ESSAYS IN APPLIED MICROECONOMICS

BY

NICHOLAS L. WILSON

B.A., REED COLLEGE, 1999

MPA/ID, HARVARD UNIVERSITY, 2001

A.M., BROWN UNIVERSITY, 2004

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE DEPARTMENT OF ECONOMICS AT BROWN UNIVERSITY

PROVIDENCE, RHODE ISLAND
MAY 2008

This dissertation by Nicholas L. Wilson is accepted in its present form

by the Department of Economics as satisfying the

dissertation requirement for the degree of Doctor of Philosophy

Date 4-14-08 ______________________

Kaivan Munshi, Advisor

Recommended to the Graduate Council

Date 4/14/08 ______________________

Andrew Foster, Reader

Date 4/14/08 ______________________

David Weil, Reader

Approved by the Graduate Council

Date __________ ______________________

Sheila Bonde, Dean of the Graduate School

VITA

Nicholas L. Wilson was born on August 8, 1977 in Kansas City, Missouri. He received his B.A. in Economics from Reed College in 1999, his MPA in International Development from Harvard University in 2001, and his A.M. in Economics from Brown University in 2004. He has taught economics at Brown University, Park College, and Harvard University and has conducted field research in Bangladesh and Zambia.

ACKNOWLEDGMENTS

I would like to thank Andrew Foster, Kaivan Munshi, and David Weil for their advice and support. In addition, my research benefited from numerous discussions with Daniel Bennett, Jason Goldman, Emily Hazan, and Alaka Holla. The chapters on HIV testing would not be possible without the cooperation of Bates Buckner at MEASURE Evaluation and many individuals in Zambia, including Bristol Cheembo and Clement Mulenga at the Virology Lab of the University Teaching Hospital, Dr. Albert Mwango at the Ministry of Health, and Nchimunya Nkombo at the Central Statistical Office.

I would like to thank Paul Koussa for excellent computing support, Angelica Spertini for tremendous administrative and emotional support, and Svetla Vitanova, without whom I would not have completed my Ph.D. Of course I would be remiss if I did not also thank my parents, Catherine and Jeff, my sister, Rachel, and my brother, Jacob, all of whom have been extremely supportive during my lengthy academic studies.

I dedicate this dissertation to two of my grandparents who passed away while I was writing it: Marion R. Simons and Logan L. Wilson.

All errors that may remain in this dissertation are solely my responsibility.

Contents

List of Figures

List of Tables

Introduction

This thesis examines several topics in applied microeconomics. The first chapter presents a model of demand for HIV testing and derives the conditions under which HIV testing will reduce risky behavior. This model captures the two main reasons why an individual would want to take a HIV test: (i) to begin antiretroviral therapy (ART) if the individual is HIV positive, and (ii) to increase the precision of belief about the marginal cost of risky behavior. This model shows that in the absence of ART, the individuals that are most willing to take a HIV test are individuals that have the least precise beliefs about their own HIV status (i.e., individuals in the middle of the distribution of past risky behavior) and are expected to demonstrate the greatest reductions in post-testing risky behavior. Introducing ART into the model means that individuals with the highest subjective probability of being HIV positive should demonstrate the greatest increases in testing demand between the pre-ART period and the ART era.

The second chapter examines empirical evidence on the predictions of the model using newly assembled data from Zambia, one of the highest HIV prevalence countries in the world. I show that testing behavior in Zambia prior to the introduction of ART is consistent with the prediction of the model: the individuals most likely to test are the individuals expected to demonstrate the greatest reductions in post-testing risky behavior. The observed testing response to the introduction of free ART, however, is not consistent with the second main prediction of the model. Specifically, a large group of individuals whose testing decisions appear to be particularly responsive to the introduction of ART are in fact expected to demonstrate only small or modest reductions in post-testing risky behavior. This finding is both a puzzle and - from the perspective of maximizing the prevention impact of ART-induced testing - evidence of an inefficiency. I provide an explanation for this puzzle and provide evidence rejecting alternative hypotheses. In addition, I estimate the structural parameters of the model and use these estimates to simulate several counterfactual ART policies. The simulation results indicate that the magnitude of the inefficiency is quite large: eliminating the inefficiency while holding constant the quantity of ART supplied would increase the number of new infections avoided due to ART-induced testing by more than ten-fold.

The third chapter examines ethnic (migrant) labor market networks in the American Midwest when it was first being settled in the middle of the 19th century, the endogenous emergence of a sense of local identity that helped support the functioning of these networks, and how this sense of local identity may have persisted across multiple generations and affect economic outcomes in the late 20th century. While fractionalization may have adversely may adversely affect the performance of secular institutions, ethnic competition in the labor market would at the same time have strengthened within-group loyalty and parochial institutions. These values and their complementary institutions, notably the church, could have mutually reinforced each other over many overlapping generations, long after the networks themselves had ceased to be salient. Counties with greater ethnic fractionalization in 1860 are indeed associated with steadily increasing participation in select religious denominations historically dominated by migrants all the way through the twentieth century. Complementing this result, individuals born in high fractionalization counties are significantly less likely to select into geographically mobile professional occupations and, hence, to migrate out of their county of birth, despite the fact that these counties are indistinguishable from low fractionalization counties in terms of local public good provision and economic activity today.

Chapter 1

A Model of Demand for HIV Testing: Descriptive and Theoretical Analysis

1.1 Introduction

Policymakers believe that HIV testing is an important intervention in the HIV/AIDS pandemic. Proponents of HIV testing presume that on average individuals that learn whether or not they are infected will respond to this new information by reducing risky behavior. However, HIV testing rates are low in much of the world and a vast majority of HIV positive individuals are not aware that they are infected. For example, among the 85 percent of the world population infected with HIV/AIDS that reside in Sub-Saharan Africa only 10 percent know their HIV status (WHO 2006a).

One of the main mechanisms by which policymakers hope to increase demand for HIV testing is increased availability of subsidized antiretroviral therapy (Global HIV Prevention Working Group 2004). The availability of antiretroviral therapy (ART) gives individuals a direct incentive to take a HIV test: if they take a HIV test and are HIV positive then they may begin therapy, reducing morbidity and prolonging life. However, the incentive effects of ART may be heterogeneous and many individuals may not choose to test in response to the introduction of ART. Moreover, those that test because of ART may not demonstrate substantial behavior change subsequent to testing.

Chapters 1 and 2 examine the effect of ART availability on demand for HIV testing and the effect of ART-induced testing on demand for risky behavior. In Chapter 1, I present a model of demand for HIV testing and for risky behavior. Initially I assume that the conditions under which individuals have an incentive to test and testing reduces risky behavior hold. In particular, I assume that demand for risky behavior is concave in the prior probability of being HIV positive. Under these conditions, I show that prior to the introduction of ART it is the riskiest individuals (i.e., individuals with the highest prior probabilities of being HIV positive) that have the greatest incentive to test in the setting examined in the current analysis. The riskiest individuals are also those individuals whose testing decisions should be most responsive to the introduction of ART. I show that under these same conditions the expected reduction in risky behavior associated with taking a HIV test in this setting should be increasing in the prior probability an individual is HIV positive. Thus, ART-induced testing has the potential to substantially reduce the spread of HIV/AIDS.

Using newly assembled data from before and after the introduction of ART in Zambia I examine the empirical evidence on these predictions. Testing behavior among women and men prior to the introduction of ART is consistent with the prediction that individuals with the highest prior probabilities of being HIV positive are those that should be most likely to test. The change in testing behavior among women after the introduction of ART is consistent with the prediction that individuals with the highest prior probabilities of being HIV positive are those that should demonstrate the greatest response to the availability of ART. However, the change in testing behavior among men after the introduction of ART is not consistent with this prediction. Specifically, testing decisions among older men appear to be particularly responsive to the introduction of ART despite the fact that older men are one of the lowest HIV prevalence groups in Zambia. Moreover, testing decisions among men in the middle of the age distribution are not particularly responsive to the introduction of ART despite the fact that these men are one of the higher HIV prevalence groups. I interpret this as evidence of a non-random rationing mechanism determining the allocation of ART among HIV positive individuals (in favor of older males) and provide evidence rejecting alternative hypotheses.

In Chapter 2, I estimate the structural parameters of the model and show that the conditions for individuals to have an incentive to test and for testing to reduce risky behavior are indeed satisfied in this setting. In addition, as part of the estimation process I recover the ART rationing rule. Using the parameter estimates and the information on the rationing rule I simulate the effect of ART availability on testing demand and the effects of ART-induced testing on demand for risky behavior and on the spread of HIV/AIDS. I also examine the effects of eliminating the non-random rationing mechanism and/or expanding the supply of ART. Simulation results show that under the existing policy ART availability increased testing demand by approximately 2 percentage points and ART-induced testing reduced the incidence of HIV by approximately 1 percent. Expanding the supply of antiretroviral drugs without eliminating the non-random rationing mechanism would only have small effects on testing rates and risky behavior. In contrast, eliminating the non-random rationing mechanism while holding fixed the existing supply of antiretroviral drugs would increase testing rates by approximately 25 percentage points and reduce the incidence of HIV by approximately 15 percent.

These chapters contributes to the small body of existing economic literature on the behavioral effects of HIV testing and of ART availability.[1] Philipson and Posner (1995) examine the effects of public subsidies for HIV testing on risky behavior and the incidence of HIV. The authors present a trade model with asymmetric information where HIV testing is a method of quality verification. An individual that is possibly HIV positive may show a negative HIV test to their potential partner to verify that they are in fact HIV negative and induce the partner to engage in unprotected sex. Although the model presented in the current analysis shares this focus on uncertainty about HIV status, I focus on uncertainty about one's own status and not the status of one's partner.

Boozer and Philipson (2000) examine the conditions under which HIV testing will reduce risky behavior and estimate the effect of learning the test result on subsequent behavior. Unlike the other literature in this area, the authors emphasize the role of prior beliefs about one's own HIV status in conditioning the effect of HIV testing on risky behavior. In particular, they argue that HIV testing should only affect behavior among the subset of individuals for whom the result of the test differs from the prior belief. The model presented in the current analysis is similar to the analytic framework discussed in Boozer and Philipson (2000). An important difference lies in the empirical analysis. Boozer and Philipson (2000) use a sample of individuals who all take a HIV test to estimate the reduced form relationship between (i) the interaction of prior beliefs and the result of the test and (ii) subsequent behavior. In contrast, I estimate the structural parameters of the risky behavior demand function, allowing for simulations of the effects of alternative ART policies.

Thornton (2006) examines the effect of randomly assigned financial incentives (and randomly placed counseling sites) on the propensity to return to receive the results of a HIV test among a sample of individuals who were asked to provide blood samples for a HIV test. Using the financial incentive and distance to the counseling site as exogenous sources of variation in the effective price of learning one's results, the author finds that demand is very sensitive to price. Moreover, the author finds that an exogenous change in whether an individual learns the result of a HIV test has only

[1] There is more literature on the behavioral effects of HIV testing than on the behavioral effects of ART availability and I focus on the former in the discussion of the existing literature. Lakdawalla *et al* (2006) is one of the few economics articles on the behavioral effects of ART availability. The authors argue that technological advances in the efficacy of antiretroviral therapy have increased the incidence of HIV in the United States. By lowering the health cost of acquiring HIV, improvements in ART have reduced the marginal cost of risky sexual behavior and hence increased the equilibrium quantity demanded (yielding an associated increase in rates of HIV transmission). Although this effect may be operative in the setting I examine in the current analysis, I focus on the effect of ART-induced testing on demand for risky behavior.

a small effect on subsequent choices about risky sexual behavior. A fundamental insight shared by the current analysis and Boozer and Philipson (2000) may explain the seemingly small effect of HIV testing on risky sexual behavior in Thornton (2006): the effect of HIV testing on subsequent choices about risky behavior depends on the individual's prior belief about their HIV status. If individuals are approximately evenly distributed between those that (i) believe they are HIV positive but are in fact HIV negative and (ii) believe that they are HIV negative but are in fact HIV positive, then HIV testing may have little effect on subsequent behavior.

1.2 Setting

Zambia has one of the highest rates of HIV prevalence in the world. According to the anonymous HIV testing component of the 2001 Zambia Demographic and Health Survey (ZDHS), the HIV prevalence rate among adults age 15-49 is 14.5 percent. Moreover, high HIV prevalence in Zambia is not a recent development. Data from antenatal clinics in the sentinel surveillance program are consistent with the hypothesis that HIV prevalence has remained relatively stable from the early 1990s onwards (UNAIDS/WHO 2006).

As in many Sub-Saharan African countries (i.e., high prevalence countries), the highest rates of HIV prevalence are among individuals in the middle of the age distribution. Figure 1 presents the results of a locally weighted smoothed regression of an indicator variable for HIV status on age, estimated separately for females and for males in the 2001 ZDHS. As shown in Figure 1, HIV prevalence is highest among individuals in the middle of the age distribution, prevalence is slightly higher among women than among men, and the age of peak prevalence among females is less than that among males (34 and 40 years, respectively). The average age difference in sexual partners may explain this last fact. Data from the Zambia Sexual Behavior Survey (ZSBS) on the age of sexual partners show that on average males are approximately six years older than their female sexual partners.

Magruder (2006) provides an explanation for the inverted U-shape that characterizes the prevalence-age profiles of most high prevalence countries. Individuals in early adulthood engage in "marital shopping" - the search for a potential marital partner - and experience higher than average turnover in sexual partners. Moreover, the effect of this behavior on prevalence is magnified by the biology of the HIV virus: infectiousness during the first month of having acquired HIV is much higher than during the majority of the life course of the virus (Wawer *et al* 2005) and rapid turnover of sexual partners during the period of having recently acquired HIV can lead to a sharp increase in incidence and prevalence.

In an effort to address the high HIV prevalence rate, the government of Zambia began offering free voluntary counseling and testing (VCT) in 1998. Although initially rates of HIV testing were low, they are increasing. According to the ZSBS, the proportion of individuals voluntarily taking a HIV test in the 12 months prior to the survey round increased from between 5.5 and 6 percent in 2000 and 2003 to over 10 percent for women and nearly 8 percent for men in 2005.

One possible explanation for the increase in testing rates is that free antiretroviral therapy (ART) became available in Zambia around this time. The early 21st century witnessed a dramatic expansion in the provision of antiretroviral therapy in Sub-Saharan Africa. Between 2000 and the end of 2005, the number of individuals on ART in the region increased from fewer than 50,000 (Attawell and Mundy 2003) to 810,000 (WHO 2006b). Zambia is a participant in this expansion and in May 2004 the government of Zambia began to provide ART free of charge at primary health care centers (Stringer *et al* 2006).

The introduction of free antiretroviral therapy in Zambia may have increased testing demand by providing those individuals that believed they may have been HIV positive with an added incentive to test. However, other factors affecting testing rates could have changed during the period as well. For example, the number of VCT sites increased between 2000-03 and 2005.

If the introduction of ART had a causal effect on testing demand, then we should expect to see an inverse relationship between distance to the nearest ART site and the change in testing rates between 2000-03 and 2005. The Japanese International Cooperation Agency 2004 Health Facilities Census (HFC) provides global positioning system (GPS) information for each of the ART sites open at the time of the 2005 survey round of the ZSBS. In addition, the ZSBS includes the same information for the precise location of the rural households in 2000-2005 survey rounds. For the urban households in the ZSBS, I compute the GPS coordinates of the centroid of the Standard Enumeration Area (SEA) of residence; urban SEAs are relatively small, so each individual resides relatively close to the centroid. Using the GPS data from the HFC and the ZSBS I calculate the distance between each household and each ART site, and for each household I construct a measure of the distance to the nearest ART site.

Figure 2 presents the results of a locally weighted smoothed regression of an indicator variable equal to one if the individual voluntarily took a HIV test in the 12 months prior to the survey month on distance to the location of the nearest ART site, estimated separately for the pre-ART and ART periods. In the pre-ART period, individuals closer to sites that eventually provided ART were more likely to test. Nonetheless, the increase in testing rates associated with the ART period is greatest for those individuals closest to ART sites. Of course an important omitted variable in this analysis is the distance to the nearest VCT site; proximity to a VCT site affects the testing decision and is positively correlated with proximity to an ART site. However, health clinics in which ART was introduced in 2004 and 2005 tended to be sites that had previously provided VCT services and the pre-ART regression line reflects this fact. (The expansion in VCT between the pre-ART and ART periods occurred largely in areas that were not close to ART sites and hence may explain the increase in testing rates among individuals further from ART sites in 2005.)

Although it appears that the introduction of subsidized ART caused an increase in aggregate testing demand, there exists substantial heterogeneity in the change in testing demand associated with this intervention. Figure 3 shows the testing-age profile for women prior to the availability of ART and after the introduction of ART. Prior to the availability of ART, women in the middle of the age distribution were most likely to take a HIV test. These women were also those with the highest HIV prevalence. After ART was introduced in Zambia testing increased among women of all ages, but the increase was largest for the women in the middle of the age distribution. These differences by age are consistent with the idea that ART availability increased testing demand and the incentive effect of ART on testing demand is greatest for those individuals with the highest probability of being HIV positive.

For males, the pattern is noticeably different. Figure 4 shows the testing-age profile for men prior to the availability of ART and after the introduction of ART. As with females, in the "pre" period it is men in the middle of the age distribution that were most likely to take a HIV test. Likewise, these men were also those with the highest HIV prevalence. In the "post" period, however, it is the oldest males that are most likely to take a HIV test. It appears that the idea that individuals with the highest probability of being HIV positive should have testing decisions that are most responsive to ART availability cannot explain the observed patterns of testing demand among males.

Although much of the change between 2000-03 and 2005 in the testing-age profiles for females and for males may be attributed to the introduction of ART, it may be that the explanation for the large increase in testing rates among older males (and perhaps that among women in the middle of the age distribution) lies elsewhere. One change that paralleled the introduction of ART in Zambia was a continued expansion in the number of VCT sites. Nonetheless, it is unlikely that the expansion in the number of testing sites caused the observed changes in the testing-age profiles for women and for men. For the expansion in VCT sites to explain these changes, the expansion should have decreased the distance to the nearest VCT site differentially by age. Moreover, the age differential in the expansion intensity should have differed by gender.

The ideal test of this hypothesis would be to regress the change in the distance to the nearest VCT site on measures of each individual's age, age squared, and gender as well as interactions between age (and age squared) and gender. Unfortunately, data on which of the VCT sites open in 2005 existed prior to the introduction of ART in Zambia do not exist. In lieu of such data, we may examine the relationship between distance to the nearest VCT site and age in 2005. If the change in the distance to the nearest VCT site is correlated with age and gender, then we should expect to see evidence of spatial clustering by age/gender and possibly evidence of substantial differences by age in distance to the nearest VCT site in 2005. The HFC includes GPS information for VCT sites and for each respondent in the 2005 ZSBS I construct a measure of the distance to the nearest VCT site. Figure 5 presents the results of a locally weighted smoothed regression of distance to the nearest VCT site on age, estimated separately for females and males. There are small differences by age and gender in the average distance to the nearest VCT site, but these differences do not readily explain the changes in the testing-age profiles (e.g., older males and males in the middle of the age distribution appear to reside at similar distances from VCT sites). Moreover, the overall flatness of the distance-age profiles suggests that individuals are not spatially clustered by age. Thus, it appears that the expansion in the number of VCT sites between 2003 and 2005 does not explain the large increase in testing rates among older males.

The likely explanation for the large testing response among older males associated with the introduction of ART is that there exists a non-random rationing mechanism for allocating ART. Currently, the supply of antiretroviral drugs in Zambia is insufficient to ensure that all individuals that are HIV positive and require ART are guaranteed coverage. If supply-side constraints mean that priority for access to ART depends on an individual's age and gender or factors correlated with these characteristics, then we would expect to see the observed changes in testing-age profiles deviate from those predicted by a simple model of testing demand where the incentive effect of ART is increasing in the prior probability that the individual is HIV positive.

Social norms in Zambia are consistent with the notion that older males receive priority for access to ART. Although there may exist variation across the more than seventy ethnolinguistic groups in Zambia, in much of traditional Zambian society younger men are subordinate to older men (Bond 1982). The oldest resident male is general identified as the household head (Scudder 1962). Among the Tonga ethnic group, one of the largest in Zambia, an elder is called *mupati* or "big person" (Colson and Scudder 1981). Among the Bemba ethnic group, another of the largest ethnic groups in Zambia, males are not considered to be fully adult members of the community until they are married and have children. Older males, or *bakalamba*, control decision-making not just within their kin group, but also more broadly within the community. They are figures of authority and largely control village decision-making (Epstein 1981).

Although scant, direct evidence on the process determining the allocation of ART among HIV positive individuals is also consistent with the idea that older males receive priority in this process. Two of the main groups that are documented as receiving priority for access to ART are richer individuals and individuals with political connections (e.g., civil servants) (Jones 2004, Bwalya 2006). Despite the lack of direct evidence on age-based preferences, it is likely that older men represent a disporportionate share of these two groups.

Thus, although it appears that ART availability in Zambia increased testing demand, the change in demand was conditioned by a non-random ART allocation mechanism. To frame the analysis of the effects of this mechanism on demand for HIV testing and on demand for risky sexual behavior, the next section presents a simple model of demand for HIV testing and risky behavior. I then estimate the parameters of the model and use these to quantify the magnitude of these effects.

1.3 Model

The changes in testing behavior associated with the introduction of ART indicate that there exists a non-random rationing mechanism determining the allocation of ART among HIV positive individuals. In particular it appears that priority for access to ART is given to older males at the expense of males in the middle of the age distribution. To examine this finding more rigorously, I present a model of demand for HIV testing and for risky behavior. I match the model to the data and find that the model fits pre-ART testing behavior quite well but fails to completely match the change in testing behavior after the introduction of ART. Thus, I expand the model to capture the effect of ART rationing and recover the rationing rule from the observed testing behavior. Then I use the parameter estimates and the information on the rationing rule to simulate the effect of ART availability on testing demand and demand for risky behavior under several ART policies.

1.3.1 Basic setup

In this model an individual chooses whether to take a HIV test and chooses how much risky behavior in which to engage. There are two components to expected utility: the pleasure the individual derives from engaging in risky behavior (e.g., unprotected sex) and the displeasure associated with acquiring HIV. Taking a HIV test means that the individual learns his HIV status with certainty and allows him to avoid *ex post* errors in forming beliefs about the marginal cost of risky behavior. In addition, when ART is available he may be able to begin treatment if he tests HIV positive.

The individual is endowed with a prior probability of being HIV positive, $p_0 \in [0, 1]$. Given this probability, the individual forms expectations about the net benefit of testing and decides whether to take a HIV test. After deciding whether or not to take a HIV test and learning the results if he chooses to test (and beginning treatment if he is HIV positive and eligible), the individual chooses an amount of risky behavior in which to engage. The amount of risky behavior in which the individual chooses to engage is summarized by the probability that he is HIV positive after engaging in this risky behavior if in fact he was HIV negative prior to this choice. Let p_1 denote this probability.

1.3.2 Demand for risky sexual behavior

Figure 6 provides a graphical representation of the mechanics of the model. For the moment, consider the upper half of this figure. On the x-axis is the exogenous prior probability the individual is HIV positive, p_0. On the y-axis is the choice of risky behavior, p_1. The concave function, $p_1(p_0)$, represents the risky behavior demand function if the individual does not take a HIV test. Although the $p_1(p_0)$ depicted in Figure 6 is increasing and concave in p_0, this function need not be increasing nor concave in p_0 in practice and this model does not assume this to be the case. If it is increasing and concave in p_0, then the model indicates a role for HIV testing in reducing risky behavior. I will explain why this is the case momentarily.

Suppose the individual knows with certainty that he is HIV negative (i.e., $p_0 = 0$). Then he will choose an amount of risky behavior given by $p_1(0)$. This is the y-intercept for the $p_1(p_0)$ function. Now suppose that the individual knows with certainty that he is HIV positive (i.e., $p_0 = 1$). Then he will choose an amount of risky behavior given by $p_1(1)$. Note that $p_1(0)$ and $p_1(1)$ are bounded below and above by zero and one but they need not equal these values.

In this model, taking a HIV test and receiving the results means that the individual knows his HIV status with certainty. Thus, if the individual takes a HIV test then he will choose one of two possible choices of risky behavior: $p_1(0)$ or $p_1(1)$. The result of the test determines which choice he will make. In expectation, the individual's choice of risky behavior if he were to take a HIV test is simply the linear combination of these two points: $(1-p_0)p_1(0)+p_0p_1(0)$. This is the linear function in the upper half of Figure 6.

It should now be clear that the expected change in choice of risky behavior associated with taking a HIV test is the difference between the two functions plotted in the upper half of Figure 6. That is, the expected change in risky behavior associated with taking a HIV test is given by

$$(1 - p_0)p_1(0) + p_0 p_1(1) - p_1(p_0). \tag{1.1}$$

If demand for risky behavior (i.e., $p_1(p_0)$) is concave in p_0, then the sign of this expression is negative and in expectation HIV testing will reduce risky behavior. Conversely, if $p_1(p_0)$ is convex, then the sign of this expression if positive and in expectation HIV testing will increase risky behavior. To fix ideas, I will restrict the following analysis to the case where $p_1(p_0)$ is concave and in expectation HIV testing will reduce risky behavior.

Let p^* denote the value of p_0 that maximizes the expected reduction in risky behavior associated with taking a HIV test. For $p_0 < p^*$, the expected reduction in risky behavior associated with taking a HIV test is increasing in p_0. For $p_0 \geq p^*$, the expected reduction in risky behavior associated with taking a HIV test is decreasing in p_0.

1.3.3 Demand for HIV testing

Now consider the lower half of Figure 6. Again, on the x-axis is the exogenous prior probability the individual is HIV positive, p_0. On the y-axis is expected utility. There are two components to expected utility in this model: the utility derived from engaging in (risky) sex and the utility loss associated with acquiring HIV. The utility derived from the choice of risky sexual behavior, $p_1(p_0)$, is denoted $V(p_1(p_0))$. The utility loss associated with acquiring HIV is denoted θ. I assume that utility is additively separable in choice of risky sexual behavior and health status.

The convex function in Figure 6, $E[U(\textit{not test})]$, is the expected utility of not taking a HIV test. Additive separability means that I may write the expected utility of not testing as

$$E[U(\textit{not test})] = V(p_1(p_0)) - [1 - (1 - p_0)(1 - p_1(p_0))]\theta \tag{1.2}$$

where $[1 - (1 - p_0)(1 - p_1(p_0))]$ is the probability that the individual is HIV positive conditional on the prior probability of being HIV positive, p_0, and on the probability that he acquires HIV through his choice of risky behavior, $p_1(p_0)$.

Although Figure 6 depicts the expected utility of not taking a HIV test as being decreasing and convex in p_0 this need not be the case in practice and this model does not assume this to be true. If it is decreasing in p_0, then the increase in the expected utility cost of becoming HIV positive due to a increase in p_0 (and the subsequent choice to increase p_1) outweighs the utility gain (if any) from engaging in a greater amount of risky behavior, $p_1(p_0)$, that results from the increase in

p_0. The convexity of $E[U(not\ test)]$ follows from the fact that the marginal cost of additional risky behavior is decreasing in the prior probability of being HIV positive. As p_0 increases, the expected utility of not testing falls because the probability that the individual is HIV positive and incurs the associated utility cost is higher. On the other hand, the loss in expected utility from an increase in p_0 is partly offset by the increase in choice of risky behavior (given the current assumption that $p_1(p_0)$ is increasing in p_0). However, the increase in choice of risky behavior, p_1, itself increases the probability that the individual acquires HIV. The magnitude of this latter effect (although always non-negative) is decreasing in p_0 and hence $E[U(not\ test)]$ is convex.

To understand the expected utility of testing in the absence of ART, first consider the case where the individual takes a HIV test and finds that he is HIV negative. Because he knows that he is HIV negative with certainty, he will choose to engage in an amount of risky behavior given by $p_1(0)$ and the only risk of acquiring HIV is because of this choice. Thus, conditional on testing HIV negative his expected utility is

$$V(p_1(0)) - p_1(0)\theta - c \tag{1.3}$$

where c is the utility cost of taking a HIV test. Now consider the case where the individual takes a HIV test and finds that he is HIV positive. Because he knows that he is HIV positive with certainty, he will choose to engage in an amount of risky behavior given by $p_1(1)$ and is certain of incurring the utility cost of acquiring HIV. Thus, conditional on testing HIV positive his expected utility is

$$V(p_1(1)) - \theta - c. \tag{1.4}$$

The expected utility of testing in the absence of ART is the linear combination of these two conditional expectations, where the weights are the prior probabilities of being HIV negative and of being HIV positive, respectively. That is, the expected utility of testing in the absence of ART, denoted $E[U(test)]$ in Figure 6, is

$$E[U(test)] = (1 - p_0)[V(p_1(0)) - p_1(0)\theta - c] + p_0[V(p_1(1)) - \theta - c]. \tag{1.5}$$

Assuming $V(p_1(0)) - p_1(0)\theta > V(p_1(1)) - \theta$ (i.e., the individual would prefer to engage in $p_1(0)$ and acquire HIV with probability $p_1(0)$ than to engage in $p_1(1)$ and acquire HIV with certainty), implies that $\frac{\partial E[U(test)]}{\partial p_0} < 0$.

With expressions for the expected utility of not testing and the expected utility of testing in the absence of ART, it is possible to examine predictions about the relationship between the testing decision and the prior probability of being HIV positive. In the absence of ART, the model predicts that demand for HIV testing should be greatest among individuals in the middle of the p_0 distribution. As long as the cost of testing is sufficiently high, individuals with prior probabilities of being HIV positive that are either very low or very high will not choose to take a HIV test and the only individuals that test will be those in the middle of the p_0 distribution. The prediction is apparent in the lower half of Figure 6: the expected utility of testing is greater than that of not testing only for individuals in the middle of the p_0 distribution. The intuition underlying this result is that the expected change in risky behavior associated with taking a HIV test is relatively small for individuals with very high (or very low) values of p_0. It is individuals that are in the middle of the p_0 distribution that have the largest expected change in demand for risky behavior associated with taking a HIV test and hence the greatest willingness to pay for the test.

The introduction of ART (without rationing) does not affect the risky behavior demand function or the expected utility of not testing, but it does affect the expected utility of testing. Specifically, the expected utility of testing increases by the prior probability that the individual is HIV positive times the net utility benefit of beginning ART conditional on being HIV positive. Let α denote the net utility of beginning ART conditional on being HIV positive. Then the expected utility of testing when there is ART, denoted $E[U(test)]$ *with ART* in Figure 6, is

$$E[U(test)] \; with \; ART = (1 - p_0)[V(p_1(0)) - p_1(0)\theta - c] + p_0[V(p_1(1)) - \theta - c + \alpha]. \tag{1.6}$$

Suppose the individual takes a HIV test and finds that he is HIV negative. Given the set-up of the model, to receive the utility benefit of beginning ART the individual needs to be HIV positive before engaging in his choice of risky behavior, p_1. Thus, conditional on testing HIV negative his expected utility when there is ART is identical to that if there were no ART (i.e., $V(p_1(0)) - p_1(0)\theta - c$). Now suppose the individual takes a HIV test and finds that he is HIV positive. Then the individual will receive the treatment benefit of testing with certainty and his expected utility conditional on testing HIV positive when treatment is available is given by

$$V(p_1(1)) - \theta - c + \alpha. \tag{1.7}$$

The expected utility of testing when there is ART is the linear combination of these two conditional expectations where the weights are the prior probabilities of being HIV negative and of being HIV positive, respectively. Assuming $V(p_1(0)) - p_1(0)\theta > V(p_1(1)) - \theta + \alpha$ (i.e., the individual would prefer to engage in $p_1(0)$ and acquire HIV with probability $p_1(0)$ than to engage in $p_1(1)$ and acquire HIV (and begin ART) with certainty), implies that $\frac{\partial E[U(test)]\ with\ ART}{\partial p_0} < 0$.

The model predicts that when ART is available (and there is no rationing) demand for testing should be lowest among individuals with the lowest prior probabilities of being HIV positive. For an individual with a low value of p_0, the expected treatment benefit of testing is relatively low and the expected change in demand for risky behavior associated with taking a HIV test is also relatively low. For an individual with a high value of p_0, the expected treatment benefit of testing is relatively high but the expected change in demand for risky behavior associated with taking a HIV test is relatively low. Whether individuals with high values of p_0 or individuals with values of p_0 in the middle of the p_0 distribution have the greatest demand for HIV testing when ART is available depends on the magnitude of the treatment benefit of testing relative to the benefit associated with the expected change in demand for risky behavior if the individual were to take a HIV test.

1.3.4 Change in risky sexual behavior

Linking the upper and lower halves of Figure 6 yields several important insights about the effect of HIV testing on risky sexual behavior. First, consider the state of the world in which ART is not available. The individual chooses to take a HIV test if and only if $E[U(test)] \geq E[U(not\ test)]$. That is, the individual will test if and only $p_0 \in A$. The associated expected change in risky behavior for this individual is $(1 - p_0)p_1(0) + p_0 p_1(1) - p_1(p_0)$.

Now suppose that ART is available. As in the absence of ART, if $p_0 \in A$ then the individual will choose to take a HIV test. Moreover, if $p_0 \in A_1$ or $p_0 \in A_2$ then the individual will also test. If individuals are uniformly distributed on the unit interval, the number of individuals in the interval A_1 is greater than those in the interval A_2 and hence the number of individuals induced to test with $p_0 > max\{p_0 \mid p_0 \in A\}$ is greater than those induced to test with $p_0 < min\{p_0 \mid p_0 \in A\}$. Because the introduction of ART increases the mean p_0 among those individuals taking a HIV test as compared to the "no ART" case, the per-person reduction in risky behavior associated with HIV testing when ART is available is greater than that in the absence of ART.

1.3.5 Effect of ART rationing

We may use Figure 6 to examine the effects of ART rationing on demand for HIV testing and for risky behavior. In particular, we may examine the effect a rationing mechanism that favors individuals that happen to have relatively low prior probabilities of being HIV positive (e.g., males age 40-59). As compared to the no-rationing case, such a rationing mechanism would serve to reduce the expected treatment benefit of testing among individuals with relatively high values of p_0 while affecting the expected treatment benefit of testing among low-p_0 individuals less. Because it reduces the expected treatment benefit of testing more among individuals expected to demonstrate larger reductions in post-testing risky behavior, a rationing mechanism of this sort reduces the prevention impact of ART-induced testing.

1.4 References

Attawell, Kathy and Jackie Mundy. 2003. Provision of Antiretroviral Therapy in Resource-Limited Settings: A Review of Experience up to August 2003. MIMEO.

Bond, George. 1982. Education and Social Stratification in Northern Zambia: The Case of the Uyombe. *Anthropology and Education Quarterly*, 13(3): 251-267.

Boozer, Michael and Tomas Philipson. 2000. The Impact of Public Testing for Human Immunodeficiency Virus. *The Journal of Human Resources*, 35(3): 419-446.

Colson, Elizabeth and Thayer Scudder. 1981. "Old Age in Gwembe District, Zambia". in Pamela Amoss and Stevan Harrell, eds. 1981. *Other Ways of Growing Old*. Stanford, CA: Stanford University Press.

Epstein, A. L. 1981. *Urbanization and Kinship*. New York, NY: Academic Press.

Global HIV Prevention Working Group. 2004. *HIV Prevention in the Era of Expanded Treatment Access*.

Jones, Peris. 2004. Who is Getting or Will Get Anti-Retroviral Treatment in Zambia? Equity, Access, and Fairness in the Governance of "Scaling Up" HIV/AIDS Medication. MIMEO.

Lakdawalla, Darius; Neeraj Sood; and Dana Goldman. 2006. HIV Breakthroughs and Risky Sexual Behavior. *The Quarterly Journal of Economics*, 121(3): 1063-1102.

Magruder, Jeremy. 2006. Marital Shopping and Epidemic AIDS. MIMEO.

Philipson, Tomas and Richard Posner. 1995. A Theoretical and Empirical Investigation on the Effects of Public Health Subsidies for STD Testing. *The Quarterly Journal of Economics*, 110(2): 445-474.

Scudder, Thayer. 1962. *The Ecology of the Gwembe Tonga*. Manchester, England: Manchester University Press.

Stringer, Jeffrey; Isaac Zulu; Jens Levy; Elizabeth Stringer; Albert Mwango; Benjamin Chi; Vilepe Mtonga; Stewart Reid; Ronald Cantrell; Marc Bulterys; Michael Saag; Richard Larlink; Alwyn Mwinga; Tedd Ellerbrock; and Moses Sinkala. 2006. Rapid Scale-Up of Antiretroviral Therapy at Primary Care Sites in Zambia: Feasibility and Early Outcomes. *Journal of the American Medical*

Association, 296(7): 782-793.

Thornton, Rebecca. 2006. The Demand for and Impact of Learning HIV Status: Evidence from a Field Experiment. MIMEO.

UNAIDS/WHO. 2006. Epidemiological Fact Sheets on HIV/AIDS and Sexually Transmitted Infections.

Wawer, Maria; Ronald Gray; Nelson Sewankambo; David Serwadda; Xianbin Li; Oliver Laeyendecker; Noah Kiwanuka; Godfrey Kigozi; Mohammed Kiddugavu; Thomas Lutalo; Fred Nalugoda; Fred Wabwire- Mangen; Mary Meehan; and Thomas Quinn. 2005. Rates of HIV-1 Transmission Per Coital Act, by Stage of HIV-1 Infection, in Rakai, Uganda. *The Journal of Infectious Diseases*, 191: 1403-9.

World Health Organization (WHO). 2006a. WHO News Release. August 2006.

World Health Organization (WHO). 2006b. *Progress on Global Access to HIV Antiretroviral Therapy: A Report on "3 by 5" and Beyond*.

1.5 Appendix

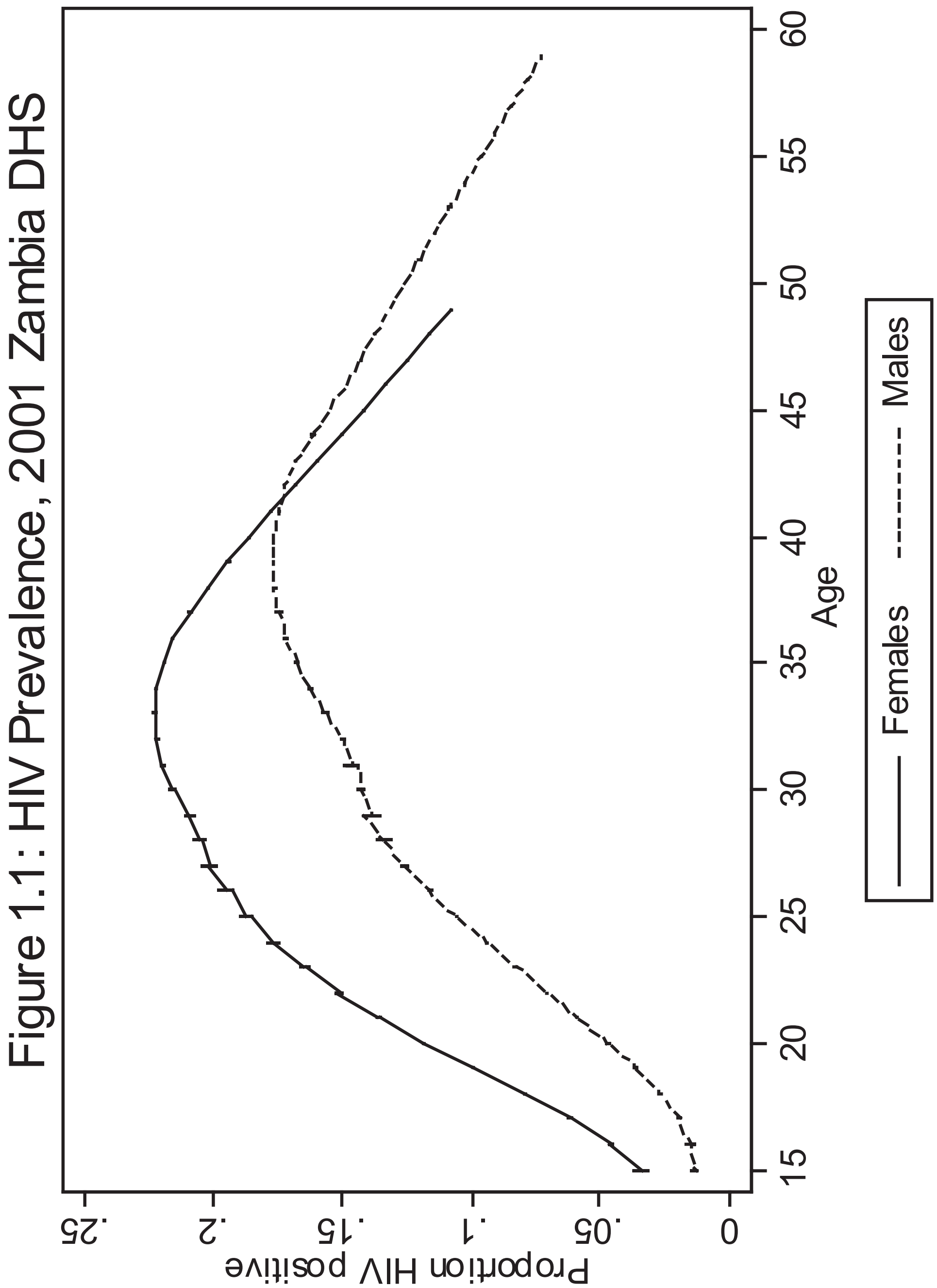


Figure 1.1: HIV Prevalence, 2001 Zambia DHS

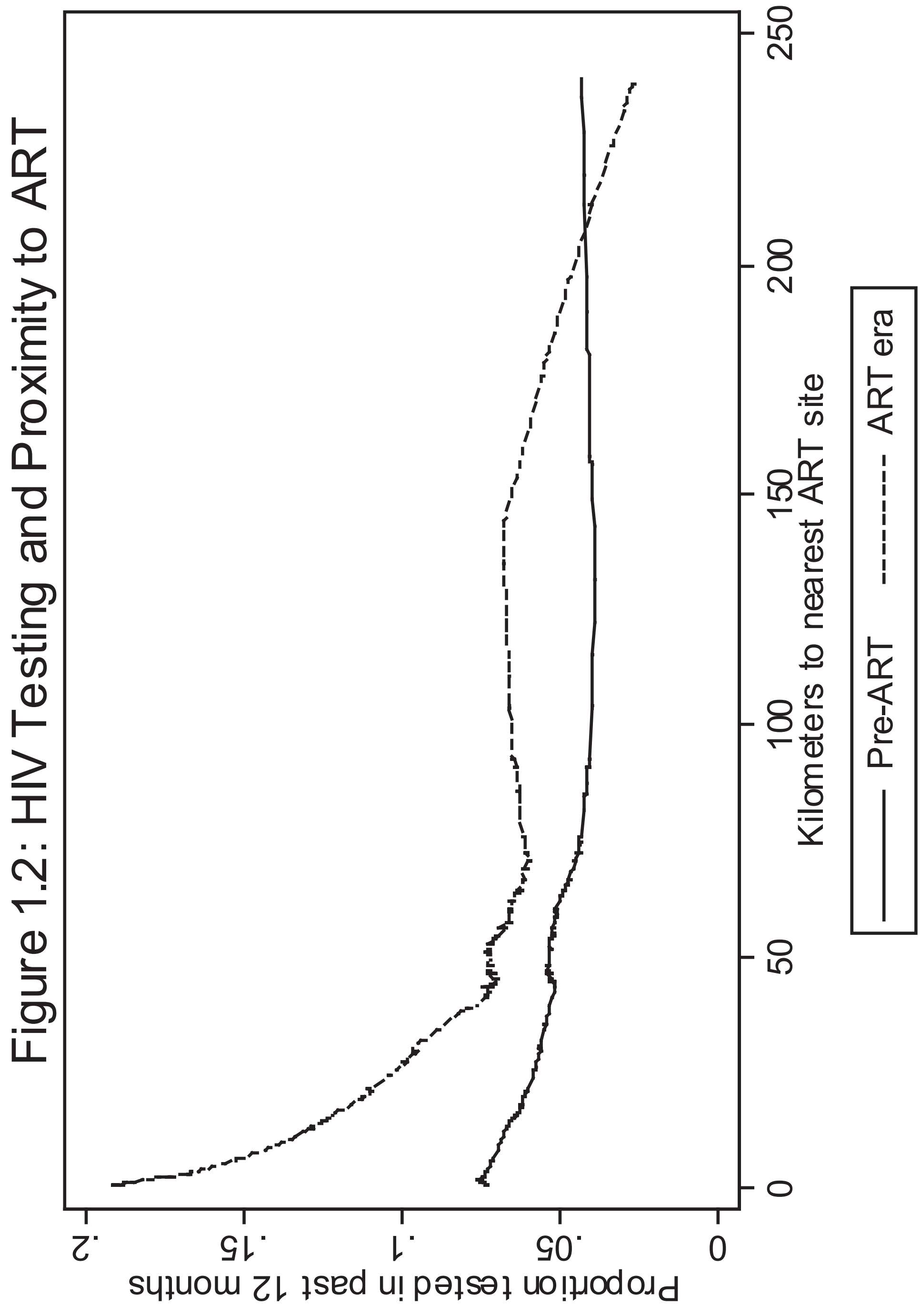


Figure 1.2: HIV Testing and Proximity to ART
Proportion tested in past 12 months
0
.05
.1
.15
.2
Kilometers to nearest ART site
0
50
100
150
200
250
Pre-ART
ART era

Figure 1.3: HIV Testing and Age, Females ZSBS

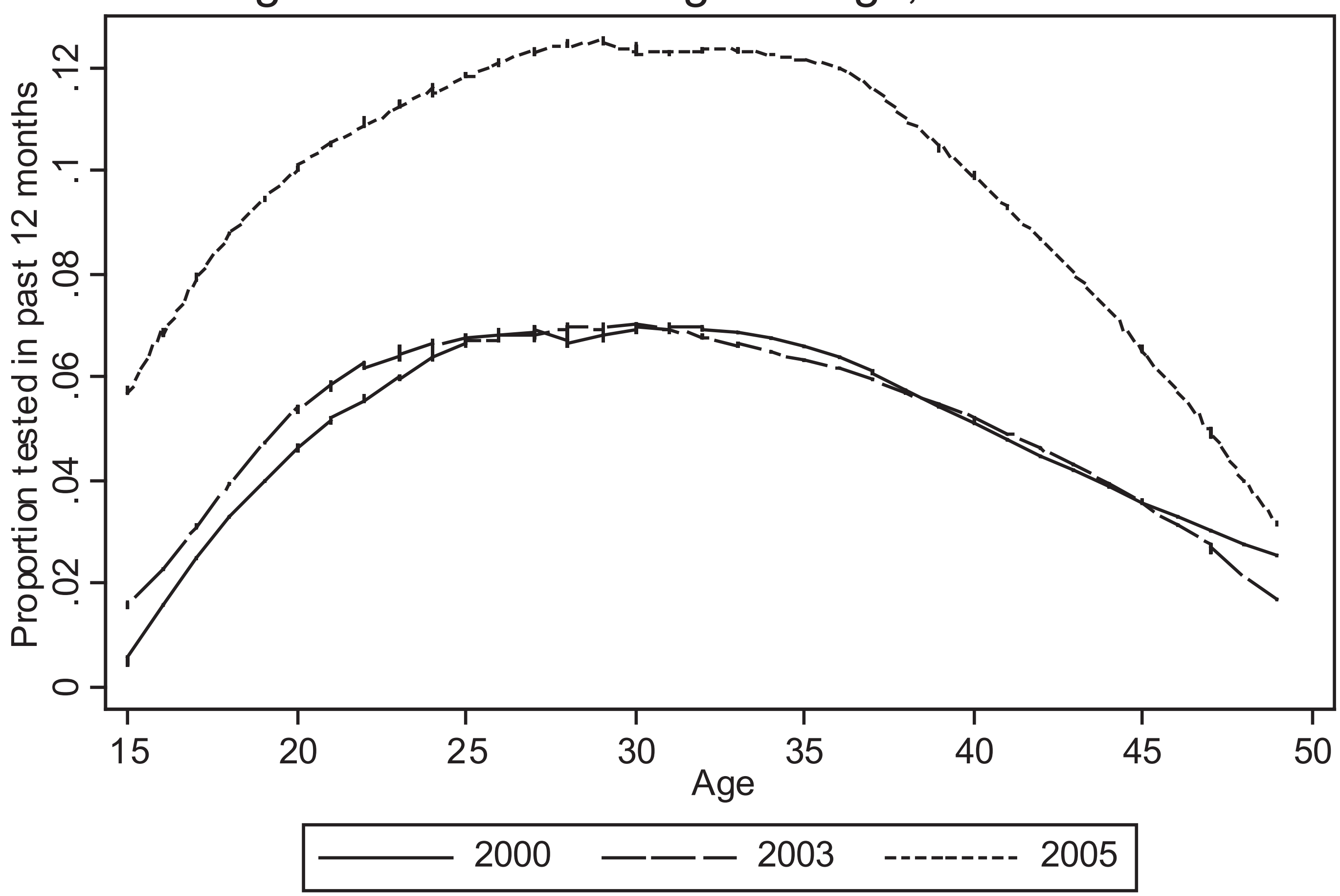


Figure 1.4: HIV Testing and Age, Males ZSBS

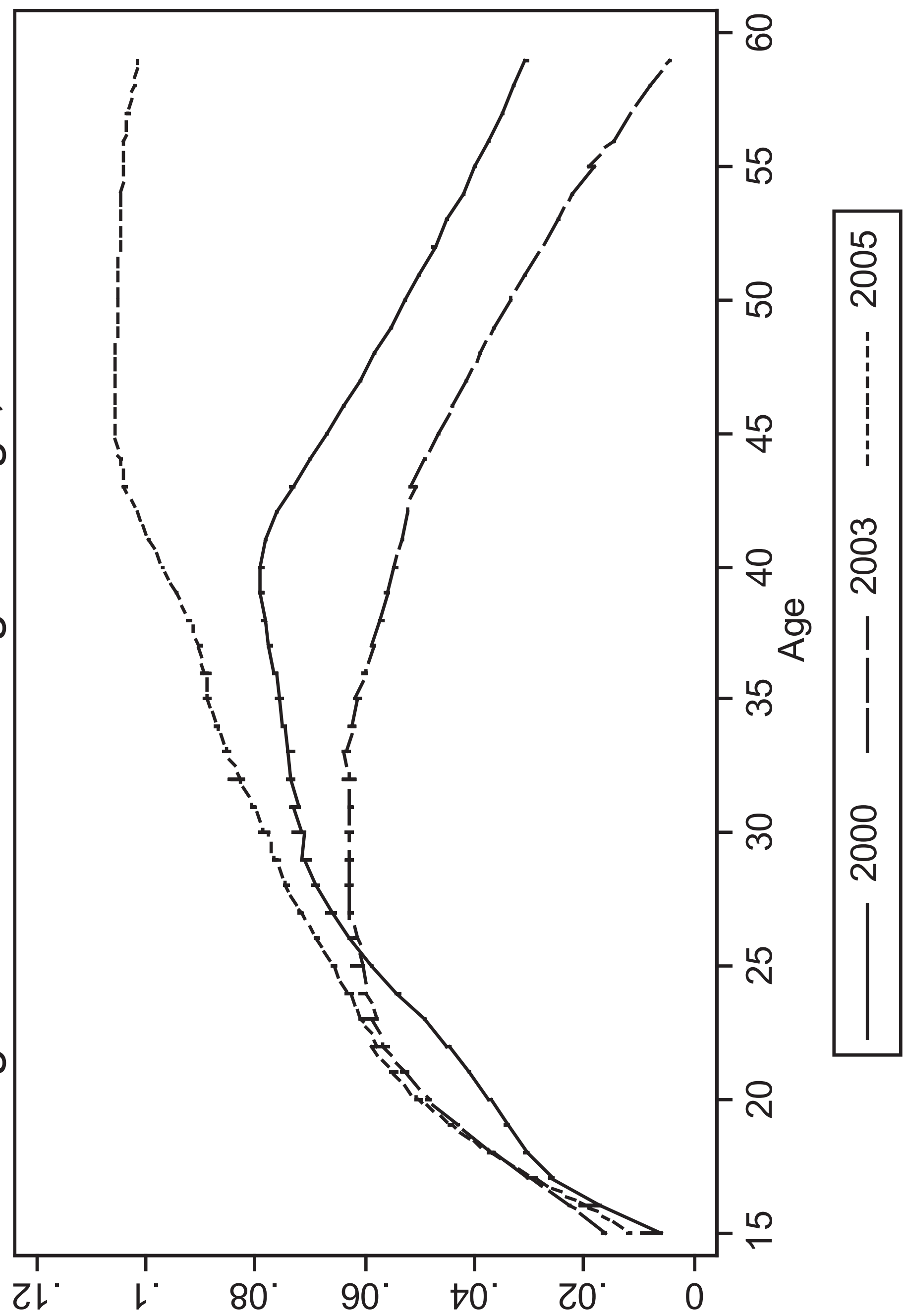


Figure 1.5: Proximity to VCT and Age, 2005 ZSBS

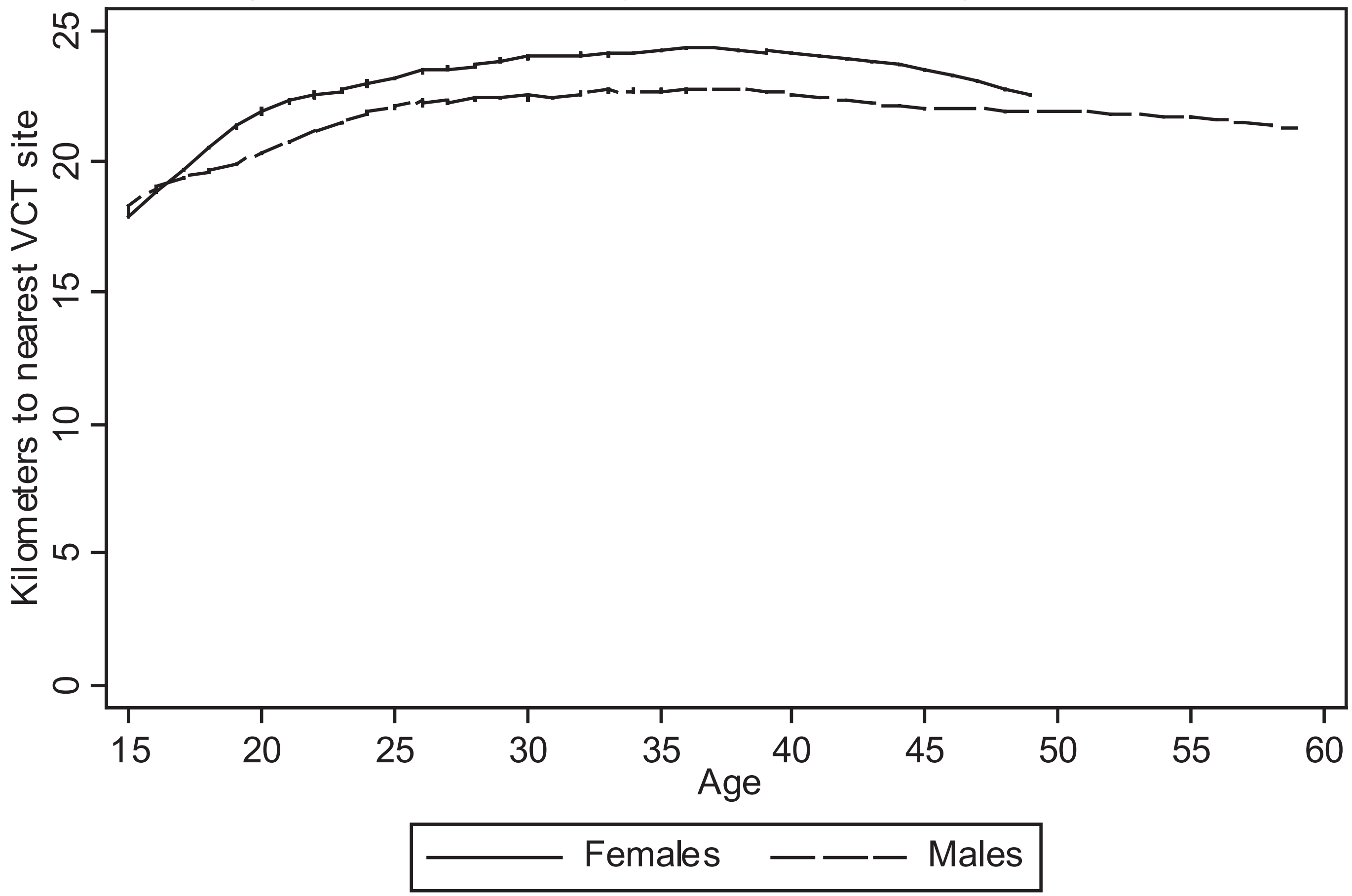
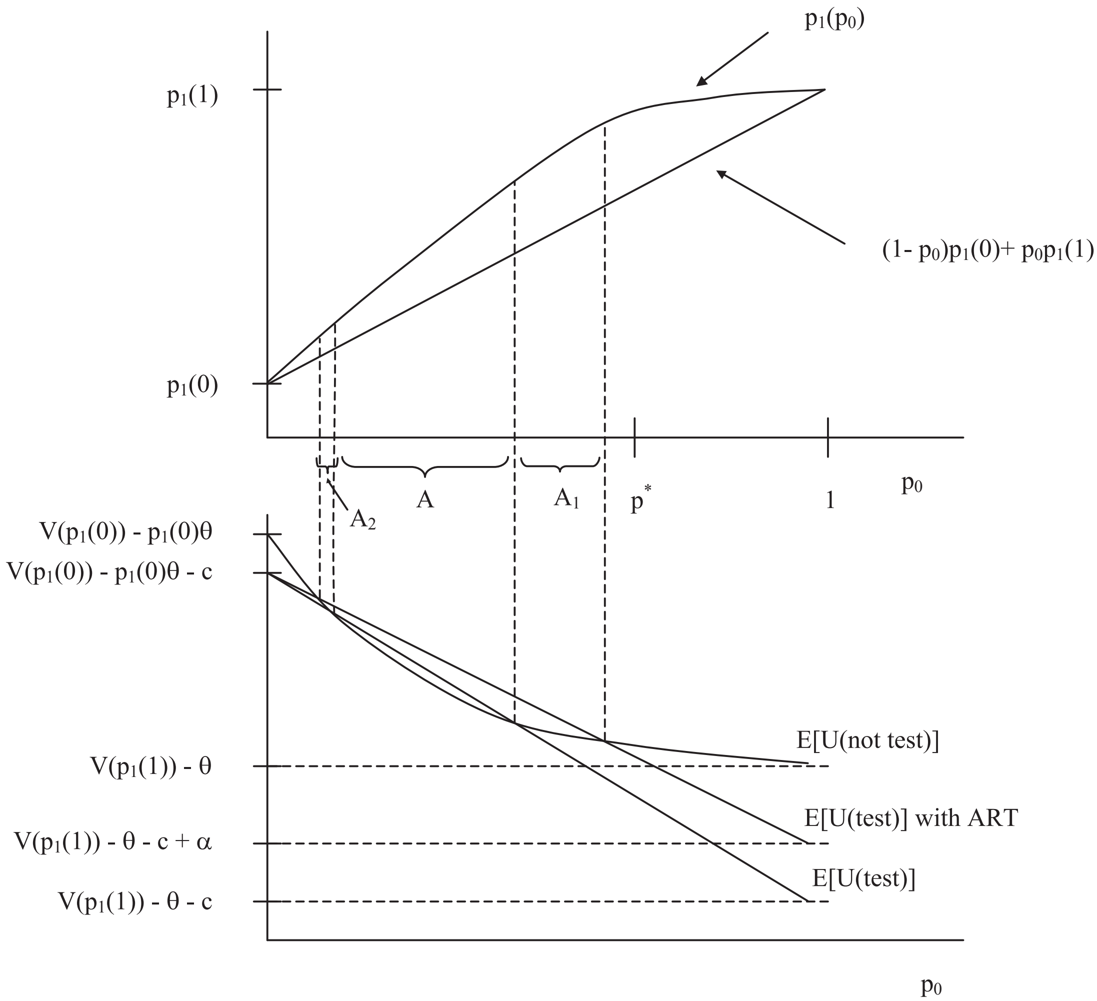


Figure 1.6: Demand for HIV Testing and for Risky Behavior



Chapter 2

A Model of Demand for HIV Testing: Structural Estimation and Results

2.1 Data and estimation

2.1.1 Data

I use data on the outcomes of individuals' testing decisions (i.e., whether or not they chose to take a HIV test) and data on HIV prevalence by demographic group to estimate the structural parameters of the model. I do not observe choices about post-testing risky behavior. In addition, I do not observe the results of the voluntary tests reported in the ZSBS.

The 2001 Zambia DHS (ZDHS) includes the results of anonymous HIV testing linked to information on the age, gender, and province of residence of each individual tested. The sample of individuals for whom a HIV test was requested was defined to be all men and all women in one-half of households selected for the men's survey module in the ZDHS. In total, approximately 5,000 individuals were asked to provide a blood sample. Among women approached to provide a sample, 15.4 percent refused and 5.3 percent were either absent or the result is missing. Among men approached to provide to a sample, 14.8 percent refused and 12.1 percent were either absent or the result is missing. The resulting sample consists of testing results for 3,949 observations.

Because the sample size for any given age-gender-province interaction is small, individuals are grouped into five-year age ranges (e.g., 15-19 years old, 20-24 years old, etc.). The ZDHS includes women age 15-49 and men age 15-59 and there are nine provinces in Zambia, so this procedure yields 144 groups with HIV prevalence ranging from zero percent to 35.3 percent. For each group I treat the proportion that were found to be HIV positive as the initial probability of being HIV positive, p_0.[1] I assume that the age-profile of HIV prevalence is static for a given gender-province interaction over the period covered by the repeated cross-sectional data used in this analysis, the Zambia Sexual Behavior Survey (ZSBS).

The data on whether an individual voluntarily chose to take a HIV test come from the ZSBS. I use the three most recent rounds of the ZSBS: 2000, 2003, and 2005. The sample sizes for each of these survey rounds are 3316, 4471, and 4208 individuals, respectively. To match the testing data to the prevalence data the testing decisions of individuals in the ZSBS are aggregated by demographic group defined as the interaction of five-year age group, gender, and province of residence. Moreover, for each demographic group I aggregate the two pre-ART survey years (i.e., 2000 and 2003). I then calculate the proportion of individuals that report taking a HIV test in the 12 months prior to the survey date, denoted $test_i$. This procedure yields 288 observations.

2.1.2 Estimation

For a group of individuals with the homogeneous belief they are HIV positive equal to p_{0i}, the proportion predicted to choose to take a HIV test in the pre-ART period is given by

$$\widehat{test}_i = \frac{1}{1 + e^{-(E[U(test)]_i - E[U(not\ test)]_i)/\delta}} \tag{2.1}$$

where δ is a smoothing parameter that is to be estimated along with the other parameters of the model. The empirical specification for $E[U(test)])$ includes as additive terms an indicator variable equal to one if the observation is a male demographic group, denoted $male$, and an indicator variable,

[1] Although the available Zambian data do not include information on the self-reported likelihood of being HIV positive, there are data from Malawi that do include this information. The Malawi Diffusion and Ideational Change Project (MDICP) includes information on respondents' subjective beliefs of being HIV positive as recorded on a scale of "no likelihood", "low", "medium", and "high". *A priori*, it is reasonable to think that beliefs about HIV status in Zambia should be similar to those in Malawi. Malawi is adjacent to Zambia, landlocked (as is Zambia), and has a HIV prevalence rate of approximately 14 percent (roughly equal to that in Zambia.) Over the age ranges available in the 2001 ZDHS, the belief-age profiles for females and males in the 2001 survey round of the MDICP appear to be similar to the prevalence-age profiles in the 2001 ZDHS: both for females and for males, it is individuals in the middle of the age distribution that tend to report the greatest likelihood of being HIV positive. One difference between the belief-age profiles in the 2001 survey round of the MDICP and the prevalence-age profiles in the 2001 ZDHS is that young women age 15-19 in the MDICP report beliefs of being HIV positive that are as high as those reported by women in the middle of the age distribution. However, in the 2001 survey round of the MDICP there are only 24 women between the ages of 15 and 19 as compared to 124 in the five-year age group with the second fewest number of observations. Moreover, I fail to reject the hypothesis that the mean belief among women ages 15-19 is significantly different from that among women ages 20-24.

denoted *post*, equal to one if the the observation comes from the 2005 survey round of the ZSBS.

In the ART era, the predicted proportion testing in group i is generated by Equation (8) as well, but with an expanded expression for $E[U(test)]$ *with ART* substituted for $E[U(test)]$. As discussed previously, the changes in the testing-age profiles discussed in Section 2 differ from those predicted by the theoretical model. The model predicts that the individuals who demonstrate the greatest testing response to the introduction of ART should have prior probabilities of being HIV positive that are at least as large as those among individuals that were more likely to test prior to the introduction of ART. In contrast, the observed testing response among males do not fit this prediction. One explanation for this puzzle is that there exists a non-random rationing mechanism for allocating ART. In this case, we may expect that the probability that an individual is favored in this process is correlated with his demographic characteristics.

To capture this possibility, I estimate the probability that the individual in group i will receive ART conditional on p_{0i}, denoted $p_{0i}^{'}$, by including measures of each individual's age, age squared, gender, and interaction terms in the empirical specification for the expected utility of testing when ART is available. Specifically, I include the righthand side of following expression as an additive term in the empirical specification for $E[U(test)]$ *with ART* (i.e., Equation 6)

$$p_{0i}^{'} = \frac{1}{1 + e^{X_i \beta}} \tag{2.2}$$

where X_i is a vector whose elements include age, age squared, gender, and interaction terms. In addition, I examine the relative importance of this rationing mechanism by estimating weights for the effects of the prior probability of being HIV positive (i.e., p_0) and of the probability of meeting the non-random rationing criteria (i.e., $p_0^{'}$). These weights are denoted $(1 - \sigma)$ and σ, respectively. Thus, the expanded expression for $E[U(test)]$ *with ART* that I use in the empirical specification is

$$\begin{aligned} E[U(test)]\ with\ ART &= (1 - p_0)[V(p_1(0)) - p_1(0)\theta - c] + p_0[V(p_1(1)) - \theta - c] \\ &+ (1 - \sigma)\alpha p_0 + \sigma \alpha p_0^{'}. \end{aligned} \tag{2.3}$$

Because the ultimate product of the empirical analysis is a set of policy simulations, I need to parameterize the model. In particular, I need to specify a functional form for the utility derived from choice of risky behavior, $V(p_1)$. I choose a functional form for $V(p_1)$ that is sufficiently flexible to allow a variety of shapes for the risky behavior demand function, $p_1(p_0)$. Specifically, the utility derived from choice of risky behavior is given by

$$V(p_1) = \frac{(p_1 - \frac{b}{e})^2}{ep_1 - 1}. \tag{2.4}$$

The flexibility of this specification means that I am able to let the data confirm that the risky behavior demand function is in fact increasing and concave in p_0, as was assumed in the discussion of the model in Section 3.

To estimate the structural parameters of the model, I minimize a simple weighted sum of squared residuals

$$\sum_{i=1}^{288} \frac{pop_i}{pop}(test_i - \widehat{test_i})^2 \tag{2.5}$$

where the weights are given by the proportion of the total adult population (pop) that is in p_0 group i (pop_i).

2.2 Results

2.2.1 Parameter estimates

Table 1 provides the parameter estimates that are the solution to the weighted non-linear least squares minimization problem. Column (1) lists the estimates for the baseline model with no additional controls. The baseline model includes the variables indicated by the theoretical model, as well as an indicator variable for the "post" period and an indicator variable for male. Substituting these parameters into Equation (2) and examining the first and second derivatives verifies that the expected utility of not testing is decreasing and convex in the prior probability of being HIV positive. In addition, the utility cost of testing is positive (i.e., $c > 0$), utility is decreasing in the probability that the individual eventually acquires HIV (i.e., $\theta > 0$), males appear to be slightly more likely to test than females, and there is evidence of a secular increase in testing rates unrelated to the introduction of ART. However, neither of these two latter effects are significantly different from zero.

The parameter estimates indicate that individual behavior is broadly consistent with a model in which ART is not allocated to all HIV positive individuals, but is instead rationed based on the individual's age and gender (or factors associated with these characteristics). In the baseline specification, the estimated σ equals 0.97. In other words, the probability that an individual responds to the availability of ART by testing is increasing in the probability that he is HIV positive but is also correlated with the individual's demographic characteristics conditional on p_0. Of course the fact that $\sigma < 1$ indicates that individual behavior is consistent with a model where the probability that an individual is HIV positive also affects the testing decision irrespective of these demographic characteristics. Nonetheless, this effect appears to be small and it appears that there exists a non-random rationing mechanism that does affect testing behavior.

If province-level unobserved heterogeneity in testing demand is correlated with variables in the model, then the estimates presented in Column (1) are biased. For example, individuals in high-prevalence provinces may face lower costs of testing than individuals in low-prevalence provinces (perhaps because policymakers may place more VCT sites in high-prevalence areas). To address this concern (as well as other concerns) about time-invariant factors, Column (2) presents parameter estimates from a model including province dummies as well as the two controls included in the baseline model.

As shown in Column (2), the basic results of the model do not change after it is re-estimated with province dummies. The expected utility of not testing is still decreasing and convex in the prior probability of being HIV positive, θ and c still have the expected signs, and individual behavior remains consistent with a model in which there exists substantial rationing of treatment based on age.

As previously mentioned, the number of VCT sites increased between the pre- and post periods. If the intensity of this expansion differed across provinces and this difference was correlated with variables in the model, then we may be concerned that the estimates in Columns (1) and (2) are biased. To address this possibility, I interact the "post" dummy with the province dummies and include these in the expanded empirical specification. Column (3) presents the results of this exercise. Again, the basic results do not change.

2.2.2 Predicted testing-age profiles

Comparing the predicted and observed testing rates is a simple method of assessing the fit of the empirical specification discussed above. Figure 7 presents predicted and observed testing-age profiles for females and for males in the pre-ART period and in the ART era. The predicted testing-age profile in a given period in Figure 7 is the result of a locally weighted smoothed regression of the predicted testing rate for group i on age. The observed testing-age profile in a given period in Figure 7 is the result of a locally weighted smoothed regression of an indicator variable equal to one if an individual tests on age. For a given gender there are three panels in Figure 7, one panel for each of the three empirical specifications discussed above.

The empirical specification appears to generate predicted values that are relatively close to the observed HIV testing rates. The predicted testing rates tend to be within one percentage point or less of the observed testing rate. The mean residuals are largest for the youngest females in the ART era and for the youngest males in the pre-ART period. In general, the empirical specification generates predicted testing-age profiles in the pre-ART period that take the inverted U-shape that characterizes the observed testing-age profiles and predicted testing-age profiles in the ART era that take the distinct shapes that characterize the observed testing-age profiles for females and males.

2.2.3 Predicted risky behavior demand function

Figure 8 presents the predicted risky behavior demand function, $p_1(p_0)$, from the empirical specification that includes the full set of controls. Although the empirical specification does not impose that $p_1(p_0)$ be increasing or concave in p_0, the estimated risky behavior demand function is in fact increasing and concave. The estimated $p_1(p_0)$ function implies that if an individual knew with certainty that they were HIV negative then they would choose an amount of risky sexual behavior in which to engage that would imply that they would acquire HIV with probability 0.140. On the other hand, if an individual believed with certainty that they were HIV positive then they would choose an amount of risky sexual behavior in which to engage that would imply that they would acquire HIV with probability 0.574 if they were in fact not HIV positive. The prior probability of being HIV positive that maximizes the expected reduction in risky behavior associated with taking a HIV test is 0.460. However, even at its maximum the expected reduction in risky behavior associated with taking a HIV test in terms of the change in the probability the reference individual acquires HIV - 0.047 - is relatively small.

2.3 Policy scenarios

This section uses the parameter estimates from the preceeding section to examine the behavioral and welfare implications of five alternative ART policies: (i) no ART, (ii) the status quo ART policy, (iii) expanding the supply of ART, (iv) eliminating the non-random rationing mechanism while holding constant the quantity of ART supplied, and (v) simultaneously eliminating the non-random rationing mechanism while expanding the supply of ART. I employ three metrics to examine the performance of each of these policies. The first metric is the predicted demand for HIV testing (i.e., the HIV testing rate). The second metric is the mean predicted choice of risky behavior. The third metric is the simulated number of new infections. The first two metrics are relatively straightforward to calculate. The simulated number of new infections, however, requires particular discussion.

I simulate the effect of testing on new infections as follows. Each individual in a given demographic group may be characterized along two dimensions: whether they take a HIV test or not and whether they are HIV positive or not. In a given demographic group, whether an individual chooses to take a HIV test or not (as well as the result of the test) affects the individual's choice of risky behavior. Only individuals that are HIV negative are at risk of acquiring a new infection.

The number of new infections among individuals taking a HIV test is given by

$$\sum_{i=1}^{N} n_i \widehat{test}_i (1 - p_{0i}) \widehat{p_1}(0) \tag{2.6}$$

where n_i is the number of individuals in group i and N equals 144, the number of demographic groups. For demographic group i, the product of the first three terms in this expression yields the number of individuals in this group that are at-risk of acquiring HIV. Because these individuals are HIV negative and the results of the HIV tests reveal this fact to each of the individuals, they choose to engage in an amount of risky behavior summarized by $\widehat{p_1}(0)$; this is the probability that they acquire HIV.

The number of new infections among individuals not taking a HIV test is given by

$$\sum_{i=1}^{N} n_i (1 - \widehat{test}_i)(1 - p_{0i}) \widehat{p_1}(p_{0i}). \tag{2.7}$$

Again, the product of the first three terms in this expression yields the number of individuals in demographic group i that are at-risk of acquiring HIV. Because these individuals choose not to take a HIV test, they engage in an amount of risky behavior summarized by $\widehat{p_1}(p_0)$; this is the probability that they acquire HIV. Total new infections are given by the sum of these two expressions.

This method of simulating the number of new infections has one major shortcoming. Namely, it ignores new infections among partners of the reference individuals. Thus, it is important to recognize that this method measures the expected number of new infections in a given population with error.

2.3.1 No ART

The first policy scenario I examine is a scenario where there is no ART. Applying the three metrics to this case provides a baseline for understanding the impact of each of four other ART policies. I calculate the predicted HIV testing rate in the "no ART" scenario by shutting down the treatment benefit to testing (i.e., setting α equal to zero), setting *post* equal to one, and generating predicted values using Equation (11). The first row of Table 2 presents the results of this calculation for each of the three empirical specifications. Model (1) is the baseline empirical specification with controls for "post" (i.e., the ART era) and for gender. Using the parameter estimates from this specification, the predicted HIV testing rate in the state of the world where there is no ART is 6.1 percent. Model (2) includes province dummies as well as the controls included in Model (1). As Column (2) of Table 2 indicates, the predicted testing rate does not change substantially, consistent with the fact that the parameter estimates did not change substantially. The simulation results indicate that the predicted HIV testing rate is 5.6 percent. Model (3) includes an indicator variable for the "post" period interacted with the province dummies as well as all of the controls in Model (2). Again, neither the parameter estimates nor the predicted testing rate changes dramatically. Using the parameter estimates from Model (3), I estimate that the predicted HIV testing rate in the absence of ART would be 5.5 percent.

The second metric is the mean choice of risky behavior. The first row of Table 3 presents the mean choice of risky behavior under the "no ART" scenario for each of the three empirical specifications. The parameter estimates from Model (1) imply that the mean choice of risky behavior would be 0.1361. This implies that the representative individual would choose to engage in an amount of risky behavior that would lead them to acquire HIV with probability 0.1361 if they were in fact HIV negative at the time of this choice. Models (2) and (3) imply substantially higher mean choices of risky behavior: 0.1651 and 0.1554, respectively. The riskier behavior implied by Models (2) and (3) is partly due to the fact that fewer individuals are predicted to test (and hence learn with certainty that they are HIV negative). In addition, the estimated risky behavior demand function given by the parameter estimates associated with each of these empirical specifications predicts riskier behavior conditional on a negative test result than does Model (1).

The third metric is the simulated number of new infections. Table 4 presents the results of this simulation; the first row of this table represents the "no ART" scenario. According to the parameter estimates from Model (1), the number of new infections in the state of the world where there is no ART appears relatively large: 927,000 new infections. Of course the theoretical model allows for one chance to take a HIV test and one chance to make a decision about future risky behavior. Thus, it may make sense to think about the simulated number of new infections as representing the number of new infections in the population over the course of the average lifetime of an individual in the population. Given that on average an individual between the age of 15 and 59 in Zambia is expected to live approximately 25 more years (WHO 2007), then the simulated annual incidence is nearly 40,000. As a point of comparison, the annual incidence of new adult HIV cases in Zambia is approximately 50,000. Consistent with the higher mean choice of risky behavior in Models (2) and (3) than in Model (1), the simulated number of new infections in Models (2) and (3) - 1,000,000 and 940,000, respectively - are greater than that in Model (1).

2.3.2 Status quo ART policy

The second policy scenario I consider is the status quo ART policy. To calculate the three metrics under this policy scenario, I set *post* equal to one and make no additional changes. The second row in Table 2 presents the predicted proportion of individuals choosing to take a HIV test under this scenario. Depending on the empirical specification, between 7.5 and 8.4 percent of individuals are predicted to take a HIV test. This represents an increase in testing demand of between 2 and 2.3 percentage points (an increase of nearly 40 percent), regardless of the empirical specification.

Although the effect is not nearly as large, ART provision under the status quo ART policy is also predicted to reduce mean choice of risky behavior. As presented in the second row of Table 3, Models (1) through (3) imply that the mean choice of risky behavior would fall by between 0.07 and 0.30 percent (or thousandths of a percentage point).

The second row of Table 4 presents the the simulated number of new infections under the status quo ART policy. Depending on the empirical specification, the number of new infections ranges from 916,000 to 989,000. These figures represent a decline in the incidence of HIV of between 0.96 and 1.19 percent. In general the effect of ART-induced testing under the status quo ART policy appears to be modest at best. However, comparing these figures with the performance of the alternative policies should yield a more complete understanding of the relative magnitude of this effect.

2.3.3 Expanding the supply of ART

The third policy scenario I examine is expanding the supply of ART while retaining the existing source of inefficiency in the process determining the allocation of ART. If the quantity of antiretroviral drugs available in Zambia is not sufficient to meet the demand for ART, then treatment providers may ration ART among the pool of eligible individuals without regards to each individual's identity. I consider the policy of expanding supply to ensure that all individuals that would be allocated ART under the existing non-random rationing mechanism are able to begin treatment. Expanding the supply of ART should yield an increase in the prevention impact of ART-induced testing as compared to the status quo ART policy. Because expanding the supply of ART increases the probability that an individual will receive the payoff associated with taking a HIV test conditional on the prior probability the individual is HIV positive (and on the probability that they would be allocated ART under the non-random rationing mechanism), it should increase the proportion of individuals testing. The cost of this policy is simply the increase in drug expenditures (and associated labor costs) required to cover these additional patients.

It appears that the quantity of antiretroviral drugs in Zambia is not sufficient to meet the demand for ART. I calculate the ratio of the number of individuals on ART at the time of the 2005 round of the ZSBS (i.e., 30,000) to the number of individuals that are predicted to test in the 2005 round of the ZSBS and to be eligible for ART. Let q denote this ratio. Then

$$q = \frac{30,000}{\sum_{i=1}^{N}((1-\widehat{\sigma})\widehat{p_{0i}} + \widehat{\sigma}\widehat{p_{0i}}')\widehat{test}_i n_i} \tag{2.8}$$

where N equals 144, the number of demographic groups. Consistent with the existence of excess demand for ART, I find that this ratio is substantially less than one ($q = 0.59$). If there are individuals that did not take a HIV test in the 12 month interval but were still eligible for ART and able to demonstrate their status to the health authorities (e.g., from the results of a test taken prior to the beginning of the 12 month interval), then the q given by Equation (15) is going to understate the extent of rationing of ART using a random lottery. On the other hand, not all HIV positive individuals are recommended to begin ART and so the q given by Equation (15) may overstate the extent of rationing of ART using a random lottery.

To calculate the expected level of testing associated with ART availability under the expanded supply scenario, first I substitute the parameter estimates from Table 1 (except the estimated α) into the predicted testing equation and set *post* equal to one. Then I substitute $\frac{1}{0.59}\widehat{\alpha}$ for α in the predicted testing equation (where $\widehat{\alpha}$ is the estimated α). Because the effect of rationing without regards to individuals characteristics is not separately identified from the benefit of beginning ART (i.e., α), eliminating the rationing of ART using a random lottery is equivalent to increasing α. The third set of rows of Table 2 present the results of this exercise. Depending on the empirical specification, I predict that expanding the supply of ART while retaining the existing non-random ART rationing mechanism should increase testing rates to between 9.2 and 10.3 percent. This represents an increase in testing rates as compared to the status quo ART policy of between 1.7 and 1.9 percentage points.

The third set of rows in Table 3 presents the mean choice of risky behavior under the "expand supply" scenario. Although the mean choice of risky behavior in this scenario is smaller than that under either the “no ART” or the “status quo ART” scenarios, the reduction in mean risky behavior is still small in an absolute sense. As compared to the “no ART” scenario, mean risky behavior is between 0.15 and 0.60 percent smaller under the “expand supply” scenario.

The simulation results fail to provide support for the idea that the continued roll-out of ART (in the absence of other interventions designed to eliminate the non-random allocation process) will induce substantial reductions in the incidence of new HIV infections. As shown in the third set of rows in Table 4, the number of new infections avoided under this policy as compared to the state of the world without ART is only between 1.81 and 2.27 percent. In relative terms, however, this policy is much more effective than the status quo ART policy: simply expanding the supply of ART to eliminate the random lottery increases the number of new infections avoided by at least 88 percent.

2.3.4 Eliminating the source of inefficiency

The fourth policy scenario is to eliminate the existing source of inefficiency in the ART allocation process while holding constant the quantity of ART supplied. Under this policy, all individuals that are HIV positive are eligible for ART and age and gender no longer affect the ART allocation decision conditional on HIV status. One way to implement this policy would be to mandate that the enrollment-age profile at ART sites in each province match the province’s prevalence-age profile. Because the existing supply of ART is held constant, eliminating the inefficiency would require no additional expenditure on antiretroviral drugs. The tradeoff or cost of this policy is that individuals that are currently favored in the ART allocation process would face a lower probability of beginning ART conditional on being HIV positive. Moreover, because the quantity of ART supplied is held constant while the number of individuals eligible for ART increases, the probability that an HIV positive individual that takes a HIV test is allocated ART in the random lottery decreases.

Eliminating the source of inefficiency in the process determining the allocation of ART has the potential to yield a large increase in the number of new infections avoided. The model provided in this paper indicates that the expected reduction in risky behavior associated with taking a HIV test is increasing in the initial probability of being HIV positive over the range relevant in the current setting. However, empirical evidence indicates that current ART policy effectively favors populations (i.e., older males) with prior probabilities of being HIV positive that are much lower than the maximum prior probability observed in the Zambia DHS. Thus, eliminating the source of the inefficiency will increase the number of new infections avoided due to ART-induced testing in part because the composition of the pool of individuals choosing to take a HIV test will shift toward individuals expected to demonstrate relatively large reductions in post-testing risky behavior. In addition, there are many more individuals in groups that are predicted to voluntarily test at high rates in the absence of the source of inefficiency (e.g., males in the middle of the age distribution) than there are in the group (i.e., older males) that currently tests at a high rate. Hence, eliminating the source of inefficiency will increase the number of new infections avoided due to ART-induced testing in part because the number of individuals testing after the inefficiency is addressed would be much greater than the number that test under the status quo ART policy.

I operationalize this policy experiment by generating predicted testing values using the parameter estimates from Table 1 (except the estimated α), setting $post = 1$, $\sigma = 0$, and recalibrating q. I recalibrate q to reflect the fact that if the policy eliminates the non-random rationing mechanism then the number of individuals that respond to ART by testing will be greater. The new value of q is 0.3. Because the quantity of ART supplied is held constant in this policy, the increase in testing demand means a decrease in the chance of being selected in the random lottery. Therefore, I substitute $0.3\widehat{\alpha}$ for α in the predicted testing equation. Again, because the effect of a random lottery determining the allocation of ART among HIV positive individuals is not separately identified from the benefit of beginning ART (i.e., α), the effect of decreasing the probability an individual is selected in the lottery is equivalent to decreasing the benefit of beginning ART.

Regardless of the metric under consideration, the simulated results of this policy provide clear evidence of the cost of the non-random rationing mechanism identified in this research. Predicted testing demand is between 33.8 and 36.2 percent, a more than three-fold increase over the that under the status quo ART policy. The percent reduction in mean risky behavior as compared to the "no ART" policy is between 0.88 and 3.39, at least eleven times greater than the reduction under the status quo ART policy. As compared to the status quo policy, the number of new infections avoided due to ART-induced testing would be more than ten times as large if the non-random rationing mechanism were eliminated: between 139,000 and 161,000 new infections avoided.

2.3.5 Eliminating the inefficiency while expanding the supply of ART

The fifth policy scenario is to expand the supply of ART in combination with eliminating the existing source of inefficiency. In particular, I consider an expansion in the supply of ART that is sufficient to ensure that all individuals that are HIV positive would be able to begin ART. This policy yields the both the mechanical benefit of increasing ART-induced testing by eliminating the random lottery for ART as well as the composition benefit of inducing additional testing among individuals expected to demonstrate greater post-testing reduction in risky behavior. Of course this policy requires the greatest increase in expenditure. The increase in expenditure on antiretroviral drugs (and the associated labor costs) is greater than that under the policy of expanding the supply of ART while retaining the existing source of inefficiency. Moreover, eliminating the source of inefficiency may itself be costly.

To simulate the effect of this new policy, I set $post = 1$, $\sigma = 0$, replace α with $\frac{1}{0.59}\widehat{\alpha}$, and substitute the parameter estimates from Table 1 for the remaining parameters. The simulation results indicate that testing demand under this policy would be between 52.1 and 54 percent, a more than seven-fold increase as compared to the "no ART" policy. The mean choice of risky behavior would be between 5.51 and 8.96 percent smaller than under the "no ART" policy. In terms of new infections avoided, this policy would have very large effects: the number of new infections would be at least 23 percent (28 percent) smaller than that under the "no ART" scenario ("status quo" scenario).

2.4 Conclusion

The debate about appropriate HIV/AIDS policy often features an analytic framework in which interventions are classified as treatment interventions or as prevention interventions (see Canning 2006). Antiretroviral drugs and care for opportunistic infections (e.g., tuberculosis) are usually seen as treatment interventions. HIV testing and condoms are usually seen as prevention interventions. However, many proponents of HIV testing argue that antiretrovirals are not only a treatment intervention and that the availability of ART will induce substantial increases in testing demand and subsequent reductions in risky behavior.

The results of the current analysis indicate that the prevention impact of antiretrovirals under the existing ART policy in Zambia is modest at best. The observed changes in testing rates associated with the introduction of ART are not consistent with the prediction of a model of testing demand where the incentive effect of ART is increasing in the prior probability that the individual is HIV positive. Instead, there appears to exist a source of inefficiency in the process determining the allocation of ART. Because of this inefficiency, the number of new infections avoided due to ART-induced testing is relatively small; those individuals largely favored in the ART allocation process are predicted to demonstrate only small reductions in risky behavior subsequent to testing. Simulation results indicate that ART-induced testing under the existing ART policy reduces the number of new infections by approximately 1 percent.

Eliminating the inefficiency in the process determining the allocation of ART yields substantial increases in the prevention impact of ART. Simulation results indicate that this intervention would increase the number of new infections avoided due to ART-induced testing by more than ten-fold. Moreover, this intervention requires no additional expenditure on antiretroviral drugs. In comparison, expanding the supply of ART without eliminating this inefficiency would only double the number of new infections avoided due to ART-induced testing while requiring a substantial increase in expenditure on antiretroviral drugs. The results of this paper confirm the emphasis among policymakers on the central role of ART-induced HIV testing in reducing the spread of HIV/AIDS and suggest that future research examine the process determining the allocation of ART in regions experiencing a rapid change in the availability of ART.

2.5 References

Canning, David. 2006. The Economics of HIV/AIDS in Low-Income Countries: The Case for Prevention. *Journal of Economic Perspectives*, 20(3): 121-142.

2.6 Appendix

Figure 2.1: Predicted and Observed Testing-Age Profiles

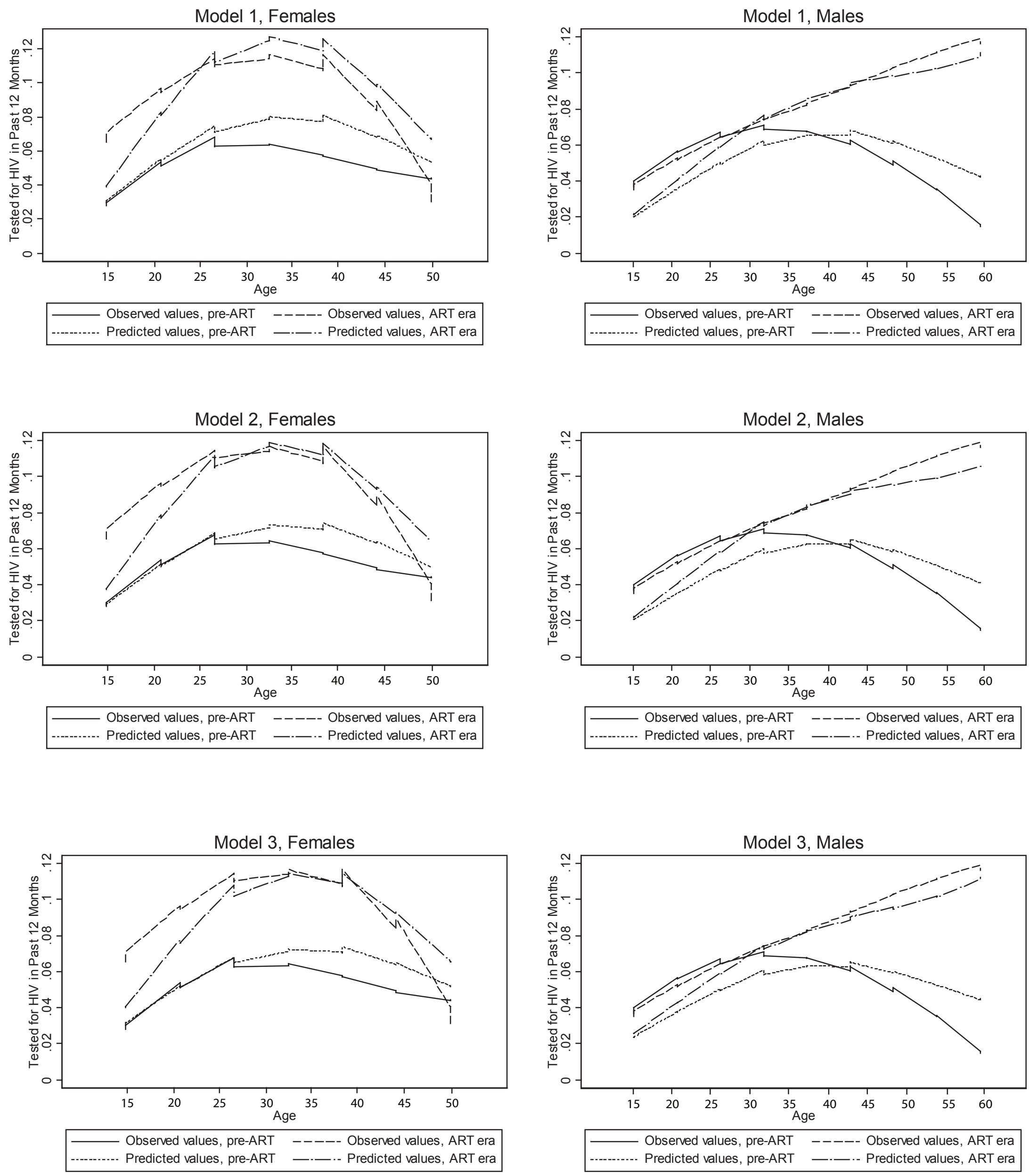


Figure 2.2: Estimated Risky Behavior Demand Function

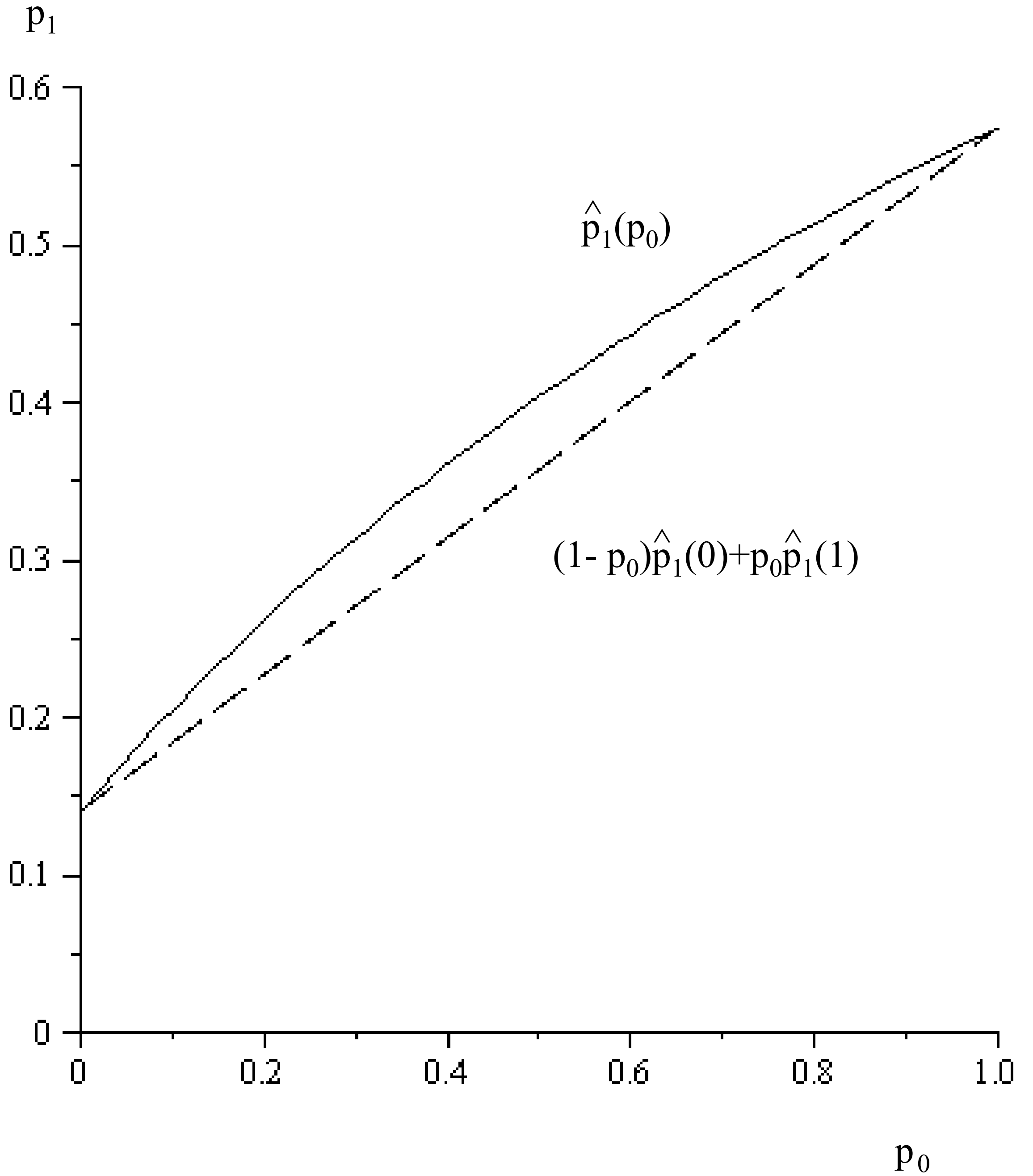


Table 2.1: Parameter Estimates

Dependent variable:	tested in past 12 months		
	(1)	(2)	(3)
b	1.686	1.673	1.690
	(0.137)	(0.132)	(0.141)
e	0.541	0.549	0.543
	(0.151)	(0.159)	(0.158)
c	0.084	0.075	0.077
	(0.031)	(0.027)	(0.035)
θ	0.862	0.850	0.824
	(0.095)	(0.098)	(0.092)
α	0.499	0.488	0.482
	(0.235)	(0.249)	(0.237)
σ	0.971	0.964	0.973
	(0.143)	(0.146)	(0.144)
δ	0.021	0.022	0.021
	(0.021)	(0.022)	(0.019)
post	0.001	0.001	0.001
	(0.034)	(0.036)	(0.031)
male	0.002	0.003	0.002
	(0.037)	(0.038)	(0.037)
province controls	no	yes	yes
post*province controls	no	no	yes
Observations	288	288	288

Note: Tested in past 12 months is a binary variable indicating took HIV test in 12 months period preceeding survey.
"Post" is an indicator variable equal to one if the observation comes from the 2005 survey round of the ZSBS.
"Male" is an indicator variable equal to one if the observation is a male demographic group.

Table 2.2: Testing Rates Under Alternative Policy Scenarios

Empirical specification:	model 1	model 2	model 3
	(1)	(2)	(3)
Policy scenario			
No ART	6.1	5.6	5.5
Status quo	8.4	7.8	7.5
Percent change compared to "No ART"	38	39	36
Expand supply	10.3	9.7	9.2
Percent change compared to "No ART"	69	73	67
Eliminate inefficiency	36.2	33.8	33.5
Percent change compared to "No ART"	493	504	509
Expand supply and eliminate inefficiency	54.0	52.3	52.1
Percent change compared to "No ART"	785	834	847

Note: Testing rates reported as percent of females age 15-49 and males age 15-59 choosing to take a HIV test.

Table 2.3: Mean Choice of Risky Behavior Under Alternative Policy Scenarios

Empirical specification:	model 1	model 2	model 3
	(1)	(2)	(3)
Policy scenario			
No ART	0.1361	0.1651	0.1554
Status quo	0.1360	0.1646	0.1550
Percent change compared to "No ART"	-0.07	-0.30	-0.26
Expand supply	0.1359	0.1641	0.1546
Percent change compared to "No ART"	-0.15	-0.61	-0.51
Eliminate inefficiency	0.1349	0.1595	0.1508
Percent change compared to "No ART"	-0.88	-3.39	-2.96
Expand supply and eliminate inefficiency	0.1286	0.1503	0.1423
Percent change compared to "No ART"	-5.51	-8.96	-8.43

Note: Mean choice of risky behavior is mean probability individual acquires HIV through choice of risky behavior, where all individuals treated as if HIV negative at time of choice.

Table 2.4: New Infections Under Alternative Policy Scenarios

Empirical specification:	model 1	model 2	model 3
	(1)	(2)	(3)
Policy scenario			
No ART	927	1,000	940
Status quo	916	989	931
Percent change compared to "No ART"	-1.19	-1.10	-0.96
Expand supply	906	979	923
Percent change compared to "No ART"	-2.27	-2.10	-1.81
Eliminate inefficiency	766	850	801
Percent change compared to "No ART"	-17.37	-15.00	-14.79
Expand supply and eliminate inefficiency	686	768	723
Percent change compared to "No ART"	-26.00	-23.20	-23.09

Note: New infections reported in thousands.

Chapter 3

Identity, Parochial Institutions, and Career Decisions: Linking the Past to the Present in the American Midwest[1]

3.1 Introduction

The textbook economic model assigns individuals to occupations on the basis of their ability, the opportunities they have to invest in human capital, and the types of jobs that are subsequently made available to them in the labor market. Outside of economics, social scientists have long argued that factors seemingly unrelated to economic incentives and constraints can also determine career choices. Weber's (1930) writings on the Protestant ethos are the classic reference in this literature. More recently, sociologists have described how working class communities can shape the aspirations their members (Gans 1962). Our objective in this paper is to document the presence of such non-economic carrer motivations in the U.S. labor market, explore the reasons why such motivations could arise, and understand why they might persist across many generations.

We start with the idea that social groups can endogenously instill values in their members that increase loyalty and reduce the propensity to exit, particularly when they are vulnerable to

[1]This is joint work with Kaivan Munshi.

acculturation (Bisin and Verdier 2000) or in competition with other groups (Sumner 1906). The negative consequences of population heterogeneity, typically measured by fractionalization along various economic and social dimensions, are well documented in the economics literature.[2] While ethnic or racial fractionalization may adversely affect the performance of *secular* institutions, we argue that such fractionalization can at the same time strengthen within-group loyalty and *parochial* institutions.

In the setting that we consider – the American Midwest – the labor market was the primary domain in which ethnic groups historically interacted. Migrant workers had to bear most of the fluctuations in labor demand in the early stages of industrialization in the United States (Hoerder 1991). This is precisely the situation in which labor market networks are most useful and there is a wealth of historical evidence documenting the role played by ethnic migrant networks in finding jobs for their members at this time. With the arrival of the railroads around 1850, the Midwest in particular witnessed a large influx of European migrants in response to the new occupational opportunities that became available. Historians describe the efforts made by migrant communities to establish "occupational beachheads" and "toe-holds" in new locations, and to subsequently work hard to maintain these coveted positions once they were established. We will show that the cost to the network from the exit of one of its members and, hence, efforts to instill a sense of loyalty to the local community would have been greatest in fractionalized labor markets with many ethnic networks competing for coveted permanent jobs.

Once such a loyalty to the local community was instilled, how and why was it maintained over many generations in the future? One institution that could have maintained local values from one generation to the next is the family. Bisin, Topa, and Verdier (2004), for example, describe how parents in minority communities in the United States socialize their children more strictly to prevent them from assimilating. In the current context, however, it is not clear why parents acting on their own would place the welfare of the community above their own children's desire to be mobile. Given the high rates of inter-county migration in the Midwest (close to 60 percent) that we document below, few families would have remained in the same local area from the time of initial settlement in any case. A more promising mechanism to explain the persistence of local values relies on a central role for the church. The migrant church was historically the focal point for economic activity in the new community, providing a domain in which information could flow and commitments could be enforced (Gjerde 1991). While all churches would have instilled denomination-specific beliefs

[2] Alesina, Baqir, and Easterly (1999), for example, document the negative relationship between racial fractionalization and public good provision in the United States.

in their congregations, migrant churches in highly fractionalized labor markets should, in addition, have put special effort into instilling a loyalty to the *local* community among their members. It is the long-term effect of these efforts, *within* particular denominations historically dominated by the migrants, that we attempt to identify in this paper.

Labor market networks and civic institutions organized around European origin-countries are largely irrelevant in the Midwest today (Gans 1979, Alba 1990). However, it would be incorrect to assume that these institutions disappeared without a trace. The church, in particular, remains at the center of the typical Midwestern community, providing important forms of support for its members such as social activity and assistance when they were sick or infirm (Elder and Conger 2000). Well-functioning churches require loyal and committed members. If members' inputs in the church are complements, as suggested by Iannaccone (1998), and if there is sufficient overlap across generations, then the first-generation members of well-functioning migrant churches, with a strong local loyalty, would have had a greater incentive to instill this value in the generation that followed. This process could have repeated itself from one generation to the next, in the church and the broader community that formed around it, linking initial settlement patterns to a loyalty to place today.

Loyalty to place or local identity (Hunter 1975, Guerson, Stueve and Fischer 1977, Hummon 1990) is well-described by the following quote from a resident of Bloomington, Indiana who moved to that city to live with his wife who was born nearby, "... my chief ambition, I discovered during our early years in Bloomington, was not to make a good career but to make a good life. And such a life as I came to understand it, meant being a husband and a father first, and an employee second; it meant belonging to a place rather than to a profession ... So as I came to recognize my children's need and my own need for a firm home place, I came to understand my community's need for citizens who stay put. Most of what I valued in Bloomington was the result of efforts by people who loved the place, either because they grew up here and chose to stay, or because they landed here and chose to remain" (Sanders 2007: 67-68).

Two aspects of local identity emerge from the preceding quote. First, local identity is associated with values that stress loyalty to a specific place. As in Akerlof and Kranton (2000), we expect that individuals born and socialized in communities with a strong local identity will incur a particularly high utility cost when they leave their communities. In the U.S. labor market, this implies that individuals endowed with a strong sense of local identity will be less likely to choose professional occupations, which are associated with significantly greater career-related moves than non-professional occupations. Second, we see that the core value associated with local identity, notably a commitment to stay, must be shared by other members of the community. Strong identity and tight-knit communities are mutually reinforcing and, as discussed above, one cannot persist without the other (Bénabou and Tirole [2006] make the same argument to explain the persistence of mutually reinforcing collective beliefs and political equilibria).

The homogenization of production that is characteristic of the Midwest only began around 1880. Thus, between 1850 and 1880 there was a window of time during which this region was rapidly settled and during which there was substantial fluctuation in labor market conditions and, hence, ethnic rivalry across local areas. These initial conditions are crucial because once community institutions and a level of individual commitment were established, these mutually reinforcing characteristics could have remained in place over many generations in the future. We consequently exploit variation in initial conditions *within* the Midwest to identify the effect of historical networks, and the identity and parochial institutions they engendered, on career decisions and geographical mobility long after the networks themselves had disappeared.

The framework we have outlined generates the prediction that counties with greater ethnic rivalry at the time of settlement should be associated with stronger local identity and better functioning churches in select denominations historically denominated by the migrants. This implies that individuals born in those counties *and* belonging to those migrant denominations should be less likely to enter a more mobile professional career today. Because occupational choice and religious affiliation are jointly determined, we cannot estimate the effect of historical fractionalization on current career decisions separately for members of migrant and non-migrant denominations. What we do instead is to test the implied hypotheses that (i) higher fractionalization should be associated with a lower supply of professionals and lower mobility on average in the county, and (ii) that higher fractionalization should be associated with greater church participation in select migrant denominations alone.

To test the first hypothesis we combine data from the U.S. census and the National Longitudinal Survey of Youth 1979 (NLSY79). A major advantage of situating the analysis in the Midwest is that individual level data on occupations and ethnicity can be obtained from the census going back to the time when this region was starting to develop. We focus on the 1860 census when rapid growth was just commencing and construct a measure of ethnic fractionalization, one minus the Herfindahl concentration index, within each broad occupational category in each county in that census year. Averaging over all occupational categories in each county we arrive at a measure of ethnic fractionalization in 1860 that is positively correlated with ethnic rivalry in the labor market at that time and, hence, with the strength of local identity in the county as it subsequently emerged.

Matching the county-level measure of ethnic fractionalization with individual data from the NLSY79, we find that individuals born in high fractionalization counties are significantly *less* likely to be employed in professional occupations and are significantly *less* likely to have migrated out of their birth county in 2000.

The values that we believe are responsible for the career choices described above could not have persisted in a local community over so many generations without institutional support. The second testable hypothesis is that high fractionalization counties should have been associated with better functioning migrant churches and related parochial institutions. If these institutional features persisted over time, supporting and being supported by local identity, then we would expect to see greater participation in religious denominations that were historically dominated by migrants, notably the Lutherans and the Catholics, today. Using data from the Census of Religious Bodies, available at roughly ten-year intervals from 1860 to 2000, we successfully verify this important prediction. The share of Lutherans and Catholics in the population is significantly larger in high fractionalization counties by 1870 and, most importantly, this gap grows steadily wider over the course of the twentieth century. In contrast, participation in other denominations is slightly *lower* in the high fractionalization counties over the entire period. We verify that these cross-denominational patterns in religious participation hold up with the NLSY data as well.

Our interpretation of the results described above is that local institutions and individual values, determined endogenously by the labor market when the Midwest was first being settled in 1860, have subsequently persisted over time and continue to shape career decisions one hundred and forty years later. Alternative explanations assume that the ability endowment in the population, investments in human capital, or access to particular types of jobs today vary with fractionalization in 1860. For example, although ethnic networks and specific European ancestry are no longer directly relevant in the Midwest economy, ethnic fractionalization in 1860 could potentially be correlated with particular features of the economy at that time, which had persistent effects and determine the *demand* for professional labor today. Similarly, the well documented negative correlation between fractionalization and public good provision could have given rise to poorly functioning schools in high fractionalization counties.

We later show that if identity is salient and the demand for professional labor is uncorrelated with historical fractionalization, then high fractionalization counties will supply too few professionals and too many non-professionals as long as there is *ex ante* uncertainty in the demand for different types of jobs. Professional labor must move into these counties and non-professional labor must move out of them *ex post* in competitive equilibrium. It follows that individuals *residing* in high fractionalization counties will be just as likely to hold professional jobs as individuals residing in low fractionalization counties, once the labor market clears. It is only individuals *born* in high fractionalization counties who should be less likely to be professionals. The alternative hypothesis, based on differences in the demand for professional labor across counties, predicts that individuals *born and residing* in high fractionalization counties should be less likely to be professionals. We verify that adult workers residing in high fractionalization counties are indeed as likely to be professionals as those residing in low fractionalization counties, ruling out the most obvious alternative explanation for our results.

Our results could also be obtained if the incoming migrants in the high fractionalization counties had lower ability, and this ability differential persisted across generations. Human capital transmission is less of a concern with our analysis since we are going back at least five generations; even if the ability distribution of the incoming migrants varied systematically with fractionalization in 1860, this heterogeneity would have long since disappeared, given the high rates of inter-county migration in the Midwest.[3] A more relevant concern is that fractionalization at the time of initial settlement could be correlated with social heterogeneity many decades later in the twentieth century when the school system was being established in the Midwest. Goldin and Katz (1999) document a negative relationship between social heterogeneity and early investments in higher education. At the same time, they show that greater church participation was associated with larger investments in human capital, and so the net effect of 1860 fractionalization on education expenditures is ambiguous. The NLSY provides information on AFQT scores and educational attainment. Information on local public good provision, including education expenditures, can also be obtained at the county level in 1990. Reassuringly, we find no relationship between historical fractionalization and current investments in schooling or educational attainment. The only exception is that individuals born in high fractionalization counties are significantly more likely to choose college majors that lead to less mobile careers.

This paper makes a contribution to the rapidly expanding literature on culture and religion in economics as well as a more established literature on institutions. While the literature on religion has tended to focus on beliefs or ideologies and their effect on individual behavior and aggregate outcomes (Barro and McCleary 2003, Bénabou and Tirole 2006), we model the church as a collective

[3]The empirical strategy that we employ is related to the strategy adopted by Fernandez and Fogli (2007) in a very interesting recent paper. Fernandez and Fogli establish that female labor force participation in the origin country affects labor force participation and fertility rates among second-generation female migrants in the United States. They rely on the assumption that cultural traits will be transmitted across space and so cultural heterogeneity will continue to manifest itself in an economic environment – the U.S. labor market – that does not distinguish between social groups. The obvious alternative explanation, which they take care to rule out, is that human capital rather than culture is being transmitted across generations. With our strategy, the chief concern is that there is something about the *place*, other than identity and its complementary institutions, that has persisted over time, which the preceding test on the county of residence, rather than the county of birth, helps rule out.

institution that was fortuitously well-positioned to reduce coordination and commitment problems historically and which continues to play that role today. Instead of focussing on differences in occupational choice and mobility across denominations or between church participants and non-participants, we study how local economic conditions shaped the orientation of individual churches *within* particular migrant denominations. Our analysis thus provides empirical support for Greif's (2006) view that institutions cannot be simply described by a set of formal rules, but are in fact more complex arrangements whose formation and evolution are determined by history and context. Finally, this paper complements recent research that investigate the relationship between culture and growth (Barro and McCleary 2003, Tabellini 2005, Fernandez and Fogli 2007). Although local identity cannot be observed directly, our analysis goes beyond previous studies to theoretically and empirically describe the formation and the persistence of this particular cultural trait. This trait is seen to have a large effect on important economic decisions in a region as homogeneous as the American Midwest, long after the forces that gave rise to it had ceased to be salient, providing evidence for a role for culture that is not subject to the usual critiques of cross-country or cross-regional analyses.

3.2 The Institutional Setting

3.2.1 The Settling of the Midwest

The Midwest first began to be settled in the early nineteenth century with the expansion of the national canal system. The Erie Canal linking the Hudson to Lake Erie was completed in 1825 and numerous inter-regional and intra-regional canals were built over the next two decades (Fishlow 2000). However, it was only with the arrival of the railroad that the Midwest took off on a steeper growth trajectory. Before 1850 the Midwest had less than one thousand miles of track, but almost ten thousand were added by 1857 (Meyer 1989).

Improved rail transportation stimulated industrialization and the Midwest's share of national manufacturing increased rapidly between 1860 and 1920, with almost half of this increase occurring in the 1860's (Meyer 1989). This increase in economic activity led, in turn, to an increase in the demand for labor. In 1810, approximately 6 percent of the labor force (outside the southern states) resided in the Midwest. By 1860, this share had increased to 41 percent, with a further increase to 51 percent by 1880, after which regional growth converged to the national average (Margo 1999).

In this paper, the Midwest is comprised of the states of Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin (Missouri, the only pre-Civil War slave state in the Midwest, is excluded from the analysis). Using county-level census data we see that the number of incorporated counties increases sharply from 1850 to 1860 and then flattens out by 1880 in Figure 1. Information on railroads, obtained from the Historical Map Archive at the University of Alabama, indicates that the number of these counties with a railroad also increases steeply over the 1850-1870 period, growing thereafter at a slower rate.[4] The rapid expansion of the railroad system and the economic activity that accompanied it led to a steep increase in the population of the Midwest as well as an influx of foreign migrants. Using county-level census statistics, the total population in our seven Midwestern states grew from less than 5 million in 1850 to 20 million in 1900. The number of foreign-born migrants nearly tripled between 1850 and 1860, reaching close to 20 percent of the population.

Where did these migrants come from? Individual-level data, including characteristics such as age, sex, occupation, and country of birth, are publicly available from the Population Census each decade from 1850 to 1930. We use the 1-in-100 sample from the 1860-1900 IPUMS to study changes in the migrant population in these critical early decades in Table 1. The English (13 percent), the Irish (25 percent), and the Germans (32 percent), dominated the migrant population in the Midwest in 1860, just after the first wave of migration described above, with no other ethnic group accounting for more than a 3 percent share of the migrants in that year. Subsequently, the English and the Irish were displaced by the Germans and the Scandinavians over the 1860-1900 period. Notice that the Italians, Poles, and Slavs continue to be insignificant in 1900, although they would display a substantial presence in Midwestern cities such as Cleveland, Chicago, and Pittsburgh by the first quarter of the twentieth century.

What jobs did the migrants occupy? Table 2 reports the occupational distribution of the migrants from the IPUMS sample in 1860, 1880, and 1900. Although agriculture was the dominant sector in this period, the share of farm employment declines from 62 percent in 1860 to 48 percent in 1900, with manufacturing operatives and laborers accounting for much of the increase in non-farm employment. These trends are consistent with the growth of the manufacturing sector described above and they are similar for the foreign-born migrants and the native workers (not reported).

The apparent similarity in the occupational distribution for migrants and natives masks differences in the type of jobs they had access to within each occupational category. Labor markets in the nineteenth century could be divided into three segments: a stable segment with permanent employment, an unstable segment with periodic short-term unemployment, and a marginal but highly

[4] Railroad maps were used to construct a county-level binary variable indicating whether any part of a railroad ran through the county in a given year. Railroad maps were unavailable in some census years in which case we used maps that were closest in vintage to those census years (the discrepancy never exceeded three years).

flexible segment characterized by spells of long-term and short-term employment (Gordon, Edwards, Reich 1982). Migrants being newcomers to the U.S. market typically ended up in the unstable and marginal segments, where the uncertain labor demand naturally provided an impetus for the formation of ethnic networks that helped their members find jobs (Hoerder 1991). Based on the discussion above and Figure 1, we would expect these networks to have been particularly important during the rapid expansion phase around 1860, with the influx of migrants and the opening up of new labor markets in both agriculture and manufacturing. The initial conditions in the empirical analysis, measuring ethnic rivalry in the labor market at the time of settlement, will consequently be measured using data at this point in time. Although not reported, regressions that treat 1850 as the starting point provide qualitatively similar results.

3.2.2 Ethnic Labor Networks in the Midwest

Accounts by contemporary observers and a rich social history literature indicate that friends and kin from the origin community in Europe played an important role in securing jobs for migrants in the Midwest in the nineteenth century and the first quarter of the twentieth century. Early historical studies used census data, which provides fairly detailed occupational and ethnic information, to identify ethnic clusters in particular locations and occupations. Gordon, Edwards, and Reich (1982) note that although foreign-born workers comprised just over 20 percent of the labor force in 1870, they accounted for 43 percent of the iron and steel operatives, 43 percent of the woolen mill workers, and 63 percent of the miners. Nearly a quarter of railroad workers were Irish, a third of the miners were British, and about half the workers in the baking and confectionary business were German. While such clustering suggests that underlying ethnic networks were channeling their members into particular occupations, it could simply reflect the fact that migrants arrived with specific skills. Hutchinson's (1956) analysis of 1950 census data, however, indicates that clustering continued even among the migrants' children, with the concentration of particular ethnic groups in some industries actually increasing from the first to the second generation.

Although census data are a useful source of information, they do not provide details of the migration process and its connection to ethnic networks in the United States. Over the past four decades, however, social historians have linked parish registers and county data in specific European sending communities to census and church records in the United States to construct the entire chain of migration from those communities as it unfolded over time. This remarkable research effort has documented the formation of new settlements in the Midwest by pioneering migrants, the subsequent channelling of migrants from the origin community in Europe to these settlements, as well as the movement of groups from the original settlement to new satellite colonies elsewhere in the United States. As with the census data, this research identifies occupational and geographic clustering, but at a dissagregate level. Over 45 percent of the Swedish emigrants from the parish of Rätvik eventually settled in Isanti County, Minnesota (Ostergren 1976). Two-thirds of the emigrants from Balestrand located in Norway Grove, Wisconsin in the first decade of migration from that Norwegian community (Gjerde 1985). And one-third to one-half of the German emigrants from Westerkappeln settled in Duden County, Missouri (Kamphoefner 1987). Although less detailed origin-country information is available for southern European migrants, similar ethnic clustering in particular neighborhoods of Midwestern cities has been documented for Polish, Italian, and Slovak immigrants from specific sending regions (Alexander 1991, Bodnar, Simon, and Weber 1982).

A possibly stronger indicator of the importance of migrant networks is the maintenance of ethnic ties over successive moves *within* the United States. Italians moved from Southern Illinois to the "Italian Hill" in St. Louis when coal mining operations were reduced in the 1920's and Slavs moved from mines in Western Pennsylvania to Detroit's growing automobile industry in the same decade (Bodnar 1985). Norwegians from Balestrand initially settled in Norway Grove, Wisconsin, but over time they established six satellite settlements in Wisconsin, Minnesota, Iowa, and Illinois (Gjerde 1985). A similar pattern has been documented for Norwegians immigrants from Fortun, who initially settled in Vernon and Crawford Counties, Missouri, but later established satellite communities throughout the Midwest (Gjerde 1997).

While the preceding descriptions of ethnic clustering are informative, ethnic concentration within specific departments or firms in a local industry provides possibly the strongest evidence that labor networks were active. Nearly all three thousand employees of the Peninsular Gas Company in Detroit in 1900 were Polish, and Croatians held only three jobs in Indiana's oil refineries: stillman helper, fireman, and still cleaner (Bodnar 1985). Italians in Pittsburgh's steel industry dominated the carpentry, repair, and rail shops. And, relying on friends and relatives, Poles established occupational niches at the Jones and Laughlin and Oliver Mills on Pittsburgh's Southside, Heppenstalls and the Pennsylvania Railroad in Lawrenceville, and at the Armstrong Cork Company and the H.J. Heinz Plant. As John K. a Polish immigrant put it, "The only way you got a job [was] through somebody at work who got you in" (Bodnar, Simon, and Weber 1982: 56).

What kept ethnic networks in place so far from their origin locations? It has been argued that "[migrants] from varying regions [in the origin country] formed a community based on common nationality and religion centered on the central cultural institution – the church" (Gjerde 1991:

176). The building of a church was one of the first organized actions in the migrant community once it arrived in an area (Barton 1975, Bodnar, Simon, and Weber 1982). Churches provided both economic and social support to their members. Information about jobs and potential land transactions flowed within the congregation and the church also served as a public arena in which members who had reneged on their obligations could be sanctioned.

Given the variety of economic opportunities in the United States, individuals and small groups drawn from the same parish in Europe often had an incentive to move and seek employment elsewhere. The stability of the *local* community in the United States, based on a common national origin rather than narrower social affiliations, was thus essential for the viability of the labor market network. One strategy to maintain stability would have been to instill a sense of loyalty to this community. The discussion that follows will describe how efforts to engender a sense of local identity by the church and the institutions that formed around it might have varied across Midwestern counties when the first wave of migrants arrived in the region.

3.2.3 Ethnic Fractionalization Across the Midwest

"You take in the erection department – it was mostly all Slavs ... Not Slovaks, it was Polish ... We didn't have Lithuanians there and the Russians were not involved there ... Now if a Russian got his job in a shear department ... he's looking for a buddy, a Russian buddy. He's not going to look for a Croatian buddy. And if he see the boss looking for a man he says, 'Look, I have a good man,' and he's picking out his friends." (Polish immigrant in Pittsburgh, quoted in Bodnar, Simon, and Weber 1982:62).

Numerous historians have described the efforts made by ethnic groups to establish a "toe-hold" (Thistlethwaite 1991) or a "beachhead" (Bodnar, Simon, and Weber 1982) in particular industries or establishments when they first settled in an area. The preceding quote suggests, in addition, that once a network had established a presence in the labor market, it was essential to *maintain* that presence. The discussion that follows will describe the labor market conditions under which migrant communities would have had the greatest incentive to restrict exit and, hence, instill a sense of local identity among their members. An explanation for the persistence of this identity over multiple generations, based on the complementary role of the church and other related institutions, is postponed to the next section.

Consider a market with N migrant workers drawn from M communities. The number of workers in each community is exogenously determined and we do not require the labor market to be in equilibrium. The workers are competing for a fixed number of coveted permanent jobs, which provide a total surplus R. Tullock's (1980) canonical model of rent seeking can be conveniently adapted to this setting to describe the share of total surplus captured by community i

$$S_i = \frac{n_i^\alpha}{n_i^\alpha + \sum_{j=1}^{M-1} n_j^\alpha}$$

where n_i is the number of workers belonging to community i, n_j is the corresponding number of workers from each community $j \neq i$, and $\alpha > 0$.[5] The cost to community i from the exit of of its members at the margin can be described by the expression

$$R\frac{dS_i}{dn_i} = R\frac{\alpha n_i^{\alpha-1} \sum_{j=1}^{M-1} n_j^\alpha}{\left(n_i^\alpha + \sum_{j=1}^{M-1} n_j^\alpha\right)^2}.$$

Assuming that all networks are of equal size, $n_i = n_j = N/M$,

$$R\frac{dS_i}{dn_i} = \frac{R\alpha}{N}\left(1 - \frac{1}{M}\right),$$

which is increasing in M. Holding N and R constant, an increase in the number of communities is associated with a greater marginal cost of exit regardless of whether the returns to network size are increasing ($\alpha > 1$) or decreasing ($\alpha < 1$) at the margin. Self-interested individuals do not internalize the cost imposed on the rest of the network by their exit. It follows that communities in labor markets with greater ethnic competition should have invested more in instilling a sense of local identity as a way of aligning individual incentives more closely with the community's welfare.

Because networks vary in size in practice, we use a standard measure of fractionalization, defined as one minus the Herfindahl index of ethnic concentration, to measure competition.[6] Following the discussion above, we assume that networks were most active in the migrant population and that they were organized around the country of origin. These ethnic networks would have competed directly with each other within (broad) occupational categories. The IPUMS provides the country of birth and the occupation (where relevant) for each sampled individual. Using these data, we compute ethnic fractionalization within the occupational categories specified in Table 2 in each Midwest county in 1860. The weighted average of the occupation-specific statistics, where the weight is

[5] Tullock's specification is identical to the equation above except that the number of workers is replaced by the investment in rent seeking.

[6] The Herfindahl index of ethnic concentration is computed as the sum of the squared share of each ethnicity in the occupational category. It is easy to verify that for the special case with networks of equal size, the fractionalization measure is equal to $1 - N^2/M$, which is increasing in the number of communities M, for a given number of migrant workers N, as above.

measured by the share of migrants in the occupation, then provides us with an overall measure of ethnic fractionalization in the county. For example, suppose that there are two occupations and two ethnicities in the county, with complete occupational segregation along ethnic lines. Our measure of fractionalization will be zero in this case, correctly reflecting the absence of ethnic competition in this labor market. In contrast, if the two ethnicities are of equal size and evenly distributed across the two occupations, our measure of fractionalization will increase to 0.5. Figure 2 plots the fractionalization measure, which has a mean of 0.5 and a standard deviation of 0.2, across the seven Midwest states. Counties that were not incorporated and those without foreign-born migrants in 1860 are unshaded in the Figure. Notice that there is substantial variation in the fractionalization measure, which is useful for the statistical tests that follow.

While we focus on the effect of ethnic fractionalization in strengthening within-group solidarity, fractionalization could at the same time undermine the functioning of secular institutions. An important advantage of using the Midwest for the analysis is that this region was settled recently enough that accurate measures of ethnic rivalry can be constructed with census data. At the same time, public services were provided at a rudimentary level in 1860 and the first investments in secondary schooling, for example, occurred more than 50 years later (Goldin and Katz 1999). While initial conditions may have determined subsequent labor in-flows and social heterogeneity, it is possible that the relevant determinants of public good provision many decades later in the twentieth century were largely unrelated to fractionalization in 1860. We will later verify that 1860 fractionalization is indeed uncorrelated with local public good provision, including education expenditures, in 1990. The major long-term effect of 1860 fractionalization seems to have been to shape the orientation of a particular set of parochial institutions that started to be established around that time.

3.2.4 Historical Fractionalization and Current Economic Conditions

To test the hypothesis that local identity and its complementary parochial institutions shape career choices, historical fractionalization must be uncorrelated with the demand for professionals, intrinsic ability in the population, and local public good provision today. We show in this section that ethnic fractionalization in 1860 is uncorrelated with measures of economic activity, racial and ethnic fractionalization, and expenditures on local public goods in 1990 once we control for a few important characteristics of the 1860 economy. More stringent tests ruling out a link between 1860 fractionalization and current economic conditions are discussed in Section 5.

Labor market networks organized around the European-origin countries in Table 2 are no longer active in the American Midwest. Nevertheless, fractionalization in 1860 could have been correlated with particular features of the economy at the time of initial settlement that had persistent effects. To explore this possibility, we proceed to understand what determined fractionalization in the first place. In a rapidly expanding Midwest economy, some of the variation in fractionalization across counties was no doubt a consequence of accidental initial settlement by ethnic groups in particular locations, which fueled the arrival of more migrants as networks crystallized. At the same time, fractionalization would have been determined by the demand for labor, with more ethnic groups attracted to rapidly growing areas. We have already discussed the importance of transportation links in the development of the Midwest and Table 3 consequently investigates the effect of railroads and distance to canals and a Great Lakes harbor on fractionalization in 1860.[7] Counties with a railroad running through them and counties that are closer to a canal or harbor have significantly higher fractionalization in Table 3, Column 1. Counties close to a harbor had a greater proportion of the workforce engaged in manufacturing and a smaller proportion in agriculture in 1860 in Table 3, Columns 2-3. Improved transportation, more generally, is associated with a larger population in 1860 in Column 4.

Notice that the pattern of coefficients in Table 3, Column 1 with fractionalization as the dependent variable matches perfectly with the corresponding pattern in Column 4 with county population as the dependent variable. Counties with superior transportation infrastructure, which were more populated and presumably growing more rapidly, were more fractionalized in 1860. The population of the county in 1860 could have determined subsequent agglomeration in economic activity, with long-term implications for the growth of the local economy. Not surprisingly, we see in Table 4, Columns 1-3 that while 1860 fractionalization has an insignificant effect on 1990 agriculture share and manufacturing share, it is significantly and positively associated with current county population as well as population density (not reported). More urban counties will be more racially diverse, and it follows that 1860 fractionalization is significantly and positively correlated with racial fractionalization in 1990 in Column 5.[8] In contrast, 1860 fractionalization is completely uncorrelated with

[7]Data on the distance to the nearest canal (or navigable river) and the nearest Great Lakes harbor is obtained from Jordan Rappaport's website at the Kansas City Federal Reserve Bank. The distance is computed in each case from the county centroid.

[8]The manufacturing share in 1990 is defined as the share of the civilian labor force employed in manufacturing in that year. The agriculture share in 1990 is computed using the farm population and the total population in the county in that year. All these statistics, as well as the area of each county used to compute the population density, are obtained from the 1994 County Data Book, compiled by the U.S. Bureau of the Census. Racial fractionalization is computed from the 1990 IPUMS as one minus the Herfindahl index of racial concentration, using the same five racial groups as in Alesina, Baqir, and Hoxby (2004). Ethnic fractionalization is computed from the 1990 IPUMS as one minus the Herfindahl index of racial concentration, using the same 16 white ethnic groups as in Alesina, Baqir, and Hoxby.

(white) ethnic fractionalization in Column 4, consistent with the idea that few individuals living in the Midwest today could trace an unbroken line of descent to European ancestors arriving in the *same* county at the time of initial settlement.

We expect the fractionalization effect to disappear once important features of the local economy that would directly determine economic activity today, such as manufacturing share, agriculture share, and particularly population in 1860, are accounted for. The fractionalization coefficient is indeed small and insignificant in Columns 7-11 of Table 4, once the county controls are included. In contrast, agricultural counties in 1860 continue to be agricultural in 1990 and more populated counties in 1860 remain larger in 1990. Notice, however, that manufacturing in 1860 does not have a persistent effect. Although the factory system began to replace artisan shops by 1820, production continued to be largely organized in workshops managed by labor contractors who hired their own employees until 1870 (Gordon, Edwards, and Reich 1982). The heavy manufacturing that characterized the Midwest economy in the twentieth century, with its emphasis on the iron and steel industry, only came at the turn of the century (Meyer 1989). Recently it has been argued that the surge of foreign immigration in the second half of the nineteenth century provided the impetus for the factory system and the subsequent industrialization of the Midwest (Kim 2007). Whatever the explanation, it is clear that the pattern of manufacturing around 1860, spread throughout the Midwest in small towns, had little connection with the heavy manufacturing, concentrated in large cities, that followed in the twentieth century. This is presumably why 1860 characteristics have such little power in predicting the share of manufacturing in the county in 1990.

For our purpose, what is important is that once we control for a few features of the nineteenth century economy, fractionalization in 1860 has no effect on characteristics of the economy today, such as the share of manufacturing and urbanization (measured by total population or by population density), that are associated with the demand for professional labor. All of the results that follow will consequently condition for these important features of the 1860 economy. We will verify that all the results hold up without the 1860 controls as well. Finally, notice that more ethnically fractionalized counties in 1860 have significantly *lower* religious fractionalization in 1990, with and without the controls in Column 6 and Column 12.[9] This intriguing result, which stands conspicuously apart from the other regressions reported in Table 4 will be clarified in the discussion on the church and the persistence of identity that follows in Section 3.

We complete this section by describing the relationship between local public good provision in

[9]Religious fractionalization is computed from the 1990 Census of Religious Bodies as one minus the Herfindahl index of religious concentration, using the same 18 religious denominations as in Alesina, Baqir, and Hoxby (2004).

1990 and ethnic fractionalization in 1860 in Table 5. Local government expenditures per capita in 1990 are regressed on ethnic fractionalization, manufacturing share, agriculture share, and county population in 1860 in Column 1. Retaining the same set of regressors, the dependent variable is the share of expenditure allocated to education, health, police, roads, and welfare in Columns 2-6.[10] The expenditure shares for these five public goods add up to 0.79, with the education share being as large as 0.6. The other shares range from 0.03 to 0.07. Given that the standard deviation of the fractionalization variable is 0.2, it is evident from the point estimates that 1860 fractionalization has a negligible effect on public good provision today. Using the same shares, Alesina, Baqir, and Easterly (1999) document a negative and significant relationship between contemporaneous racial fractionalization and the allocation of resources to certain public goods, particularly education, roads, and welfare. We have already shown that 1860 ethnic fractionalization is uncorrelated with 1990 racial fractionalization in Table 4, and not surprisingly we see in Table 5 that historical ethnic fractionalization has no effect on current local public good provision, including education expenditures.

3.3 The Church and the Persistence of Identity

In a recent paper, Akerlof and Kranton (2005) describe how firms and other economic organizations can instill a sense of identity or loyalty among their workers to solve agency problems. In our setting, the natural organization to instill loyalty in the community would have been the local church. The church was among the first institutions to be established when immigrants arrived in an area (Hoerder 1991, Barton 1975, Bodnar, Simon, and Weber 1982). The church congregation provided many forms of mutual assistance including credit, insurance, job referrals, business information, and social support (Gjerde 1985, 1997, Alexander 1991). Indeed, it has been argued that immigrants participated in church communities to benefit from these economic and social services, instead of being drawn to the church by a particular belief or ideology (Bodnar 1985).

Despite these material attractions, exit from the local church and labor market was always a threat in nineteenth century America, especially in areas that had been recently settled. Migration to the United States, and subsequent internal migration, was typically organized around small groups drawn from the same parish or province in Europe. These groups were too small to maintain a viable church congregation and labor network when they first arrived in an area and so migrant churches in the Midwest typically brought together many regional groups with the same national origin.

[10] These data are obtained from the Annual Survey of Governments, 1990. Alesina, Baqir, and Easterly (1999) use data drawn from the same source.

Inter-regional conflict was common, and there was always the possibility that groups within the church would move again to take advantage of new opportunities that became available elsewhere (Gjerde 1997). Under these circumstances, the migrant church had two distinct objectives: (i) to instill denomination-specific beliefs in the congregation, and (ii) to instill a loyalty to the local community that transcended narrower social affiliations. We saw earlier that the negative externality associated with individual exit was greatest in ethnically fractionalized labor markets. This implies that migrant churches should have made a special effort to instill a sense of local identity in such markets. In contrast, we have no prediction for the effect of fractionalization on religious beliefs or the orientation of churches belonging to non-migrant denominations.

If local identity played a role in reducing exit in the nineteenth century, how and why did it persist long after the ethnic labor networks it supported ceased to be salient? Our explanation for this persistence is based on the observation that churches continue to provide important forms of social support to their members. Church activities include Sunday school service, youth groups, pot-lucks, informal home parties, and food, visits, and other forms of support when members of the congregation are ailing or infirm. The church also lies at the center of a cluster of inter-linked civic institutions, including the school and various voluntary organizations. Life in a Midwest community revolves around these institutions, which bring families, friends, and multiple generations together on a regular basis (Elder and Conger 2000). We focus on the church in the analysis that follows because of its central position and because church participation is available at the county level, by denomination, from 1860 till 2000.[11]

There are complementarities associated with church inputs; if the rest of the congregation commits time and effort to the church, then the returns to the individual's inputs would increase as well (Iannaccone 1998). The presence of these complementarities introduces two problems. First, self-interested individuals will devote a level of inputs that is sub-optimal because they do not internalize the benefits that the rest of the congregation derives from their actions. Second, a coordination problem can arise since individuals will only make career choices that are associated with a high level of commitment to the church if they expect other members of their generation, who are simultaneously making these choices, to do likewise. A strong sense of local identity reduces each of these problems,

[11]The number of civic and social associations in the county and the corresponding number of religious organizations can be obtained from the County Business Patterns, 1990. The number of not-for-profit organizations in the county can be obtained from National Center for Charitable Statistics Core Files, 1990. Controlling for manufacturing share, agriculture share, and county population in 1860, ethnic fractionalization in that year has no effect on any of these variables. Note, however, that these statistics are based on the *number* of organizations, whereas what we require are measures of *participation*. For example, we will see below that 1860 fractionalization has a significant and positive effect on religious participation in 1990 despite the fact that it is uncorrelated with the number of religious organizations.

serving as a commitment device and increasing the level of church inputs in equilibrium. Given that church inputs are complements and that generations overlap, it is easy to verify that the first generation in a migrant church with a committed congregation would have been more willing to bear the cost of instilling a strong sense of local identity in the generation that followed. Local identity and church inputs could have subsequently reinforced each other in this way over many overlapping generations.[12]

Local identity and church inputs cannot be observed directly. However, both the Census of Religious Bodies (CRB) and the NLSY provide information on the number of church participants. The preceding discussion indicates that migrant churches in high fractionalization counties should have been relatively well-functioning, with a high level of church inputs, at the time of initial settlement and over many future generations. These churches would, in turn, have attracted more members *and* supported higher rates of attendance in the congregation. Restricting attention to denominations historically dominated by the migrants, the testable prediction is that high fractionalization counties should be associated with greater church participation overall. Moreover, the effect of fractionalization on church participation must be sustained over time, reflecting the underlying role of the church in maintaining local identity. This prediction distinguishes our model, with identity as a commitment device, from the model of religious cults proposed by Iannaccone (1992) in which self-sacrifice and strict norms of behavior are used to screen out free-riders and ensure higher levels of participation among those that remain. While this alternative mechanism may induce high levels of participation at the intensive margin, it results in small congregations and is typically short-lived.

Based on the country of origin of the incoming migrants, reported in Table 1, most of the migrant churches would have been Lutheran or Catholic. Regressing the population-share of different denominations, computed with data from the CRB in 1860, on the share of migrants in that year, counties with a greater share of migrants are indeed disproportionately Lutheran and Catholic.[13] A strong test of our framework is that the predictions derived above should apply to those migrant

[12]The persistence in local identity that we describe is related to a paper by Bénabou and Tirole (2006) linking investments in human capital at the household level with the political equilibrium at the macro level. In their model, children choose a level of effort (schooling) based on their belief about the returns to this effort. If a sufficiently large number of households invest in effort, they will form a pivotal voting block and set a low tax rate, generating, in turn, a high return to effort that reinforces the initial beliefs. Thus, two political equilibria can arise; a low-effort equilibrium with substantial redistribution (high taxes) and a high-effort equilibrium with little redistribution (low taxes). Bénabou and Tirole fix the equilibrium by assuming that children have imperfect willpower and so will underinvest in effort if left to themselves. Their parents, who provide them with information about the returns to effort, will consequently systematically inflate these returns. This "ideological" position leads parents in the high-effort equilibrium to (optimally) ignore negative signals about the returns to effort, allowing particular equilibria and the collective beliefs that support them to persist over many generations. In this paper we use local identity rather than ideology to fix the level of church inputs and sustain a social equilibrium over many generations.

[13]Apart from the Germans and the Irish, the English were also an important migrant group in 1860. The English would have been disproportionately Anglican (Episcopalian). This denomination accounts for just 3 percent of church participants in 1860 and is never a significant force in the Midwest. Inclusion of the Episcopalians, not surprisingly, has little effect on the religious participation results reported below.

denominations only. In particular, higher fractionalization counties should be associated with a greater share of participating Lutherans and Catholics in the population around the time of initial settlement and this effect of fractionalization should be sustained over time. In contrast, the model has no prediction for the relationship between fractionalization and participation in other denominations, providing us with a useful falsification test.

The relationship between ethnic fractionalization and religious participation can be tested with data from the CRB, which has been conducted at roughly ten-year intervals from 1860 to 2000. This census was conducted as part of the population census from 1860 to 1890, with census enumerators collecting information from individual churches in each county. Subsequently, the U.S. Bureau of the Census conducted the CRB separately from the population census in 1906, 1916, 1926, and 1936. Starting from 1952, the National Council of Churches of Christ undertook the responsibility of conducting the CRB, with subsequent census rounds in 1972, 1980, 1990 and 2000.

The 1860-1890 census rounds collected information on the number of church seats by denomination in each county. From 1890 onwards, information was collected on the number of members directly, and from 1972 onwards the number of adherents was collected as well. Despite these changes in the management of the CRB and the measure of religious participation, we uncover clear changes in the mix of denominations as well as the effect of 1860 fractionalization on religious participation over time.

Table 6 reports changes in the mix of denominations and overall participation rates in our Midwestern counties over the 1860-2000 period. To take account of the fact that the measure of participation was changing over time, we report statistics in the first and the last census-year that each measure was used. Thus, participation is measured by the number of church seats from 1860 to 1890, by the number of members from 1890 to 1952, and by the number of adherents from 1972 to 2000.[14] The participation rate is then computed as the number of participants divided by the contemporaneous population in the county. Five denominations – Baptist, Catholic, Lutheran, Methodist, and Presbyterian – account for roughly 80 percent of church participants over the 1860-2000 period. Among these denominations, the Lutherans grow rapidly in popularity over the 1860-1890 period and the 1890-1952 period, remaining stable thereafter. In contrast, the Methodists and the Presbyterians decline steadily over time. There is no clear trend among the Baptists and the Catholics.

The inability of the Baptist church to increase its share of church participants contrasts with the

[14] Although the number of members was also collected in the 1972-2000 census rounds, this statistic is not available for Catholics, a major denomination in our Midwestern counties, in these rounds.

surge in popularity of this denomination elsewhere in the United States, as documented by Finke and Stark (1992). The Midwest stands apart from the rest of the country in that the "traditional" denominations, particularly the Catholics and the Lutherans, continue to dominate, accounting for over half of all church participants by 2000. Even without the Baptist surge, religious participation increased steadily over time in the Midwest, rising over the 1860-1890 and 1890-1952 periods. Based on the statistics in the most recent 1972-2000 period, 55-60 percent of the population in our Midwestern counties are church adherents.[15]

We next proceed to estimate the relationship between ethnic fractionalization in 1860 and religious participation in the county in each round of the CRB. The religious participation regression includes the same set of county-level controls as the regressions in Table 4, Columns 7-12 and Table 5. Because the regression is estimated over many census years, we simply report the 1860 fractionalization coefficient, together with the 95 percent confidence band, in each census year in Figure 3. This coefficient is less precisely estimated in the early census years, but grows steadily larger, while remaining statistically significant, all the way through to 2000.[16]

A stronger test of our framework is that its predictions should only apply to participation in migrant denominations. We consequently proceed to estimate two separate regressions in each census year; the first regression has the share of Lutherans and Catholics in the population as the dependent variable and the second regression has the share of all other denominations as the dependent variable. The 1860 fractionalization coefficient, with the corresponding 95 percent confidence band, is reported for each regression in each census year in Figure 4. We now see that the share of Lutherans and Catholics in the population is significantly greater in high fractionalization counties by 1870 and that the fractionalization effect gets steadily larger over time. Although we do not report results separately by denomination, this pattern is obtained for both the Catholics and the Lutherans. A one standard deviation decline in 1860 fractionalization would increase the population share of Lutherans and Catholics in the county by four percentage points (22 percent) in 2000. In contrast, the fractionalization coefficient is much smaller and *negative*, but remains significant and stable over

[15]Drawing on a number of different sources, Iannaccone (1998) reports church participation rates of 70 percent or even higher in the United States. Our statistic may be lower because it is based on a census of churches rather than the population, or because we divide by the entire population in the county when computing participation rates.

[16]The 1860 fractionalization coefficient with standard error in parentheses is -0.66(0.39). This outlying coefficient is omitted from Figure 3 and Figure 4 that follows to clarify changes in the fractionalization coefficient over time. As noted, church seats and the number of members are both available in the 1890 CRB. We use the first statistic to measure church participation in Figure 3 and Figure 4 because it is more in line with trends in the fractionalization coefficient over time. Although not reported, the coefficient on 1860 agriculture share is also positive and significant in the religious participation regressions and grows larger over time. Unlike ethnic fractionalization, which soon ceased to be directly relevant, recall from Table 4 that agricultural counties in 1860 remained disproportionately agricultural in 1990. It is well known that farming communities tend to be more religious and community-oriented and so the persistent effect of the 1860 agriculture share may simply reflect the persistence of agricultural activity in particular areas over time.

time in the companion regression. These results, taken together, provide a simple explanation for the negative correlation between ethnic fractionalization in 1860 and religious fractionalization in 1990 that we reported in Table 4.[17]

Our explanation for the pattern in Figure 4 is that historical economic conditions gave rise to particularly well-functioning Lutheran and Catholic churches in the high fractionalization counties. An alternative explanation for this pattern is that there are increasing returns to scale in church performance; Lutheran and Catholic churches were large to begin with in the high fractionalization counties, simply because there were more migrants from those denominations in the area, and subsequently grew larger over time. To rule out this explanation we experimented with augmented specifications that included (i) the share of Lutherans and Catholics, and (ii) a full set of 32 ethnic shares (which sum up to the share of migrants in the population), in 1860 by county, as regressors. The initial share of Lutherans and Catholics, as well as some of the ethnic shares do have persistent effects. However, the fractionalization coefficient remains stable over the course of the twentieth century with both the augmented specifications.[18]

Initial conditions in 1860 would have determined subsequent in-migration by diverse ethnic groups and the formation of new Catholic and Lutheran churches in the decades that followed. The analysis in this paper is restricted to estimating the (overall) effect of initial conditions in 1860 on later outcomes and Figure 4 thus incorporates additions to the stock of churches and to the migrant population over time.[19] The increase in the share of participating Lutherans and Catholics that we observe could in that case simply reflect growth at this extensive margin rather than a widening of the gap in church attendance (the intensive margin) between low and high fractionalization counties. The mechanism that we have proposed to explain the persistence of local identity requires that differences in church participation at the *intensive* margin be maintained across counties. The CRB data do not allow us to separate church participation along the intensive and the extensive margin. However, the NLSY collects information on both religious affiliation (inherited at birth) as well as church attendance, which can be compared to shed light on these different marginal effects.

[17]The pattern of coefficients reported in Figure 4 would also be obtained if the county controls were omitted from the participation regression, although the point estimates would be somewhat smaller. We are using three different measures of church participation over the 1860 period in Figure 4. One concern would be that the change in participation over time is mechanically driven by change in these measures. Notice, however, that participation increases steadily over time and not in three distinct jumps.

[18]In a related test, we also included religious fractionalization in 1860 as a regressor to allow for the possibility that competition between churches could have affected participation rates (Finke and Stark 1992, Gruber 2005). We find that religious fractionalization in 1860 has no effect on future church participation, leaving the estimated coefficients in Figure 4 unchanged.

[19]The advantage of focussing on the initial conditions is that all networks had the same vintage at that point in time and so ethnic rivalry can be measured relatively easily. To estimate the effect of social heterogeneity in later decades we would need to consider the entire history of in-migration and account for differences in vintage and, hence, market power across ethnic networks.

The NLSY consists of a nationally representative sample of American high school seniors in 1979 who were subsequently interviewed annually from 1979 to 1994 and biennially thereafter. The NLSY collects information on the respondents' county of birth, the religious denomination that they were raised in, and church attendance in the 1979 and 2000 rounds. To be consistent with the results using the CRB, we aggregate the individual data to the county level. Table 7, Columns 1-2 regress religious affiliation, separately for the migrant and non-migrant denominations, on 1860 fractionalization and the county-controls. The dependent variable in these regressions is measured as the share of respondents born in each county who were raised in the migrant denominations and the non-migrant denominations, respectively. Table 7, Columns 3-4 regress church participation, separately for the migrant and non-migrant denominations in 1979 and 2000, on 1860 fractionalization and the county-controls. The dependent variable is now measured as the share of the respondents born in each county who were raised in the relevant denominations and report attending church.

Individuals born in high fractionalization counties are more likely to be raised in migrant denominations and less likely to be raised in non-migrant denominations in Columns 1-2, although the fractionalization coefficient is not significant at conventional levels. This result indicates that some of the variation in participation with fractionalization in Figure 4 can be explained by growth in the migrant denominations at the extensive margin. Restricting attention to individuals who actually attend church, we see in Columns 3-6 that the fractionalization coefficient retains the same sign, but is now very precisely estimated for the migrant denominations. Notice that the fractionalization coefficient in Column 5 (the 2000 round) is very similar to the corresponding coefficient in Figure 4.

Church attendance and religious affiliation are jointly determined and so we cannot estimate the relationship between attendance (the intensity of participation) and fractionalization separately for individuals raised in the migrant denominations. However, the results in Table 7 can be used to test (and reject) the weaker hypothesis that attendance is uncorrelated with fractionalization in the migrant denominations. Suppose that a fraction θ_i of individuals raised in those denominations in county i attend church. The value of the dependent variable in the participation regression, Column 3 and Column 5, will then be θ_i multiplied by the dependent variable in the religious affiliation regression in Column 1. It follows that the fractionalization coefficient in the participation regression will be simply $E(\theta_i) < 1$ multiplied by the corresponding coefficient in the religious affiliation regression if θ_i is uncorrelated with fractionalization.[20] What we see instead is that the coefficient

[20] Let the share of the population affiliated with the Lutheran and Catholic church in county i be a_i. Let X_i denote the level of 1860 fractionalization in that county. Ignoring the additional controls and the constant term, we are effectively estimating the following regressions:

$$a_i = \beta X_i + \epsilon_i$$

in Column 3 (the 1979 round) is 50 percent larger than the corresponding coefficient in Column 1, while the coefficient in Column 5 (the 2000 round) is roughly the same size. The hypothesis that the intensity of participation is uncorrelated with fractionalization is clearly rejected.[21]

3.4 Local Identity and Career Decisions

Having described a mechanism through which local identity could have persisted over many generations, we now proceed to study its economic consequences. We hypothesize that individuals belonging to migrant denominations and born in high fractionalization counties should be endowed with a stronger sense of local identity and will consequently shade their career choices away from more mobile professional occupations. If the demand for professional occupations does not vary with fractionalization, as suggested by the results in Table 4 with the 1860 county-controls, then labor supply must respond to clear the market.

One supply-response strategy would be for individuals belonging to the non-migrant denominations in high fractionalization counties to select disproportionately into the professional occupations. A second strategy would be for individuals from the migrant denominations to continue to select into the professional occupations but to forego the relatively large returns to moving that are associated with those occupations. Finally, a third strategy would be for individuals born in other counties and trained as professionals to move (at a cost) into the high fractionalization counties. Once we introduce individual heterogeneity in the population and assume that the cost of becoming a professional is decreasing in ability, it follows under reasonable conditions that all three strategies will be adopted in equilibrium. This implies that individuals born in high fractionalization counties will be less likely to become professionals *on average*. Professionals must move into those counties and non-professionals must move out to clear the market.

If the objective is to avoid moving from the county of birth, why are non-professionals systematically over-supplied in the high fractionalization counties? The model developed in this section and described in greater detail in the Appendix shows that this result will be obtained in competitive

$$\theta_i a_i = \gamma X_i + u_i,$$

where ϵ_i, u_i are mean-zero disturbance terms. It follow that $\hat{\gamma} = E(\theta_i)\hat{\beta}$ if $cov(\theta_i, X_i) = 0$.

[21] The dependent variable in Columns 3-6 includes all individuals who attend church multiple times in the year, roughly matching the definition of adherents in the CRB. We experimented with a more stringent measure of participation – multiple times per month – without qualitatively changing the results. The fractionalization coefficient with the non-migrant denominations continues to be negative, small, and imprecisely estimated. The corresponding coefficient for the migrant denominations in 1979 declines slightly to 0.23 and is no longer significant at conventional levels, while the coefficient in 2000 remains significant and is just slightly smaller (0.18) than the point estimate in Column 5. We continue to reject the hypothesis that the fractionalization coefficient is smaller with participation rather than religious affiliation as the dependent variable, despite the fact that participation rates with the more stringent measure fall by over 15 percentage points and are now below 0.7.

equilibrium as long as the third strategy is active and there is *ex ante* uncertainty in the demand for different types of labor across counties.

To focus on the third supply-response strategy, we now ignore heterogeneity in ability and variation in local identity within counties. There are two counties in this economy with N individuals born in each county in each period. Two types of jobs are available: professional and non-professional. Individuals who are *ex ante* identical live for two periods, working in the second period of their lives. Those individuals who choose to occupy more productive professional jobs must invest in training in the first period of their lives. Individuals who expect to end up in non-professional jobs incur no such cost. On average, sN professional jobs and $(1-s)N$ non-professional jobs are demanded in each county in each period. However, these counties also face positive and negative demand shocks with equal probability, which separately shift the demand for professional and non-professional labor by ϵsN and $\epsilon(1-s)N$ respectively, but leave the total number of professional jobs $2sN$ and non-professional jobs $2(1-s)N$ constant across the two counties in all periods.[22]

Once demand shocks are introduced, labor must flow across counties at the beginning of each period to clear the market. In addition, we assume that professional jobs are also associated with mid-career moves. This additional dimension of mobility in professional occupations is based on the idea that the labor market for a school teacher or a secretary tends to be local. In contrast, professionals such as university professors or management consultants are continually re-sorting across local and regional labor markets as new opportunities arise and fresh cohorts of workers enter. A professional can advance her career considerably with such a move if it becomes available and we will later see that professionals are indeed much more likely to migrate from their county of birth. We model this characteristic of professional occupations by assuming that the opportunity to enhance overall productivity by switching a professional working in a particular county with a professional working in the other county arrives with probability P in each period.

Individuals dislike moving, particularly those with a strong local identity. Let the cost of moving be C_1 for individuals born in county 1 and $C_2 < C_1$ for individuals born in county 2 with weaker local identity. This is the only difference between workers in the two counties. We assume that all workers are employed and that professional workers always take advantage of the productivity enhancing career opportunities when they arise, in which case total output in this economy remains constant across all states of the world. Notice that this assumption effectively rules out the second labor-response strategy described above. Although wages for professionals and non-professionals will

[22]Within a county, demand shocks for professionals and non-professionals could be positively or negatively correlated. We could also allow the size of the shocks to vary, $\overline{\epsilon}$ for the professionals and $\underline{\epsilon}$ for the non-professionals, without changing the results reported below.

adjust across the two locations to clear the labor market in practice, the competitive labor allocation can be conveniently derived as the solution to the Central Planner's problem of minimizing expected training and moving costs across both locations.

Let the supply of professional labor in county 1 be x_1. We show in the Appendix that expected cost is a piece-wise linear function of x_1, as described in Figure 5, which is minimized at $x_1^* = sN - \epsilon sN$ for $P < 1/2(C_1 + C_2)/(C_1 - C_2)$ and at $x_1^{**} = sN - \epsilon sN - \epsilon N(1 - 2s)$ for $P > 1/2(C_1 + C_2)/(C_1 - C_2)$. The equilibrium supply of professional labor in county 1, with stronger local identity, falls short of the expected demand sN for all values of P. Moreover, the share of professionals supplied by county 1 is strictly less than the corresponding share in county 2.[23]

The intuition for this result is that as long as there is *ex ante* uncertainty in the type of jobs demanded, some individuals will have to move in both counties with positive probability to clear the market. By allocating a surplus of non-professionals to county 1, the Central Planner increases the probability of such moves, but trades this off against the lower probability of having individuals born in county 1 move *ex post* on the job. We show in the Appendix that the supply of professionals in county 1 could also fall short of the expected demand without uncertainty, but only if $P > (C_1 + C_2)/(C_1 - C_2)$.

In our set up, individuals *born* in county 1 are less likely to be professionals because they incur a greater cost when they move ($C_1 > C_2$). However, this result could also be obtained if $C_1 = C_2$ and the demand for professional labor is lower in county 1, $s_1N < s_2N$. The second prediction of the model, which allows us to rule out this alternative demand-side explanation, is that once the labor market has cleared, the share of professionals *residing* in the two counties should on average be the same (sN). In contrast, if the demand for professional labor is lower in county 1, then individuals born and subsequently residing in that county should be less likely to be professionals.

3.5 Empirical Analysis

3.5.1 Individual Data

To test the predictions from the previous section we need information on the individual's career choice, county of birth, and county of residence (post-employment). The NLSY is the only large-scale data set that we are aware of that includes this information. The data set includes basic information on the respondent's age, gender, race and, most importantly, county of birth. The

[23] For $P < 1/2(C_1 + C_2)/(C_1 - C_2)$, the share of professionals is $s - \epsilon s$ in county 1 and $s + \epsilon s$ in county 2. For $P > 1/2(C_1 + C_2)/(C_1 - C_2)$, the corresponding shares are $s - \epsilon(1 - s)$ and $s + \epsilon(1 - s)$.

Armed Forces Qualification Test (AFQT), which is designed to provide an unbiased measure of the individual's intelligence, was administered to all respondents in 1979. Subsequent survey rounds collected contemporaneous information on educational attainment, employment, occupation, income, and county of residence. We will study occupational choice and other outcomes related to that economic decision at two points in time – 1994 and 2000 – when the respondents were old enough to be settled in their careers and to have made some job-related moves. Occupational choices from the NLSY in these years will be matched to census data on historical fractionalization, both in the individual's county of birth and the contemporaneous county of residence, to test the predictions of the model.

Table 8 reports descriptive statistics for the individuals in our sample, who were on average 18 years old in 1979 and so around 33 years old in 1994 and 39 years old in 2000. Occupational categories in the NLSY (up to the 2000 round) are based on the 1970 codes from the census. Professional occupations are defined to include relevant codes listed under the Professional, Technical, and Kindred Workers category [1-196]. Job-related geographical mobility is the chief property that distinguishes professional and non-professional occupations in our model. We consequently exclude technical occupations and other occupations where career moves are unlikely to be important from this category.[24]

Based on this occupational classification, 9 percent of the respondents hold professional jobs, with little change from 1994 to 2000. 56 percent of all respondents had migrated out of their birth-county by 1994, with an increase to 59 percent by 2000. Consistent with the assumption that professional occupations are associated with greater mobility, 75 percent of the professionals and 53 percent of the non-professionals had migrated out of their birth-county by 1994 (these differences are significant at the 5 percent level and similar in 1994 and 2000).

Individuals in the sample are on average 33 years old by 1994 and should be established in the labor market. Nevertheless, employment levels continue to increase over time, from 81 percent in 1994 to 92 percent in 2000. Conditional on being employed, annual income (in 2000 dollars) also increases from 28,000 in 1994 to 33,000 in 2000. These changes in employment and income are presumably life-cycle effects, but they could, in principle, be due to selective attrition since this is a longitudinal survey. Notice, however, that racial composition and the proportion of women in the sample are very stable over the 1994-2000 period. Thus, we do not observe selective attrition from the sample, at least with respect to two important demographic characteristics that are associated

[24] These occupations include Librarians [32], Nurses, dieticians, and therapists [74-76], Religious workers [86,90], Social and recreation workers [100,101], Teachers, except college and university [141-145], and Technicians [150-174].

with income and employment.

3.5.2 Fractionalization in the County of Birth and Occupational Choice

The first prediction from the model of occupational choice incorporating local identity is that individuals born in counties with greater ethnic fractionalization in 1860 should be less likely to hold professional jobs. Including race, gender, and age as regressors (although their omission would not affect the results) we see in Table 9, Columns 1-2 that individuals born in high fractionalization counties are indeed less likely to hold professional jobs in the 1994 and 2000 rounds of the NLSY. The individuals in our sample are drawn from 150 of the approximately 400 Midwestern counties that were incorporated and had attracted foreign migrants by 1860. While it thus seems unlikely that a few outlying counties are driving the results, we nevertheless report nonparametric estimates of the relationship between occupational choice and historical fractionalization in Figure 6. We see that the probability that the individual is a professional declines steadily with fractionalization, both in 1994 and in 2000, verifying the robustness of our results.[25]

Our test of the hypothesis that identity shapes occupational choice relies on the assumption that fractionalization is uncorrelated with the demand for professional workers in the county. Recall from Table 4, however, that historical fractionalization had a positive and significant effect on current economic characteristics, measured by population in the county as well as racial and ethnic fractionalization, but that this effect disappeared once important characteristics of the 1860 economy were included as regressors. More urban counties will have a greater demand for professional workers, shifting *up* the coefficient on the fractionalization variable in Table 9, Columns 1-2. The regressions that follow will thus include the same 1860 characteristics as in Tables 4 and 5 – population, manufacturing share and agriculture share – to control for variation in the *demand* for professional workers across counties.

The prediction that high fractionalization counties should supply fewer professional workers is derived conditional on the total *surplus* that was historically available to migrants in the labor market R and the total number of migrant workers N. Everything else equal, an increase in R would increase the cost of exit at the margin, encouraging efforts to instill local identity historically and lowering the propensity to enter professional occupations today. Superior transportation infrastructure in 1860 was associated with high fractionalization and a larger county population in Table 3 and

[25]The nonparametric regressions are estimated using the Epanechnikov kernel smoothing function. Less than 5 percent of the observations in Table 9, Columns 1-2 are drawn from counties with fractionalization below 0.45 and so too much weight should not be placed on the extremely steep initial decline in Figure 6. Although all the parametric regressions that follow will use the full set of counties, we verified that the fractionalization coefficient is unchanged when the sample is restricted to individuals drawn from counties in the 0.45-0.8 range.

so we expect fractionalization to be positively correlated with R. It follows that exclusion of R from the occupation choice regression in Table 9, Columns 1-2 would have shifted the coefficient on fractionalization *down*. The same 1860 characteristics that we described above could be used to control for R, but since the professional demand effect and the surplus effect work in opposite directions, the effect of their inclusion on the fractionalization coefficient would now be ambiguous. By the same argument, we can no longer predict the sign of the coefficients on the 1860 characteristics when they are included in the occupation choice regression. For example, larger 1860 population is associated with urbanization today and, hence, an increased demand for professional jobs. At the same time, a larger 1860 population is associated with a larger R, stronger identity and, hence, a reduced supply of professional workers. Finally, we would want to account for the number of migrants N when estimating the effect of fractionalization on occupational choice. This variable is highly correlated with county population (the correlation is 0.9) and so will be omitted from the regressions that follow. 1860 population captures the professional demand effect, the surplus effect, and the effect of N on occupational choice.

We see in Table 9, Columns 3-4 that the coefficient on 1860 fractionalization becomes more negative and is more precisely estimated once the additional county-level controls are included in the occupational choice regressions. Based on the estimates in Column 4, a one standard deviation decline in ethnic fractionalization would increase the probability of holding a professional job from 9 percent to as much as 15 percent. Individuals born in counties with a greater share of manufacturing and agricultural jobs in 1860 are also less likely to hold professional jobs. Population in 1860, in contrast, has no effect on occupational choice in both the 1994 and the 2000 rounds. Finally, women and Non-Whites are significantly less likely to hold professional jobs in Columns 1-4.

Figure 4 traces a continuous link between initial conditions in 1860 and participation in select religious denominations historically dominated by the migrants over the subsequent 140 years. Table 9 brings the different pieces of the empirical analysis together by documenting the relationship between 1860 fractionalization and occupational choice in 2000. We close this section by discussing a number of tests that we conducted to verify the robustness of the key occupational choice result.

1. Alternative construction of the fractionalization variable: First, we computed ethnic fractionalization in 1860 with men only. Both men and women participated in the workforce in the nineteenth and early twentieth centuries, with ethnic networks channeling women into jobs as well. Bodnar (1980), for example, cites a 1930 study of two thousand foreign-born women, most of whom reported that they had secured their first job through social connections. Nevertheless, we might expect labor networks to have been organized along gender lines within ethnic groups, with male networks occupying a dominant position in the labor market and in the communities they were drawn from. The coefficient on the alternative fractionalization measure (computed using men alone) continues to be negative and significant, although it is smaller in size.

Inspection of the occupation categories in Table 2 indicates that farmers made up 50 percent of the workforce in 1860. Ethnic competition may have been less relevant in this category and so we recomputed the county-level fractionalization statistic placing zero weight on fractionalization among the farmers as a second robustness test.[26] The fractionalization coefficient becomes slightly larger in absolute magnitude and is more precisely estimated.

As a third test, we constructed the fractionalization variable using alternative occupational classifications. The county-level statistic that we use in all the regressions reported in the paper is computed as the weighted average of ethnic fractionalization within each of the 11 occupational categories listed in Table 2. We measured labor market rivalry within a coarser set of four occupational categories as well as within a finer set of 139 occupations, without changing any of the results.[27]

2. Alternative construction of the occupation variable: First, we expanded the set of professional occupations by including individuals assigned to the "managers and administrators" code [245] within the broader Managers and Administrators, except Farm category [201-245]. Most of the specific occupations listed in this broad category do not conform to our definition of a professional occupation and the "managers and administrators" classification does not provide much information on the actual nature of jobs that it covers. However, 74.7 percent of the individuals assigned to that category migrated out of the county of their birth, which is close to the level of migration for the other occupations that we include in the professional category. Inclusion of the managers and administrators in the professional classification increased the share of professionals in the sample to 17.7 percent. Although these results need to to interpreted with caution since the additional professional occupations are selected on the basis of an outcome (migration) rather than their fundamental characteristics, we nevertheless verified that the fractionalization coefficient in

[26] We include all individuals who report being employed, regardless of their age, when computing the fractionalization statistic since there were no age restrictions on employment at that time. Apart from the 11 broad occupational categories in Table 2, some individuals in the census were also assigned to an undefined occupational category. Women were disproportionately represented in this category, which presumably covers home production and other informal activities. As with the farmers, we assume that networks were less relevant in this category and assign it zero weight when computing ethnic fractionalization in all the regressions that we report. Nevertheless, we verified that assigning a weight to this category based on its share of the migrant workforce in the county had no effect on the estimated fractionalization coefficient in the occupational choice regression.

[27] The four aggregate categories are white collar, agriculture, manufacturing, and service and laborers. These categories correspond to the broad headings in Table 2 except that Blue collar, nonfarm is divided into manufacturing (craftsman, operative) and service and laborers (household service, service, laborer non-farm). The 139 disaggregate occupations are defined at the three-digit census code level. 84 percent of the listed occupations in the 1860 data are drawn from just 11 of these occupations.

the occupational choice regression with this expanded definition of professional activity remained negative and significant.

As a second test we ran the occupational choice regression in other years – 1993, 1996, 1998 – generating a pattern of fractionalization coefficients very similar to the point estimates obtained in 1994 and 2000.

3. Alternative specification of the occupation regression: First, we replaced total population with the migrant population and, separately, with the workforce, in the occupational choice regression. These variables are all highly correlated, and not surprisingly the results were unchanged.

Second, we included the share of migrants in the workforce, and the interaction of this variable with ethnic fractionalization, as additional regressors. This specification allows for the possibility that ethnic fractionalization had a larger effect in counties where migrants made up a larger share of the total population. However, both these variables had an insignificant effect on occupational choice, leaving the (uninteracted) fractionalization coefficient unchanged.

Third, we included the population share of each ethnicity in the 1860 census, computed at the level of the county, as additional regressors to allow for the possibility that individuals with particular ethnic ancestry continue to be concentrated in specific occupations today. Given the high rate of inter-county migration, we do not expect the ethnic shares in 1860 to be relevant today and not surprisingly the fractionalization coefficient was hardly affected by the inclusion of these additional regressors.

3.5.3 Fractionalization and Outcomes Related to Occupational Choice

An individual born in a county with higher ethnic fractionalization in 1860 is less likely to select into a professional occupation in 1994 and 2000. Our interpretation of this result is that individuals born in high fractionalization counties identify strongly with their local communities and so wish to avoid the spatial mobility that comes with professional occupations. The regression results reported in Table 10, Columns 1-2 indicate that individuals from high fractionalization counties are indeed significantly less likely to migrate from the county of their birth.[28] On average, around 58 percent of the individuals in the sample migrate from the county of birth. The point estimates indicate that a one standard deviation increase in fractionalization reduces migration by 8 percentage points

[28]Recall that one supply-response strategy was for individuals from migrant denominations to continue to select into the professional occupations but to forego the relatively large returns to moving that are associated with those occupations. Although it is tempting to estimate the relationship between migration and fractionalization separately for the professionals, note that occupational choice is jointly determined with migration in our framework. By the same argument, we cannot estimate the relationship between occupational choice and fractionalization separately by religious denomination.

(a 14 percent decline). Among the other regressors, none of the 1860 county-level controls significantly affect migration, but Whites are significantly more likely to move. Age is also (mechanically) positively associated with migration.

Does the effect of fractionalization on occupational choice and migration that we have just described have financial consequences? The results in Table 10, Columns 3-4 indicate that while individuals born in counties with a greater share of manufacturing and agriculture in 1860 are significantly more likely to hold a job by 2000, employment levels do not vary significantly with ethnic fractionalization in 1994 or 2000. Whites and males are, not surprisingly, significantly more likely to be employed, although the importance of these individual characteristics declines over time.

In contrast with the results for employment, the income regressions reported next in Table 10, Columns 5-6 indicate that high fractionalization is associated with significantly lower income (in 2000), conditional on being employed. Average annual income in 2000 was 33,000 dollars, and so our estimates indicate that a one standard deviation increase in fractionalization would have reduced income by 2,300 dollars (a 7 percent decrease). A greater share of manufacturing and agriculture in 1860 is associated with lower income in 2000, with Whites and males earning significantly more in 1994 and 2000.

3.5.4 Alternative Explanations

1. Fractionalization and the demand for professionals: The second prediction of the model is that historical fractionalization in the county of residence should have no effect on occupational choice. This implies that non-professional workers must flow out of the high fractionalization counties, while professional workers flow in to clear the labor market. This prediction rules out the possibility that differences in the demand for professional labor across counties are driving our results.

To test this prediction we regress occupation on fractionalization in the county of *residence* in Table 11. The fractionalization coefficient is small and insignificant in 1994 and 2000, with and without the county controls in Columns 1-4. It actually changes sign and is positive in three of the four specifications. Individuals working in high fractionalization counties are at least as likely to be professionals as individuals working in low fractionalization counties. It is only individuals *born* in those counties who are less likely to select into those occupations.[29]

[29]One remaining concern is that the mismatch between the supply and the demand for professional labor in the high fractionalization counties could have been driven by unexpected changes in the demand for professionals in those counties rather than a strong local identity. Suppose, for example, that the demand for professionals was systematically lower in high fractionalization counties historically and then increased in the 1980's, with the restructuring of the U.S.

2. Fractionalization and educational attainment: Could variation in individual ability or the quality of the school system have generated the observed variation in the supply of professionals across counties? High fractionalization counties had superior transportation infrastructure in 1860 and so were most likely growing relatively rapidly at that time. It is entirely possible that particular types of migrants were attracted to those counties, although we would expect migrants drawn to areas with many competing ethnic groups to have been positively selected on innate talent. Given the high rates of inter-county migration, we would in any case expect few families to have maintained an unbroken line of descent to the present day in the same county. We noted earlier that while 1860 fractionalization may have been positively correlated with relevant measures of social heterogeneity many decades later when the secondary school system was being established, fractionalization was also positively correlated with religious participation. The net effect of fractionalization on investments in education was thus seen to be ambiguous, and reassuringly, 1860 fractionalization was uncorrelated with education expenditures in 1990 in Table 5.

To provide direct support for the idea that individuals born in high fractionalization counties are no less capable and no less prepared to invest in the human capital that is necessary to secure professional jobs, we regress measures of educational attainment obtained from the NLSY on 1860 fractionalization in Table 12. We see in Table 12, Columns 1-3 that fractionalization, and all the 1860 county-controls for that matter, have no effect on AFQT scores, high school completion, and college completion.[30] In contrast, individual characteristics have a strong effect on AFQT scores and college completion.

Why are individuals born in high fractionalization counties as likely to attend and complete college even though they expect to end up in non-professional occupations? To explore this result further, we construct a binary variable that indicates whether the college major would have channeled the individual into a less mobile occupation.[31] Individuals born in high fractionalization counties and females, not surprisingly, are significantly more likely to choose such majors in Table 12, Column 4. On average, 39 percent of the individuals in the sample who complete college choose majors that are

economy, after our NLSY cohort had made their career choices. To rule out this possibility, we regressed manufacturing share, agriculture share, and county population in 1970, obtained from the 1972 County Data Book, on the same 1860 characteristics as in Table 4, Columns 7-9. 1860 fractionalization has no effect on economic conditions in 1970, just as we saw earlier with the same conditions in 1990.

[30] The fact that fractionalization is uncorrelated with education also rules out an alternative channel through which fractionalization could affect religious participation (see Sacerdote and Glaeser [2001] for an analysis of the complex relationship between education and religious participation).

[31] The non-professional (less mobile) majors include agriculture and natural resources, accounting, hotel and restaurant management, marketing and purchasing, transportation and public utilities, real estate, insurance, secretarial studies, personnel management, labor and industrial relations, communications, data processing, education, health professions (except allied health, other, veterinary medicine specialties, speech pathology and audiology), home economics, library science, psychology for counseling, rehabilitation counseling, community services, parks and recreation management, social work and helping services, law enforcement, administration of justice, and theology.

associated with lower mobility. The point estimates indicate that a one standard deviation increase in fractionalization would increase the likelihood of such a major being chosen by four percentage points (a 10 percent increase). Once again, we uncover evidence that suggests that individuals born in high fractionalization counties are making conscious choices that restrict their future mobility.

3.6 Conclusion

This paper draws a connection between ethnic labor networks in the American Midwest when it was first being settled, the local identity that emerged endogenously to support these networks and then persisted over many generations, and occupational choice today. Individuals born in counties with greater ethnic fractionalization in 1860, which we expect to be associated with stronger local identity and better functioning parochial institutions, are significantly less likely to select into professional occupations, which come with greater geographical mobility, in 2000.

The results in this paper are relevant to the ongoing debate on "economic institutions" (Acemoglu, Johnson and Robinson [2001, 2002]) versus "culture" (Barro and McCleary 2003, Tabellini 2005, Fernandez and Fogli 2007) as determinants of growth. We find that a local cultural trait generates significant variation in occupational choice within a relatively homogenous region – the American Midwest – and it is possible that cultural effects of this sort would be even larger across countries or regions with very different histories. At the same time, culture cannot be sustained without institutional support. Social institutions such as the church and the family help sustain cultural traits, which in turn keep these institutions alive. As long as these institutions continue to be useful, cultural traits can persist long after the economic circumstances that gave rise to them have ceased to be relevant. While the economics literature has focussed much of its attention on economic and political institutions, social institutions, with their complementary cultural traits, might have important effects on growth as well.

3.7 References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5): 1369-1401.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2002. Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *Quarterly Journal of Economics*, 117(4): 1231-1294.

Akerlof, George A. and Rachel E. Kranton. 2000. Economics and Identity, *Quarterly Journal of Economics* 105(3): 715-53.

Akerlof, George A. and Rachel E. Kranton. 2005. Identity and the Economics of Organizations, *Journal of Economic Perspectives* 19(1): 9-32.

Alba, Richard D. 1990. *Ethnic Identity: The Transformation of White America*. New Haven and London: Yale University Press.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. Public Goods and Ethnic Division, *Quarterly Journal of Economics* 114(4): 1243-84.

Alesina, Alberto, Reza Baqir, and Caroline Hoxby. 2004. Political Jurisdictions in Heterogeneous Communities, *Journal of Political Economy* 112(2): 348-396.

Alexander, June Granatir. 1991. Moving into and out of Pittsburgh: Ongoing Chain Migration, in *A Century of European Migrants, 1830-1930*, eds. Rudolph J. Vecoli and Suzanne M. Sinke, Urbana: University of Illinois Press.

Barro, Robert J. and Rachel M. McCleary. 2003. Religion and Economic Growth Across Countries, *American Sociological Review* 68: 760-781.

Barton, Josef J. 1975. *Peasants and Strangers: Italians, Rumanians, and Slovaks in an American City, 1890-1950*. Cambridge, Massachusetts: Harvard University Press.

Bénabou, Roland and Jean Tirole. 2006. Belief in a Just World and Redistributive Politics, *Quarterly Journal of Economics* 121(2): 699-746.

Bisin, Alberto and Thierry Verdier. 2000. "Beyond the Melting Pot": Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits, *Quarterly Journal of Economics* 115(August): 955-988.

Bisin, Alberto, Georgio Topa, and Thierry Verdier. 2004. Religious Intermarriage and Socialization in the United States, *Journal of Political Economy* 112(3): 615-664.

Bodnar, John. 1980. Immigration, Kinship, and the Rise of Working-Class Realism in Industrial America, *Journal of Social History.* 45-65.

Bodnar, John. 1985. *The Transplanted: A History of Immigrants in Urban America.* Bloomington: Indiana University Press.

Bodnar, John, Roger Simon, and Michael P. Weber. 1982. *Lives of Their Own: Blacks, Italians, and Poles in Pittsburgh, 1900-1960.* Urbana: University of Illinois Press.

Elder Jr., Glen H. and Rand D. Conger. 2000. *Children of the Land: Adversity and Success in Rural America.* Chicago and London: University of Chicago Press.

Fernandez, Raquel and Alessandra Fogli. 2007. Culture: An Empirical Investigation of Beliefs, Work, and Fertility. New York University Department of Economics mimeograph.

Finke, Roger and Rodney Stark. 1992. *The Churching of America, 1776-1990.* New Brunswick, NJ: Rutgers University Press.

Fishlow, Albert. 2000. Internal Transportation in the Nineteenth and Early Twentieth Centuries, in *The Cambridge Economic History of the United States*, eds. Stanley L. Engerman and Robert E. Gallman, Cambridge, England: Cambridge University Press.

Gans, Herbert. 1962. The Urban Villagers: Group and Class in the Life of Italian-Americans. New York: The Free Press of Glencoe.

Gans, Herbert. 1979. Symbolic Ethnicity: The Future of Ethnic Groups and Cultures in America, *Ethnic and Racial Studies* 2(January): 1-20.

Gerson, Kathleen, C. Ann Stueve, and Claude S. Fischer. 1977. Attachment to Place, in *Networks and Places: Social Relations in the Urban Setting*, eds. Claude S. Fischer et al., New York: The Free Press.

Gjerde, Jon. 1985. *From Peasants to Farmers: The Migration from Balestrand, Norway to the Upper Middle West.* Cambridge, England: Cambridge University Press.

Gjerde, Jon. 1991. Chain Migrations from the West Coast of Norway, in *A Century of European Migrants, 1830-1930*, eds. Rudolph J. Vecoli and Suzanne M. Sinke, Urbana: University of Illinois Press.

Gjerde, Jon. 1997. *The Minds of the West: Ethnocultural Evolution in the Rural Middle West, 1830-1917*. Chapel Hill: University of North Carolina Press.

Goldin, Claudia and Lawrence Katz. 1999. Human Capital and Social Capital: The Rise of Secondary Schooling in America, 1910-1940, *Journal of Interdisciplinary History* 29 (March): 683-723.

Gordon, David M., Richard Edwards, and Michael Reich. 1982. *Segmented Work, Divided Workers: The Historical Transformation of Labor in the United States.* Cambridge England: Cambridge University Press.

Greif, Avner. 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge England: Cambridge University Press.

Gruber, Jonathan. 2005. Religious Market Structure, Religious Participation, and Outcomes: Is Religion Good for You? NBER Working Paper 11377.

Hoerder, Dirk. 1991. International Labor Markets and Community Building by Migrant Workers in the Atlantic Economies, in *A Century of European Migrants, 1830-1930*, eds. Rudolph J. Vecoli and Suzanne M. Sinke, Urbana: University of Illinois Press.

Hummon, David M. 1990. *Commonplaces: Community Ideology and Identity in American Culture.* Albany: State University of New York Press.

Hunter, Albert. 1975. The Loss of Community: An Empirical Test Through Replication, *American Sociological Review* 40(5): 537-552.

Hutchinson, E. P. 1956. *Immigrants and Their Children 1850-1950* New York: John Wiley and Sons, Inc.

Iannaccone, Laurence R. 1992. Sacrifice and Stigma: Reducing Free-Riding in Cults, Communes, and Other Collectives, *Journal of Political Economy* 100(2): 271-297.

Iannaccone, Laurence R. 1998. Introduction to the Economics of Religion, *Journal of Economic Literature* 36: 1465-1496.

Kamphoefner, Walter D. 1987. *The Westfalians: From Germany to Missouri*. Princeton: Princeton University Press.

Kim, Sukkoo. 2007. Immigration, Industrial Revolution and Urban Growth in the United States, 1820-1920: Factor Endowments, Technology and Geography, *NBER Working Paper* No. 12900.

Margo, Robert A. 1999. Regional Wage Gaps and the Settlement of the Midwest, *Explorations in Economic History* 36: 128-143.

Meyer, David R. 1989. Midwestern Industrialization and the American Manufacturing Belt in the Nineteenth Century, *Journal of Economic History* 49(4): 921-937.

Sacerdote, Bruce and Edward L. Glaeser. 2001. Education and Religion, *NBER Working Paper* No. 8080.

Sanders, Scott Russell. 2007. To Stay or Move On: The Artist's Dilemma. *Bloom*, February/March: 65-68.

Sumner, William G. 1906. *Folkways*. New York: Ginn.

Tabellini, G. 2005. Culture and Institutions: Economic Development in the Regions of Europe, CESifo Working Paper No. 1492.

Thistlethwaite, Frank. 1991. Migration from Europe Overseas in the Nineteenth and Twentieth Centuries, in *A Century of European Migrants, 1830-1930*, eds. Rudolph J. Vecoli and Suzanne M. Sinke, Urbana: University of Illinois Press.

Tullock, Gordon. 1980. Efficient Rent Seeking, in *Toward a Theory of the Rent-Seeking Society*, eds. James M. Buchanan, Robert D. Tollison, and Gordon Tullock, College Station: Texas A&M Press.

Weber, Max. 1930. *The Protestant Ethic and the Spirit of Capitalism*. London, England: Allen and Unwin.

3.8 Appendix: Labor Market Equilibrium

3.8.1 Population and Production Technology

Two types of jobs are available in this economy: professional and non-professional. Individuals who are *ex ante* identical live for two periods, working in the second period of their lives. Those individuals who choose to occupy professional jobs incur a training cost C_e in the first period of their lives. Individuals who expect to end up in non-professional jobs incur no such cost. The expected output obtained from a professional worker who takes advantage of career opportunities when they arise is $\overline{\theta}$ and the output obtained from an non-professional worker with certainty is $\underline{\theta} < \overline{\theta}$.

There are two counties in this economy with N individuals born in each county in each period. On average, sN professional jobs and $(1-s)N$ non-professional jobs are demanded in each county in each period. However, these counties also face demand shocks, with two states of the world occurring with equal probability:

State 1: $sN + \epsilon sN$ professional and $(1-s)N + \epsilon(1-s)N$ non-professional jobs in county 1.
$sN - \epsilon sN$ professional and $(1-s)N - \epsilon(1-s)N$ non-professional jobs in county 2.
State 2: $sN - \epsilon sN$ professional and $(1-s)N - \epsilon(1-s)N$ non-professional jobs in county 1.
$sN + \epsilon sN$ professional and $(1-s)N + \epsilon(1-s)N$ non-professional jobs in county 2.

Notice that these demand shocks are skill neutral, in the sense that the probability of receiving a shock is the same for professional and non-professional workers within each county. The shocks are also positively correlated for professional and non-professional workers within a county. We could relax each of these assumptions without changing any of the results that follow.[32] In addition, the opportunity to enhance overall productivity by switching a professional working in a particular county with a professional working in the other county arrives with probability P in each period.

Let the cost of moving be C_1 for individuals born in county 1 and $C_2 < C_1$ for individuals born in county 2 with weaker local identity. We assume that all workers are employed and that professional workers always take advantage of the productivity enhancing career opportunities when they arise, in which case total output in this economy remains constant across all states of the world: $2N\left[s\overline{\theta} + (1-s)\underline{\theta}\right]$. The competitive labor market equilibrium can be obtained in that case as the solution to the Central Planner's problem of minimizing training and moving costs across both locations.

[32]Specifically, we allowed shocks to be perfectly negatively correlated for professional and non-professional workers within each county. We also allowed the size of the shocks to vary; $\overline{\epsilon}$ for professionals and $\underline{\epsilon}$ for non-professionals.

3.8.2 The Central Planner's Problem

Let the supply of professional labor in county 1 be x_1. From the structure of the demand shocks it then follows that the supply of professional labor in county 2 will be $2sN - x_1$ and that the supply of non-professional labor in county 1 will be $N - x_1$. We derive x_1 as the solution to the Central Planner's cost minimization problem. Expected cost turns out to be a piece-wise linear function of x_1 and so it will be convenient to solve what is essentially a linear programming problem in three regimes:

Regime 1: $x_1 \in [sN - \epsilon sN, sN + \epsilon sN]$

The supply of professional labor in each county is sufficient to satisfy the minimum demand in that county but does not exceed the maximum demand.

The labor flow in each state of the world can then be derived as:

Flow in state 1: $sN + \epsilon sN - x_1$ professional labor from county 2 to county 1.

$(1 - s)N + \epsilon(1 - s)N - (N - x_1)$ non-professional labor from county 2 to county 1.

Flow in state 2: $x_1 - [sN - \epsilon sN]$ professional labor from county 1 to county 2.

$(N - x_1) - [(1 - s)N - \epsilon(1 - s)N]$ non-professional labor from county 1 to county 2.

The Central Planner chooses x_1 to minimize expected cost

$$E(C) = 2sNC_e + \frac{1}{2}\epsilon N (C_1 + C_2) + Px_1C_1 + P(2sN - x_1)C_2, \tag{3.1}$$

where the second term on the right hand side is the cost associated with movement at the start of the period and the last two terms reflect movement of professional labor during the period. Because $C_1 > C_2$ it is easy to verify that $E(C)$ is increasing linearly in x_1 and will be minimized at $x_1^* = sN - \epsilon sN$. Labor flows with $x_1 = x_1^*$ are then obtained as:

Flow in state 1 ($x_1 = x_1^*$): $2\epsilon sN$ professional labor from county 2 to county 1.

$\epsilon N(1 - 2s)$ non-professional labor from county 2 to county 1.[33]

Flow in state 2 ($x_1 = x_1^*$): No flow of professional labor.

ϵN non-professional labor from county 1 to county 2.

Regime 2: $x_1 \in [sN - \epsilon sN - \epsilon N(1 - 2s), sN - \epsilon sN]$

We now reduce the supply of non-professional labor in county 2, but at most to the point where no non-professional labor flows to county 1 in state 1. In our set up, any reduction in non-professional

[33] To generate a positive labor flow we require $s \leq 1/2$. This is a reasonable assumption since just a small fraction of jobs (9 percent in our data) are professional. For the case with asymmetric shocks, $\overline{\epsilon}$ for the professionals and $\underline{\epsilon}$ for the non-professionals, the corresponding condition is $s \leq 1/(1 + \overline{\epsilon}/\underline{\epsilon})$.

labor supply in county 2 must lead to a reduction in professional labor in county 1 by the same amount. It then follows that the supply of professional labor in county 1 will no longer be sufficient to meet even the minimum demand in that county, while the supply of professional labor in county 2 will exceed the maximum demand in that county. Labor flows at the beginning of the period will necessarily increase, with an accompanying increase in moving costs, but we will see that this may be outweighed by the reduced cost of recounty for professional workers from county 1.

The labor flow in each state is derived as:

Flow in state 1: $sN + \epsilon sN - x_1$ professional labor from county 2 to county 1.

$(1-s)N + \epsilon(1-s)N - (N - x_1)$ non-professional labor from county 2 to county 1.

Flow in state 2: $sN - \epsilon sN - x_1$ professional labor from county 2 to county 1.

$(N - x_1) - [(1-s)N - \epsilon(1-s)N]$ non-professional labor from county 1 to county 2.

The expected cost can then be expressed as:

$$E(C) = 2sNC_e + \frac{1}{2}\left[\epsilon N + sN(1-\epsilon) - x_1\right](C_1 + C_2) + Px_1C_1 + P(2sN - x_1)C_2. \tag{3.2}$$

Collecting terms, $E(C)$ is declining linearly in x_1 and continues to be minimized at $x_1^* = sN - \epsilon sN$ if $P < 1/2(C_1+C_2)/(C_1-C_2)$. However, the local minimum is obtained at $x_1^{**} = sN - \epsilon sN - \epsilon N(1-2s)$, if the sign of the inequality is reversed. This will be the case if P and $C_1 - C_2$ are sufficiently large. Substituting x_1^{**} in equation (3.2) above, it is easy to verify that the term in square brackets is greater than ϵN, the term corresponding to it in equation (3.1), which implies that moving costs at the beginning of the period increase when going from x_1^* to x_1^{**}. However, moving costs during the period decrease with the reduction in x_1, and this effect dominates under the conditions on P and $C_1 - C_2$ derived above. Labor flows with $x_1 = x_1^{**}$ are obtained as:

Flow in state 1 $(x_1 = x_1^{**})$: ϵN professional labor from county 2 to county 1.

No flow of non-professional labor.

Flow in state 2 $(x_1 = x_1^{**})$: $\epsilon N(1-2s)$ professional labor from county 2 to county 1.

$2\epsilon N(1-s)$ non-professional labor from county 1 to county 2.

Regime 3: $x_1 \in [0, sN - \epsilon sN - \epsilon N(1-2s)]$

We now reduce the supply of professional labor in county 1, with an accompanying increase in non-professional labor, even further so that professional labor flows from county 2 to county 1 and non-professional labor flows in the opposite direction in both states of the world.

Labor flows are now derived as:

Flow in state 1: $sN + \epsilon sN - x_1$ professional labor from county 2 to county 1.

$(N - x_1) - [(1-s)N + \epsilon(1-s)N]$ non-professional labor from county 1 to county 2.

Flow in state 2: $sN - \epsilon sN - x_1$ professional labor from county 2 to county 1.

$(N - x_1) - [(1-s)N - \epsilon(1-s)N]$ non-professional labor from county 1 to county 2.

The corresponding expected cost expression is obtained as:

$$E(C) = 2sNC_e + \frac{1}{2}\left[2(sN - x_1)\right](C_1 + C_2) + Px_1C_1 + P(2sN - x_1)C_2. \tag{3.3}$$

It is straightforward to verify that $E(C)$ is unambiguously decreasing in x_1 and, hence, is minimized at $x_1^{**} = sN - \epsilon sN - \epsilon N(1-2s)$.

3.8.3 Equilibrium Labor Allocation

As described in Figure 5 and derived above, $E(C)$ is increasing in x_1 to the right of $sN - \epsilon sN$ and decreasing in x_1 to the left of $sN - \epsilon sN - \epsilon N(1-2s)$ for all values of P. For $x_1 \in [sN - \epsilon sN - \epsilon N(1-2S), sN - \epsilon sN]$, $E(C)$ is declining in x_1 for $P < 1/2(C_1 + C_2)/(C_1 - C_2)$, whereas $E(C)$ is increasing in x_1 when the sign of the inequality is reversed. The global minimum is consequently obtained at $x_1^* = sN - \epsilon sN$ for $P < 1/2(C_1 + C_2)/(C_1 - C_2)$ and at $x_1^{**} = sN - \epsilon sN - \epsilon N(1-2s)$ for $P > 1/2(C_1 + C_2)/(C_1 - C_2)$. The supply of professional labor in county 1 falls short of the expected demand sN for all values of P. Without uncertainty, the supply of professionals in county 1 could still fall short of the expected demand, but only if $P > (C_1 + C_2)/(C_1 - C_2)$.[34]

[34]Without uncertainty in labor demand, sN professional and $(1-s)N$ non-professional jobs are available in each county in each period. Let $x_1 \in [0, sN]$ measure the supply of professional workers in county 1. It then follows that $sN - x_1$ professional workers would flow from county 2 to county 1 and $(N - x_1) - (1-s)N$ non-professional workers would flow in the opposite direction at the beginning of each period. Using the same notation as above, the Central Planner chooses x_1 to minimize

$$E(C) = 2sNC_e + (sN - x_1)(C_1 + C_2) + Px_1C_1 + P(2sN - x_1)C_2.$$

It follows that $x_1 = sN$ if $P < (C_1 + C_2)/(C_1 - C_2)$. $x_1 = 0$ and professional labor is under-supplied in county 1 if the sign of the inequality is reversed.

Figure 3.1: Early Growth in the Midwest

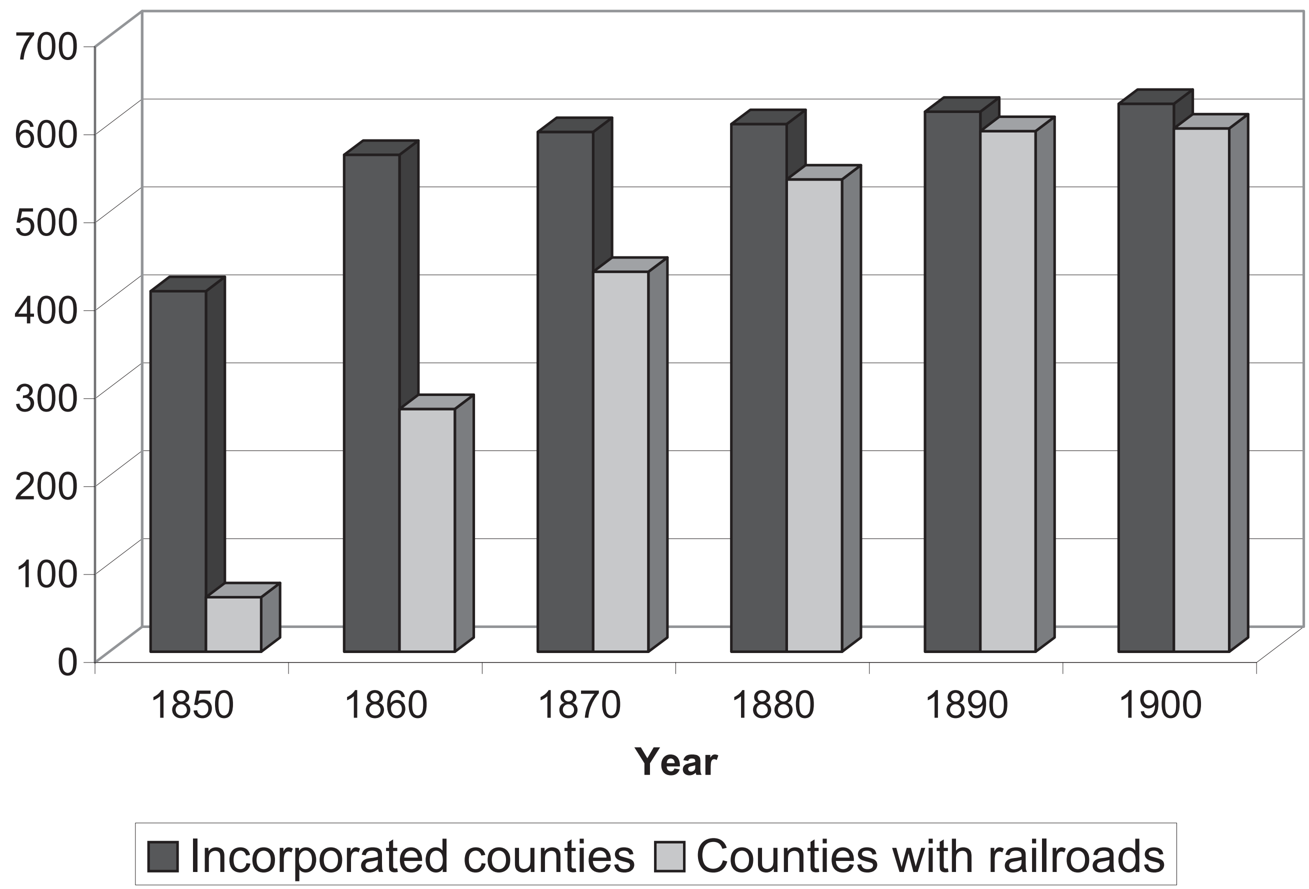


Figure 3 . 2: Ethnic Fractionalization in 1860

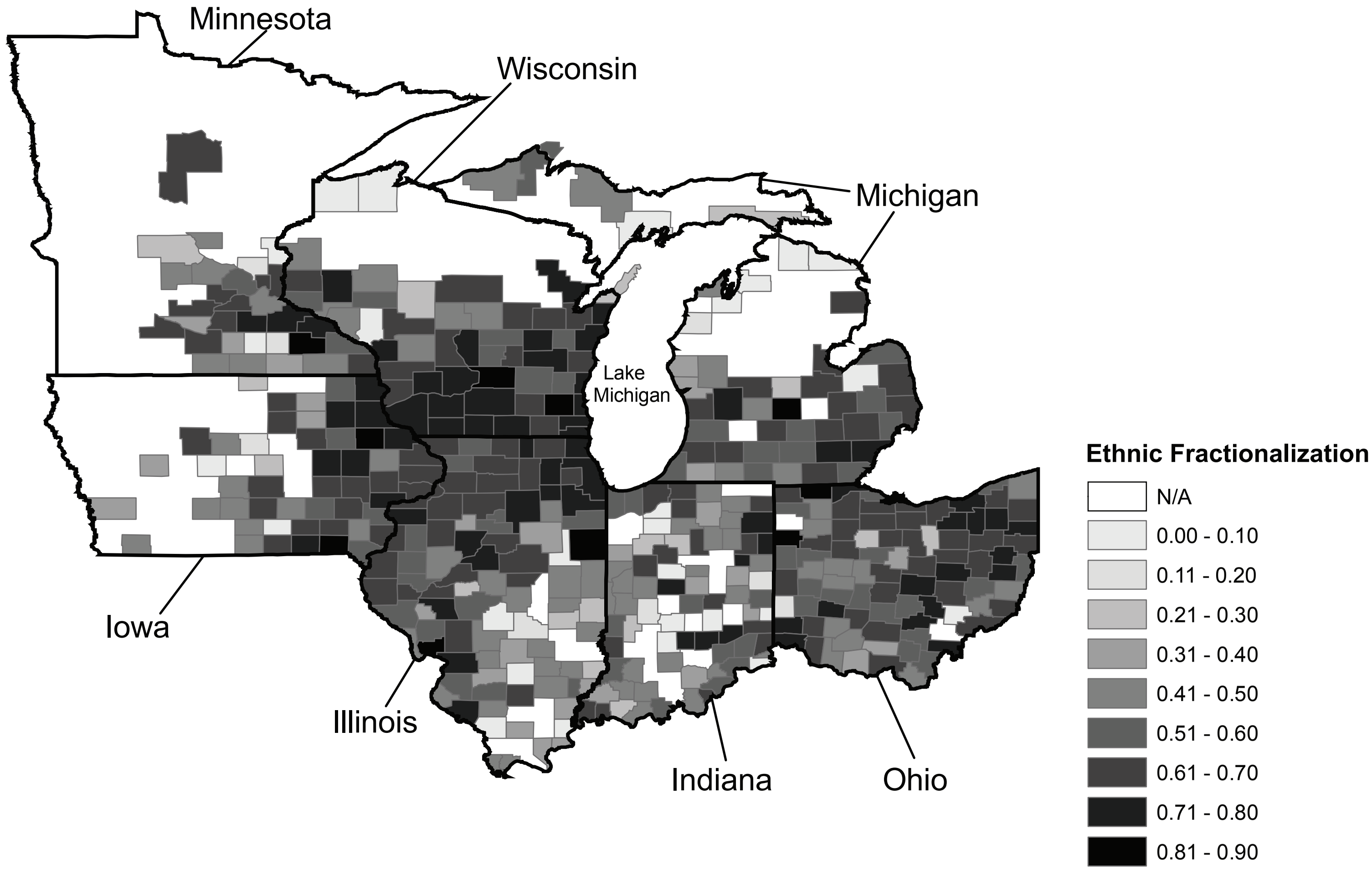


Figure 3.3: Religious Participation

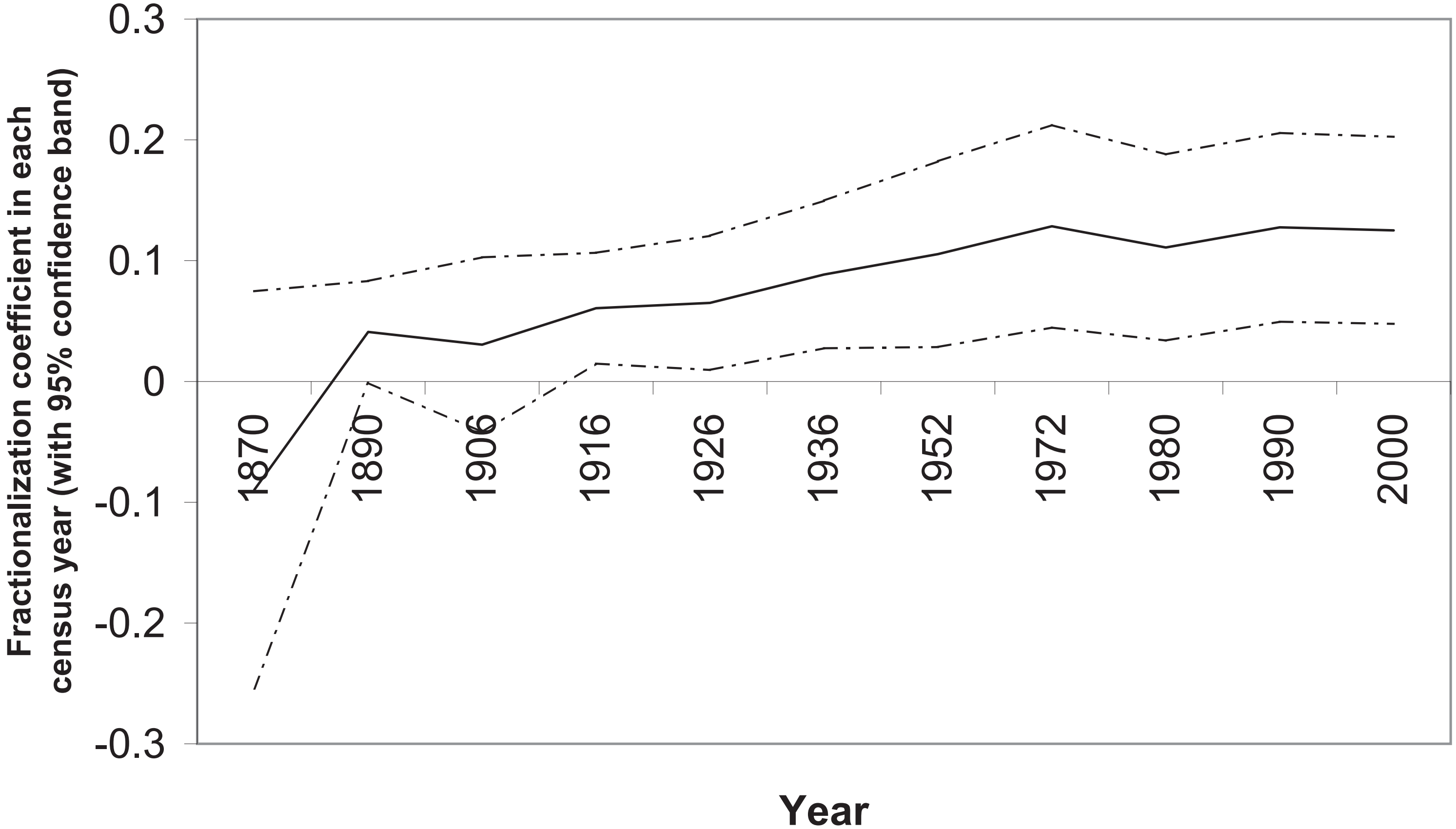


Figure 3.4: Religious Participation by Denomination

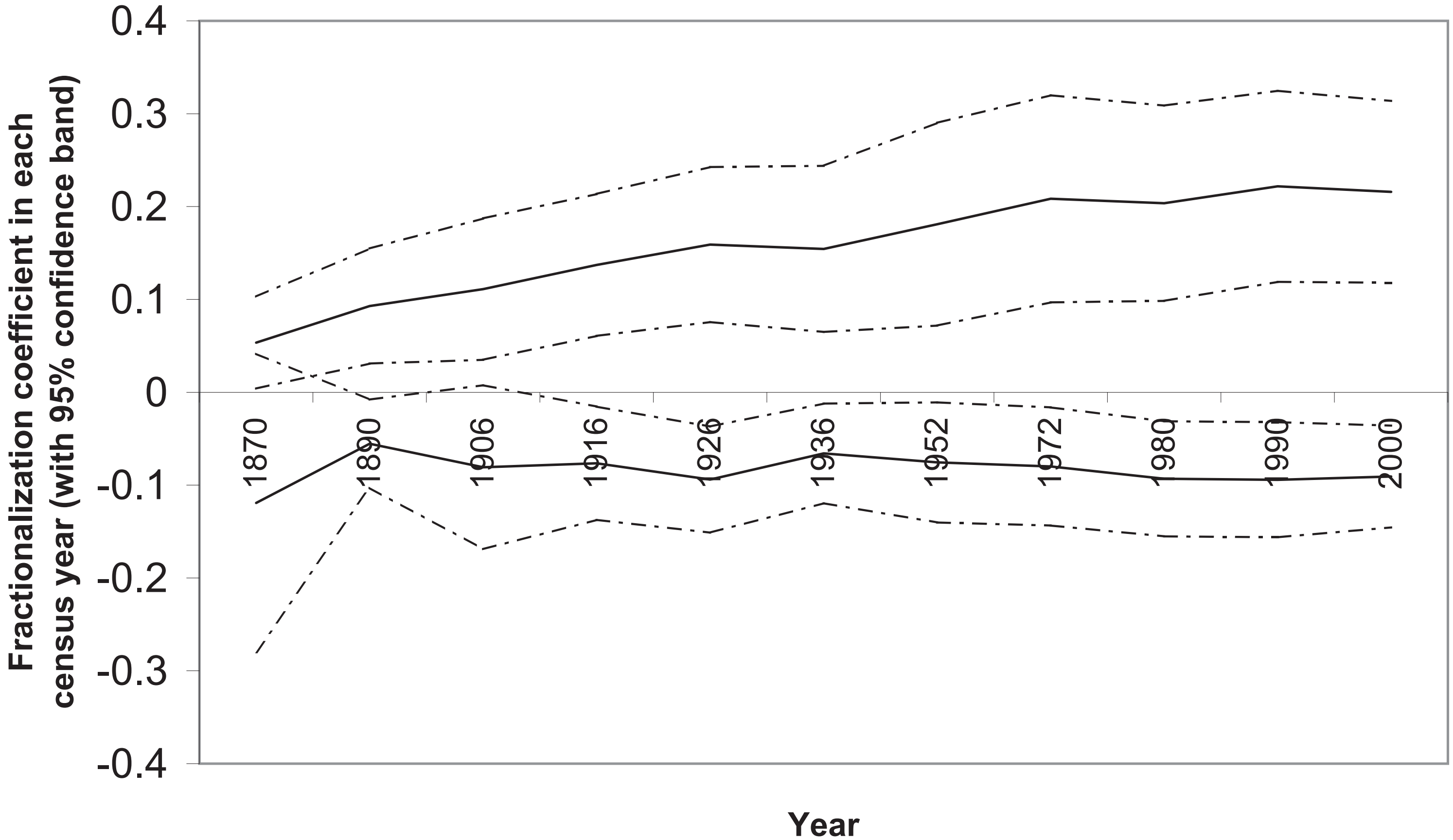


Figure 3.5: The Central Planner's Problem

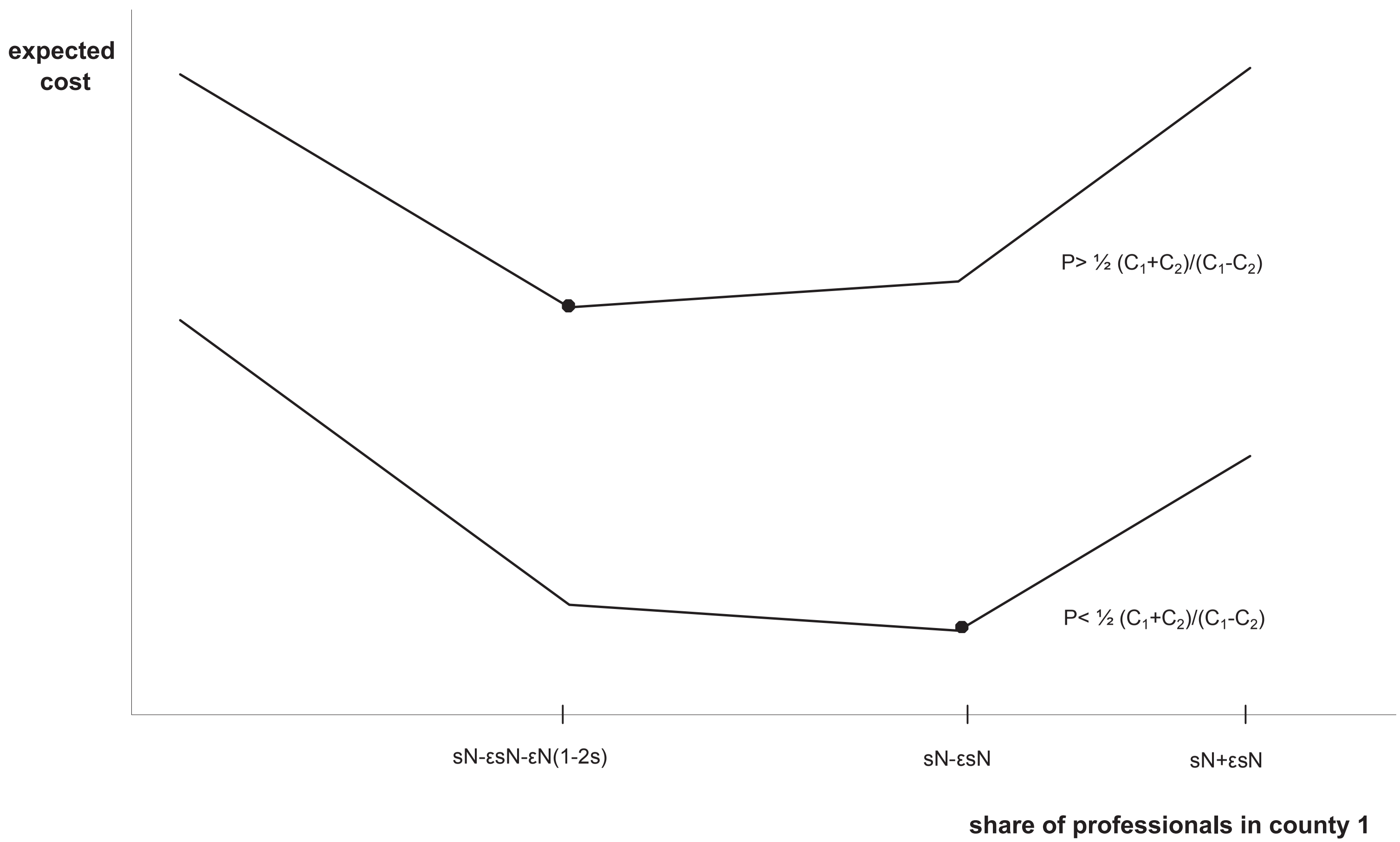


Figure 3.6 : Historical Fractionalization and Current Occupational Choice

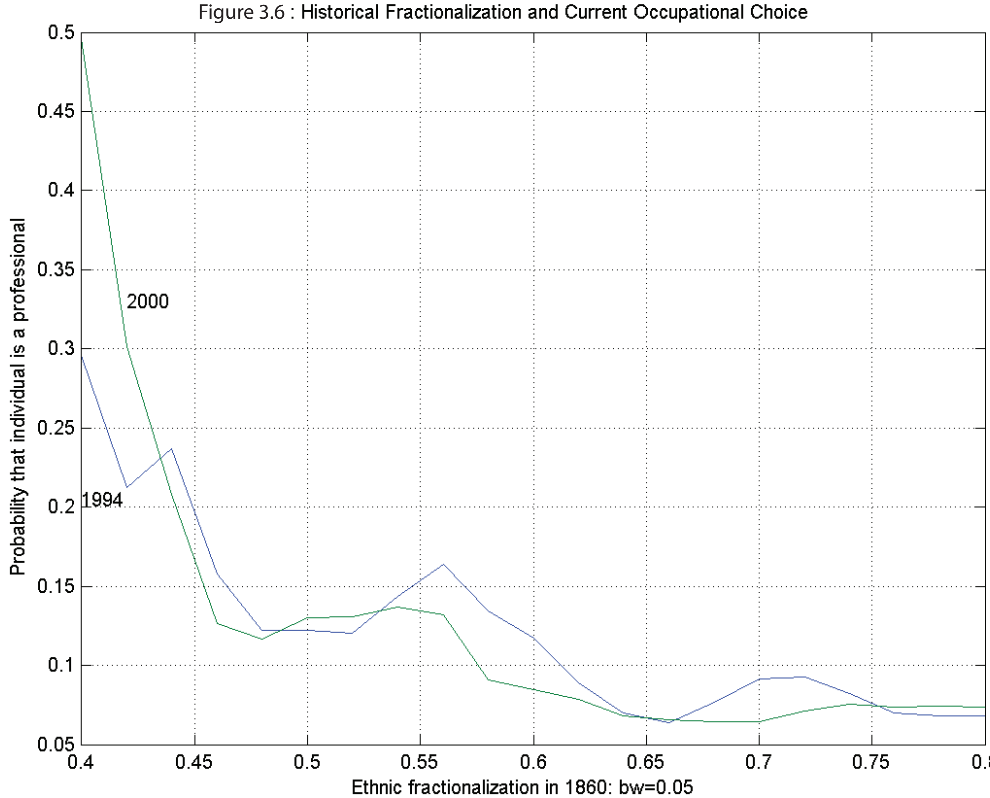


Table 3.1: Ethnic Distribution, 1860-1900

Census year:	1860	1880	1900
	(1)	(2)	(3)
Scandinavia			
Danish	0.01	0.02	0.02
Finish	0.00	0.00	0.01
Norwegian	0.03	0.07	0.07
Swedish	0.02	0.06	0.10
British Isles			
English	0.13	0.11	0.09
Irish	0.25	0.19	0.11
Scottish	0.03	0.03	0.02
Welsh	0.01	0.01	0.01
Western Europe			
Belgian	0.00	0.01	0.01
Dutch	0.01	0.01	0.01
French	0.03	0.02	0.01
German	0.32	0.37	0.41
Italian	0.00	0.00	0.01
Swiss	0.02	0.02	0.02
Eastern Europe			
Czech	0.00	0.01	0.02
Hungarian	0.00	0.00	0.00
Polish	0.00	0.01	0.02
USSR	0.00	0.00	0.01
Other	0.14	0.05	0.04
Total	1.00	1.00	1.00

Source: IPUMS 1:100 sample, including all foreign-born individuals.

Table 3.2: Occupational Distribution, 1860-1900

Census year:	1860	1880	1900
	(1)	(2)	(3)
White collar			
Professional	0.04	0.04	0.05
Manager	0.04	0.04	0.06
Clerical	0.00	0.01	0.02
Sales	0.01	0.02	0.03
Farm			
Farmer	0.50	0.41	0.31
Laborer, Farm	0.12	0.17	0.17
Blue collar, nonfarm			
Craftsman	0.10	0.08	0.09
Operative	0.05	0.08	0.09
Household Service	0.05	0.05	0.05
Service	0.00	0.01	0.02
Laborer, Non-Farm	0.09	0.10	0.12
Total	1.00	1.00	1.00

Source: IPUMS 1:100 sample, including all foreign-born individuals who report that they are employed and report an occupational category.

Table 3.3: Transportation Infrastructure and County Characteristics, 1860

Dependent variable:	ethnic fractionalization	manufacturing share	agriculture share	population
	(1)	(2)	(3)	(4)
Railroad through county, 1860	0.049	0.001	-0.013	0.102
	(0.022)	(0.008)	(0.013)	(0.012)
Distance to canal, 1890	-0.668	0.169	-0.034	-0.469
	(0.191)	(0.097)	(0.125)	(0.102)
Distance to Great Lakes harbor	-0.252	-0.066	0.100	-0.136
	(0.073)	(0.029)	(0.048)	(0.056)
Observations	401	401	401	401

Note: Robust standard errors in parentheses.

Distance to canal and distance to Great Lakes harbor measured in thousands of kilometers.

Fractionalization is one minus the the Herfindahl index of ethnic concentration, averaged across occupational categories.

Manufacturing share and agriculture share in 1860 computed using IPUMS.

Population divided by 100,000.

Table 3.4: Fractionalization in 1860 and County Characteristics in 1990

Year:	1990											
Dependent variable:	agri share	manufac share	pop	ethnic frac	racial frac	religious frac	agri share	manufac share	pop	ethnic frac	racial frac	religious frac
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Fractionalization, 1860	-0.022	0.024	2.590	0.045	0.072	-0.080	0.003	0.018	-0.269	-0.027	0.011	-0.124
	(0.014)	(0.019)	(0.792)	(0.106)	(0.020)	(0.033)	(0.015)	(0.020)	(0.854)	(0.113)	(0.017)	(0.034)
Manufacturing share, 1860	--	--	--	--	--	--	-0.090	-0.056	1.636	0.176	0.084	-0.257
							(0.038)	(0.053)	(1.063)	(0.309)	(0.055)	(0.110)
Agriculture share, 1860	--	--	--	--	--	--	0.136	0.057	-0.962	0.049	-0.080	-0.119
							(0.027)	(0.037)	(0.746)	(0.225)	(0.030)	(0.074)
Population, 1860	--	--	--	--	--	--	-0.087	0.030	11.716	0.326	0.246	0.124
							(0.031)	(0.026)	(5.372)	(0.110)	(0.052)	(0.076)
Observations	437	437	437	437	437	437	437	437	437	437	437	437

Note: Robust standard errors in parentheses.
Fractionalization is one minus the Herfindahl index of ethnic concentration, averaged across occupational categories.
Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.
Population is divided by 100,000.
Manufacturing share in 1990 defined as share of civilian labor force employed in manufacturing.
Agriculture share in 1990 is computed using farm population and total population in county.
Ethnic fractionalization in 1990 is one minus the Herfindahl index of (white) ethnic concentration based on 16 ethnicities.
Racial fractionalization in 1990 is one minus the Herfindahl index of racial concentration based on 5 racial groups.
Religious fractionalization in 1990 is one minus the Herfindahl index of religious concentration based on 18 denominations.

Table 3.5: Fractionalization in 1860 and Local Government Expenditure in 1990

Year:	1990					
Dependent variable:	total expenditure per capita	education share	health share	police share	roads share	welfare share
	(1)	(2)	(3)	(4)	(5)	(6)
Fractionalization, 1860	0.143	-0.055	-0.011	0.014	0.019	0.006
	(0.145)	(0.047)	(0.026)	(0.004)	(0.012)	(0.009)
Manufacturing share, 1860	0.432	0.065	-0.019	0.004	-0.050	-0.034
	(0.469)	(0.155)	(0.060)	(0.014)	(0.030)	(0.029)
Agriculture share, 1860	-0.042	0.020	0.047	-0.011	-0.0001	-0.025
	(0.410)	(0.106)	(0.048)	(0.009)	(0.026)	(0.031)
Population, 1860	0.168	-0.167	-0.005	0.026	-0.046	0.000
	(0.160)	(0.047)	(0.016)	(0.007)	(0.020)	(0.010)
Observations	437	437	437	437	437	437

Note: Robust standard errors in parentheses.
Fractionalization is one minus the Herfindahl index of ethnic concentration, averaged across occupational categories.
Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.
Population is divided by 100,000 and population density is measured in thousands per square mile.
Total expenditure per capita is measured in thousands of dollars (1990).
Shares in columns (2) through (6) are computed as fraction of total expenditure.

Table 3.6: Distribution of Denominations and Religious Participation, 1860-2000

	religious participation					
Measure of religious participation:	church seats		church members		church adherents	
Census year:	1860	1890	1890	1952	1972	2000
	(1)	(2)	(3)	(4)	(5)	(6)
Distribution of denominations						
Baptist	0.14	0.11	0.08	0.06	0.06	0.09
Catholic	0.11	0.13	0.29	0.27	0.30	0.33
Lutheran	0.05	0.12	0.14	0.22	0.21	0.20
Methodist	0.38	0.27	0.21	0.18	0.19	0.13
Presbyterian	0.14	0.08	0.06	0.05	0.05	0.03
Other	0.18	0.28	0.22	0.22	0.19	0.22
Total	1.00	1.00	1.00	1.00	1.00	1.00
Proportion religious	0.59	0.67	0.30	0.50	0.58	0.54

Source: Census of Religious Bodies.

Proportion religious is computed as the number of church seats divided by the population 1860-1890, the number of church members divided by the population 1890-1952, and the number of church adherents divided by the population 1972-2000.

Table 3.7: Fractionalization in 1860 and Religious Participation in NLSY79

Dependent variable:	share of population raised in church denomination		share of population participating in church denomination			
Year:	--		1979		2000	
Church denomination:	migrant	non-migrant	migrant	non-migrant	migrant	non-migrant
	(1)	(2)	(3)	(4)	(5)	(6)
Fractionalization, 1860	0.205	-0.235	0.292	-0.124	0.210	-0.116
	(0.160)	(0.159)	(0.155)	(0.161)	(0.112)	(0.129)
Manufacturing share, 1860	0.456	-0.373	0.637	-0.537	0.226	0.138
	(0.359)	(0.360)	(0.348)	(0.378)	(0.302)	(0.370)
Agriculture share, 1860	0.292	-0.244	0.195	-0.377	0.153	-0.088
	(0.313)	(0.304)	(0.321)	(0.291)	(0.227)	(0.247)
Population, 1860	-0.276	0.273	-0.314	0.198	-0.090	0.133
	(0.135)	(0.126)	(0.136)	(0.134)	(0.087)	(0.089)
Observations	222	222	222	222	222	222

Note: Standard errors in parentheses.
Fractionalization is one minus the Herfindahl index of ethnic concentration, averaged across occupational categories.
Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.
Population divided by 100,000.
Migrant denominations are Catholic and Lutheran.
Partipation defined as attend church at least "several times" in year prior to interview date.

Table 3.8: NLSY79 Descriptive Statistics

	year	
	1994	2000
	(1)	(2)
Professional	0.10	0.10
	(0.01)	(0.01)
Migrated out of county of birth	0.56	0.59
	(0.01)	(0.01)
Employed	0.81	0.92
	(0.01)	(0.01)
Income	27.80	33.06
	(0.51)	(0.58)
White	0.79	0.79
	(0.01)	(0.01)
Female	0.50	0.51
	(0.01)	(0.01)
Age	33.36	39.38
	(0.05)	(0.05)

Note: Standard errors in parentheses.
Professional occupations are relevant codes in the Professional, Technical and Kindred Workers category.
All variables except income and age are binary.
Income is measured in thousands of dollars (2000).

Table 3.9: Fractionalization in the County of Birth and Occupational Choice

Dependent variable:	professional			
Year:	1994	2000	1994	2000
	(1)	(2)	(3)	(4)
Fractionalization, 1860	-0.146	-0.163	-0.253	-0.316
	(0.083)	(0.088)	(0.096)	(0.109)
White	0.031	0.045	0.047	0.059
	(0.022)	(0.019)	(0.020)	(0.020)
Female	-0.065	-0.023	-0.064	-0.021
	(0.020)	(0.017)	(0.020)	(0.016)
Age	0.0003	0.003	0.001	0.003
	(0.004)	(0.004)	(0.004)	(0.004)
Manufacturing share, 1860	--	--	-0.213	-0.521
			(0.172)	(0.213)
Agriculture share, 1860	--	--	-0.272	-0.431
			(0.144)	(0.160)
Population, 1860	--	--	0.018	0.011
			(0.017)	(0.025)
Observations	1209	1122	1209	1122

Note: Standard errors in parentheses are clustered at the county level.
Fractionalization is one minus the Herfindahl index of ethnic concentration, averaged across occupational categories.
White, female, and age are individual-level characteristics.
Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.
Population divided by 100,000.
Professional is a binary variable indicating whether the individual is employed in a professional occupation.
Professional occupations are relevant codes in the Professional, Technical, and Kindred Workers category.

Table 3.10: Fractionalization and Outcomes Related to Occupational Choice

Dependent variable:	migrated		employed		income	
Year:	1994	2000	1994	2000	1994	2000
	(1)	(2)	(3)	(4)	(5)	(6)
Fractionalization, 1860	-0.388	-0.414	0.114	0.037	1.659	-11.722
	(0.155)	(0.138)	(0.108)	(0.065)	(4.804)	(6.183)
Manufacturing share, 1860	-0.071	-0.242	0.078	0.362	-1.195	-26.379
	(0.305)	(0.260)	(0.230)	(0.136)	(10.268)	(12.231)
Agriculture share, 1860	-0.171	-0.317	0.031	0.253	-8.197	-21.016
	(0.311)	(0.222)	(0.164)	(0.107)	(7.808)	(9.359)
Population, 1860	0.021	0.065	-0.015	0.013	-0.210	0.449
	(0.039)	(0.040)	(0.026)	(0.015)	(1.041)	(1.691)
White	0.232	0.284	0.153	0.043	5.773	4.791
	(0.048)	(0.052)	(0.024)	(0.021)	(1.850)	(1.567)
Female	-0.014	-0.014	-0.155	-0.059	-11.986	-14.476
	(0.017)	(0.024)	(0.021)	(0.015)	(1.122)	(0.988)
Age	0.013	0.015	0.001	0.003	0.764	0.211
	(0.005)	(0.005)	(0.004)	(0.004)	(0.208)	(0.312)
Observations	1598	1437	1614	1332	1251	1122

Note: Standard errors in parentheses are clustered at the county level.
Fractionalization is one minus the the Herfindahl index of ethnic concentration, averaged across occupational categories.
Fractionalization is measured in the county of birth.
Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.
Population is divided by 100,000.
White, female, and age are individual-level characteristics.
Migrated is a binary variable that indicates whether the individual resides outside the county of birth.
Employed is a binary variable that indicates whether the individual currently holds a job.
Income is measured in thousands of dollars (2000).

Table 3.11: Fractionalization in the County of Residence and Occupational Choice

Dependent variable:	professional			
Year:	1994	2000	1994	2000
	(1)	(2)	(3)	(4)
Fractionalization, 1860	0.080	0.088	-0.053	0.067
	(0.073)	(0.070)	(0.078)	(0.075)
White	0.073	0.066	0.090	0.063
	(0.016)	(0.019)	(0.015)	(0.020)
Female	-0.062	-0.031	-0.062	-0.031
	(0.016)	(0.017)	(0.016)	(0.017)
Age	-0.0041	0.002	-0.004	0.002
	(0.004)	(0.004)	(0.004)	(0.004)
Manufacturing share, 1860	--	--	-0.321	-0.210
			(0.136)	(0.188)
Agriculture share, 1860	--	--	-0.183	-0.075
			(0.111)	(0.111)
Population, 1860	--	--	0.055	-0.001
			(0.016)	(0.017)
Observations	1205	1130	1205	1130

Note: Standard errors in parentheses are clustered at the county level.
Fractionalization is one minus the Herfindahl index of ethnic concentration, averaged across occupational categories.
White, female, and age are individual-level characteristics.
Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.
Population divided by 100,000.
Professional is a binary variable indicating whether the individual is employed in a professional occupation.
Professional occupations are relevant codes in the Professional, Technical, and Kindred Workers category.

Table 3.12: Fractionalization and Educational Attainment

Dependent variable:	AFQT score	high school completion	college completion	non-professional major
	(1)	(2)	(3)	(4)
Fractionalization, 1860	-0.535	0.016	-0.095	0.207
	(7.457)	(0.100)	(0.104)	(0.107)
Manufacturing share, 1860	11.570	0.197	0.003	0.306
	(17.828)	(0.160)	(0.241)	(0.224)
Agriculture share, 1860	-3.015	-0.017	-0.261	0.327
	(14.672)	(0.143)	(0.193)	(0.229)
Population, 1860	-2.769	-0.052	-0.018	0.017
	(1.981)	(0.023)	(0.023)	(0.040)
White	22.608	0.010	0.083	-0.029
	(1.626)	(0.024)	(0.018)	(0.048)
Female	-2.728	0.000	-0.039	0.283
	(1.248)	(0.017)	(0.020)	(0.022)
Age	2.880	0.006	0.005	-0.002
	(0.332)	(0.003)	(0.004)	(0.005)
Observations	2187	2023	2023	1239

Note: Standard errors in parentheses are clustered at the county level.
Fractionalization is one minus the Herfindahl index of ethnic concentration, averaged across occupational categories.
Fractionalization is measured in the county of birth.
Manufacturing share in 1860 and agriculture share in 1860 are computed using IPUMS.
Population is divided by 100,000.
White, female, and age are individual-level characteristics.
AFQT is the score on the Armed Forces Qualification Test.
High school completion is a binary variable indicating whether the individual completed high school, including GED.
College completion is a binary variable indicating whether the individual completed a four-year college/university degree.
Non-professional major is a binary variable indicating whether the individual's college major is associated with a non-professional occupation.