back to the economic significance of the crowding out effect, we present Table 3.15. As in Table 3.8, we present the 1980 level and 1980 to 2007 changes for the 10 fastest declining occupations for Blacks. Using our estimated coefficient of -.052 in the "crowding out" regression and the observed change in Hispanic concentration in each of these occupations, we estimate the implied decline for Blacks and the share of the decline that can be explained by the increased concentration of Hispanic workers. We find large effects. The median amount of decline explained by increased Hispanic concentration in these occupation is 42% of the total decline.

3.6 Conclusion

Over the past few decades the U.S. labor market has been flooded with lower skilled immigrants. While foreign born workers who were High School Dropouts more than tripled between 1980 and 2007, native born HSD workers decreased in numbers. The labor market literature found little empirical evidence to the claim that immigrants depress wages of similarly skilled workers. In this paper, we access whether the impact of immigration is mitigated by occupational transition of natives. In particular, we focus on the labor market outcomes of Black workers because they are over represented among HSD and they are likely to be affected by low skilled immigration.

The results of the shift-share analysis show that the wage increase was 46% higher than what it would have been without occupational mobility. Similarly, 32% of the wage gain is attributable to the between effect indicating the Blacks are able to re-specialize to avoid competition with Hispanics by shifting their occupations. The point estimates from the regression results are negative, statistically significant and large in magnitude. A 10% point increase of Hispanics in an occupation would decline the occupational importance of Black by approximately 5% point. We also find strong evidence that the dominant theme in changes to occupational distribution of Black and Hispanic are labor demand shocks that effect each group similarly.

The analysis in the above sections describes the correlations between the *equilibrium* occupational distributions for Black and Hispanic labor. Without a suitable instrument, this correlation measures both how Black labor supply is affected by changes in the labor supply of Hispanic, as well as how Black labor supply is affected by demand shocks common to the two groups. For future work we would like to use an Instrumental variable to isolate the causal effect of shifts in Hispanic labor supply on the occupational distribution of Black workers. A variable that affects Hispanic occupational labor supply but is uncorrelated with Black labor demand (or Hispanic labor demand, to be safe) will serve as a relevant and valid instrument.

To summarize our results we find three things. One, occupation reallocation has been important for Black wages. Two, there seems to be substantial (or significant) crowding out effect of Hispanics on Black occupational choice. And three, there is a strong and positive correlation in the occupational importance for the two groups, suggesting that common labor demand shocks have affected both groups and explain most of the occupational reallocation over the past three decades.

Table 3.1: Total Share of Immigrant Population in each Skill Category

	High School Dropout	High School Graduate	Secondary College	College Graduate
1980	0.12	0.05	0.06	0.08
1990	0.22	0.07	0.08	0.10
2000	0.34	0.10	0.10	0.14
2005	0.42	0.14	0.11	0.16
2006	0.42	0.15	0.11	0.17
2007	0.44	0.15	0.11	0.17

Note: For education category the following criterion is used. A worker is a: High School Dropout if education < 10 years, High School Graduate if education $= 10$ years, Secondary College if education > 10 and education < 14 and College Graduate if education > 13 years. Source: Public Use Microdata Sample and American Community Survey.

Table 3.2: Mean wages for natives in the working age by education (2007 dollars)

	High School Dropout	High School Graduate	Secondary College	College Graduate
1980	14.51	15.64	15.94	23.33
1990	13.52	15.30	16.86	25.51
2000	13.77	16.07	18.07	28.28
2005	13.50	16.03	18.66	30.47
2006	13.02	15.45	17.73	28.77
2007	13.32	15.74	17.97	29.28
Pct change 1980 2007	-8.19	0.65	12.73	25.55

Note: For education categories the following criterion is used: High School Dropout if education < 10 years, High School Graduate if education $= 10$ years, Secondary College if education > 10 and education $<$ 14 and College Graduate if education > 13 years. Pct Change refers to Percentage change from the year 1980 to 2007. Source: Public Use Microdata Sample and American Community Survey.

Table 3.3: Prior Elasticity Estimates

Author	**Labor Market**	**Variation**	**Elasticity**	**Outcome**
Altonji & Card	U.S.	Lagged Shares	-1.2	wage
Card (1990)	Miami	Boarlift	≈ 0	wage & empl
Kugler & Yuksel (2008)	U.S.	Hurricane	$+.062$ to $+.125$	wage
Hunt (1992)	France	Pied Noir Repatriation	$-.82$	wage
Angrist & Kugler (2003)	E.U.	Balkan Wars	$-.02$ to $-.07$	empl
Friedberg (2001)	Israel	Perestroika	$+.718$	wage
Borjas (2005)	U.S. Ph.D.s	Cohort	$-.31$	wage
Borjas (2003)	U.S. low-skilled	Cohort	$-.947$	wage

Note: This table summarizes the results found by other select studies. Outcome columns shows the variables effected by the exogenous variation as explored in each study, wage and or employment(empl). Source: Public Use Microdata Sample and American Community Survey.

Table 3.4: Summary Statistics

Variable	**Mean**	**(Std. Dev.)**	**Min.**	**Max.**	**N**
Age	37.963	(12.33)	18	64	19134375
Hourly Wage	17.148	(9.311)	4.205	55.619	19134375
Highest Grade Completed	13.191	(2.818)	0	22	19134375
Male Indicator Variable	0.522	(0.5)	0	1	19134375
Immigrant Indicator Variable	0.111	(0.314)	0	1	19134375
African-American Indicator Variable	0.095	(0.293)	0	1	19134375
Hispanic Indicator Variable	0.09	(0.286)	0	1	19134375
White Indicator Variable	0.771	(0.42)	0	1	19134375

Note: The summary statistics is across all the years in the sample. Hourly Wage is in 2007 dollars. Male Indicator variables = 1 if the worker is a male, 0 otherwise. Immigrant Indicator variables = 1 if the worker is an immigrant, 0 otherwise. African-American Indicator variables = 1 if the worker's race is African-American (Black), 0 otherwise. Hispanic Indicator variables = 1 if the worker's race is Hispanic, 0 otherwise. White Indicator variables = 1 if the worker's race is White, 0 otherwise. Source: Public Use Microdata Sample and American Community Survey.

Table 3.5: Most Common Occupations

Most Common Occupations 1990 basis (3 digit code)	**Freq.**	**Percent**	**Cum.**
Executives and Managers (22)	105,264	8.52	8.52
Secretaries and Stenographers (313)	43,609	3.53	12.05
Primary School Teachers (156)	37,487	3.03	15.08
Supervisors and Proprietors (243)	35,661	2.88	17.96
Truck and Tractor Drivers (804)	34,380	2.78	20.74
Nursing aides and Orderlies(447)	29,614	2.40	23.14
Retail Sales Clerk(275)	28,366	2.29	25.43
Freight and Material Handler (29)	27,422	2.22	27.65
Registered Nurses (95)	26,383	2.13	29.78
Cashiers (276)	24,951	2.02	31.80

Note: Freq refers to frequency. Cum refers to cumulative percentage. Column 1 gives the name of the most occupation with the maximum number of workers over all years in descending order. Column 2 gives the absolute number of workers in each occupation over all years. Column 3 gives total number of workers in each occupation as a percentage of the all workers over all years. Column 4 refers to the cumulative percentage obtained from column 3. Source: Public Use Microdata Sample and American Community Survey.

Table 3.6: Largest Black Occupations in 1980

	Rank	Blacks	Black (%)	Hisps	Hisp (%)	Wage Black	Wage Hisp
Freight and Material Handlers	1	502,960	5.87	296,020	5.89	13.46	13.03
Nursing aides and orderlies	2	363,940	4.25	81,740	1.63	11.45	10.99
Janitors	3	358,560	4.19	153,460	3.05	12.00	12.13
Truck and tractor drivers	4	326,940	3.82	162,280	3.23	15.47	15.58
Secretaries and Stenographers	5	323,920	3.78	191,300	3.81	12.80	12.27
Fabricators and Assemblers	6	323,300	3.78	235,140	4.68	15.49	13.62
Machine operators	7	316,860	3.70	212,900	4.23	14.86	13.38
Housekeepers, maids and butlers	8	310,640	3.63	99,440	1.98	10.26	9.93
Executives and Managers	9	288,700	3.37	198,100	3.94	18.16	17.71
Cooks, variously defined	10	219,680	2.57	94,040	1.87	10.76	10.93

Note: Hisp refers to Hispanic Workers. Column 1 refers to number of Black workers in a particular occupation in 1980, Column 2 reports the number of Black workers as a percentage of all workers in each occupation in 1980, Columns 3 and 4 do the same for Hispanic workers, Column 5 and 6 report the wages for Blacks and Hispanics respectively for each corresponding occupation. Source: Public Use Microdata Samples and American Community Surveys.

Table 3.7: Largest Black Occupations in 2007

	Rank	Blacks	Black (%)	Hisps	Hisp (%)	Wage Black	Wage Hisp
Nursing aides and Orderlies	1	907,713	6.13	462,864	2.47	12.62	11.61
Executives and Managers	2	784,662	5.30	854,648	4.57	23.56	22.36
Truck and Tractor drivers	3	554,912	3.74	674,248	3.60	15.67	14.96
Freight and Material Handlers	4	535,148	3.61	723,108	3.86	12.84	11.47
Retail sales clerks	5	497,582	3.36	579,015	3.09	14.45	14.47
Cashiers	6	497,259	3.36	463,148	2.48	10.69	10.45
Cooks, variously defined	7	439,608	2.97	713,821	3.82	10.98	10.38
Janitors	8	383,605	2.59	603,357	3.22	12.49	11.52
Secretaries and Stenographers	9	378,977	2.56	395,635	2.11	16.41	15.53
Customer service reps	10	372,623	2.51	329,781	1.76	14.43	14.04

Note: Hisp refers to Hispanic workers. Column 1 refers to number of Black workers in a particular occupation in 2007, Column 2 reports the number of Black workers as a percentage of all workers in each occupation in 2007, Columns 3 and 4 do the same for Hispanic workers, Column 5 and 6 report the wages for Blacks and Hispanics respectively for each corresponding occupation. Source: Public Use Microdata Samples and American Community Survey.

Table 3.8: Largest Decrease in Employment from 1980-2007

	Rank	AA 1980	AA Change	H 1980	H Change
Machine operators	1	3.70	-2.37	4.23	-2.68
Freight and Material Handlers	2	5.87	-2.26	5.89	-2.02
Housekeepers, Maids and Butlers	3	3.63	-2.22	1.98	0.06
Fabricators and Assemblers	4	3.78	-2.12	4.68	-2.52
Janitors	5	4.19	-1.60	3.05	0.17
Secretaries and Stenographers	6	3.78	-1.23	3.81	-1.69
General office clerks	7	2.30	-1.22	1.86	-0.89
Textile sewing machine operators	8	1.37	-1.18	2.55	-2.10
Farm workers	9	0.87	-0.71	3.42	-1.08
Production supervisors	10	1.32	-0.68	1.96	-1.25

Note: AA refers to African American (Black) workers, AA 1980 refers to share of Blacks as a percentage of total workers in an occupation in 1980, AA change refers to the change in occupational share from 1980 to 2007 for Black workers, H refers to Hispanic workers, similar definitions follow for Hispanics. Source: Public Use Microdata Sample and American Community Survey.

Table 3.9: Wage Changes for Largest Decrease in Employment from 1980-2007 for Blacks

	Rank	AA 1980	2007	Change	H 1980	2007	Change
Machine operators	1	14.86	14.69	-0.17	13.38	12.70	-0.68
Freight and Material Handlers	2	13.46	12.84	-0.63	13.03	11.47	-1.57
Housekeepers, Maids and Butlers	3	10.26	10.98	0.72	9.93	10.32	0.39
Fabricators and Assemblers	4	15.49	15.01	-0.48	13.62	12.72	-0.89
Janitors	5	12.00	12.49	0.49	12.13	11.52	-0.61
Secretaries and Stenographers	6	12.80	16.41	3.61	12.27	15.53	3.26
General office clerks	7	12.79	14.94	2.15	12.09	13.70	1.61
Textile sewing machine operators	8	10.00	12.34	2.34	9.85	9.19	-0.66
Farm workers	9	10.44	10.78	0.35	10.68	9.58	-1.10
Production supervisors	10	18.75	19.07	0.33	17.53	17.57	0.04

Note: AA refers to African American (Black) workers, AA 1980 refers to the wage of Black workers in an occupation in 1980, AA change refers to the change in wage in a particular occupation from 1980 to 2007 for Black workers, H refers to Hispanic workers, similar definitions follow for Hispanics. Source: Public Use Microdata Sample and American Community Survey.

Table 3.10: Largest Decrease in Employment from 1980-2007 for Black Men

	Rank	AAM 1980	AAM Change	HM 1980	HM Change
Machine operators	1	4.75	-2.93	4.69	-2.90
Freight and Material Handler	2	8.82	-2.69	7.17	-2.55
Fabricators and Assemblers	3	4.36	-2.15	4.59	-2.08
Janitors	4	5.85	-1.99	3.97	-0.50
Farm workers	5	1.32	-1.02	4.45	-1.47
Production supervisors	6	2.07	-1.02	2.60	-1.67
Metal and Plastic Workers	7	1.26	-0.95	1.41	-1.12
Precision Metal Workers	8	1.13	-0.74	1.52	-1.13
Construction laborers	9	2.82	-0.70	2.60	3.43
Housekeepers, Maids and Butlers	10	1.04	-0.52	0.68	-0.29

Note: AAM refers to African American (Black) men, AAM 1980 refers to share of Black men in an occupation in the year 1980, AAM change refers to the change in occupational share from 1980 to 2007 for Black men, HM refers to Hispanic men. similar definitions follow for Hispanics. Source: Public Use Microdata Sample and American Community Survey.

Table 3.11: Wage Changes for Largest Decrease in Employment from 1980-2007 for Black Men

	Rank	AAM 1980	2007	Change	HM 1980	2007	Change
Machine operators	1	16.26	15.56	-0.71	14.85	13.36	-1.50
Freight and Material Handlers	2	13.99	13.11	-0.88	13.89	12.10	-1.80
Fabricators and Assemblers	3	17.21	15.73	-1.48	15.57	13.62	-1.95
Janitors	4	12.49	13.02	0.53	12.59	12.27	-0.32
Farm workers	5	10.40	10.59	0.18	10.82	9.76	-1.05
Production supervisors	6	19.93	19.55	-0.38	19.06	18.72	-0.33
Metal and Plastic Workers	7	17.85	15.52	-2.33	15.75	13.49	-2.26
Precision Metal Workers	8	18.07	16.88	-1.19	17.80	16.37	-1.43
Construction laborers	9	14.20	15.03	0.82	15.10	13.13	-1.96
Housekeepers, Maids and Butlers	10	11.96	12.64	0.68	11.96	10.95	-1.01

Note: AAM refers to African American (Black) men, AAM 1980 refers to wage of Black men in an occupation in 1980, AAM change refers to the change in occupational wage from 1980 to 2007 for Black men, HM refers to Hispanic men. Similar definitions follow for Hispanics. Source: Public Use Microdata Sample and American Community Survey.

Table 3.12: Largest Decrease in Employment from 1980-2007 for Blacks aged 18 - 35

	Rank	AA 1980	AA Change	H 1980	H Change
Secretaries and Stenographers	1	5.16	-3.13	4.69	-2.72
Machine operators	2	3.68	-2.53	4.03	-2.68
Fabricators and Assemblers	3	4.13	-2.51	4.77	-2.78
Freight and Material Handler	4	6.30	-1.71	6.29	-2.25
General office clerks	5	2.82	-1.66	2.11	-1.01
Janitors	6	3.24	-1.49	2.41	-0.17
Textile sewing machine operators	7	1.51	-1.42	1.99	-1.67
Housekeepers, Maids and Butlers	8	1.81	-0.86	1.35	0.12
Production Inspectors and Testers,	9	1.21	-0.71	1.21	-0.63
Metal and Plastic Operator	10	0.83	-0.70	0.96	-0.84

Note: AA refers to African American (Black) workers, AA 1980 refers to the share of Blacks aged 18-35 in an occupation in 1980, AA change refers to the change in occupational share from 1980 to 2007, H refers to Hispanics. Similar definitions follow for Hispanics. Source: Public Use Microdata Sample and American Community Survey.

Table 3.13: Average wage of Black workers weighted by actual and counterfactual occupational distribution

	Mean Wage (Weighted by actual occ distribution)	Mean Wage (Weighted by 1980 occ distribution)
BLACK		
1980	14.37	14.37
1990	15.24	14.99
2000	16.45	15.90
2005-2007	16.45	15.79
HISPANIC		
1980	13.98	13.98
1990	14.21	14.13
2000	14.71	14.51
2005-2007	14.65	14.46
WHITE		
1980	16.10	16.10
1990	17.04	16.80
2000	18.58	17.95
2005-2007	19.22	18.47
ALL		
1980	15.80	15.80
1990	16.61	16.42
2000	17.88	17.40
2005-2007	18.27	17.75

Note: Occ refers to occupation. Column 2 gives the mean wage over all occupations of Black workers weighted by their occupational share in the respective years (actual). Column 3 gives the mean wage over all occupations of Black workers weighted by their occupational share in the year 1980 (counterfactual)for each year. For the year 2005, 2006 and 2007, an average wage was taken. Source: Public Use Microdata Sample and American Community Survey.

Table 3.14: OLS and Fixed Effect Regression Analysis

	(1)	(2)	(3)	(4)
	Occ Imp. of Blacks	Occ Imp. of Blacks	Change in Occ Imp. of Blacks	Change in Occ Imp. of Blacks
Occ Conc. of Hispanics in year ==1980	-0.0450			
	(11.25)**			
year==1990	-0.0002	-0.00100		
	(0.53)	(3.08)**		
year==2000	-0.00100	-0.00200		
	(1.22)	(6.10)**		
year==2005	0.00100	-0.00100		
	(1.67)	(5.04)**		
year==2006	0.00100	-0.00100		
	(1.94)	(4.90)**		
year==2007	0.00100	-0.00100		
	(2.12)*	(4.96)**		
Occ Imp. of Hispanics		0.778		
		(40.13)**		
Change in Occ Conc. of Hispanics			-0.0520	
			(2.76)**	
Change in Occ Imp. of Hispanics				0.770
				(6.57)**
Constant			0.00200	-0.00100
			(0.73)	(0.87)
Observations	1670	1670	277	277
Number of Occupation, 1990 basis	281	281		
* significant at 5%; ** significant at 1%				
R-squared			0.110	0.540
Robust t statistics in parentheses				
Absolute value of z statistics in parentheses				

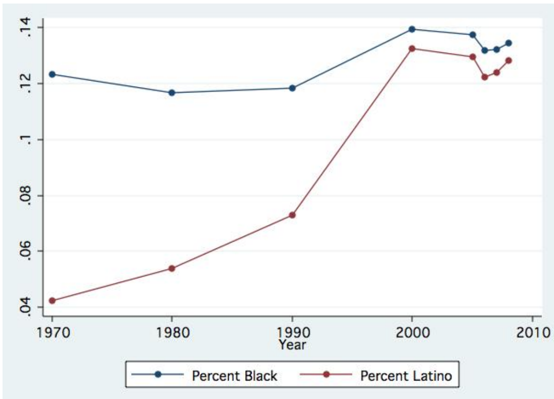
Note: Occ Imp refers to Occupational Importance, Occ Conc refers to Occupational Concentration. Columns 1 and 2 are in levels. Columns 3 and 4 are in changes between 1980 and 2007. No. of Obs include for all the years. Columns 1 and 3 are regressions of Black Occ Imp ($S2_{ijt}$) on Hispanic Occ Conc ($S1_{i'jt}$). Columns 2 and 4 refer to the regression of Black Occ Imp ($S2_{ijt}$) on Hispanic Occ Imp ($S2_{i'jt}$). Source: Public Use Microdata Sample and American Community Surveys.

Table 3.15: Largest Decrease in Employment from 1980-2007 for Blacks

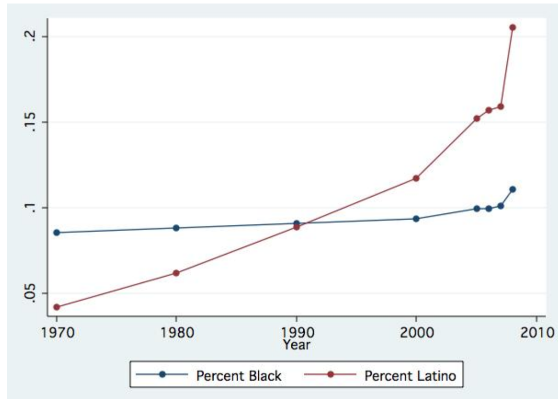
	Rank	AA 1980	AA Change	Change in concentration of Hispanics	Percentage of Crowding Out
Machine operators	1	3.70	-2.37	-0.6233	0.26
Freight and Stock Handler	2	5.87	-2.26	-0.7572	0.33
Housekeepers, maids and butlers	3	3.63	-2.22	-1.3861	0.63
Fabricators and Assemblers	4	3.78	-2.12	-0.5635	0.27
Janitors	5	4.19	-1.60	-0.9751	0.61
Secretaries and Stenographers	6	3.78	-1.23	-0.2949	0.24
General office clerks	7	2.30	-1.22	-0.4787	0.39
Textile sewing machine operators	8	1.37	-1.18	-1.2271	1.04
Farm workers	9	0.87	-0.71	-1.7410	2.46
Production supervisors	10	1.32	-0.68	-0.4250	0.62

AA refers to African American (Black) workers, AA 1980 refers to the share of Black workers in an occupation in the year 1980, AA Change refers to the change in the share of Black workers as a percentage of all workers in an occupation from 1980 to 2007. Column 5 refers to the change in Concentration of Hispanics in each occupation from the year 1980 to 2007. Column 6 refers to the percentage of crowding out explained by the change in concentration of Hispanics in a particular occupation. Source: Public Use Microdata Sample and American Community Survey.

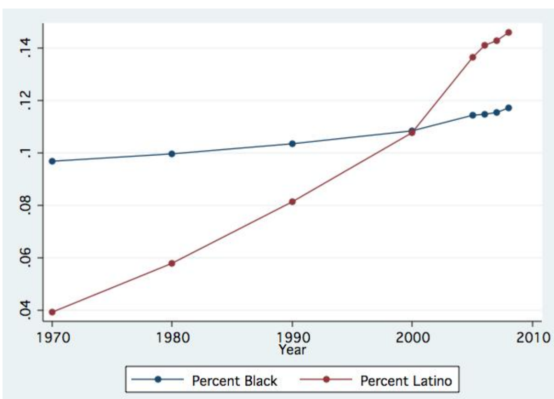
Figure 3.1: Increase in Group sizes



(a) Females



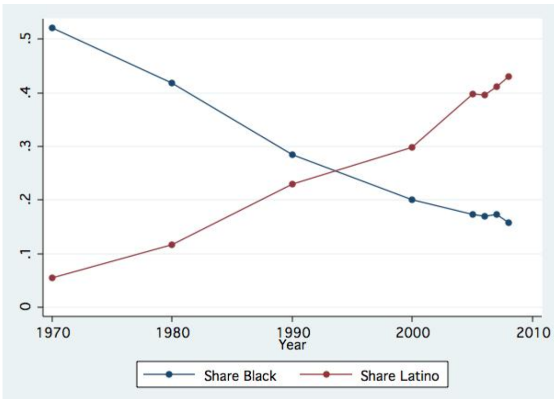
(b) Males



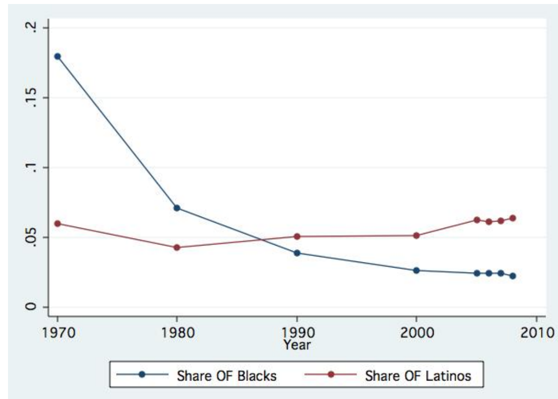
(c) Hispanics

Note: Figure (a) shows the increase in percentage of Hispanic and Black female workers over the period 1980 - 2007. Figure (b) shows the same for male workers in both groups. Figures (c) refers to the total percentage of Black and Hispanic workers in the US labor market during the period 1980 - 2007. Source: Public Use Microdata Sample and American Community Survey.Source: Author's calculations. Source: Author's calculations.

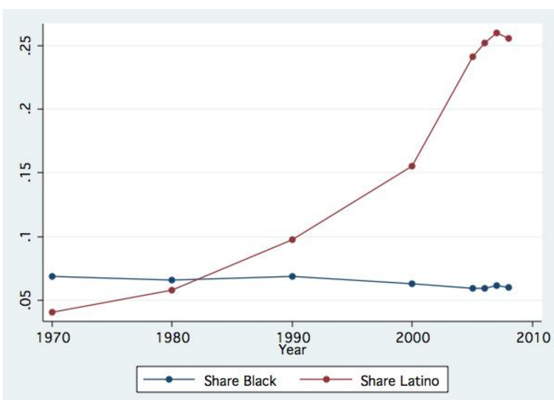
Figure 3.2: Change Consistent with Crowding Out



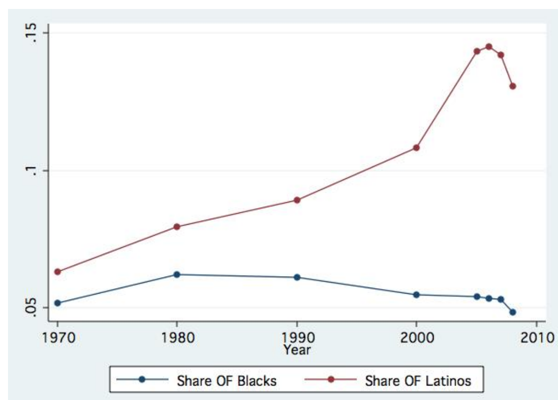
(a) Concentration of Female Housekeepers



(b) Importance of Female Housekeepers



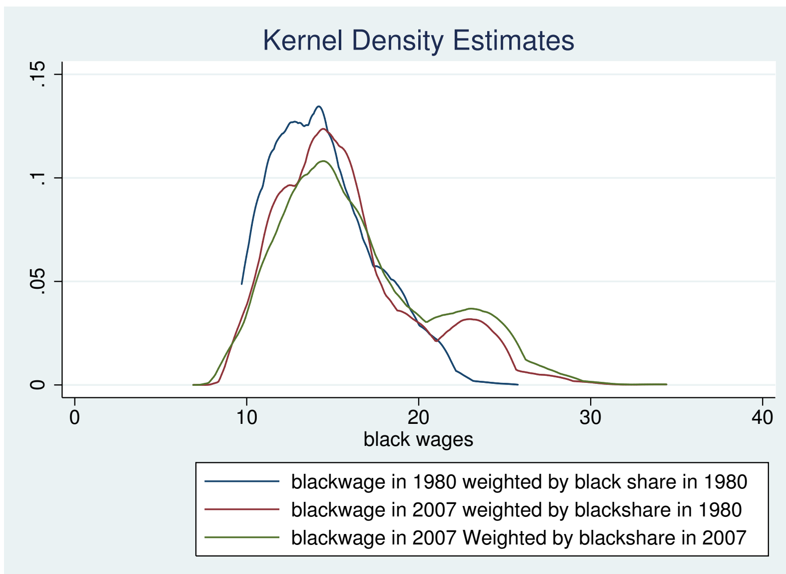
(c) Concentration of Male Construction Workers



(d) Importance of Male Construction Workers

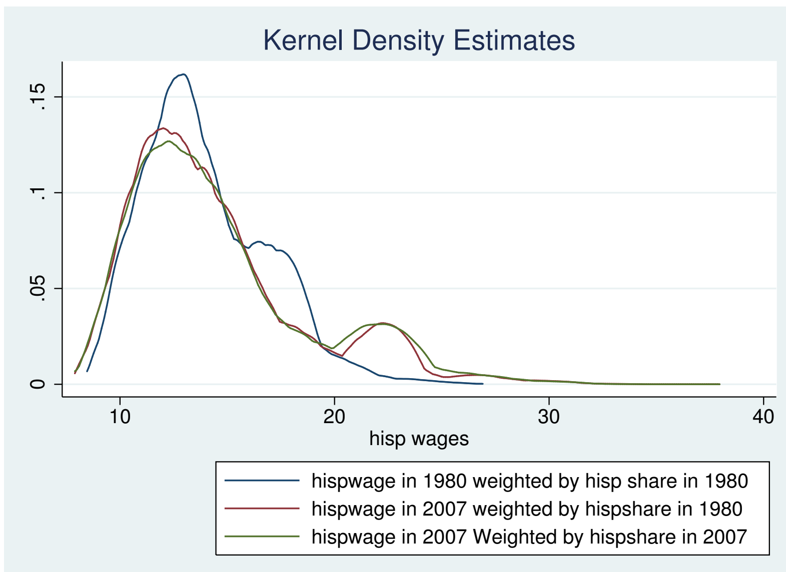
Note: Figures (a) and (b) plot the occupational concentration and importance respectively of females working as housekeepers for both Blacks and Hispanics over the period 1980 - 2007. Figures (c) and (d) do the same for Black and Hispanic men working as construction workers. Occupational Importance refers to the share of an occupation amongst all workers of a race, for example out of all the Black workers how many work as housekeepers? Occupational Concentration refers to the share of a particular race in an occupation. For example out of all housekeepers how many are Blacks? Source: Public Use Microdata Sample and American Community Survey.Source: Author's calculations.

Figure 3.3: Actual and Counterfactual Black wages



Note: The figure here plots the mean wage over all occupations in 1980 and 2007 of Black workers weighted by their occupational share in the year 1980 and 2007 respectively (actual). It also plot the mean wage over all occupations of Black workers in 2007 weighted by their occupational share in the year 1980 (counterfactual). Source: Public Use Microdata Sample and American Community Survey.

Figure 3.4: Actual and Counterfactual hisp wages



Note: The figure here plots the mean wage over all occupations in 1980 and 2007 of Hispanic workers weighted by their occupational share in the year 1980 and 2007 respectively (actual). It also plot the mean wage over all occupations of Hispanic workers in 2007 weighted by their occupational share in the year 1980 (counterfactual). Source: Public Use Microdata Sample and American Community Survey.

Chapter 4

Conclusion

This dissertation is a compilation of three independent empirical chapters. While the first two chapters on Infant Mortality Rates and the impact of fertility on womens labor supply are pressing issues in most developing countries, the third chapter is focused on the impact of immigration on occupational reallocation and ethnic composition. Though developing countries are less plagued with immigrations issues, research done in the third chapter uses common labor-market models and theories which could be applied to any developing nation. For example one might look at the ethnic composition and its impact on occupational reallocation for the case of India. The empirical methods used in the last chapter can be generalized for use in developing countries context. Below are the concluding remarks on each chapter.

The first Chapter titled Infant Mortality Rates in India: District-Level Variations and Correlations examines the correlates of infant mortality in India using district-level data from the 1991 and 2001 Census of India. India being a large country with 28 states and 7 union territories is heterogeneous in demographic, economic and social factors across each state. The States are further divided into 593 districts. The number of districts varies across each state. Using a panel data set of 666 districts, this paper tried to determine which of socio and or economic factors play an important role in reducing infant mortality rates. The explanatory variables used are male and female literacy, male and female labor force participation, the level of poverty, urbanization and other socio-economic variables. We use quantile regression analysis to determine which of these factors impact infant mortality. Quantile regression is preferred over OLS because it allows us to estimate models for the conditional median function, and the full range of other conditional quintile functions and therefore provides a more complete statistical analysis of the stochastic relationship among random variables. The analysis brings out the powerful influence of women's characteristics on infant mortality, especially literacy and labor force participation. Increases in both of these variables significantly reduce child mortality at the district level. Improvements in male laborers in non-agricultural work and reductions in poverty also reduce child mortality, but their quantitative impact is weak in comparison. Further the non-parametric analysis reinforces the results found in the parametric section. They indicate that the action or the impact of the covariates is strongest in the districts which lie in the center of the conditional distribution, rather than those at the extreme. This analysis allows us to determine in which districts the impact of additional target policies would yield the greatest reduction in infant mortality.

Our explanation and analysis in terms mortality decline is just one possible way of looking at the issue of high infant mortality. While we have managed to construct a detailed picture of the present scenario by drawing evidence from diverse available sources, in the absence of comparable data on mortality rates across years and across parities, more detailed information on health facilities and also the extent of sex selection across districts our conclusions remain but a starting point for more detailed analysis.

Our analysis is based on a limited data set and indicates at best certain trends. While son preference is high in India this and other studies indicates resource scarcity not to be the causal factor for discrimination against female children and infants and may even lead to a relative better treatment in the absence of biases. It thus becomes imperative to delve deeper into the causes and correlates of mortality, which again maybe disparate in different regions of incidence through micro surveys and deeper analysis. As a note of caution, while evaluating the implications of the analysis it is also important to keep in mind that fertility and IMR are not determined exclusively by socio-economic conditions and the role of biological factors also have an important role to play. In our analysis we have tried to understand the comparative static effects of socio-economic factors affecting IMR assuming that the biological factors may not be regionally different or at least randomly so.

The study in this thesis has concerned itself with analyzing IMRs wherever data has been available, for the Census years 1991 and 2001. However, in this concluding section it would be important to mention, that women in India are vulnerable at different age cohorts, and high IMR, is an aspect of this vulnerability measured by a lower relative survival rate at the onset of life.

In our second paper titled Same-sex siblings and Womens Labor Force Participation and their affect on mother's labor supply in South Africa., we look at the labor-supply consequences of childbearing for women in South Africa. Fertility is thought to have an important impact on the welfare of women, children, and men. Public policies that can help individuals reduce unwanted fertility are expected to improve the well being of their families and society. But there is relatively little empirical evidence of these connections from fertility to family well-being and to inter-generational welfare gains, traced out by distinct policy interventions. Associations do not clarify the underlying causal relationships which should be the foundation for program evaluations (Moffitt, 2005). As a result it is difficult to evaluate the consequences of such welfare programs, such as family planning, which in turn may influence lifetime behavior of households and their arrangements.

Implementation of a particular policy (e.g. voluntary family planning) plays a crucial role in the responsiveness individuals to a policy-induced change in fertility. Mothers would respond to having fewer children depending on what the alternatives are available in terms of activities. For example If the mother has access to self employment opportunities or schooling options, she could consider having fewer children. Important implications of having fewer children could translate in the form of more productive work by mothers outside of the family or providing their children with better education and health care, or even being able to save more in order to have a better lifestyle in old age.

Using parental preferences for a mixed sibling-sex composition we construct instrumental variables (IV) estimates of the effect of childbearing on labor supply of all women aged 15-35 years having more than two children. The data used for the study is the 10% household sample from the 2001 census. The covariate of interest in the labor supply model is the indictor *More than two children*. Demographic variables include mother's age,

age of mother at first birth, years of schooling, indicators for race and an urban dummy. Labor-supply variables include hours worked per week, worked for pay and total income. Unlike previous studies which restrict their sample to include only female household heads or spouse of male household heads, this study expands the sample size to include all child-bearing women in the household aged 15-35. The IV estimates that exploit the fertility consequences of sibling sex do not perform well compared to the OLS estimates showing that more children lead to lowering of female labor supply. While OLS estimates exaggerate the causal effect of children, 2SLS estimates gave mixed reviews. IV estimates give a stronger indication of causality, they certainly come with a higher price and that is lower predictive power.

Married women with siblings of same sex lowered the probability of working for pay by 5 percentage points upon having a third child. The number of hours worked per week decreased by 4 with an addition of a third child. The instruments, same sex and Two boys and Two girls were unable to explain the impact of family size on income earned by mothers. One of the reasons for the poor predictive power could be the poor quality and insufficient income data.

While a number of empirical studies have reviewed the causal link between fertility and mother's labor market supply, most of the studies are constrained in their analysis of low-income countries due to lack of data. Trade-off between the quantity of children a women bears and the quality of those children is a common feature of high-income common societies. However this inverse relationship could also be caused by unobserved heterogeneity in people's preferences and constraints. Therefore it is important to evaluate fertility-reducing programs in predominantly rural low-income countries to assess whether the impact of such programs contribute to improving the mother's health and the welfare of the child. What would also be interesting and insightful it to look at the productivity of women at home and in the labor market. Such an analysis would help confirm if having fewer children really makes the mother more productive and allocate her time to labor market activities. Confirmation of such a hypothesis could help women specially in poor agricultural societies.

Finally, the overall aim of research on fertility is to overcome challenges in the provision and assessment of population programs that help reduce fertility and improve family outcomes. Policies should be implemented such that they cost-effective way and benefit the least advantaged individuals, generally the rural poor.

The third paper is on The effect of Immigration on Ethnic Composition and Occupational Reallocation. In this paper we aim to assess whether the impact of immigration is mitigated by occupational transition of natives. Being over represented among HSDs, we focus on the labor market outcomes for Black workers. We use data from the 5% public use sample of the census (1980, 1990 and 2000) as well as the 1% sample of the population from the American Community Survey (2005, 2006 and 2007) to estimate the effect of occupational reallocation on the wages of Black workers as well as the effect of immigration on reallocation. A shift-share analysis reveals that occupational transitions caused wages for Blacks to rise by 46% more than they would have with a static occupational distribution. However, we find that these occupational shifts were due to crowding out effect of Hispanics on Black occupations: a 10 percentage point increase in the share of workers in an occupation who are Hispanics leads to a 5 percentage point decrease in the share of Black workers in that occupation. This is significantly large to explain substantially, occupations that declined in importance for Blacks during the period of study. We find a strong correlation between importance of occupations to Hispanic and Blacks, suggesting that most occupational transition for these two groups has not only been driven by outside factors such as trade and technological change, but that these shocks are affecting the two groups similarly.

The analysis in the above sections describes the correlations between the *equilibrium* occupational distributions for Black and Hispanic labor. Without a suitable instrument, this correlation measures both how Black labor supply is affected by changes in the labor supply of Hispanic, as well as how Black labor supply is affected by demand shocks common to the two groups. For future work we would like to use an Instrumental variable to isolate the causal effect of shifts in Hispanic labor supply on the occupational distribution of Black workers. A variable that affects Hispanic occupational labor supply but is uncorrelated with Black labor demand (or Hispanic labor demand, to be safe) will serve as a relevant and valid instrument.

Bibliography

Abadie, A. (2003), 'Semiparametric Instrumental Variable Estimation of Treatment Response Models', *Journal of Econometrics* **113**, 231–63.

Aguero, J. M. & Marks, M. S. (2008), 'Motherhood and Female Labor Force Participation: Evidence from Infertility Shock', *American Economic Review: Papers and Proceedings* **98**(2), 500–04.

Alston, L. (1981), 'Tenure choice in southern agriculture', *Explorations in Economic History* **18**, 211–232.

Altonji, J. & Card, D. (1989), 'The effects of immigration on the labor market outcomes of less-skilled natives', *Princeton Industrial Relations Section Working Paper* **256**.

Altonji, J. G. & Card, D. (1991), 'Immigration, trade, and the labor market', *John. M. Abowd and Richard B. Freeman edts.* pp. 201–234.

Amuedo-Dorantes, C., Bansak, C. & Raphael, S. (2007), The impact of amnesty on labor market outcomes: A panel study using the legalized population survey.

Angelucci, M. (2005), 'Us border enforcement and the net inflow of illegal migration', *IZA Discussion Paper* **1642**.

Angrist, J. (1991*a*), 'Grouped-Data Estimation and Testing in Simple Labor Supply Models', *Journal of Econometrics* **47**(2-3), 243–66.

Angrist, J. (1991*b*), 'Treatment Effect Heterogeneity in Theory and Practice', *The Economic Journal* **114**(494), C52–C83.

Angrist, J. D. & Kugler, A. D. (2003), 'Protective or counter-productive? labour market institutions and the effect of immigration on eu natives', *The Economic Journal* **113**(488), F302–F331.

Angrist, J. & Evans, W. (1998), 'Children and their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size', *American Economic Review* **88**(3), 450–77.

Angrist, J., Imbens, G. & Rubin, D. (1996), 'Identification of Causal Effects Using Instrumental Variables', *Journal of the American Statistical Association* **91**, 444–455.

Angrist, J. & Krueger, A. (1995), 'Split Sample Instrumental Variables Estimates of the Return to Schooling', *Journal of Business and Economic Statistics* **13**, 225–235.

Arnold, F., Choe, M. & Roy, T. (1998), 'Son preference, the family-building process and child mortality in India', *Population Studies* **52**(3), 301–315.

Ashenfelter, O. & Heckman, J. (1974), 'The Estimation of Income and Substitution Effects in a Model of Family Labor Supply', *Econometrica* **42**(1), 73–85.

Autor, D. & Dorn, D. (2008), 'Inequality and specialization: The growth of low-skilled jobs in the united states', *Mimeo* .

Bardhan, P. (1974), 'On life and death questions', *Economic and Political Weekly* **9**(32), 1293–1304.

Bardhan, P. (1982), 'Little girls and death in India', *Economic and Political Weekly* **17**(36), 1448–1450.

Bardhan, P. (1988), 'Sex disparity in child survival in rural India ', *Rural Poverty in South Asia* pp. 473–480.

Barrett, J., Dawber, A. and others (1985), 'South African Women on the Move'.

Basu, A. (1990), 'Cultural influences on health care use: two regional groups in India', *Studies in Family Planning* **21**(5), 275–286.

Basu, A. & Das, G. M. (2001), 'Family Systems and Preferred Sex of Children', *International Encyclopedia of Social and Behavioral Sciences* .

Becker, G. (1985), 'Human Capital, Effort, and the Sexual Division of Labor', *Journal of Labor Economics* **3**(1), S33–58.

Becker, G., Landes, E. & R., M. (1977), 'An Economic Analysis of Marital Instability', *Journal of Political Economy* **85**(6), 1141–87.

Becker, G. & Lewis, G. (1973), 'On the Interaction Between the Quantity and Quality of Children', *Journal of Political Economy* **81**(2), S279–88.

Becker, G. & Tomes, N. (1977), 'Child Endowments and the Quantity and Quality of Children', *Journal of Political Economy* **84**(4), S143–62.

Ben-Porath, Y. & Welch, F. (1976), 'Do Sex Preferences Really Matter?', *Quarterly Journal of Economics* **90**(2), 285–307.

Ben-Porath, Y. & Welch, F. (1980), *On Sex Preferences and Family Size*, JAI Press.

Blau, F. & Grossberg, A. (1992), 'Maternal Labor Supply and Children's Cognitive Development', *Review of Economics and Statistics* **74**(3), 474–81.

Blumberg, R. (1991), 'Income under female vs. male control: Hypothesis from a theory of gender stratifications and data from the third world', *Gender, family, and economy: The triple overlap* pp. 97–127.

Borjas, G. (2003), 'The labor curve is downward sloping: Reexamining the impact of immigration on the labor market', *Quarterly Journal of Economics* pp. 1335–1374.

Borjas, G. (2005), 'Native internal migration and the labor market impact of immigration', *NBER Working Paper* **11610**.

Borjas, G. (2006), 'Native internal migration and the labor market impact of immigration', *Journal of Human Resources* **41**(2), 221–258.

Borjas, G. J. (1987), 'Immigrants, minorities, and labor market competition', *Industrial and Labor Relations Review* **40**(3), 382–392.

Boserup, E., Kanji, N., Tan, S. & Toulmin, C. (2007), *Woman's role in economic development*, Earthscan/James & James.

Boustan, L. P. & Margo, R. A. (2007), 'Going postal: What black employment in the postal service reveals about the increasing cost of racial segregation, 1940-2000', *Working Paper* .

Bronars, S. & Grogger, J. (1994), 'The Economic Consequences of Unwed Motherhood: Using Twins as a Natural Experiment', *American Economic Review* **84**(5), 1141–56.

Browning, M. (1992), 'Children and Household Economic Behavior', *Journal of Economic Literature* **30**(3), 1434–75.

Butcher, K. & Card, D. (1991), 'Immigration and wages: Evidence from the 1980's', *The American Economic Review* **81**(2), 292–296.

Butcher, K. & Case, A. (1994), 'The Effect of Sibling Sex Composition on Women's Education and Earnings', *Quarterly Journal of Economics* **109**(3), 531–63.

Butz, W. & M., W. (1979), 'The Emergence of Countercyclical U.S. Fertility', *American Economic Review* **69**(3), 318–28.

Cain, G. (1990), *Labor Force Participation of Married Women*, Oxford University Press.

Caldwell, J. (1979), 'Education as a factor in mortality decline an examination of Nigerian data', *Population studies* **33**(3), 395–413.

Card, D. (1990), 'The impact of the mariel boatlift on the miami labor market', *Industrial and Labor Relations Review* **43**(2), 245–257.

Card, D. (2001), 'Immigrant inflows, native outflows, and the local market impacts of higher immigration', *Journal of Labor Economics* **19**(1), 22–64.

Card, D. (2005), 'Is the new immigration really so bad?', *The Economic Journal* **115**(507), F300–F323.

Card, D. (2007), 'How immigration affects u.s. cities', *CReAM Discussion Paper* **11/07**.

Card, D. & Lewis, E. G. (2005), 'The diffusion of mexican immigrants during the 1990s: Explanations and impacts', *NBER Working Paper* **W11552**.

Card, David & DiNardo, John (2000), 'Do immigrant inflows lead to native outflows?', *The American Economic Review* **90**(2), 360–367.

Card, David & Krueger, Alan B. (1994), 'Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania', *The American Economic Review* **84**(4), 772–793.

Card, David & Krueger, Alan B. (2000), 'Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania: Reply', *The American Economic Review* **90**(5), 1397–1420.

Chaudhury, Rafiqul, H.,Gunasekera,P. and Others (2006), 'District-level Variations in Infant Mortality in Sri Lanka: A Challenge to Achieving the Millennium Development Goals on Child Survival', *Regional Health Forum* **10**(1).

Coleman, M. & Pencavel, J. (1993), 'Trends in Market Work Behavior of Women Since 1940', *Industrial and Labor Relations Review* **46**(4), 653–76.

Das Gupta, M. (1987), 'Selective discrimination against female children in rural Punjab, India', *Population and development review* **13**(1), 77–100.

Das Gupta, M. e. a. (2003), 'Why is Son Preference so Persistent in East and South Asia? A cross-country study of China, India, and the Republic of Korea', *Journal of Development Studies* **40**(2), 153–187.

Dasgupta, I., Palmer-Jones, R. & Parikh, A. (2000), *Between cultures and markets: an eclectic analysis of juvenile gender ratios in India*, University of Nottingham, Centre for Research in Economic Development and International Trade.

Debroy, B. & Bhandari, L. (2003), 'District-level deprivation in the New Millennium', *New Delhi: Konark* .

Desai, S. & Jain, D. (1994), 'Maternal employment and changes in family dynamics: The social context of women's work in rural South India', *Population and Development Review* **20**(1), 115–136.

Dooley, M. (1993), 'Labor Supply and Fertility of Married Women: An Analysis with Grouped and Individual Data from the 1970 U.S. Census', *Journal of Human Resources* **17**(4), 499–532.

Dyson, T. & Moore, M. (1983), 'On kinship structure, female autonomy, and demographic behavior in India', *Population and Development Review* **9**(1), 35–60.

Economic Commission for Africa (2001), 'Transforming Africas Economics'.

Friedberg, R. M. (2001), 'The impact of mass migration on the israeli labor market', *The Quarterly Journal of Economics* **116**(4), 1373–1408.

Friedberg, Rachel M. & Hunt, Jennifer (1995), 'The impact of immigrants on host country wages, employment and growth', *The Journal of Economic Perspectives* **9**(2), 23–44.

Fuchs, V. (1989), 'Women's Quest for Economic Equality', *Journal of Economic Perspectives* **3**(1), 25–41.

Gangdharan, J. & Rosenbloom, J. L. (1992), 'Effects of Child-bearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment', *National Bureau of Economic Research* .

Goldin, C. (1990), *Understanding the gender gap*, Oxford University Press.

Government of India (2000), 'National Population Policy 2000'.

Government of India (2006), 'Sample Registration Bulletin', **40**(1).

Greene, W. H. (2003), *Econometric Analysis, 3rd edition*, Prentice-Hall, New Jersey.

Gronau, R. (1973), 'The Effect of Children on the Housewife's Value of Time', *Journal of Political Economy* **81**(2), S168–99.

Gronau, R. (1977), 'Leisure, Home Production, and Work - The Theory of the Allocation of Time Revisited', *Journal of Political Economy* **85**(6), 1099–124.

Gronau, R. (1988), 'Sex-Related Wage Differentials and Women's Interrupted Careers - The Chicken or the Egg', *Journal of Labor Economics* **6**(3), 277–301.

Guillermo, C. & Galiani, S. (2007), 'Fertility and Female Labor Supply in Latin America: New Causal Evidence', *Labor Economics* **14**(3), 465–73.

Gupta, M. (1990), 'Death clustering, mothers' education and the determinants of child mortality in rural Punjab, India', *Population studies* **44**(3), 489–505.

Gupta, M. & Bhat, P. (1997), 'Fertility decline and increased manifestation of sex bias in India', *Population Studies* **51**(3), 307–315.

Gupta, R. & Attari, S. (1994), 'Gender bias in rural economy of m.p', *Journal of Indian School of Political Economy* **6**(1).

Haines, M. (1995), 'Socio-economic differentials in infant and child mortality during mortality decline: England and Wales, 1890-1911', *Population studies* **49**(2), 297–315.

Hanson, G. H. & Spilimbergo, A. (1999), 'Illegal immigration, border enforcement, and relative wages: Evidence from apprehensions at the u.s.-mexico border', *The American Economic Review* **89**, 1337–1357.

Heckman, J. (1974), 'Shadow Prices, Market Wages and Labor Supply', *Econometrica* **42**(4), 679–94.

Heckman, J. & MaCurdy, T. (1980), 'A Life-Cycle Model of Female Labor Supply', *Review of Economic Studies* **47**(1), 47–74.

Heckman, J. & Walker, J. (1990), 'The Relationship Between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data', *Econometrica* **58**(6), 1411–441.

Hobcraft, JN and McDonald, JW and Rutstein, SO (1985), 'Demographic determinants of infant and early child mortality: A comparative analysis', *Population Studies* **39**(3), 363–385.

Hotz, V. & Miller, R. (1988), 'An Empirical Analysis of Life Cycle Fertility and Female Labor Supply', *Econometrica* **56**(1), 91–118.

Imbens, G. W. & Angrist, J. D. (1994), 'Identifications and Estimation of Local Average Treatment Effects', *Econometrica* **62**(2), 679–94.

India, R.G. (1999), 'Compendium of India's Fertility and Mortality Indicators 1971-1997 based on the Sample Registration System (SRS)', *Delhi: Controller of Publication* .

International Institute of Population Science (1998), 'Demographic and Health Survey, India 1998/99 Final Report.', (SR81).

International Institute of Population Studies 1995 (1995), 'National Family Health Survey - 2'.

Jacob, M. (1962), 'Labor Force Participation in Married Women: A Study of Labor Supply', *in H. Gregg Lewis ed,. Aspects of Labor Economics* pp. 63–105.

Jacobsen, J. (1994), *The Economics of Gender*, Basil Blackwell.

Jain, A. (1985), 'Determinants of regional variations in infant mortality in rural India', *Population Studies: A Journal of Demography* pp. 407–424.

Juster, F. & Stafford, F. P. (1991), 'The Allocation of Time: Empirical findings, Behavioral Models, and Problems of Measurement', *Journal of Economic Literature* **28**(2), 471–522.

Kishor, S. & Parasuraman, S. (1998), *Mother's employment and infant and child mortality in India*, International Institute for Population Sciences.

Koenker, R. & Bassett Jr, G. (1978), 'Regression quantiles', *Econometrica* **46**(1), 33–50.

Koenker, R. & K., H. (2001), 'Quantile Regression', *Journal of Economic Perspectives* **15**(4), 143–156.

Korenman, S. & Neumark, D. (1992), 'Marriage, Motherhood, and Wages', *Journal of Human Resources* **27**(2), 233–55.

Kugler, A. & Yuksel, M. (2008), 'Effects of low-skilled immigration on u.s. natives: Evidence from hurricane mitch', *IZA Discussion Paper* .

Lehrer, E. (1992), 'The Impact of Children on Married Women's Labor Supply: Black-White Differntials Revisited', *Journal of Human Resources* **27**(3), 422–44.

Leung, S. (1991), 'A Stochastic Dynamic Analysis of Parental Sex Preferences and Fertility', *Quarterly Journal of Economics* **106**(4), 1063–88.

Liddle, J. & Joshi, R. (1989), *Daughters of Independence: Gender, caste, and class in India*, Rutgers Univ Pr.

Meenakshi, J., Ray, R. & Gupta, S. (2000), 'Estimates of poverty for SC, ST and female-headed households', *Economic and Political Weekly* **35**(31), 2748–2754.

Miller, B. (1982), 'Female labor participation and female seclusion in rural India: A regional view', *Economic Development and Cultural Change* pp. 777–794.

Miller, B. (1997), *The endangered sex: Neglect of female children in rural North India*, Oxford University Press Delhi.

Moffit, R. (1984), 'Profiles of Fertility, Labour Supply and Wages of Married Women: A Complete Life-Cycle Model', *Review of Economic Studies* **51**(2), 263–78.

Morgan, P., Lye, D. & Condran, G. (1988), 'Sons, Daughters, and the Risk of Marital Disruption', *American Journal of Sociology* **94**(1), 110–29.

Mosley, W. & Chen, L. (2003), 'An analytical framework for the study of child survival in developing countries. 1984.', *Bulletin of the World Health Organization* **81**(2), 140.

Mroz, T. (1987), 'The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions', *Econometrica* **55**(4), 765–99.

Murthi, M. & Guio, A. (1995), 'Mortality, fertility, and gender bias in India: A district-level analysis', *Population and Development Review* **21**(4), 745–782.

Myrianthopoulos, N. (1970), 'An Epidemiologic Survey of Twins in a Large Prospectively Studied Population', *American Journal of Human Genetics* **22**(6), 611–29.

Nag, M. (1983), 'Impact of social and economic development on mortality: Comparative study of Kerala and West Bengal', *Economic and Political Weekly* **18**(19), 877–900.

Nakamura, A. & Nakamura, M. (1992), 'The Econometrics of Female Labor Supply and Children', *Econometric Reviews* **11**(1), 1–71.

New Delhi Ministry of Health and Family Welfare (2000), 'National Population Policy 2000'.

Newey, W. & West, K. (1987), 'Hypothesis Testing with Efficient Method of Moments Estimation', *International Economic Review* **28**(3), 777–87.

Olusi, J. (1997), 'Enhancing Female Participation in African Agricultural Transformation: The Nigerian Experience', *Journal of Opinion* **25**(2).

Pagan, A. & Ullah, A. (1999), *Nonparametric econometrics*, Cambridge Univ Pr.

Pandey, A., Choe, M., Luther, N., Sahu, D. & Chand, J. (1998), 'Infant and child mortality in India', *National Family Health Survey Subject Report* **11**.

Peri, G. & Ottaviano, G. I. (2005), 'Rethinking the gains from immigration: Theory and evidence from the u.s.', *UCD Working Paper* .

Peri, G. & Ottaviano, G. I. (2007), 'The effects of immigration on u.s. wages and rents: A general equilibrium approach', *CReAM Discussion Paper* .

Peri, G. & Sparber, C. (2007), 'Task specialization, comparative advantages, and the effects of immigration on wages', *NBER Working Paper* **13389**.

Peri, G. & Sparber, C. (2009), 'Task specialization, immigration and wages"', *American Economic Journal, Applied Economics* .

Ram, F., Shekhar, C. & Mohanty, S. (2005), 'Human Development: Strengthening District Level Vital Statistics in India', *International Institute for Population Sciences, Mumbai* .

Region, S.A. (2004), India Attaining the Millennium Development Goals in India How Likely and What Will It Take to Reduce Infant Mortality, Technical Report 30266-IN.

Robinson, J. (1988), 'Who's Doing the Housework', *American Demographics* **10**(12).

Rosenzweig, M. & Schultz, P. (1985), 'The Demand for and Supply of Births: Fertility and its Life Cycle Consequences', *The American Economic Review* **75**(5), 992–1015.

Rosenzweig, M. & Schultz, T. (1982), 'Market opportunities, genetic endowments, and intrafamily resource distribution: Child survival in rural India', *The American Economic Review* **72**(4), 803–815.

Rosenzweig, M. & Wolpin, K. (1980*a*), 'Life-Cycle Labor Supply and Fertility: Causal Inferences From Household Models', *Journal of Political Economy* **88**(2), 328–48.

Rosenzweig, M. & Wolpin, K. (1980*b*), 'Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment', *Econometrica* **48**(1), 227–40.

Rosenzweig, M. & Wolpin, K. (2000), 'Natural 'Natural Experiments' in Economics', *Journal of Economic Literature* **38**(4), 827–874.

Rwomire, A. (1992), *African Women and Children*, Praeger Publishers.

Schultz, P. (1981), *Economics of population*, Addision-Wesley.

Schultz, P. (1990), 'Testing the Neoclassical Model of Family Labor Supply and Fertility', *Journal of Human Resources* **25**(4), 599–634.

Schultz, P. T. (1978), 'The Influence of Fertility on Labor Supply of Married Women: Simultaneous Equation Estimations', *Research in Labor Economics* .

Sen, A. (1987), 'Africa and India: what do we have to learn from each other?', *WIDER Working Papers (UNU)* .

Sen, A. (1998), 'Mortality as an indicator of economic success and failure', *The Economic Journal* **108**(446), 1–25.

Sinha, S. (2007), Essays on economic development and the survival of girls in India, PhD thesis, UNIVERSITY OF CALIFORNIA, RIVERSIDE.

Smith, J. & Ward, M. (1984), *Women's wages and work in the twentieth century*, Rand Corporation.

Stafford, F. P. (1987), 'Women's Work, Sibling Competition, and Children's School Performance', *American Economic Review* **77**(5), 972–80.

Statistics South Africa (2001), 'Stages in the Life Cycle of South Africans'.

The World Bank (2004), Attaining the Millenium Development Goals in India: How Likely and What Will It Take To Reduce Infant Mortality, Child Malnutrition, Gender Disparity and Hunger-Poverty and to Increase School Enrollment and Completion?, Technical Report 30266-IN.

The World Bank (2007), India: Achieving the MDGs in India's poor states: Reducing child mortality in Orissa, Technical Report 39855-IN.

Thomas, D. (1994), 'Like Father, Like Son; Like Mother, Like Daughter: Parental Resources and Child Height', *Journal of Human Resources* **29**(4), 950–89.

United Nations Development Program (1997), 'Human Development Report'.

Waterhouse, J. (1950), 'Twinning in Twin Pedigrees', *British Journal of Social Medicine* (5), 197.

Westoff, C., Potter, R. & Sagi, P. (1963), *The third child: A study in the prediction of fertility*, Princeton University Press.

Williamson, N. (1963), *Sons or daughters: A cross-cultural survey of parental preferences*, Princeton University Press.

Willis, R. (1987), 'What Have We Learned from the Economics of the Family?', *American Economic Review* **77**(2), 68–81.