Three Essays in Labor Economics

by

Todd Andrew Sorensen

A Dissertation Submitted to the Faculty of the

Department of Economics

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

In the Graduate College

The University of Arizona

2 0 0 7

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation

prepared by Todd Andrew Sorensen

entitled Three Essays in Labor Economics

and recommend that it be accepted as fulfilling the dissertation requirement for the

Degree of Doctor of Philosophy

____ Date: 23 April 2007
Price V Fishback

____ Date: 23 April 2007
Ronald L Oaxaca

____ Date: 23 April 2007
Manuela Angelucci

____ Date: 23 April 2007
Gregory S. Crawford

____ Date: 23 April 2007
Alfonso Flores-Lagunes

____ Date: 23 April 2007
Keisuke Hirano

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: 23 April 2007
Dissertation Director: Price V Fishback

____ Date: 23 April 2007
Dissertation Director: Ronald L Oaxaca

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Signed: Todd Andrew Sorensen

Acknowledgments

First and foremost, I would like to thank my two main advisors, Dr. Price V. Fishback and Dr. Ronald L. Oaxaca. Their guidance and constant support has been crucial in helping me develop the skills needed to succeed in the PhD program and in life in general.

Thanks to the other members of my committee as well. Dr. Keisuke Hirano and Dr. Alfonso Flores-Lagunes have been invaluable in giving me feedback on the econometric models in my dissertation. Dr. Manuela Angelucci's graduate labor course introduced me to much of the literature that taught me how to ask and answer interesting questions in labor economics. Dr. Greg Crawford has been valuable in helping me with the IO models used in two chapters of this dissertation. I would also like to thank Dr. Daniel Ackerberg, who during his two years on the faculty here successfully instilled in his students a very unique and invaluable intuitive understanding of applied econometrics. Dr. John Drabicki has greatly helped me develop my skills as a teacher.

I've benefited tremendously from my interactions with the other graduates students in the program. I thank Carmen Carrión-Flores, Wafa Hakim, Quasi Shariar and Hironori Otsubo, the other students in my cohort, for their help and support. It's been quite the half decade we've spent together. Carmen's constant support during and since the first year has been indispensable

Tracy Regan, Simona Lup Tick and Todd Neumann were all both peers and mentors. The enthusiasm of Adrian Stoian, Chung Choe, Sonam Gupta, Eylem Gervek and Dina Shatnawi, the other labor economists in the program, has served as a constant inspiration. It has been a privilege to have been a sounding board both for and to have had Alex Shcherbokov as a sounding board; I have collectively learned more from our conversations over the last three years than from any formal course.

The many hours of work I have put into my dissertation have been far more enjoyable on account of Doug Martsch, Joe Strummer, Mick Jones, Ben Gibbard, Thom Yorke, Lou Reed and Rivers Cuomo. Thanks for your work.

Thanks to the department chair, Dr. Mark Walker, the department staff, Lana Sooter, Mary Flannery, and Carole Merly, and the department as a whole for all their help during my time here.

A very special thanks to my parents for their support during my time in graduate school, and for everything they taught me while I was growing up that enabled me to make it to graduate school. They also taught me to value education and intellectual curiosity. This has allowed me to enjoy my last twenty one years of school.

Table of Contents

Table of Contents—*Continued*

List of Figures

List of Tables

Abstract

This dissertation consists of three essays in labor economics. The first essay models how migrants crossing the border between the United States and Mexico respond to increases in border enforcement. We model a potential migrants' joint decision of whether to cross the border and, if so, where to cross the border using a random utility function. Our model allows us to calculate the migrants' substitution patterns: does more enforcement primarily on one part of the border primarily deter migrants from crossing the border altogether, or simply divert them to other parts of the border? We find that a substantial proportion of migrants are indeed diverted. These findings should serve as a caveat to policy makers who seek to address immigration reform issues primarily through tightening the border.

The second chapter models the internal migration decisions of U.S. households during the period 935 to 1940. We measure the impact of spending on New Deal programs on migration patterns. Using a model of random utility similar to that in prior chapter, we find that more public works and relief spending in a region made it more attractive to potential migrants, while additional spending on the Agricultural Adjustment Administration (AAA) made the locale less attractive. The structural nature of our model allows us to compute counterfactual estimates to assess the overall impact of these programs. We find that regional disparities in spending on public works and relief programs we responsible for nearly 20% of long distance moves made between regions during this period.

In the third chapter, we decompose the gap between mean sentences for males and females in the U.S. criminal justice system into the portion that can be explained by differences in the average severity of the crime committed by males and females and the portion explained by differences in how males and females who commit the same crime are treated. We find that differences in characteristics of the defendant can

explain only half of the gap between mean male and females sentences, suggesting that women receive more lenient treatment in the U.S. criminal justice system.

Chapter 1
Introduction

My dissertation consists of three papers in labor economics. The first two chapters involve questions related to the location choice of a migrant, while the third seeks to explain the difference in male and female outcomes in the federal criminal justice system by using decomposition analysis.

Over the last two years, immigration reform has become one of the most prominent political debates in Congress. Many involved in this debate propose to stem the flow of illegal immigration into the United States by significantly increasing the intensity of enforcement along the U.S.-Mexican border, while others believe that increases in enforcement along any particular stretch of the border will do nothing more than to shift the crossing paths of migrants into another less tightly patrolled sector of the border.

In this chapter, we use a discrete choice model to estimate how migrants respond to an increase in border enforcement along one sector of the border. We are able to decompose their responses into deterrence (the intended consequence: migrants responding by staying home) and diversion (an unintended consequence: migrants responding by crossing elsewhere).

There is anecdotal evidence of a strong diversion effect: the major increases in border enforcement around San Diego and El Paso during the 1990s were followed by major increases in crossings in the Tucson sector of the border. A heavy increase in security also occurred around major ports of entry following the attacks of September 11th, 2001. In the succeeding summers, the numbers of migrants dying in isolated (and less heavily patrolled) stretches of the Arizona desert spiked.

The results of this decomposition have very clear policy implications. If the effect of increased enforcement is primarily deterrence, a policy that focuses on putting more border enforcement resources into the most heavily crossed sectors of the border will be effective. However, if there is a significant diversion effect, many illegal immigrants will simply change where they cross, and local level estimates of the effectiveness of enforcement will be misleading.

The effect of changes in border enforcement on the decisions of migrants is an understudied area in the economics literature. While others have used time-series variation to establish that migrants do indeed cross less frequently when enforcement increases, we are the first to use an econometric model to address the diversion vs. deterrence issue.

One important issue for all econometric studies in this area is to address the potential endogeneity between crossing and enforcement. The instruments we develop for use in this paper are based on a "political supply" approach. For example, we use the number of House of Representatives districts overlapping with each border sector as an instrument: while the size of the delegation representing the border sector should not directly affect the migrant's utility of crossing in that sector, the increased levels of enforcement which result from a larger political lobby should.

In addition to our instrumental variables approach, our panel data allows us to control for the time invariant characteristics of border crossings. Prior studies have been unable to control for these effects. We also face the challenge of implementing an instrumental variables estimator in the context of a discrete choice model. Our estimation strategy comes from the Petrin and Train (2002) control function approach.

After addressing these estimation issues, were are able to obtain consistent estimates of the parameters of our structural model, allowing us to compute meaningful counterfactuals. Our primary finding is that there is indeed a strong diversion effect, and that the magnitude of the deterrence effect of border enforcement generally very small. This result should serve as a caveat to policy makes which wish to address illegal immigration by increasing the intensity of enforcement along the parts of the border which see the heaviest traffic.

The second paper develops an econometric model of internal migration during the period of 1935 to 1940. This chapter estimates how individual migrants valued local level spending on New Deal grants when deciding where to locate. We find that New Deal spending public works and relief grants increased net migration to a given locale. On the other hand, spending on the Agricultural Adjustment Administration decreased net migration, presumably because taking land out of production decreased local labor demand. This finding is consistent with an earlier study by Fishback, Horrace, and Kantor (2006), but the nature of our model allows us to address a more broad set of questions than the prior paper.

We decompose the effects of program spending on migration into three categories: the effect of spending on keeping households in their origin (retention), the effect of pulling non-migrants out of their origin (creation), and the effect of causing migrants to substitute away from an alternative destination (diversion). Counterfactual estimation indicate that Agricultural Adjustment Act spending had a small overall impact on rates of mobility in the United States, while spending on public works and relief grants was responsible for a significant proportion of long distance moves in the United States during these 5 years.

As in the previous chapter, a household's choice of where to locate is based on a random utility model. By observing households "voting with their feet", we can identify whether program spending made households more or less likely to locate in a given destination.

My third chapter addresses an important question in the law and economics literature. It has been established that women receive lighter prison sentences than their male counterparts. However, few econometric studies have been undertaken to determine whether this difference can be attributed to differences in the characteristics (such as severity of the offense committed) of male and female convicts, or whether this difference is indeed due to females being treated more leniently. After carefully modeling the sentence determination process, we apply a generalized decomposition methodology, finding that nearly half of the gap in sentence length can be attributed towards leniency to women.

In attempting to model the data generating process, we encountered a number of econometric issues that are not typically addressed in the literature. Thus, we believe that our model provides a more valid estimate of the causes of the gender sentence differential than previous work.

The first issue we incorporate into our model is the institution of plea bargaining. Clearly individuals will only plead guilty to a charge if they believe they will receive a lighter sentence. Accordingly, separate regressions for all individuals who were sentenced after pleading guilty (the "plea regime") and individuals who were sentenced after being convicted in a trial (the "trial regime") would address this issue, given the assumption of exogenous selection into regimes. However individuals most likely self-select into the regime which will minimize their expected sentence length. To take account of this, we estimate a mover stayer-model using a full information maximum likelihood estimator.

One additional estimation issue that is addressed in our paper is the large proportion of sentences which involve no prison time. We address this problem by treating sentences involving no prison time as limit observations, adding a Tobit like element to each regime of our model.

After estimating the parameters of our model, we are able to decompose sentence differentials. Because we do not possess sufficient data on female convicts to estimate this rather cumbersome model separately for women, we apply an insight from Oaxaca and Ransom (1994) to identify each term in the decomposition while only estimating the model for male convicts. Finally, we make a methodological contribution to the decomposition literature by addressing the nature of estimation error in a non-linear model, which makes the assumption that the fitted value at the average value of the independent variables will equal the mean of the dependent variable.

Chapter 2

The Effects of Border Enforcement on Migrants' Border Crossing Choices: Diversion or Deterrence?

2.1 Introduction

[1]Illegal immigration and immigration reform have recently risen to the forefront of public policy debate. Current estimates put the stock of unauthorized immigrants in the United States at 10.3 million, 5.9 million from Mexico alone. Proponents of a tougher policy towards illegal immigration favor stronger enforcement of the U.S.-Mexican border to deter further illegal entry into the United States. A recently signed bill authorizes 700 more miles of fencing along the 2000 mile wide U.S.-Mexico border. The President also plans to add 6,000 agents to the U.S. Border Patrol (Arizona Daily Star, September 26th 2006). Current and proposed measures involve spending substantial amounts of taxpayer resources. While the US government currently spends $2.2 billion annually on border enforcement, the construction of the proposed physical barriers is estimated to cost $2 billion to $5 billion (Hanson, 2006).

Accordingly, any change in policy should be accompanied by a better understanding of how migrants react to increases in border enforcement. In this paper we add to a small but growing literature on border enforcement by addressing a crucial policy question: does increasing enforcement along one part of the border deter unauthorized crossers from coming at all, as intended, or merely divert these crossers to other parts of the border?

There are two major consequences of ignoring the diversion effect. From a policy point of view the existence of a strong diversion effect would suggest that border enforcement is not as effective as it may appear at first. To achieve desired levels

[1]This is joint work with Carmen Carrión-Flores

of deterrence at a national level, enforcement would need to be strengthened across the board and not just in the places that currently experience high rates of illegal crossings. During the 1990's the U.S. Border Patrol significantly increased the intensity of enforcement in urban crossing zones, starting with El Paso and San Diego. While these operations were deemed successful at the local level, it is estimated that a ten-fold increase in the number of border patrol agents would be needed in order to apply the same intensity of enforcement across the entire border (Arizona Daily Star Sep 26th 2006). Clearly, when policy makers conduct cost-benefit analysis of border security measures, the magnitude of the diversion effect will have a significant impact on the costs that will need to be incurred to achieve given border security goals.

The El Paso and San Diego operations ("Operation Hold the Line" and "Operation Gatekeeper", respectively) were followed by a large spike in the number of migrant deaths in the isolated desert stretches of the border. In his best selling book *The Devil's Highway*, Luis Alberto Urrea documents the attempted crossing of 26 migrants from southern Mexico, 14 of whom died in the summer heat of the Arizona desert. While such large groups of deaths remain relatively rare, it is well established that migrant crossing deaths have increased significantly in the years following Gatekeeper and Hold the Line. A report by the General Accounting Office documents that deaths along the border doubled between 1995 and 2005 (GAO, 2006). In the presence of a strong diversion effect, these deaths can be viewed as an unintended consequence of increased enforcement.

Popular media and anecdotal evidence frequently suggest that there is indeed a strong diversion effect, of that this passage from an AP article is typical:

> "It's known as the water-balloon effect: Squeeze one spot and illegal immigration will bulge elsewhere along the 1952 mile frontier......A crackdown launched in 1994 [in San Diego, CA] and modeled on a similar effort in El Paso, Texas, pushed many migrants away from the border's

> two largest cities and into Arizona's mountains and deserts." (Arizona Daily Star August 8th 2006)

A few previous studies have used time-series variation to identify an overall migration response to tighter border enforcement. Aside from Angelucci (2004), we are the first to use panel data to evaluate how migrants respond to enforcement. This will allow us to provide the first econometric estimates of the magnitude of the diversion effect. The panel nature of our data also allows us to control for the effects of time invariant characteristics of border crossings on migrations decisions, as well as the effects of border wide changes in government policy over time. We also depart from the majority of the previous literature in our use of individual level data on migration decisions, rather than relying upon data on the number of apprehensions made by the Border Patrol to proxy for the number of crossings taking place.

To obtain consistent estimates of effect of enforcement on migrants' decisions, we employe an instrumental variables estimator that involves both cross-sectional and time series variation. Using these instruments and Mexican Migration Project (MMP) micro level data on the crossing choice of illegal immigrants, we estimate a discrete choice model. Our estimates from this model then allow us to decompose the effect of a marginal increase in enforcement into a deterrence effect and a diversion effect. We then evaluate the effects of past border patrol operations. While we find that more border enforcement in a given sector does indeed deter migrants, it also diverts a significant number of migrants to other sectors. These findings suggest that estimates of the effect of border enforcement on migration in a particular geographic area that ignores the diversion effect may overstate the impact of enforcement on migrations into the United States.

2.2 Literature Review

A number of recent papers have sought to estimate effects of border enforcement rates on border crossings. Hanson (2006) provides a comprehensive review of the literature on illegal migration from Mexico to the United States. Hanson and Spilimbergo (1999) was one of the first papers to look at the estimates of the effects of border enforcement on apprehensions. Since enforcement is endogenous to apprehensions, they use instrumental variables for enforcement using U.S. government expenditures on national defense and the timing of U.S. presidential, gubernatorial and Senate elections. OLS estimates will be biased upwards, but IV estimation shows a strong positive causal relationship between the number of migrants apprehended by the Border Patrol and the number of labor hours used by the agency to catch those crossing illegally across the border.

Hanson and Spilimbergo (2001) model the importance of political lobbying on border enforcement. They found that price shocks to sectors employing large numbers of undocumented workers are negatively correlated with the level of border enforcement. These results are evidence that employers lobby behind the scenes to weaken enforcement when prices and wages rise. Hanson, Robertson and Spilimbergo (2002), meanwhile, find that increased border enforcement has little impact on labor markets on either the United States or Mexican side of the border. This is consistent with two hypotheses: either enforcement has little impact on migration levels, or immigration has little impact on wages.

Using data on migrant deaths, Cornelius (2001) found that border enforcement has rechanneled flows of unauthorized migrants towards more hazardous areas and discouraged unauthorized migrants already in the United States from returning to their places of origin. However, Cornelius did not find evidence that the strategy is deterring or preventing significant numbers of new entries, particularly given the absence of a serious effort to curtail employment of unauthorized migrants through work site enforcement.

Orrenius and Zavodny (2003) examined whether mass legalization programs reduce future undocumented immigration. They found that directly after the 1986 Immigration Reform and Control Act, the apprehensions of persons attempting to cross the U.S.-Mexico border illegally declined, but the amnesty program did not change long-term patterns of undocumented immigration from Mexico.

Bean et al (1994) studied the effect of "Operation Hold the Line" in the Border Patrol's El Paso sector. This operation marked a sharp change in the enforcement strategy, shifting the focus of enforcement from internal checkpoints to line watch. They considered this change to be an exogenous shock to the level of border enforcement. In contrast to other findings, the increase in enforcement intensity reduced the number of apprehensions. However, they also found evidence that half of the decrease in the flow of migrants, as measured by the level of apprehensions, was offset by an increase in flows to other border sectors.

Carrión-Flores (2005) found that increased border enforcement may increase the trip duration of migrants engaging in repeat migrations. Angelucci (2004) develops a model of return migration, and estimated the impacts of increased border enforcement on both the probability of a migrant undertaking a trip from Mexico to the United States, and of migrants already living in the United States returning to Mexico. She found that while increased border enforcement discourages migrants from crossing into the United States, it may discourage the return to Mexico of migrants already in the United States, thus increasing trip duration.

2.3 Model

Consistent with Sjaastad (1962), our model of migration represents that of a utility maximizing individual. A potential migrant living in Mexico chooses between migrating illegally to the United States or remaining in Mexico. Should he choose to migrate illegally, he will also have to decide through which of the nine US Border Patrol sectors to cross. The utility to the i^{th} individual remaining at home (choice o) in origin l during time period t is normalized to 0

$$U_{iotl} = 0 \text{ for } j = 0 \tag{2.1}$$

The utility to the individual of crossing into the United States through Border Patrol sector j is

$$U_{ijtl} = \alpha \cdot Enf_{jt} + X_{ijtl}\beta + W_{itl}\gamma + \xi_{jtl} + \rho \cdot \zeta_{itl} + (1-\rho) \cdot \varepsilon_{ijtl} \ \text{ for } j > 0 \tag{2.2}$$

The Enf_{jt} variable is a measure of the intensity of border enforcement at crossing j during year t. The term α is the marginal disutility to the migrant of this enforcement. The expression $X_{jtl}\beta$ represents the utility of the choice attributable to the observable characteristics of the crossing choice, X_{jtl}, and the marginal utilities to those characteristics, β. The term $W_{itl}\gamma$ represent the utility of the choice attributed to observable characteristics constant across choices that involve migrating, W_{itl}, and the marginal utilities to those characteristics, γ. The ξ_{jtl} term is the utility of the choice observed by the migrant but unobserved by the econometrician.

Without the addition of the σ and ζ terms, we would be assuming that random utility draws are $i.i.d$ across j; i.e. a high draw of random utility to crossing through any one sector contains no information about the draws corresponding to crossing through other sectors. Clearly, this is an unreasonable assumption. Crossing through any sector involves accessing U.S. labor markets, and individuals with a strong preference for access to U.S. labor markets (rather than not migrating) through one channel will likely have a strong preference to access these labor markets through another channel.

We relax this assumption with the addition of a random utility term that is constant across all choices that involve migration. This random utility model is based on Cardell's (1997) paper on the existence of a distribution necessary to rationalize McFadden's (1973) model. Let $\upsilon_{ijt} = \rho \cdot \zeta_{itl} + (1-\rho) \cdot \varepsilon_{ijlt}$. The covariance between random utility terms in the set of choices involving migration is now clearly non-zero:

$$E(\upsilon_{ijtl} \cdot \upsilon_{iktl}) = \rho^2 \cdot \sigma_\zeta^2 + (1-\rho^2) \cdot \sigma_\varepsilon^2 \ \ \forall\ j = k \text{ where } j > 0 \tag{2.3}$$

$$E(\upsilon_{ijtl} \cdot \upsilon_{iktl}) = \rho^2 \cdot \sigma_\zeta^2 \ \ \forall\ j \neq k \text{ where } j, k > 0 \tag{2.4}$$

The random utility to not migrating remains uncorrelated with the random utility of choices within the set of choices involving migration.

$$E(\varepsilon_{iotl} \cdot \upsilon_{ijtl}) = 0\ \forall\ j > 0 \tag{2.5}$$

The variance-covariance matrix, $\sum\limits_{JxJ}$, for the random utility terms is:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & \cdot\ \cdot\ \cdot\ \cdot\ \cdot & 0 \\ 0 & \rho^2\sigma_\zeta^2 + (1-\rho)^2\sigma_\varepsilon^2 & \rho^2\sigma_\zeta^2 & \rho^2\sigma_\zeta^2 & \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \rho^2\sigma_\zeta^2 \\ 0 & \rho^2\sigma_\zeta^2 & \rho^2\sigma_\zeta^2 + (1-\rho)^2\sigma_\varepsilon^2 & \rho^2\sigma_\zeta^2 & \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \rho^2\sigma_\zeta^2 \\ 0 & \rho^2\sigma_\zeta^2 & \rho^2\sigma_\zeta^2 & \cdot & & \\ \cdot & \cdot & \cdot & & \ddots & \\ \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & & & \cdot \\ 0 & \rho^2\sigma_\zeta^2 & \rho^2\sigma_\zeta^2 & & & \rho^2\sigma_\zeta^2 + (1-\rho)^2\sigma_\varepsilon^2 \end{pmatrix}$$

Note that the degree of the correlation between the random utility of choices in the migration choice set is measured by ρ.

Define Y_{ijtl} as a binary variable taking on the value of 1 if individual i from origin l chooses choice j in time period t, and 0 otherwise. Y_{ijtl} will then take the value of 1 if $U_{ijtl} > U_{iktl}\ \forall\ j \neq k$. Cardell (1997) shows that for a random variable ε that takes the Type I Extreme Value distribution, there exists a distribution for the ζ such that the random variable $\rho \cdot \zeta + (1-\rho) \cdot \varepsilon$ will also be distributed Type I Extreme Value. The probability that crossing j will be the utility maximizing choice is then the product of the probability that the individual will choose to migrate, multiplied by the probability that j will be the utility maximizing crossing choice,

$$\begin{aligned}
\Pr(Y_{ijtl} &= 1) = p_{jtl} = p_{jtl|mig} * p_{mig\ tl} && (2.6)\\
&= \frac{\exp(\frac{\delta_{jtl}}{1-\rho})}{\sum_{j=1}^{9}\exp(\frac{\delta_{jtl}}{1-\rho})} \cdot \frac{\exp(w_{itl}\gamma)\cdot[\sum_{j=1}^{9}\exp(\frac{\delta_{jtl}}{1-\rho})]^{1-\rho}}{1+\exp(w_{itl}\gamma)\cdot[\sum_{j=1}^{9}\exp(\frac{\delta_{jtl}}{1-\rho})]^{1-\rho}} \text{ for } j>0 && (2.7)\\
\Pr(Y_{iotl} &= 1) && (2.8)\\
&= p_{otl} = \frac{1}{1+\exp(w_{itl}\gamma)[\sum_{j=1}^{9}\exp(\frac{\delta_{jtl}}{1-\rho})]^{1-\rho}} \text{ for } j=0 && (2.9)
\end{aligned}$$

where $\delta_{jtl} = \alpha \cdot Enf_{jt} + X_{jtl}\beta$[2]. Note that these probabilities are equal to those of a standard conditional logit (CLOGIT) model when ρ takes the value of 0.

This model provides a natural framework for answering our research question: how does an increase in border enforcement affect the border crossing choices of migrants? A negative and statistically significant value of α tells us that an increase in border enforcement at crossing j does indeed push migrants away from sector j. However it is not immediately clear whether the migrant will be pushed to another border sector $k \neq j$, or deterred from crossing altogether. To answer this question we turn to the implied substitution patterns of the model. Below, we consider the marginal effects of enforcement on the probability of choosing a given option. As ε_{ijlt} is distributed *i.d.d.* across the i, these changes in probabilities for a single individual can also be interpreted as changes in the shares of the population making the choice.

Substitution away from a crossing with respect to the change in enforcement at that crossing is:

$$\frac{\partial p_{jtl}}{\partial Enf_{jt}} = \alpha \cdot p_{jtl} \cdot \frac{1}{1-\rho}[1 - \rho\frac{p_{jtl}}{1-p_{otl}} - (1-\rho)p_{jtl}] \qquad (2.10)$$

[2]This result follows from the assumption that ξ_{jtl} is independent of the X and Enf variables; we relax this assumption when developing our IV estimator below.

This total effect encompasses both divergence and deterrence. The diversion effect is the cross effect between an increase in enforcement at j and the increase in the probability that the migrant will choose another crossing. The term DIV_{jtl} represents the change in the probability that the migrant from origin l in year t will choose to cross at an alternative crossing when enforcement increases in j.

$$DIV_{jtl} = \sum_{j \neq k} \frac{\partial p_{ktl}}{\partial Enf_{jt}} \tag{2.11}$$

The deterrence effect is the change in the probability of not migrating (staying in Mexico) when enforcement increases at j. Consider DET_{jtl} , the change in the probability of the migrant remaining in the origin as a result of increased enforcement at j.

$$DET_{jtl} = \frac{\partial p_{otl}}{\partial Enf_{jt}} = -\alpha \cdot p_{otl} \cdot p_{jtl} \tag{2.12}$$

Having calculated both the total effect and the deterrence effect, we can decompose the total effect into the deterrence portion and the diversion portion[3]. Because the migrants' choice set represents mutually exclusive options, the decline in the probability of migrating to j (the total effect) must equal the increases in the probabilities of all other choices.

$$\begin{aligned} -\frac{\partial p_{jtl}}{\partial Enf_{jt}} &= \frac{\partial p_{otl}}{\partial Enf_{jt}} + \sum_{j \neq k} \frac{\partial p_{ktl}}{\partial Enf_{jt}} && (2.13) \\ &= DET_{jtl} + DIV_{jtl} && (2.14) \end{aligned}$$

[3]This relationship is similar to the "market-stealing" vs. "market-expansion" analysis found in Berry and Waldfogel (2001). This analysis is of the effect of changes in radio broadcasting variety. An important question in that market is whether this increased product choice resulted in creating more listeners, or simply *diverting* (stealing) listeners from a pre-existing station to a new station.

Deterrence, as a share of the total effect, can then be expressed as

$$PCTDET_{jtl} = \frac{\frac{\partial p_{otl}}{\partial Enf_{jt}}}{\frac{\partial p_{otl}}{\partial Enf_{jt}} + \sum_{j \neq k} \frac{\partial p_{ktl}}{\partial Enf_{jt}}} = \frac{\frac{\partial p_{otl}}{\partial Enf_{j}}}{-\frac{\partial p_{jtl}}{\partial Enf_{jt}}} \tag{2.15}$$

Substituting from equations for the own effect for $\frac{\partial p_{jtl}}{\partial Enf_{jt}}$ and the substitution to the outside good for $\frac{\partial p_{otl}}{\partial Enf_{jt}}$

$$\begin{aligned} PCTDET_{jtl} &= \frac{-\alpha \cdot p_{otl} \cdot p_{jtl}}{-a \cdot p_{jtl} \cdot \frac{1}{1-\rho}[1 - \rho \frac{p_{jtl}}{1-p_{otl}} - (1-\rho)p_{jtl}]} \\ &= \frac{(1-\rho) * p_{otl}}{[1 - \rho \frac{p_{jtl}}{1-p_{otl}} - (1-\rho)p_{jtl}]} \end{aligned} \tag{2.16}$$

This term will be both decreasing and concave in ρ. For $\rho = 1$, there will be no deterrence. The model provides intuition for this: ρ measures correlation among the ε terms in the choices in the migration choice set. If there is perfect correlation between these terms, migrants will substitute only within the migration choice set, thus an increase in when a characteristic of the utility maximizing choice changes, individuals will substitute only to another choice that is in the same set of choices, i.e. all choices involving migration. Note that the value of the shares alone will determine this figure if ρ were set equal to 0 as it is in the CLOGIT model; i.e. the data has more of a hand in determining substitution patterns with the model that we employ when compared to the CLOGIT model.

The structural nature of our model also allows us to conduct policy experiments to evaluate the effects of Border Patrol policies such as Hold the Line and Gatekeeper. By predicting crossing probabilities under counterfactual enforcement levels ($\hat{p}^*_{jtl}(Enf^*_{jt})$) and then comparing them to the predicted probabilities under historical values of enforcement ($\hat{p}_{jtl}(Enf_{jt})$) we can identify a total impact of the change in policy on migration rates. Computing the same figure for changes in the share of migrants choosing to remain in the origin under counterfactual values ($\hat{p}^*_{otl}(Enf^*_{jt})$) and historical values ($\hat{p}_{otl}(Enf_{jt})$) allows us to decompose the effect into diversion and deterrence.

$$DETCF^*_{jtl} = \frac{\hat{p}^*_{otl}(Enf^*_{jt}) - \hat{p}_{otl}(Enf_{jt})}{\hat{p}^*_{jtl}(Enf^*_{jt}) - \hat{p}_{jtl}(Enf_{jt})} \tag{2.17}$$

2.4 Identification

In this section we describe the strategies that we employ to estimate the parameters of the model. First we a two stage estimator we use to estimate our model of migration crossing choice. Next, we describe the instrumental variables strategy that we employ to account for the endogenous enforcement variable. Finally, we motivate the instrumental variables used to implement the previously described identification strategy.

2.4.1 Nested Logit Model

The log likelihood function for this model is

$$\ln L(\theta|X,Z,Y) = \sum_{i=1}^{N}\sum_{j=o}^{9} 1(Y_{ijtl} = j) * \Pr(Y_{ijtl} = j)$$

Defining $D_{Mjtl} = \sum_{j=1}^{9}\exp(\frac{\delta_{jtl}}{1-\rho})$, our models give the following probabilities an individual will choose a given choice

$$p_{jtl} = \frac{\exp(\frac{\delta_{jtl}}{1-\rho})}{D_{Mjtl}} \cdot \frac{\exp(w_{itl}\gamma)\cdot[D_{Mjtl}]^{1-\rho}}{1+\exp(w_{itl}\gamma)\cdot[D_{Mjtl}]^{1-\rho}} \text{ for } j = 1,..,9 \tag{2.18}$$

$$p_{otl} = \frac{1}{1+\exp(w_{itl}\gamma)[D_{Mjtl}]^{1-\rho}} \text{ for } j = 0 \tag{2.19}$$

This likelihood function can be maximized using a full information maximum likelihood estimator. However, for computational reasons, it may be advantageous to use the following estimation strategy (Greene 2003, Wooldrige 2002). First we estimate β and α by maximizing the following likelihood function

$$\ln L(\psi|X,Z,Y) = \sum_{i=1}^{N}\sum_{j=1}^{9} 1(Y_{ijtl}=j) * \Pr(Y_{ijtl}=j) \tag{2.20}$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{9} 1(Y_{ijtl}=j) * \frac{\exp(\frac{\delta_{jtl}}{1-\rho})}{\sum_{j=1}^{9}\exp(\frac{\delta_{jtl}}{1-\rho})} \tag{2.21}$$

As noted above, the nested logit model assumes that the ε terms will be $i.i.d$ within the specified sets of choices. With the $i.i.d.$ assumption, a conditional logit model of where migrants cross, ignoring the information from the non-crossers, is a valid model; i.e. consistent but not efficient. Maximizing the following likelihood function will provide consistent estimates of $\psi = \frac{\beta}{1-\rho}$.

Using the estimates from this model, we can then compute $\hat{D}_{Mtl}$ a fitted value of D_{Mtl} using $\hat{\psi}$. Then maximizing the following likelihood will allow us to obtain consistent estimates of all the structural parameters of the model (directly giving us estimates of α and ρ, and by identifying ρ allowing us to separately identify the β terms.

$$\ln L(\theta|X,Z,Y) = \sum_{i=1}^{N}\sum_{j=o}^{9} 1(Y_{ijtl}=j) * \Pr(Y_{ijtl}=j)$$

Using the estimates of β and α, we construct $\exp(\widehat{\frac{\delta_{jtl}}{1-\rho}})$ and $\hat{D}_{Mjtl} = \sum_{j=1}^{9}\exp(\widehat{\frac{\delta_{jtl}}{1-\rho}})$, and estimate γ and ρ by maximizing the following likelihood function

$$p_{jtl} = \frac{\exp(\widehat{\frac{\delta_{jtl}}{1-\rho}})}{\hat{D}_{Mjtl}} \cdot \frac{\exp(w_{itl}\gamma)\cdot[\hat{D}_{Mjtl}]^{1-\rho}}{1+\exp(w_{itl}\gamma)\cdot[\hat{D}_{Mjtl}]^{1-\rho}} \text{ for } j = 1,..,9 \tag{2.22}$$

$$p_{otl} = \frac{1}{1+\exp(w_{itl}\gamma)[\hat{D}_{Mjtl}]^{1-\rho}} \text{ for } j = 0 \tag{2.23}$$

2.4.2 Instrumental Variables Strategy

A major issue in identifying this model is the endogeneity of border enforcement. Recall our utility function:

$$U_{ijtl} = \alpha \cdot Enf_{jt} + X_{ijtl}\beta + W_{itl}\gamma + \xi_{jtl} + \upsilon_{ijt} \quad \text{for } j > 0$$

Observations with high values of the unobserved utility to the choice, ξ_{jtl}, will attract more migrants. The Border Patrol would be expected to increase the allocation of resources to sectors during times in which they are experiencing a large number of illegal crossings. Accordingly, we would expect a positive correlation between ξ_{jtl} and Enf_{jt} standard methods of estimating α will lead to inconsistent and upward biased estimates.

Following a similar approach to binary models taken by Blundell and Smith (1989), Petrin and Train (2004) develop a strategy for obtaining consistent estimates in the context of this endogeneity problem. Consider the following decomposition of the enforcement variable:

$$Enf_{jt} = g(z_{jt}) + \mu_{jt}$$

$$\Rightarrow \mu_{jt} = enf_{jt} - g(z_{jt})$$

A function of a set of predetermined variables, z_{jt}, explains a portion of the variation in the endogenous variable; the residual term μ_{jt} contains the all the endogenous variation in Enf_{jt}. The term υ_{ijt} is the portion of the error term in the structural equation that is independent of μ_{jt}; the endogeneity problem arises from a correlation between μ_{jt} and ξ_{jtl}. If we assume that the vectors μ_{jt} and ξ_{ktl} are jointly normally distributed with no correlation between terms for which $j \neq k$ and that ξ and υ are independent in distribution then the distribution of the error term in the structural equation, $\xi_{jtl} + \upsilon_{ijt}$, conditional upon μ_{jt} will be

$$\xi_{jtl} + \upsilon_{ijt} \sim [\lambda_j \mu_{jt} + N(0, \sigma_\xi) + \upsilon_{ijt}]. \tag{2.24}$$

Estimation of a conditional logit model, after conditioning upon $\lambda_j \mu_{jt}$ will provide consistent estimates of α and β.

The μ_{jt} term is estimated by a linear regression of our set of instruments and other predetermined variables z_{jt} on the endogenous variable Enf_{jt}. Substituting the fitted values of μ_{jt} into our utility function,

$$\hat{U}_{ijtl} = \alpha \cdot Enf_{jt} + X_{ijtl}\beta + W_{itl}\gamma + \lambda_j \hat{\mu}_{jt} + \xi_{jtl} + \upsilon_{ijt} \quad \text{for } j > 0 \tag{2.25}$$

we can now proceed to estimate α and β through a standard conditional logit, also including the μ_{jt} term as a covariate and estimating the λ_j terms.

Petrin and Train compare estimates obtained using this estimator to estimates obtained using a computationally intensive approach implemented on disaggregated data by Berry, Levinson, and Pakes [BLP] (2004), based upon the estimator in BLP (1995). While the assumptions behind the control function approach have been questioned as stronger than those of the BLP estimator, Petrin and Train show that in the general case, the assumptions for each estimator are different, with neither set of assumptions necessarily being a subset of another. In their application of the estimators, the two methods provide nearly identical results.

2.4.3 Instrumental Variables

Relevant and valid instruments will be correlated with the level of enforcement in a given Border Patrol sector, but uncorrelated with the unobserved characteristics affecting the mean utility to migrants for crossing through that sector in t. For a set of potential instruments, we turn to the political process determining the level of funding in each sector. Hanson and Spilimbergo (2001) provide evidence of the importance of local politicians on the level of enforcement in the sector. Their principal finding is that border enforcement is reduced when output prices increase for goods whose production process involves the labor of undocumented immigrants as an input. This suggests that pressure groups are able to exert influence over the process determining how strictly the border is patrolled.

We exploit variation in enforcement caused by political lobbying to serve as an instrument for the observed level of enforcement. For example, variables such as the party of the Congressional representatives of border sector constituents may be correlated with the level of funding for the sector. While relevant, these instruments may also be invalid. The election of these representatives is likely a function of unobservable characteristics of the district, as a constituency that votes for a representative who is "tough on illegal migration" is likely composed of individuals who have a distaste for migration that may directly affect the utility of migration.

Suppose that all elected officials have some demand for increased funding in their home sector, if not to crack down on illegal immigration, simply to increase the amount of federal funds and jobs provided to constituents in their district. Then variables that do not measure a politicians' *desire* to increase enforcement resources, but rather their *ability* to increase local level enforcement should be valid.

We create two such "political supply" instruments to estimate our model. Our first instrument for border enforcement is the size of the Congressional delegation representing the Border Patrol sector. Specifically, we count how many Congressional districts share the sector's border with Mexico. As shown in Table 2.1 and Figure 2.1, in Arizona the Yuma county Border Patrol sector includes only Yuma county (the most western border county), which is represented by the 7th Congressional District. The Tucson sector includes the three border counties to the east of Yuma County and is represented by both the 7th and 8th Congressional districts. Thus the Tucson sector has two representatives who have an interest in securing funding for the sector, while the Yuma sector has only one. As the size of the sector's lobby in Congress grows, so should its budget and accordingly the level of enforcement. We refer to this variable as NUMREPS. Values of this instrument will vary not only across sectors, but also over time as redistricting occurs every ten years.

Our second instrument for enforcement is the strength of the state Congressional delegation representing a Border Patrol sector. We count how many members from the state's delegation are on the House Appropriations committee and then match these state delegations to Border Patrol sectors. Note that the El Paso sector includes counties in both Texas and New Mexico, while the Marfa sector includes only counties in Texas. This variable is named STATEAPP. Thus, the delegation that has an interest to procure funds for the El Paso sector is larger than the delegation representing Marfa.[4]

Validity of the instruments hinges on the assumption that these instrumental variables capture nothing that directly affects the utility of the choice to the migrant. Essentially,we need to believe that a migrant should have no direct preference for avoiding crossing through a region that has a powerful political lobby at the federal level. Thus, the only effect the political power of this region will have on the utility of the migrant to crossing there will be due to the tighter border enforcement that results from the strength of this lobby. When estimating the first stage of our model, we also include year and crossing fixed effects. We are then identifying the effect of enforcement on migrants' crossing decisions that results from changes in the power of the local political lobby, relative to the average strength of political lobbies in other crossings in a given year.

[4] Constructed using data from Congressional Quarterly Almanac.

2.5 Data and Specification

The data on the choices of migrants comes from the Mexican Migration Project (MMP), a survey constructed jointly by researchers in Mexico and the United States. Separate waves of the survey have taken place annually, starting in 1982. Using the retrospective migration history information provided by the MMP, we are able to observe whether a head of household crossed the border in any given year, and if so, through which sector they crossed the border. Our sample consists of all household heads who have already undertaken at least one temporary migration to the U.S. Experienced migrants are more likely to be informed about different levels of enforcement across the border, and thus their decisions will better help us identify how enforcement affects the location of crossing, conditional upon a migration being undertaken.

The use of this sub-sample may cause concern about the general applicability of our results and our ability to conduct meaningful policy experiments. This concern would be valid if a policy experiment were to evaluate the impact of a decrease in border enforcement on a migrants crossing decisions, as decreases in enforcement would likely induce individuals without migration experience to begin migration. However, our policy experiments involve estimating the impact of increases in enforcement. As these policy choices will decrease migration rates, inexperienced migrants, outside of our sample, will continue to not migrate. The migrants observed under the historic enforcement levels, on the other hand, will be the ones affected by increases in enforcement. Thus by estimating the parameters for experienced migrants, we are identifying the effect of increased enforcement on those whom policy changes will most likely affect.

The MMP gives data on the Municipio[5] at which the migrant crossed the U.S. border. We then match these Municipios to one of the Border Patrol's nine sectors

[5] A mutually exclusive sub-state geographic division in Mexico, similar to counties in the United States.

that span the U.S.-Mexico border. The location of crossing is then proxied by the principal crossing point in each of the nine sectors. These characteristics are described in Table 2.1. In Table 2.2 we present summary data on decisions made by individuals in our data set. We observe a total of 30,162 decisions made by a total of 1919 choosers; the average individual appeared in our data set for 15.7 consecutive years. Our data spans decisions made between 1977 and 2003, with the mean observation coming from the late 1980s. Just over 1 out of 8 observations involved the decision to cross the border into the U.S. Of these migrations, about half took place through the San Diego sector.

We include four types of explanatory variables in our model. The first set of variables involve no individual heterogeneity and vary across all choices. First and foremost of these variables is the enforcement measurement. The border enforcement data that we employ were originally used in Hanson and Spilimbergo (1999). The data provide rich information on both the number of apprehensions and line watch hours (the number of labor hours spent patrolling the border) in the nine Border Patrol sectors. Monthly data by sector are provided from 1977 to 2000. Table 3.2 provides enforcement hours per mile of border for 1980, 1990 and 2000; the table indicates that there is significant variation in border enforcement across both time and space. As our measure of enforcement, we use the log of linewatch hours divided by the mileage of border spanned by the sector.

Distance to the border is the other explanatory variable of this type. For confidentiality reasons, the MMP does not disclose the location of the home community of survey respondents; only the state of each community is given. Thus, we use distance from the state capital of the home community. A choice specific constant term is also estimated for eight of the nine choices, while the constant term for the ninth choice is normalized to zero.

The second type of variable in our model does not vary across choices that involve migrating, nor does it involve individual heterogeneity. The quality of economic opportunities in the origin should affect the probability of migrating but not the location on the border over which the migrant crosses. To proxy for this, we use the log of the population of the home community. A constant term for the utility of choosing any choice involving migration is also included.

Recall the *i.i.d* problem that motivated the use of the nested logit model: a portion of individuals' random utility draw is constant across choices involving migrating, thus these draws are not *i.i.d* across choices in the migration choice set. The nested logit model solves this problem by allowing for correlation across these draws and allowing the "migration-specific" utility constant term to vary across individuals-essentially estimating a random coefficient on the constant term.

This correlation, however, would not be a problem if the migration-specific draw for each individual were observable: conditional upon this term, the ε terms are *i.i.d.* While it is not possible to observe this draw, it may be proxied for by our third group of variables. These involve individual heterogeneity and do not vary across choices within the migration choice set. These variables, characteristics of an individual that will affect their probability of migrating but not their border crossing point, include the log of the total number of migrations the individual has under taken, the log of the number of dependents in the household, and the age of the head of household. For our total experience measure, we use the log of the count of previous migrations undertaken. Past migration experience may affect the cost of migration, as more experienced migrants would be more likely to have an established network in the destination. However, past migrations may also increase the likelihood of a future migration because they reveal that an individual having a strong taste for migration (thus controlling for some of the ζ_{itl} term). Therefore, the coefficient on this parameter should not be interpreted causally. Household heads with more dependents to provide for should be more likely to migrate. Older migrants should also be less likely to migrate, as they will have stronger ties to the origin community.

Our final group of variables are those that affect the probability of crossing along any point on the border and involve individual heterogeneity. The sole variable in this group is a measure of choice specific experience: the log of the count of prior migrations made through the particular border crossing in question.

Instruments are constructed using data from the Congressional Quarterly Almanac and Congressional District Almanac. For each Congress, we count the number of members of each state delegation who sit on the House of Representatives Appropriations Committee and assign these values to the appropriate Border Patrol sector[6]. The variable measuring the number of Representatives responsible for each sector was constructed by overlaying maps of Congressional District boundaries with those of the Border Patrol Sector. As redistricting takes place with each decennial census in the US, these variables vary across both time and space. Table 2.4 presents summary statistics of our data. The NUMREPS variable varies between 1 and 3 with a mean of 1.61 representatives per sector, while the STATEAPP variable has a mean of 3.35, a minimum of 0 and a maximum of 7. Thus, there is substantial variation across both time and space for each of these instruments. The panel nature of our data allows us to control for both time and location fixed effects in our estimation, eliminating a number of sources of potential endogeneity. Our first stage also includes year indicator variables; the effect of enforcement on crossing variation is then identified by the de-trended variation in enforcement attributed to the political supply instruments. In Table 2.4 we also see that the average observation in our data set corresponded to a migrant who was 33 years old, had 2.21 dependents, had previously migrated to the U.S. 1.13 times, lived in a community of around 60,000 people, and was on average 1037 miles from a given crossing point. We also report the log of these variables (log of the variable plus one in the cases where support of the variable included zero values).

[6] The total for the California delegation to San Diego and El Centro, the total for the Arizona delegation to the Yuma and Tucson delegation, the total for the New Mexico and Texas delegation to the El Paso sector, and the total from only the Texas delegation to the Marfa, Del Rio, Laredo and McAllen sectors

2.6 Results

2.6.1 Parameter Estimates

In Table 2.5 we present our parameter estimates. The first column gives results from the first stage linear regression of our instruments on enforcement, crossing sector fixed effects, and year fixed effects. As predicted, we find a positive and statistically significant relationship between the STATEAPP and NUMREPS variables and enforcement. The F statistic on a test of joint significance of these variables is 16.55, so there do not appear to be problems with weak instruments.

The first set of estimates of the parameters from our main equation is given in the second column. The CLOGIT model yields a negative and statistically significant estimate of the effect of enforcement on a migrant's utility. The distance coefficient is negative and statistically significant as expected, and the crossing-specific experience coefficient is positive and statistically significant also as expected. Observing our expected results gives evidence that our model of utility maximization is consistent with migrants' behavior.

In our set of characteristics that affect the probability of migration but have no effect on the place of the migration, we find the unexpected result that total number of prior migrations is negatively related with the likelihood that a migrant will make another crossing. This could be driven by a life cycle effect, where migrants plan to undertake a fixed number of migrations during their lifetime. The more migrations the individual has undertaken to date, the less likely they may be to undertake another trip. As expected, we also find that older individuals are less likely to migrate, while migrants with more dependents are more likely to set off to the U.S. Migrants from larger communities are more likely to stay at home in lieu of migrating.

The next set of results in Table 2.5 comes from our CLOGIT-IV model. Other than the inclusion of the control function, this model is identical to the CLOGIT model. As expected, the control function eliminates upward bias in the estimated effect of enforcement on a migrants' utility. While the CLOGIT model found a negative and statistically significant estimate of the coefficient on enforcement of only $-.099$, inclusion of the control function lowers the estimate of this parameter to $-.415$. The estimates of the other parameters in the model are very similar to those of the CLOGIT model. A likelihood ratio test rejects the use of the CLOGIT model in favor of the CLOGIT-IV model.

In the next three models we relax the assumption that the random utility terms are *i.i.d.* across choices in the migration choice set. As the *i.i.d.* property will still hold for choices within the migration choice set, the first step in estimating the NLOGIT model is a conditional logit model of the location choice of individuals, conditional upon observing that they have migrated. This will provide consistent, but inefficient, estimates of the α and β parameters.

The fourth column in Table 2.5 presents estimates of marginal utilities on the three characteristics which vary across choices within the migration choice set: enforcement, distance, and prior experience at a specific crossing. Here, the coefficient on enforcement is estimated to be positive and statistically significant. While we found there to be an upward bias on the estimate of this parameter in the CLOGIT model, one may wonder why the bias would be of a greater magnitude here; the CLOGIT model estimate without instrumenting was -0.099 while here our estimate is 0.427. The intuition for this result comes from the fact that we are only considering choices that involved migrating.

In the CLOGIT models we were exploiting the variation in enforcement across all choices, including staying in Mexico. Because the level of enforcement that a potential migrant face should he choose to remain in Mexico is exogenously set to zero, we lose this source of variation in the enforcement variable and are left with a variable whose variation is now comprised of more endogenous variation than before. In this model, the estimates of the effect of distance on the individuals' utility become more negative, while the estimates of the effect of prior experience is similar to the estimates obtained from the CLOGIT models.[7]

After the inclusion of the control function, the estimate of the coefficient on enforcement drops to $-.520$, which is statistically significant. Here, as was the case in the CLOGIT model, we reject a likelihood ratio test of the equivalence of the models with and without the control function. The estimates on the marginal utilities of distance and past experience are similar to those in the model without the control function.

Finally, we present estimates of the parameters in the final stage of our NLOGIT model. In this step we calculate estimates of the coefficients on variables which will affect the utility to migrating, but will not vary across choices in the migration choice set (γ) and the correlation between random utility draws for choices involving migration (ρ). This is done by computing the inclusive value parameter from the NLOGIT-IV model and including it in a binary model of migration for each observation. Our estimates of the effect of origin community size, age, and number of dependents are very similar to those estimates in the CLOGIT-IV model. The effect of total migration experience is still negative and statistically significant, but of a much smaller magnitude than in the CLOGIT-IV model. Our estimate of the ρ term is .642, suggesting that migrants view alternative border crossings as relatively close substitutes. The lower bound of our estimate is .589, furthering the evidence that the crossings are close substitutes.[8]

[7] Note that the first stage estimates of the coeffients on choice specific characterisitcs are estimates of $\frac{\alpha}{1-\rho}$ and $\frac{\beta}{1-\rho}$. Testing for signifcance of these reduced from parameters is the appropriate way to make inferences about the effect of X on migration choices, as the partial derivative of each probability with respect to a choice characteristic produces the $\frac{\beta}{1-\rho}$ term rather than simply the β term.

[8] The standard errors used to test hypothesis regarding these variables have not been adjusted to relfect the multi-step nature of this estimator.

2.6.2 Marginal Effects and Policy Experiments

When interpreting our results; recall that our data provides us with a measure of variation in enforcement rates *between* Border Patrol sectors, but not *within* Border Patrol sectors. Accordingly, individuals in our model choose in which sector they will cross the border, but not a specific point within the sector. This is important to remember when making inferences about policy: as we are using a sector wide average of enforcement intensity, our policy simulations and marginal effects identify how migrants respond to an increase in enforcement which is uniform across the entire sector. In the case of Gatekeeper and Hold the Line, this is clearly an abstraction from the actual policy change migrants encountered. Each of these operations entailed a large increase in the intensity of enforcement at the main urban crossing in their respective sectors. Enforcement levels in rural areas remained relatively constant. Thus, one would expect there to be a great deal of *within-sector diversion*, i.e. many migrants responded to the policy change by continuing to cross through the same sector, but through a different point in the sector. One would then expect that our analysis provides an upper bound to the amount of deterrence when evaluating a policy that involves increasing enforcement along only part of the border within a given sector.

Using the estimates from Table 2.5, we calculate the decomposition of the marginal effects into diversion and deterrence. In Table 2.6, we report the percentage of a marginal increase that will deter, rather than divert, migrants. This figure is computed by first calculating the decomposition for each crosser, and then averaging over individuals. When we compute the mean effect over all observations, we find that the portion of the effect of a marginal increase enforcement that is deterrence will be only 31 percent to 37 percent of the total effect; while between 63 percent and 69 percent of migrants pushed away for the border crossing will be diverted elsewhere.

Alternatively, we estimate the marginal effect of enforcement on the probability of making a given choice for those individuals whom we observe making that choice. For example, to estimate the marginal effects for a marginal increase in enforcement in the San Diego sector, we compute marginal effects and then average these over the set of all individuals who were observed to have crossed in San Diego, rather than over the sample as a whole. Here, we find a higher deterrence effect of 59 percent (or 63 percent using the lower bound of ρ). Put another way, our evidence suggests that if the Border Patrol were to enact an increase in enforcement along one part of the border sufficient to stop 100 migrants from crossing there, then around 60 of those migrants would remain in Mexico, while the other 40 would find another sector on the border to cross into the United States.

We also report the deterrence share broken down by sector. A sectorwide increase in San Diego clearly has the highest overall deterrence effect (70 percent of the total effect), while estimates of the deterrence effect for other sectors range from 30 percent to 60 percent. This result is driven by the fact that choosers are more likely to substitute to choices that are popular for the population as a whole. As San Diego is the most popular crossing sector in the choice set, migrants experiencing an increase in enforcement at San Diego will face a less desirable alternative set of choices compared to migrants crossing at other locations ,who have the option of substituting to San Diego. The lowest estimated deterrence is in the Yuma Sector of the border, at 31 percent.

Tables 2.7 and 2.8 report estimates from two policy simulations. For each policy, we evaluate the effectiveness of a major Border Patrol operation. Table 2.7 focuses on Operation Gatekeeper in the San Diego sector of the border, while Table 2.8 evaluates Operation Hold the Line, which occurred in the El Paso sector.

Operation Gatekeeper involved an 85 percent increase in enforcement in the San Diego sector of the border between 1994 and 1996. To evaluate the effects of this policy, we set the 1996 level of enforcement in San Diego equal to the 1994 level, thus eliminating the increase in enforcement associated with the operation. The second column in Table 2.7 reports the mean predicted probability of an individual crossing through the San Diego sector in 1996. The first row reports this change under the historical values of enforcement, while the second row reports the figure for a counterfactual value of enforcement that eliminates the increase in enforcement. Comparing the two numbers, we see that the effect of Gatekeeper on migration rates was a drop in the predicted migration rate from 6.6 percent of our sample to 5.7 percent. The local elasticity, which is the percentage change in *number* of migrants crossing in San Diego with respect to the percentage change in enforcement intensity, is -.159. Even at the local level, the migrants appear to be relatively unresponsive to border enforcement.

This measure of the *total effect* of enforcement in San Diego does not account for diversion. To take into account the fact that not all the migrants who stopped crossing through the San Diego sector stopped crossing all together, we must measure the change in the probability that an individual in our sample did not migrate. We see that this probability increased from 88.8 percent to 89.4 percent as a result of Gatekeeper; a .6 percent point increase. As Gatekeeper caused a .9 percent point decrease in migration through San Diego but only a .6 percent point increase in non-migration, it is clear that .3 percent of the choosers in our data set substituted to an alternative border crossing as a result of Gatekeeper. Thus, roughly 2/3 of the total effect of this policy was deterrence, while 1/3 was diversion (68.1 percent, to be exact).

In the final four columns of Table 2.7, we compute the elasticity of migration across any part of the border with respect to the policy change in San Diego. If one were to ignore the .3 percent point diversion effect, it would appear that the mean probability of migration would have dropped from 11.2 percent to 10.3 percent, giving an elasticity of -.094. However, once we account for the diversion effect by increasing our estimate of the overall migration rate by the .3 percent point diversion effect, we calculate the elasticity to be only -.064.

Table 2.8 performs similar analysis for Operation Hold the Line, which entailed a 82.6 percent increase in enforcement in the El Paso sector between 1992 and 1995. Here, we compare predicted outcomes under the levels of enforcement in 1995 under Hold the Line to the predicted outcomes in 1995 in the absence of increases in enforcement brought forth by Hold the Line. The migration rate to El Paso decreased by .18 percent points while at the same time the overall predicted probability of migration only decreased by .8 percent points: .18 percent of migrants stopped crossing in the El Paso sector as a result of Hold the Line, .1 percent of migrants substituted to other crossings while only .08 percent chose to not cross altogether. Deterrence then accounted for only 45 percent of the total effect. The local elasticity measurement is -.24, suggesting that this program was more cost effective at the local level than Gatekeeper. However, as the absolute number of crossings at El Paso was far lower than the number of crossings in San Diego, the overall effect of this increase in enforcement was smaller than Gatekeeper. Without accounting for diversion, each 10 percent increase in enforcement in El Paso would result in only a .2 percent decrease in overall migration. However, after accounting for diversion, this estimate drops to resulted in only a .08 percent decrease in overall migration, whereas each 10 percent increase in enforcement associated with Gatekeeper resulted in a .6 percent decrease in overall migration

In summary, our evaluation of these two Border Patrol operations reveals a number of interesting results. First, we find that, even at the local level, migration rates are relatively inelastic with respect to changes in enforcement. When measuring the elasticity of the overall migration rate with respect to changes in enforcement at one border crossing sector, it is apparent that local level Border Patrol operations have a very small impact on the total number of migrants choosing to cross illegally into the United States. After accounting for diversion, the estimated deterrence effect of border enforcement becomes even smaller.

2.7 Conclusion

As the debate about immigration reform and increased border security continues, it is increasingly important for policy makers to have a better understanding of the effectiveness of border enforcement, particularly when changes in enforcement usually target specific parts of the border. A popular perception is that the U.S.-Mexican border suffers from a "water-balloon" effect: when one part of the border is tightened immigrants flow through the next easiest place to cross. The effect thus nullifies a good portion of the attempt to stem the flow of illegal immigrants into the United States.

In this paper we estimate the magnitude of this *diversion* effect. We face several issues in identifying this effect. First, we must develop instrumental variables with cross-sectional variation in order to instrument for the border enforcement in the model of a migrants' crossing location decision. We are able to do so with the use of "political supply" variables, that measure the ability of a local political lobby to procure funds to increase the intensity of border patrol in the locale. Our use of enforcement data that varies both over time and space allows us to control for time invariant characteristics of each border crossing, as well as border wide changes in policy in each year. Prior studies, using only time series variation in enforcement, have been unable to hold these factors constant.

Second, to properly identify the diversion and deterrence effects, we must use a random utility model that produces reasonable substitution patterns. To do this, we model the utility of each choice in a more realistic way, taking into account the fact that migrants who have a strong preference to migrate into the United States through one border crossing likely also have a strong preference to migrating through another sector. The correlation among these preferences, identified as ρ in our model, allows us to estimate more realistic patterns of substitution between crossing choices when enforcement changes. The decomposition of the effect of border enforcement into diversion and deterrence is straightforward. Our estimates suggest that there is indeed a strong diversion effect: while border enforcement may appear very effective at the local level, a good part of this effect will be offset by an increase in migration through other sectors. Further, our analysis of Operations Gatekeeper and Hold the Line suggest that migrants are relatively unresponsive to large increases in border enforcement intensity

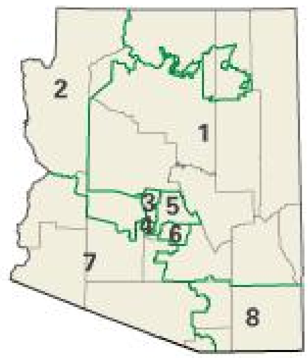

Figure 2.1: AZ Congressional Districts

Sector	Counties	Principal Crossing	Miles of Border
San Diego	San Diego (CA)	Tijuana-San Diego	66
El Centro	Imperial (CA)	Mexicali-Calexico	75
Yma	Yuma (AZ)	SL Rio Colorado-Yuma	118
Tucson	Pima, Santa Cruz, Cochise (AZ)	Nogales-Nogales	261
	Hidalgo, Luna, Dona Ana (NM)		
El Paso	El Paso, Hudspeth (TX)	Ciudad Juarez-El Paso	289
Marfa	Presidio, Brewster (TX)	Ojinaga-Presidio	420
Del Rio	Terrell, Val Verde (TX)	Del Rio-Piedras Negras	205
	Kinney, Maverick,		
Laredo	Webb, Zapata (TX)	Nuevo Laredo-Laredo	171
McAllen	Starr, Hidalgo, Cameron (TX)	Reynosa-McAllen	284

Table 2.1: Border Patrol Zones

Decisions Observed	30162
Decisions to not Migrate	26439
Total Decisions to Migrate	3723
San Diego	1774
El Centro	129
Yuma	19
Tucson	458
El Paso	286
Marfa	12
Del Rio	79
Laredo	608
McAllen	358
Number of Choosers	1919
Average Number	15.72
First Year Observed	1977
Last Year Observed	2003
Mean Year	1988

Table 2.2: Summary of Migration Decisions

Sector	**1980s**	**1990s**	**2000**
San Diego	7236	7769	30764
El Centro	1422	2849	9210
Yuma	1955	1880	2680
Tucson	550	994	7950
El Paso	928	1456	2849
Marfa	195	137	248
Del Rio	1134	1372	3217
Laredo	573	1566	4848
McAllen	574	954	5183

Hanson and Spilembergo (1999)

Table 2.3: Border Enforcement in Annual Labor Hours Per Mile

Variable	Description	Mean	Stdev	Min	Max
Enforcement	*Enforcement Per Mile*	2641.61	4327.24	81.39	34737.93
Enf	*log(Enforcement)*	7.20	1.12	4.40	10.46
Dist	*Distance from Home to*	1037.21	441.15	0	2095.63
lnDist	*log(Dist+1)*	6.81	0.66	0	7.65
Experience	*# Prior Migrations*	0.13	0.62	0	18
lnExp	*log(exper+1)*	0.07	0.26	0	2.94
Age	*Age at Time of Decision*	33.19	10.57	18	65
lnAge	*log(Age+1)*	3.49	0.30	2.94	4.19
depen	*Number of Dependents*	2.21	2.10	0	10
lnDep	*log(Depen+1)*	0.94	0.70	0	2.40
ExperTot	*Total Prior Migrations*	1.13	1.61	0.00	18.00
lnExpTot	*log(ExperTot+1)*	0.58	0.55	0	2.94
comPop	*Home Community*	59069.36	163182.00	1000.00	1189000.00
lnComPop	*log(ComPop+1)*	8.95	1.81	6.91	13.99
numReps	*Congression District Spanning Border in Sector*	1.61	0.70	1	3
stateapp	*Number of Members of State US Congressional Delegation of House Appropriations Committee*	3.35	1.77	0	7

Table 2.4: Summary Statistics

	First Stage	CLOGIT	CLOGIT-IV	NLOGIT-1st	NLOGIT-1st-IV	NLOGIT Final Stage
Enf		-0.099*	-0.415***	0.427***	-0.520**	
stateapp	0.283***					
numReps	0.051*					
log(Distance+1)		-0.258***	-0.274***	-0.664***	-0.661***	
log(# of Prior Migrations through Crossing+1)		3.331***	3.286***	3.436***	3.432***	
log(Total Prior Migrations +1)		-2.064***	-1.967***			-0.590***
log(# of dependents+1)		0.203***	0.196***			0.197***
log(age)		-0.568***	-0.526***			-0.616***
log(home community size)		-0.142***	-0.138***			-0.156***
Rho						0.642***
Choice/Group Fixed Effects	X	X	X	X	X	
Control Function			X		X	X
Year Fixed Effects	X					
Inclusive Value Parameters						X
F-stat on joint signifcance of instruments	16.55					
P-value on Test of Joint Significance of Lambda j Terms			0.000		0.000	
log likelihood		-15016.49	-14893.43	-3906.339	-3835.154	-14864.55
N	243	301620	301620	33507	33507	301620

*** denotes signifcant at the 1% level
** denotes signifcant at the 5% level
* denotes signifcant at the 20% level

Table 2.5: Parameter Estimates

	At Rho Point Estimate: .642	**At Rho Lower Bound: .589**
Mean % Deterrence (over all choosers)	0.31	0.37
Mean %DET at Crossing Chosen	0.59	0.63
San Diego	0.66	0.70
El Centro	0.43	0.48
Yuma	0.31	0.37
Tucson	0.54	0.59
El Paso	0.49	0.54
Marfa	0.32	0.38
Del Rio	0.43	0.49
Laredo	0.55	0.60
McAllen	0.54	0.59

Table 2.6: Marginal Effects: Diversion and Deterrence

					Ignoring Diversion		Accounting for Diversion		
	lnEnf	Mean Probability of Crossing in San Diego	Mean Probability of not Migrating	Local Elasticity	Mean Probability of Migration	Elasticity Estimate	Mean Probability of Migration	Elasticity Estimate	Deterrence Share
Under Gatekeeper	10.067	0.057	0.894	-0.159	0.103	-0.094	0.106	-0.064	
Without Gatekeeper	9.451	0.066	0.888		0.112		0.112		
Change	0.851	-0.009	0.006		-0.009		-0.006		0.681

Table 2.7: Evaluating Gate Keeper: Migrations through the San Diego Sector in 1996

	lnEnf	Mean Probability of Crossing in El Paso	Mean Probability of not Migrating	Local Elasticity	Ignoring Diversion: Mean Probability of Migration	Ignoring Diversion: Elasticity Estimate	Accounting for Diversion: Mean Probability of Migration	Accounting for Diversion: Elasticity Estimate	Deterrence Share
Under Hold the Line	7.7733	0.0072	0.8879	-0.2378	0.111	-0.019	0.1121	-0.0087	
Without Hold the Line	7.1710	0.0090	0.8871		0.113		0.1129		
Change	0.8263	-0.0018	0.0008		-0.0018		-0.0008		0.4581

Table 2.8: Evaluating Hold the Line: Migrations through the El Paso Sector in 1995

Chapter 3

Migration Creation, Diversion, and Retention: New Deal Grants and Migration: 1935-1940

3.1 Introduction

[1]What is the impact of federal government programs on internal migration patterns? There is a large literature in regional economics and public economics on Tiebout models where people migrate across states and counties to take advantage of differences in taxation and spending policies. Much of the literature focuses on the impact of state and local government fiscal choices and the impact of policy competition among localities.[2] Yet, the federal government's distribution of funds across districts can also be extremely important. Locations of federal highways and military bases influence patterns of migration. One of the greatest experiments with federal spending in local areas occurred during the 1930s, when a wide range of New Deal grants programs were established to combat the problems arising from the Great Depression. A large literature on the political economy of that spending shows that there was substantial variation in federal grants across states and counties (see Fishback, Kantor and Wallis, 2003 for a summary).

In a recent study Fishback, Horrace, and Kantor [FHK] (2006) examined the impact of federal grants on net migration for over 3000 counties. They find that New Deal public works and relief grants stimulated net in-migration into counties and that the Agricultural Adjustment Administration's payments to farmers to take land out of production was associated with net out-migration. The FHK study, however, was limited to the study of a summary measure, net migration, and thus was unable to

[1]Joint Work with Price Fishback, Samuel Allen and Shawn Kantor

[2]See Rhode and Strumpf (2004) for a summary of the literature and results that do not find much Tiebot sorting over long time periods in response to state and local government goods provision.

examine the flows of people from location to location. In this paper we use information from the 1940 Census on locations of households in 1935 and 1940 to estimate a model in which households are choosing among 460 state economic areas (SEAs). We build an aggregate discrete choice model from a random utility model of locational choice to estimate the impacts of spending on these programs on migration patterns.

When we combine the structure of the model with the information on flows of migration between the SEAs, we are able to derive a series of inferences about the impact of New Deal programs on the substitution patterns of migrants between locations. We compute marginal effects of program spending that are decomposed into three distinct effects. The first is migration creation. A migrant is "created" by a program when program spending causes a household that would otherwise have remained in its 1935 location to migrate to a given location. The second, a retention effect, occurs when a migrant is "retained" because program spending causes a household that would have out-migrated otherwise to remain in the 1935 location. Finally, migration diversion results when a migrating household is diverted to another location.

The structural nature of our model also allows us to estimate counterfactuals to identify the overall effect of New Deal spending on migration, an analysis that FHK (2006) could not perform. We estimate counterfactuals for the absence of each of two major categories of New Deal spending, as well as a counterfactual for the absence of all New Deal spending. The results show that New Deal public works and relief spending caused a significant increase in the number of internal migrations in the United States between 1935 and 1940, while spending on the Agricultural Adjustment Administration decreased migrations, presumably by causing farmers who would have otherwise migrated away from their origins to stay put.

3.2 New Deal Programs

We consider two categories of per capita New Deal grants that may have affected the desirability of a location: AAA farm grants and public works and relief grants. The AAA grants were made to farmers who voluntarily removed land from production for designated crops.[3] The goal of the program was to increase the incomes of farmers, both directly through benefit payments and indirectly by raising market prices to pre World War I levels (1920s levels for tobacco), through the curtailment of the output of specific crops.[4] The AAA programs likely had conflicting effects on migration. The farmers who received payments from the AAA were likely to stay in farming and, thus, less likely to migrate. In contrast, farm workers and tenants might have been pushed out by the AAA because the AAA payments led to reductions in acreage under cultivation, which was likely to lead to a decrease in the demand for farm

[3]Prior to 1936, the first version of the AAA made rental and benefit payments to farmers who removed land from the production of designated crops. After the program was struck down as unconstitutional by the U.S. Supreme Court in 1935, the AAA was redesigned, to make "soil conservation payments" through the Soil Domestic Allotment Act (SDAA). In the original AAA the benefit payments were financed from special processing taxes on the commodity being curtailed. There was a general belief that most of the burden of the processing taxes would be passed on to consumers of farm products. After the Supreme Court declared the processing taxes unconstitutional, the SDAA eliminated the processing taxes and the funds were appropriated from the general budget.

[4]The AAA was administered by the Department of Agriculture, which established state and local committees or associations of producers to help administer the act. The administration of the Act was often done through a series of programs specific to the household crops. Thus the geographic distribution of the AAA funds across counties was determined by the crop choices made prior to the AAA involvement and by the parameters set for each of the crops. For each crop the actual distribution of funds was determined by a complex interaction between federal administrators, local committees, local extension agents, and the farmers who decided to join the program. Since this was a voluntary program, farmers had to agree to sign up for the acreage reduction program. For signing up to reduce acreage, their payments were based on multiplying the national price set for acreage reduction and their average yield per acre over a base period. Thus, the program had to be made attractive enough for farmers to agree to join. The federal decision makers influenced the attractiveness of the program by the national price they set for acreage reduction and by the acreage that they asked the farmers to take out of production. In the case of tobacco and cotton the federal decision-makers added a degree of coercion to the system by levying heavy taxes on any production beyond designated limits. The local administrators influenced the attractiveness of the program through decisions on base-year yields for the household farmer and the acreage the farmers would be allowed to produce. In addition, the effort they put into marketing the program and cajoling their neighbors helped determine the sign-up rates.

workers (Alston 1981; Holley, Winston, and Woofter 1971; Saloutos 1974; Mertz 1978; Whatley 1983; Biles 1994, 39-43, Fishback, Horrace, and Kantor 2006).

New Deal funds also were distributed to local economies through public works and relief grants. Relief grants were primarily distributed under the auspices of the Federal Emergency Relief Administration (FERA) from 1933 through mid 1935, the Civil Works Administration (CWA) from November 1933 through March 1934, the Works Progress Administration (WPA) from mid 1935 through 1942, and the Social Security Administration's Aid to the Blind, Aid to Dependent Children, and Old-Age Assistance programs after 1935. The principal goal of these programs was to provide immediate relief to unemployed and low-income people, as 85 percent of the grants were used to hire the unemployed on work relief jobs. These relief jobs ranged from make-work activities to maintenance activities to the building of sidewalks, post offices, schools, local roads, and other additions to local infrastructure.

The public works grants included expenditures by the Public Works Administration (PWA), Public Buildings Administration, and the Public Roads Administration. These grants were also used largely to employ workers. Many of the workers were hired directly from the relief rolls, but the public works programs had more freedom to hire a broader class of workers who were not on relief. The public works programs were said to be more focused on building larger scale projects such as dams, roads, schools, and sanitation facilities. Both public works and relief grants were likely to attract migrants to local areas because they provided either work opportunities on federal projects or support for the unemployed. This effect was mitigated to the extent that local relief administrators imposed residency requirements

The major relief and public works programs had the potential to stimulate migration across counties, as the unemployed sought work in areas with new relief and public works projects. The economics literature on the impact of welfare benefits on locational choice in the modern era is mixed, some find that movement of low-income people is positively correlated to differences in states' welfare benefit levels (Gramlich and Laren 1984, Blank 1988, Moffit 1992), while others find a small or negligible effect (Allard and Danziger 2000; Kauffman and Kiesling 1997, and Levine and Zimmerman 1999). We should note that our measure of relief and public works spending is total spending per capita, so it combines both differences in the number of people obtaining funds and the monthly payments to recipients of emergency jobs or direct relief. There were federal efforts to establish a certain minimum level of benefits, but the eventual compromise between officials at all levels was to pay attention to prevailing wage levels. Faced with extraordinary unemployment rates, relief officials were forced to make trade-offs between providing adequate benefits and finding work for as many unemployed workers as possible (see Brown 1940, Howard 1943, Williams 1968, Wallis and Benjamin 1981). Given the large number of unemployed workers, access to benefits might have been as important as the actual level of benefits.

Since the public works and relief projects involved not only relief of economic distress, but also led to expansions in civil infrastructure that potentially promoted economic activity in a deeply depressed national economy, we might expect to see more of a migration response in the 1930s than we would for federal welfare programs in the modern era. The migration response during the Depression, however, might have been limited by a complex web of residency requirements for relief eligibility. Unlike modern federal welfare programs that have largely eliminated residency requirements since 1970 (Gramlich and Laren 1984, 490), the residency requirements of the Depression-era relief programs were determined largely by state and local governments, sometimes in ways that seemed to violate the spirit of federal statements. Donald Howard (1943, 332-7) noted that the official WPA policy as of 1939 was that eligible people could not be refused certification for work relief jobs on the basis of non-residence in the area. At the same time, the WPA did not want families moving for the "sole purpose" of obtaining a relief job. Most of the barriers to movement were erected by state and local bureaucracies, which created elaborate procedures for transferring workers' records from one state to another and required that workers reestablish their eligibility in new places, among other factors. An unemployed worker took an additional risk by moving because state and local length-of-residency requirements for direct relief and public assistance may have differed. The de facto result might have been limits on non-residents' abilities to qualify for the WPA positions. On the other hand, to the extent that work relief projects stimulated the local economy, there may have been increased private opportunities for migrants.

The FERA policies for most types of relief were similar to the later WPA policies, although the FERA explicitly provided a small portion of its funds for the transient population. Josephine Brown (1940, 250) noted that federal FERA policy forbade discrimination against non-residents, blacks, aliens, and veterans, "yet the fact remained that the actual administration of relief was in the hands of local authorities and the promulgation of a rule by the FERA was not sufficient in many cases to overcome sectional traditions and prejudices in a comparatively short time." Aware of this problem, the FERA formulated a transient program for workers with less than a year's continuous residence (Williams 1968, 172-3). The program was funded by the federal government and administered by the states. It typically provided aid to the transient unemployed who could not have obtained aid under the legal settlement or residency requirements of the states (Webb 1936, 1-4, 16). The transient program accounted for about 2 percent of the total obligations of FERA programs (Federal Works Agency, Works Progress Administration 1942, 74 and 81), so in the final analysis the impact of FERA spending on migration patterns may not have differed much from that of the WPA. [5]

3.3 Migration and the Location Choice Literature

Sjaastad (1962) viewed migration as an investment in obtaining access to a labor market with higher wages. The moving costs are treated as the fixed costs of the

[5] The Civilian Conservation Corps often moved young men across states, but we do not have county level information on the CCC and, thus, cannot measure its impact in this study.

investment while the gain in earnings is the return. A household chooses to migrate if the present value of the migration is less than the cost of undertaking it. In his adaptation of the Roy (1951) model, Borjas et al.(1992) develops a theoretical model predicting that households will sort themselves into regions paying the highest return to their skills. They find that individuals who face a mismatch between their skills and rewards to their skills are more likely to migrate, i.e. highly educated individuals living in states with a relatively low return to education will migrate to a state with higher returns to education.

Later papers in the literature apply discrete choice models to location choice among migrants. A seminal paper in this literature was Bartel (1989), which studies the secondary migration choices of international migrants. In the context of a location choice model, she estimates the effects of ethnicity, population, social welfare programs, and distance on the probability of moving to a particular destination. She finds that the level of general assistance payments in a destination is positively and significantly correlated with a migrants probability of choosing the destination.

Schaefer (1989) analyzes location choices in the southern United States in the 1850's to examine non-slave owners' preferences over the racial composition of a destination county. Jaeger (2000) studies the location choice of newly arrived international migrants. He finds that migrants are more likely to locate in areas with disproportionately high foreign born populations. Herzog and Schlottmann (1986) used a binary logit model to study how amenities weigh in the decision to migrate or stay. Cragg and Kahn (1997) also use a conditional logit model of the destination choice of migrants. The structural nature of this model allows them to calculate migrants' willingness to pay for climate characteristics.

The common thread of these papers is the estimation of a discrete choice model that leads to a structural interpretation of the parameters, allowing the researcher to answer various questions, which could not be addressed in non-structural models. In this paper we estimate structural model of migration in order to calculate effects of

New Deal spending on migration patterns which would not be computed for a linear model of net migration.

3.4 Data

The data set is built up from a variety of sources, but we will focus on two sources here. The U.S. Office of Government Reports (1940) reported the distribution of New Deal grants for public works, relief programs, and AAA payments across counties. Meanwhile, the 1940 Population Census was the first decennial census to ask respondents about their place of residence in preceding years, in this case 1935. While information on previous county of residence is not available from the IPUMS, we are able to identify the "state economic area" (SEA) where the respondent resided in 1935.[6] Our sample consists of all native households for which location variables were available, giving us a total of 337,803 observations. We observe 466 SEA's that fully cover the continental states, excluding Washington, D.C, and the Alaska and Hawaii territories. In order to measure rates of migration, we construct a Markov transition matrix using the IPUMS extract of the 1940 Census 5 percent microsample.[7]

Figures 3.1 through 3.3 show that there is a significant amount of variation in migration rates and New Deal programs that can be used to identify the effects of the New Deal grants. Figure 3.1 plots the in-migration and out-migration per 100 people in 1930 for each of the SEAs. The figure shows quite a bit of variation in both types of migration rates and only hints at the information we use in estimating the model because it does not show the flows between SEAS, which are a key part of our model.

[6] The SEAs were developed by Donald Bogue to combine “relatively homogeneous” counties within a state (See Bogue and Beale (1953) and Bogue (1951). Bogue led a research team that analyzed how individual characteristics influenced the migration patterns across these regions. For example, see Bogue, et. al. (1957).

[7] A complete description of the sources of the New Deal spending variables and demographic, geographic, and economic characteristics of SEAs can be found at Price Fishback's website, http://www.u.arizona.edu/~fishback/ in the information on the construction of varaibles for the Fishback, Horrace, and Kantor (2005, 2006) migration and retail sales papers. The county-level data underlying the SEA aggregates for this study can be found there.

Figures 3.2 and 3.3 plot the net-migration rates against public works and relief grants per capita and AAA grants per capita. Net migration rates range from -20 percent to nearly 40 percent of the 1930 population in the SEAs, while public works and relief spending per capita varies from close to zero in some SEAS to over $300 per capita in some locations. The AAA distribution of grants was not as dispersed, but there still remains an extensive spread.

3.5 Identifying Program Effects

Our evaluation of the impact of New Deal spending is built up from a structural model of household choice. We start with a model of individual household choices and then show how an aggregate discrete choice model-one that uses shares of households making a decision rather than discrete decisions of individual households-can be used to estimate the parameters of this model.

3.5.1 Household Decision Model (CLOGIT)

Each household in our sample resided in one of 466 "State Economic Areas" in 1935. Between 1935 and 1940, each of these households decided where to locate in 1940: in the 1935 location, or in one of the 465 other SEAs in the country. Consider the following utility function for households.

$$U_{ij} = X_{oj}\beta + \xi_{oj} + \varepsilon_{ij} \tag{3.1}$$

The term U_{ij} represents the utility to the i^{th} household of residing in the j^{th} location in 1940. The utility for households from origin o is determined by the characteristics of the location j to households from origin o, X_{oj}. The product $X_{oj}\beta$ represents the utility the household receives from these characteristics, where β is a vector of marginal utilities.

The purpose for including the subscript o is that characteristics of locations, such as distance, will differ across origins. For example, the distance between Southern California and Northern California is significantly lower than the distance between Florida and Northern California. If we assume that the utility of distance traveled, β_{Distance}, is negative, then

$$X_{NorCal,SoCal} \cdot \beta_{\text{Distance}} > X_{Florid,SoCal} \cdot \beta_{\text{Distance}} \tag{3.2}$$

Thus, because one of the characteristics of Southern California as a migration destination (distance), varies between two different origins, Northern California and Florida, the utility of the location will differ across origins. Note however, that this utility does not differ among households from a common origin: the term $X_{oj}\beta$ assigns the same level of utility of locating in location j to all households from origin o.

Simply specifying the utility of the location as $X_{oj}\beta$ is unreasonable in two ways. First, it assumes that all characteristics that influence the decision of the household are observed by the econometrician. This assumption can be relaxed by introducing an error term that represents the utility of the location derived from characteristics observed by the household but not by the econometrician, ξ_{oj}.

While the introduction of the ξ_{oj} term eliminates the "omnipotent econometrician" assumption, the model still has one unreasonable property. It predicts that households from a common origin will all make the same location decision because the term $X_{oj}\beta + \xi_{oj}$ does not vary among households within the origin. The common way to relax this assumption is to introduce a random error term ε_{ij} that varies across both households and locations. These ε terms can be thought of as random draws from a distribution of consumer tastes in a population. Because these draws introduce heterogeneity into the model, all households from a common origin will no longer make the same locational choice. For example, suppose household 1 has a high unobservable taste for location 6, while household 2 does not. The difference implies that $\varepsilon_{16} > \varepsilon_{26}$, thus $U_{16} > U_{26}$.

The standard assumption about these error terms is that they are identically and independently distributed across both i and j. The ε terms are assumed to be from a Type One Extreme Value Distribution, which we discuss further below. This framework allows us to address the utility maximization problem in a way that is econometrically convenient while still making reasonable assumptions about consumer preferences.

The household will choose from the set of 466 destinations the location which provides the highest level of utility. Location j will be the utility maximizing location if

$$U_{ij} > U_{ik} \ \forall \ j \neq k \tag{3.3}$$

In terms of our model, we can express this term as

$$X_{oj}\beta + \xi_{oj} + \varepsilon_{ij} > X_{ok}\beta + \xi_{ok} + \varepsilon_{ik} \ \forall \ j \neq k \tag{3.4}$$

Isolating the "idiosyncratic taste" terms ε on the left hand side

$$\varepsilon_{ij} - \varepsilon_{ik} > X_{ok}\beta - X_{oj}\beta + \xi_{ok} - \xi_{oj} \forall j \neq k \tag{3.5}$$

The probability that the household will choose location j is then

$$\begin{aligned} \Pr(U_{ij} &> U_{ik}) \ \forall \ j \neq k \\ \Pr(\varepsilon_{ij} - \varepsilon_{ik} &> X_{ok}\beta - X_{oj}\beta + \xi_{ok} - \xi_{oj}) \ \forall \ j \neq k \end{aligned} \tag{3.6}$$

The payoff to the distributional assumption about ε comes from the result that the difference between two draws from the Type One Extreme Value Distribution takes the logistic distribution. Define the indicator variable Y_{ij} as a variable that takes on a value of 1 if household i chooses location j and 0 otherwise. McFadden (1973) shows that by integrating out over the distribution of the logistic distribution, we can obtain the following probabilities:

$$\Pr(Y_{ij}=1)=\frac{\exp(X_{oj}\beta+\xi_{oj})}{\sum_{J=1}^{J}\exp(X_{oj}\beta+\xi_{oj})} \tag{3.7}$$

Using this property of the model, we can construct the following likelihood function for β:

$$\ln L(\beta|Y,X,\xi)=\sum_{i=1}^{N}\sum_{j=1}^{J}Y_{ij}\cdot\frac{\exp(X_{oj}\beta+\xi_{oj})}{\sum_{J=1}^{J}\exp(X_{oj}\beta+\xi_{oj})} \tag{3.8}$$

This is equivalent to McFadden's choice model (Conditional Logit) with one important distinction: McFadden does not consider the ξ term. If the true value of ξ were known, it could be treated as an observable. However, because ξ is an unknown stochastic term, like the ε terms, we must integrate out over its distribution:

$$\ln L(\beta|Y,X,\xi)=\sum_{i=1}^{N}\sum_{j=1}^{J}Y_{ij}\cdot\int_{-\infty}^{\infty}\frac{\exp(X_{oj}\beta+\xi_{oj})}{\sum_{J=1}^{J}\exp(X_{oj}\beta+\xi_{oj})}\partial\xi \tag{3.9}$$

One important distinction between the ξ and ε terms is that we cannot obtain an analytical integral over the distribution of both variables. Hence we would have to compute the integral over ξ using numeric methods. This is most commonly accomplished through simulation. Taking simulation draws $s=1,2...,NS$ from the distribution of ξ, we can compute the average of the likelihood function at these simulated draws and then estimate the parameters by maximizing the analytical portion of the likelihood function:

$$\ln L(\beta|Y,X,\xi)=\sum_{s=1}^{NS}\sum_{i=1}^{N}\sum_{j=1}^{J}\frac{Y_{ij}}{NS}\cdot\frac{\exp(X_{oj}\beta+\xi_{soj})}{\sum_{J=1}^{J}\exp(X_{oj}\beta+\xi_{soj})} \tag{3.10}$$

Relaxing the i.i.d. Assumption in the Household Decision Model (NLOGIT): One problematic assumption made in this type of model is that the ε terms are *i.i.d.*

across j. This assumption implies that the idiosyncratic taste for one option in the choice set contains no information about the idiosyncratic tastes for other locations. This would be a unreasonable assumption if we believe that a household with a strong taste for a location such as Southern Arizona would not also be likely to have a high taste for a similar neighboring region, such as Central Arizona. If this assumption is violated, the integration McFadden uses to obtain the probabilities used in the likelihood function is invalid, and the model is mispecified.

A common way to address this problem is to create set of choices among which we expect there to be correlated tastes. For example, individuals who have a high unobservable utility from migrating to an SEA in one state may be more likely to have a high unobservable utility draw to migrating to any other SEA within the same state. Here, to account for this issue, we group choices by state. While the model maintains the assumption that the households' ε draws are independent among locations in different states, we allow there to be correlation among locations within a given state.

Cardell (1997) develops a model based on a random utility function that allows for "within group" correlation while maintaining the assumption of *i.i.d.* draws outside of the groups. Cardell expresses the idiosyncratic utility term as a weighted average of a term that is *i.i.d* over j, and a term which is common among choices in a group g_j (a given state).

$$U_{ij} = X_{oj}\beta + \xi_{oj} + \sigma\zeta_{ig_j} + (1-\sigma)\varepsilon_{ij} \tag{3.11}$$

For notational convenience in the following example, define the total idiosyncratic utility of the choice as.

$$\varphi_{ij} = \sigma\zeta_{ig_j} + (1-\sigma)\varepsilon_{ij} \tag{3.12}$$

The natural interpretation of the term σ is a measure of how strong the "within" state correlation among the φ terms are. When σ is equal to 1, the idiosyncratic utility term is identical among all locations within the state.

If σ is equal to zero, there will be no within group correlation in the φ terms. A σ close to zero suggests that there is no correlation between unmeasured and stochastic factors in the same state; therefore, household will be almost equally likely to choose a location outside the state as one inside the state, conditional upon observables. In other words, there is little reason to nest the choices at the state level in a conditional logit.

Consider how this model allows for correlation in the φ terms for two locations in the same state:

$$\begin{aligned} Cov(\varphi_{ij}, \varphi_{ik} || k &\in g \cap j \in g]) \\ &= E\{[\sigma\zeta_{ig} + (1-\sigma)\varepsilon_{ij}][\sigma\zeta_{ig} + (1-\sigma)\varepsilon_{ik}] | k \in g \cap j \in g]\} \\ &= E\{\sigma^2 \zeta_{ig}^2\} \\ &= \sigma^2 \cdot \sigma_{\xi}^2 \end{aligned} \tag{3.13}$$

Note, however, that there is no covariance between two terms that are not in the same state:

$$\begin{aligned} Cov(\varphi_{ij}, \varphi_{ik} || k &\notin g \cap j \in g]) \\ &= E\{[\sigma\zeta_{ig} + (1-\sigma)\varepsilon_{ij}] * [\sigma\zeta_{ig} + (1-\sigma)\varepsilon_{ik}] | k \notin g \cap j \in g]\} \\ &= 0 \end{aligned} \tag{3.14}$$

The probability of observing household i choosing location j now becomes the product of two terms. The first is the probability of household i choosing location j conditional upon choosing one of the locations in the same state as j, while the second term is the probability of the household choosing any SEA in the state to which SEA j belongs. Note that for a σ value of 0, this likelihood function is identical to that of the equation 3.10.

$$\begin{aligned}\Pr(Y_{ij} &= 1) = \Pr(Y_{ij|g_j} = 1) * \Pr(Y_{ig_j} = 1) \\ &= \frac{\exp(\frac{X_{ij}\beta}{1-\sigma})}{\sum_{j \in g_j} \exp(\frac{X_{ij}\beta}{1-\sigma})} * \frac{[\sum_{j \in G_j} \exp(\frac{X_{ij}\beta}{1-\sigma})]^{1-\sigma}}{\sum_{g=1}^{G} [\sum_{j \in g_j} \exp(\frac{X_{ij}\beta}{1-\sigma})]^{1-\sigma}}\end{aligned} \tag{3.15}$$

Complications to the Household Decision Model: There are a number of significant challenges to implementing this estimator, particularly when trying to resolve problems with potential endogeneity of the New Deal programs; therefore, we identify the parameters of the random utility function using the aggregate discrete choice model, rather than household data.

One problem with using individual households as the unit of observation is the sheer number of observations. The census data we use contains observations on the location choice of 337,803 households choosing between 466 SEAs. At each iteration of the MLE estimation, this would require over 150 million computations of the location specific utility for each household. For each of these 150 million calculations, our model would also require us to numerically integrate out over the ξ term using simulation methods. Clearly, this approach to estimation involves significant computational burdens.

An additional issue we must address is the possible endogeneity of New Deal program spending. From a statistical standpoint, it would be ideal if the New Deal grants were distributed on a random basis and, thus, could serve as a natural experiment. Of course, there is plenty of evidence that the New Deal grants were not distributed in a random fashion across areas. More per capita relief grants were distributed to areas with higher unemployment and greater drops in economic activity between 1929 and 1933. Although we control for some of these factors, we still need to worry about potential endogeneity bias in the coefficients if the New Deal grants were distributed in response to unmeasured characteristics of the location that influenced the attractiveness of the location to migrants. Instrumenting for program spending in the household based model would involve fully specifying the data generating process for ξ and integrating out over its distribution. A significant advantage of the aggregate discrete choice model is that it will allow us to instrument for New Deal grants using two-stage-least-squares (2SLS).

3.5.2 Using an Aggregate Discrete Choice Model

An alternative to estimating the household discrete choice model is the use of an aggregate discrete choice model, which is free of the complications of the prior model. The seminal paper developing this methodology is Berry (1994). Over the last ten years, the "Berry Inversion" has become common in the empirical industrial organization literature.

Recall the probabilities in the likelihood function for the conditional logit model:

$$\Pr(Y_{ij} = 1) = \frac{\exp(X_{oj}\beta + \xi_{oj})}{\sum_{J=1}^{J} \exp(X_{oj}\beta + \xi_{oj})} \tag{3.16}$$

Note that the the right hand term varies across o, but not across i. In this model individuals from the same origin have equal chances of migrating to location j. Accordingly, the probability that an individual from o moves to j can also be interpreted as the share of individuals from o moving to j. Define s_{oj} as the share of households from origin o moving to location j.

$$s_{oj} = \Pr(Y_{ij} = 1) = \frac{\exp(X_{oj}\beta + \xi_{oj})}{\sum_{j=1}^{J} \exp(X_{oj}\beta + \xi_{oj})} \tag{3.17}$$

Similarly, we can define the share of households from origin o who remain in the origin, s_{oo}, as

$$s_{oo} = \Pr(Y_{io} = 1) = \frac{\exp(X_{oo}\beta + \xi_{oo})}{\sum_{j=1}^{J} \exp(X_{oj}\beta + \xi_{oj})} \tag{3.18}$$

The log of the ratio of the terms in 3.17 and 3.18 is then

$$\begin{aligned}
\ln\left(\frac{s_{oj}}{s_{oo}}\right) &= \ln\left(\frac{\frac{\exp(X_{oj}\beta+\xi_{oj})}{\sum_{j=1}^{J}\exp(X_{oj}\beta+\xi_{oj})}}{\frac{\exp(X_{oo}\beta+\xi_{oo})}{\sum_{j=1}^{J}\exp(X_{oj}\beta+\xi_{j})}}\right) \\
&= \ln\left(\frac{\exp(X_{oj}\beta+\xi_{oj})}{\exp(X_{oo}\beta+\xi_{oo})}\right) \\
&= (X_{oj}-X_{oo})\,\beta + (\xi_{oj}-\xi_{oo})
\end{aligned} \tag{3.19}$$

We observe choices made by households in 466 SEAs. Each of these households chooses between the 466 locations in their choice set. The information needed to construct our dependent variable is essentially a transformation of a Markov transition matrix, (represented below) summarizing the movement of households among 466 locations between 1935 and 1940. Rather than requiring over 150 million computations (excluding the simulations over ξ) we now have 466*465=216690. The 466 cells along the diagonal of the matrix are the s_{oo} terms used in the normalization. Note that the model is now invertible in ξ; this is a key result which will allow us to estimate the parameters of the random utility function using conventional linear methods, such as 2SLS to instrument for endogenous X variables.

Migration Markov Transition Matrix

$$\begin{pmatrix}
s_{1,1} & s_{2,1} & s_{3,1} & \cdot & \cdot & \cdot & s_{465,1} & s_{466,1} \\
s_{1,2} & s_{2,2} & s_{3,2} & \cdot & \cdot & \cdot & s_{465,2} & s_{466,2} \\
s_{1,3} & s_{2,3} & s_{3,3} & \cdot & \cdot & \cdot & s_{465,3} & s_{466,3} \\
\cdot & \cdot & \cdot & \cdot & & & \cdot & \cdot \\
\cdot & \cdot & \cdot & & \cdot & & \cdot & \cdot \\
\cdot & \cdot & \cdot & & & \cdot & \cdot & \cdot \\
s_{1,465} & s_{2,465} & s_{3,465} & \cdot & \cdot & \cdot & s_{465,465} & s_{465,466} \\
s_{1,466} & s_{2,466} & s_{3,466} & \cdot & \cdot & \cdot & s_{466,465} & s_{466,466}
\end{pmatrix}$$

Relaxing the i.i.d. Assumption in the Aggregate Discrete Choice Model: Berry (1994) also derives an inversion of the nested logit model, whose likelihood function is given in equation 3.15. In his paper, the inversion relies upon an assumption that the "outside alternative" is a group to itself. Recall that in our model groups are states and the households' "outside alternative" is to choose not to migrate. Berry's assumption does not hold in our case as the option of staying in the home SEA is just one of several choices available for the household that chooses to remain in the home state (the group containing the "outside alternative"). In Appendix A we invert the nested likelihood model without relying upon this assumption, arriving at the same result as Berry. The additional term is the log of the share of the households from origin o choosing location j, conditional upon choosing one of the SEAs in the same state as j.

$$\ln\left(\frac{s_{oj}}{s_{oo}}\right) = \left(X_{oj} - X_{oo}\right)\beta + \sigma \cdot \ln(s_{oj|g_j}) + \left(\xi_{oj} - \xi_{oo}\right) \tag{3.20}$$

Instrumenting for New Deal Spending: Because migration flows, or unobserved variables correlated with migration, might have influenced the distribution of New Deal grants, there might be endogeneity bias in the coefficients on the New Deal variables. A priori, it is difficult to predict the direction or magnitude of the endogeneity bias. If out-migration was associated with economic distress during the 1930s, local officials may have sought greater New Deal funds from the federal government to alleviate the local unemployment situation and to stave off a continuing exodus of the workforce. Roosevelt's "relief, recovery, and reform" mantra would suggest that federal officials targeted funds to alleviate such economic problems. In fact, Fleck (1999a, 1999b, 2001a) and Fishback, Kantor, and Wallis (2003) find that both relief and public works spending were positively related to unemployment in 1930. To the extent that out-migration was a symptom of unfavorable economic conditions, we might expect federal officials to have distributed more funds to areas where people were more likely to leave than to arrive. Thus, the endogeneity bias might have been negative, causing the OLS coefficient to understate a positive effect that public works and relief spending might have had in attracting migrants.

Alternatively, the endogeneity bias could have gone the other way. Increased in-migration placed greater pressure on public facilities, such as schools and sanitation and water systems, which would have encouraged local officials to lobby for New Deal projects that would have alleviated these population pressures. In addition, if migrants into a county misestimated the employment opportunities in their new homes, their arrival might have contributed to greater unemployment and the need for federal New Deal assistance. However, the tendency for local relief officials to restrict non-residents' relief certification was likely to have mitigated this effect.

It is also likely that the AAA variable is endogenous, but the direction of the bias is unclear. Unlike the relief programs, the objective of the AAA was to limit national production of various commodities to raise farm-gate prices. The parameters were designed with national prices and production in mind and, therefore, were not explicitly tied to local problems. The officials' parameter choices, however, might have been indirectly influenced by local conditions because national AAA parameters depended on the need to raise prices for specific crops. Since crop mix varied substantially across the country and the distress in specific crops may have been felt more heavily in some areas than in others, local agricultural conditions may have indirectly influenced the policy parameters that determined the distribution of AAA funds. Thus, to the extent AAA officials were seeking to raise prices by reducing production, they may have seen reductions in production caused by the out-migration of farmers as a means in itself to limit supply and, thus, saw less of a need to provide AAA funds. Under these conditions, the OLS coefficient of the AAA variable is likely biased upward. On the other hand, federal officials may have seen out-migration as a sign of distress and, thus, more reason to find ways to prop up farmers in those areas. In this case the OLS coefficient would be biased downward.

One of the key advantages of the share-based model is that we can instrument for endogenous variables using the standard Two-Stage Least Squares (2SLS) procedure.

Relevant and valid instruments are correlated with the level of New Deal spending but uncorrelated with unobservables affecting the utility of the location.

There is an extensive literature on the political economy of the supply of New Deal program spending.[8] Robert Fleck (1999a), Fishback, Haines, and Kantor (2001), and Fishback, Horrace, and Kantor (2005b, 2006) have had success using some of these political supply variables as instruments in studies of unemployment statistics, birth and death rates, retail sales growth, and net migration. Based on these studies, the political supply equation for New Deal spending includes quite a few factors that might influence the household's choice to migrate to a location. These might include the typical party affiliations, the long-term structural features of the economy, the size of the population, the ethnic structure of the population, opportunities for home ownership and other socioeconomic factors. These factors are already incorporated in the vector of attributes associated with locations in the migration share analysis. There are other factors, however, that either would not be expected to influence the attractiveness of locations to migrants, or would have influence only indirectly through incomes and the other socioeconomic factors already included as attributes in the migration analysis.

There are two political supply factors that potentially influenced public works and relief spending but would have only indirectly influenced the choice to migrate to a location, the land area in the SEA and the share of the population that voted in the 1928 election. Land area and population were key factors in the formula for distributing road funding and larger counties and states in terms of square mileage tended to receive larger Public Works Administration grants (Fleck 2001b, Fishback, Kantor, and Wallis, 2003). The size of the SEAs in square miles is determined by the

[8] For discussions of the determinants of New Deal spending, see Reading 1973; Wright 1974; Wallis 1987, 1998, and 2001; Anderson and Tollison 1991; Couch and Shughart 1998; Couch, Atkinson, and Wells, 1998; Fleck 1999a, 1999b, 2001a, 2001b; Couch and Williams 1999; Stromberg 2004. Fishback, Kantor, and Wallis 2003. The last paper summarizes the results of all of the studies and provides new estimates for a broad range of programs..

counties' geographic boundaries, which were typically set long before the 1930s, and Bogue's combination of counties into relatively homogeneous SEAs. After controlling for a wide variety of socioeconomic factors and for in-variant state effects, there is no apparent reason why migrants would have additional reasons to care about the amount of land in the county boundaries and then how many were grouped into the SEAs.

Robert Fleck (2001a) found that the relief program distribution in the South was strongly influenced by the share of the population voting, as politicians rewarded areas with more politically active voters. There is already temporal distance between the migration decisions from 1935 to 1940 and the share of the population voting in 1928. If we had estimated the model with no other correlates, we might have anticipated that the voting share would have influenced migration patterns, particularly for blacks who sought to find areas where they could vote. However, we already control for the percent black, the state policies toward voting, the mean share voting Democrat for president, retail sales per capita, and other factors that would be the actual channels through which voting activity in 1928 would have influenced migration. Further, as explained more below, we also employ a fixed effects estimation to control for the unobservable taste that each migrant has for his origin.

In addition, we include a key political supply variable that we expected influenced AAA spending but at best would have influenced migration decisions only indirectly through other variables: average farm size in 1929. Average farm size strongly influenced the AAA distribution at the county level (Fishback, Horrace, and Kantor 2006; Fishback, Kantor, and Wallis 2003). The AAA has always been know to have been tilted toward large farms. The average farm size measure is for the period prior to the New Deal and thus farmers had no opportunity to change this size in anticipation of obtaining New Deal funds. In addition, we are controlling for a variety of factors including retail sales per capita, crop output per acre, crop failures, urbanness, and state effects that would have been the channels through which migrants would have cared about farm size.

Since the true errors in the migration share equation are unknown by definition, we can never be sure that these identifying instruments are uncorrelated with the error term. We can, however, use the Hansen J-statistic to test for correlation between the identifying instruments and the estimated error term in the migration share equations.

The first-stage equations in Table 3.3 show that the coefficients of all of the identifying instruments except the percent voting variable are statistically significantly different from zero with the expected sign. Larger SEA land area was associated with higher than average per capita public works and relief spending in the SEA and larger average farm size in 1929 was associated with higher AAA spending. The coefficient of percent voting was unexpectedly negative and not statistically significant in the public works and relief equation and negative and statistically significant in the AAA equation. The F-statistics in the first stage are 5.15 for public works and relief and 7.53 for the AAA variable, both statistically significant at the 99.9% level The Hansen J-statistic shows that we cannot reject the hypothesis of no relationship between the indentifying instruments as a group and the estimated error term in the final migration share equation.

Instrumenting for the Within-Group Term: The inversion of the nested logit model introduces an additional endogenous variable into the model. The term $\ln(s_{oj|g_j})$ is endogenous if it is correlated with ξ. This will be the case when unobservables affecting the share of individuals choosing location j also affect the share of individuals who choose j conditional upon choosing one of the locations within the same state of j, that is, when $\ln(s_{oj|g_j})$ is correlated with ξ.

Berry (1994) suggests that the number of choices within the nested group might work as an instrument for the endogenous within group term. The validity of this instrument hinges upon the assumption that the number of SEAs in a state is exogenously determined by the historical drawing of geographic boundaries. Given that the number of counties was set many years before and that Bogue combined counties into SEAs based primarily on factors that we have already controlled for in the share equation, this seems like a reasonable assumption.[9]

The intuition for the relevancy of this instrument is best understood with a simple example. When there are only two SEAs a_1 and b_1 in state g_1 the probability that a_1 is the utility maximizing choice is the probability that $a_1 > b_1$: $\frac{\exp(X_{a1}\beta)}{\exp(X_{a1}\beta)+\exp(X_{b1}\beta)}$ When state g_2 contains three SEAs: a_2, b_2, and c_2, the probability that a_2 is the utility maximizing choice, holding all else constant, is the probability that $a_2 > b_2$ and $a_2 > c_2$: $\frac{\exp(X_{a2}\beta)}{\exp(X_{a2}\beta)+\exp(X_{b2}\beta)+\exp(X_{c2}\beta)}$ a smaller number: Thus, we anticipate that the $\ln(s_{oj|g_j})$ variable would be negatively related with the number of SEAs in a state.

Further Controls for Unmeasured Heterogeneity Across SEAs: There may be features of the SEAs that are observable to the migrants for which we have no measures, denoted ξ in the model. We observe households from many origins o choosing whether or not to migrate to each SEA. By taking the means for all observations for an SEA that have that SEA origin, we can use a mean-differencing approach to control the unobservable features that influence the utility of staying in that SEA (ξoo). Recall the econometric equation given for the inversion of the conditional logit.

$$\ln\left(\frac{s_{oj}}{s_{oo}}\right) = (X_{oj} - X_{oo})\,\beta + (\xi_{oj} - \xi_{oo}) \tag{3.21}$$

For each origin o, we sum this expression over locations j and scale by $\frac{1}{J-1}$ (465, the number of locations outside of the origin):

[9]This is a relatively stronger assumption to make in the context of the IO literature. It essentially means assuming that unobserved product characteristics will have no impact on entry. However, in a geographic context, it is easier to believe that high draws of the ξ term does not cause new geographic entities to "enter" the choice set.

$$\begin{aligned}
\frac{1}{J-1}\sum_{j\neq o}\ln\left(\frac{s_{oj}}{s_{oo}}\right) &= \frac{1}{J-1}\sum_{j\neq o}\left[\left(X_{oj}-X_{oo}\right)\beta+\left(\xi_{oj}-\xi_{oo}\right)\right] \\
\overline{\ln\left(\frac{s_{oj}}{s_{oo}}\right)} &= \left(\overline{X_{oj}}-X_{oo}\right)\beta+\left(\overline{\xi_{oj}}-\xi_{oo}\right) \qquad (3.22)
\end{aligned}$$

Subtracting this term from equation the 3.21 results in the econometric specification for fixed effects estimation:

$$\begin{aligned}
\ln\left(\frac{s_{oj}}{s_{oo}}\right)-\overline{\ln\left(\frac{s_{oj}}{s_{oo}}\right)} &= \left[\left(X_{oj}-X_{oo}\right)\beta+\left(\xi_{oj}-\xi_{oo}\right)\right]-\left[\left(\overline{X_{oj}}-X_{oo}\right)\beta+\left(\overline{\xi_{oj}}-\xi_{oo}\right)\right] \\
&= (X_{oj}-\overline{X_{oj}})\beta+\xi_{oj}-\overline{\xi_{oj}} \qquad (3.23)
\end{aligned}$$

The mean differencing operation results in elimination of both observables and unobservables of the origin from the specification. This operation helps obtain consistent estimates of the β parameters because one possible source of endogeneity, correlation between ξ_{oo} and X_{oo} has been eliminated.

The fixed effects estimator for the nested logit inversion is nearly identical to the fixed effect estimator for the conditional logit inversion. The only difference is that we must also include a mean differenced within group share term.

$$\ln\left(\frac{s_{oj}}{s_{oo}}\right)-\overline{\ln\left(\frac{s_{oj}}{s_{oo}}\right)}=(X_{oj}-\overline{X_{oj}})\beta+\sigma\cdot[\ln(s_{oj|g_j})-\overline{\ln(s_{oj|g_j})}]+\xi_{oj}-\overline{\xi_{oj}} \qquad (3.24)$$

Heteroscedasticity and Clustering: It can be shown that there is covariance between the error terms that share a common origin. Because of this, our estimates of the standard errors on the parameters of our model may be inconsistent. To address this problem, we cluster standard errors for observations sharing a common origin.

Dealing with "Zero Shares": One issue of concern in the estimations is the prevalence of origin-destination pairs for which no migrations are observed. This is problematic in our model, because we must take the log of the share of individuals choosing this destination, and natural log of zero is undefined. To avoid this problem, the value 0.0000001 was added to all of the shares before they were logged.

3.6 Results

The nested logit specification of the model is estimated using both Ordinary Least Squares (OLS) and Two-Stage Least Squares (2SLS). The variables used in our model are described in Table 3.2. The results are reported in the first two columns of Table 3.3.

The attractiveness of the different New Deal Grants are quite different for the AAA farm payments and the public works and relief expenditures. In all four of the reported specifications an increase in AAA spending per farm population in the SEA has a negative and statistically significant impact on the utility of living there. On the other hand, increases in public works and relief grants to an SEA increased the SEA's attractiveness. The positive effect is larger and statistically significant when instruments are used to control for endogeneity. These results are similar to the findings by Fishback, Horrace, and Kantor (2006) in analysis of net-migration between 1930 and 1940 at the county level. Their analysis was tightly constrained by the absence of information on flows between locations and jurisdictions, which limited what they could say about in-migration and out-migration decisions.

The OLS estimate of the σ parameter measuring the correlation of the ε terms for choices in a given state is .213, which is statistically significant. Although this would imply that the CLOGIT model is not the correct specification, a positive bias in the OLS estimate of σ is anticipated because the share of individuals choosing to locate in destination SEA j, conditional upon migrating to that state is expected to be positively correlated with the overall share of individuals moving to SEA j. Additionally, we would expect that unobservables that increase the share of individuals choosing j conditional upon choosing j's state would also positively effect the overall share of people choosing j. When we instrument for the within state share term, the estimate of sigma is cut sharply to a value of 0.006, that is not statistically significantly different from 0; thus we proceed using the CLOGIT functional form of the formulas used to compute marginal effects and counterfactuals.

Many of the socioeconomic features of the SEAs have the expected influence on the SEA's attractiveness. Households were more likely to choose SEAs that were nearer, that had experienced less of a downturn between 1929 and 1933, had greater economic activity in the form of retail sales purchases and crop values per capita in the peak year of 1929, experienced fewer layoffs per capita in 1930, had higher populations in 1930 and greater population growth in the 1920s. The main surprises were the positive effect of greater unemployment in 1930 and the absence of a strong effect of the Dust Bowl.

3.6.1 Marginal Effects of Program Spending on Migration Patterns

The point estimates in this model provide intuition about whether the characteristics make the SEA more or less attractive to households. Given the logistic character of the model and the large number of flows between SEAS, the point estimates cannot be read as marginal effects. The parameters of the random utility function instead can be used to examine three types of marginal effects of the spending of an additional dollar of New Deal spending per capita: retention, creation, and diversion. People who otherwise would have migrated are *retained* in origin o if the increase in program spending in o causes them to remain in o rather than to migrate. Migration to o is *created* when the increase in spending causes people who otherwise would have remained in their origin SEA j to choose instead to migrate to o. Finally, migration is *diverted* to o when the spending increase in o causes people who otherwise were migrating to another destination k to move instead to destination o. Thus, if we consider the overall effect of program spending on the level of net migration to an SEA, *retention* is a decrease in out-migration, while the sum of *diversion* and *creation* is an increase in in-migration.

Table 3.1 shows the formulas for calculating the marginal retention, creation, and diversion effects associated with a one dollar per capita increase in spending by a New Deal program. Consider the effects for the Boston SEA, denoted with subscript o in Table 3.1. The *retention* effect is associated with people who would have normally migrated from Boston but now stay in Boston due to a change in program spending. In other words, this is a reduction in out-migration as people substitute staying at home for migration. The marginal retention effect varies for each SEA based on the share of people in the SEA choosing to stay home. Specifically, the share of people retained is calculated as the coefficient β of New Deal spending per capita multiplied by the share of people from Boston staying in Boston (s_{oo}) multiplied by one minus that same share ($1 - s_{oo}$). This value is then multiplied by the number of people in Boston in 1930 (N_o) to convert the change in share into the change in the number of people retained.

The increase in the number of in-migrants Boston from any given SEA k as a result of a dollar increase in per capita program spending in Boston is calculated by multiplying the New Deal coefficient β by the share of people from SEA k moving to Boston (s_{ok}), by one minus that same share ($1 - s_{ok}$), and by the number of people in SEA k in 1930 (N_k). To get the total increase in the number of in-migrants to Boston from all other SEAS, sum the number from each SEA k across all of the SEAS except Boston.

The increase in in-migrants is the sum of the creation and diversion effects. The creation effect captures people who would not have migrated without the increase in spending, but who are pulled out of their home SEAs and choose to migrate to

Boston in response to the rise in New Deal spending there. The share of in-migrants created is found by multiplying the New Deal coefficient β by the share of people from k who choose to stay in k (s_{kk}) and by the share of people in k who choose to move to Boston (s_{ko}). By multiplying this number by the population of SEA k (N_k), we obtain the total number of migrations from k to Boston. Summing this term over all SEAs beside Boston gives the total number of migrants to Boston who would have stayed home in the absence of the additional spending in Boston.

Migration diversion to Boston consists of people who were already migrating to another SEA (say Providence) but switch to Boston in response to the rise in New Deal spending there. To compute the share of individuals from origin k who are diverted from Providence to Boston, the New Deal coefficient β is multiplied by the share of individuals from SEA k who choose to migrate to Providence and by the share of individuals from SEA k who choose to migrate to Boston. Multiplying this number by the population of SEA k, we obtain the total number of migrants from SEA k who are diverted from Providence to Boston. After calculating this figure for origin k for all SEAs other than Boston and k, we sum over all those SEAs to obtain the total number of households from origin k who are diverted to Boston. To find the total amount of diversion to Boston for people migrating out of all origin SEAs, we compute the preceding figure for all SEAs other than Boston and sum across all those origins.[10]

Since the marginal effects are determined in part by the migration shares, the marginal effects are different for each SEA. A list for each SEA is available from the authors. Table 3.4 and Figures 3.4 and 3.5 give the averages of the marginal effects on migration flows as a percentage of the population for the SEAs in each of the nine Census sub-regions for Public Works and Relief, as well as the average of the

[10] If we sum the creation effect for all SEAs and sum the retention effect for all SEAs, the two sums will be equal. This occurs because of adding up restrictions caused by the fact that internal migration flows are zero-sum. Flows out of one SEA become flows into another SEA. For a mathematical proof see Appendix B.

raw effect on population flows into these regions. Table 3.5 and Figures 3.6 and 3.7 present the same figures for the effect of the AAA.

The region where migration rates were most responsive to a marginal dollar of New Deal spending was New England. An additional dollar of public works and relief spending per capita in the average New England SEA kept about 0.23 percent of the 1930 population from out-migrating, created new in-migration from other SEAs equal to about 0.81 percent of the population, and diverted enough migrants from other SEAs to this one to raise in-migration by another 0.17 percent of the 1930 population. These marginal effects were the highest in all three categories for any region, typically many-fold larger than the marginal effects in the Pacific and Mountain regions.

A similar story is true for AAA spending, although the effects are in the opposite direction. An additional dollar of AAA spending per person on farms in New England had a retention effect of -0.38, which raised the out-migration rate by 0.38 percentage points. About 0.29 percent of the already migrating population was diverted away from the typical New England SEA to other SEAs and 1.35 percent of the population in other SEAs was prevented from migrating to the New England SEA. As was the case for the public works spending, these marginal effects are several times larger than the marginal effects in the far west regions.

These findings suggest that New England households considered the remaining SEAs to be closer substitutes to their own SEA than did households in SEAs in the far west. Actual in- and out-migration rates in New England SEAs were roughly one-third to one-fourth of the levels of the SEAs in the far west. These two sets of facts are consistent with a situation where New Englanders considered many SEAs to be similar to their own 1935 SEA and thus saw little reason to move unless there was an external policy change that would widen the differential between SEAs. In the far west, on the other hand, there was much greater variation in the unobservables that caused people to be more likely to move, so that a single dollar change in New Deal spending did not generate as much of a response.

3.6.2 Policy Simulations

In addition to estimating the marginal effects of a change in a variable on migration flows, we can also estimate what migration patterns would have looked like under various counterfactuals. In such a simulation, we can set $\mathring{X}_{oj}$ to a counterfactual value, and compute estimated out-migration flows for each SEA o to each destination j.

$$\mathring{s}_{jo} = \frac{\exp(\mathring{X}_{jo}\hat{\beta})}{\sum_{j=1}^{J} \exp(\mathring{X}_{jo}\hat{\beta})} \tag{3.25}$$

By multiplying the estimated shares by the population and then summing across all destinations we find the number of out-migrants for each SEA. And then sum across all SEAs to obtain the total number of households out-migrating for the entire country.

$$\mathring{M} = \sum_{j=1}^{J} \sum_{o=1}^{J} \mathring{s}_{jo} * N_o \tag{3.26}$$

Since all out-migrants are in-migrants elsewhere, this would also be equal to the total number of in-migrants. We then obtain the number of migrants in each of the three counterfactual settings by inserting zero for the relevant New Deal spending. The benchmark for our counterfactual analysis is the number of migrations predicted by the model at the actual levels of New Deal spending.

Table 3.6 presents the results of our counter factual estimations. While the migration rates in the IPUMS 5% sample imply a total of over 9 million total migrations between SEAs from 1935 to 1940, the model with actual New Deal spending predicts a total of 13.7 million out-migrations of households. Removing only the AAA grants leads to a modest decline of about 22,000 migrations, or a 0.16% decline in migrations. Removing public works and relief spending, on the other hand, leads to a prediction of only 10,896,388 migrations, a decrease of 20.58% from the predicted amount with all New Deal spending. Nearly a fifth of all migrations that took place in the in the United States from 1935 to 1940 were a direct result of the way in which public works and relief grants were distributed. Had these programs not existed, or had policy makers distributed funds in such a way that per-capita levels of funding were equal across SEAs, only 80% of migrants would have continued to move.

If New Deal spending on both the AAA and public works and relief had been completely eliminated, there would have been 11,440,681 migrations, which is 16 percent fewer migrations than the predicted number with the full amount of New Deal spending. Note that the highly interactive nature of the flows of households between SEAs leads to a combined effect of New Deal grant spending that is different from the sum of the separate AAA and public works and relief effects.

3.7 Conclusion

The simulations from a structural model of migration flows between state economic areas during the Second New Deal shows that the uneven distribution of New Deal grant spending helps account for approximately 20 percent of the migration flows witnessed between 1935 and 1940. The majority of the effect is associated with the broad range in spending per capita on public works and relief by the federal government. As found by Fishback, Horrace, and Kantor (2006) in studies restricted to net migration rates by county, public works and relief tended to increase migration into a state economic area and reduce out-migration from the area. The AAA programs designed to restrict farm acreage contributed to a small degree of out-migration. The overall effect of the AAA on migration was relatively small.

These effects appear to be large relative to the effects found for modern migration in response to differences in welfare spending. There are several potential reasons for these differences. First, the distribution of spending in the 1930s was much less tightly tied to matching formulas that force states to spend more to obtain more welfare funding. The negotiations between Congress and the Roosevelt administration in passing the Social Security Act of 1935 ensured that the permanent federal/state public assistance programs, like old-age assistance, aid-to-the-blind, and aid to dependent children required matching formulas. This matching grant feature imposed constraints on differences across states, as a state who sought higher expenditures had to dip into their own revenues as well. The compromise meant that the Roosevelt administration obtained much greater control over the emergency relief funds distributed under the WPA and public works programs that ended during World War II (Wallis, Fishback, and Kantor 2006). Studies at the time show that despite the presence of de jure matching formulas in the WPA, the formulas appear to have been ignored in fact (Howard, 1943). Thus, the range in spending per capita was much larger. Second, the unemployment and underemployment rates during the 1930s were several orders of magnitude larger than in the modern era and the standard of living was lower. Therefore, the same amount of federal spending in the 1930s as in the modern era was a larger share of the person's average income. Much of the spending was designed to put people explicitly back to work and in the case of public works spending in a number of cases required movement of people to relatively less populated areas. Thus, it might be more relevant to compare New Deal spending programs with the spending on interstate highways or modern public works.

There are several additional issues we would like to address. The AAA was likely to have quite different effects on laborers, croppers, and low level tenants than on farm owners. We have information on whether people were in agriculture in 1935 and also what their occupation status was in 1940. We would like to re-estimate the model by restricting the households to farm households in 1935 and to people categorized as low-skilled workers as of 1940, to see if the AAA had a strong effect on out-migration for those groups. We are working on setting up the data for that purpose.

Another issue we would like to address is how to deal with outliers for New Deal spending per capita in the parameters of the model. The SEAs for Nevada, Kitsap county in Washington, and Cape Cod and Martha's Vineyard on Massachusetts have very high public works and relief spending levels per capita, for example, and they may be introducing problems in estimating the data. We plan additional analysis without these locations to see whether they are driving some of the results.

Migration Retention by o	$N_o(1-s_{oo})*s_{oo}*\beta$
Total Increase in In-Migration to o	$\sum_{j\neq o} N_j*(1-s_{jo})s_{jo}*\beta$
Migration Diversion to o	$-\sum_{j\neq o} N_j*[\sum_{k\neq j,o} -s_{jk}s_{jo}*\beta]$
Migration Creation to o	$-\sum_{j\neq o} N_j*-s_{jj}s_{jo}*\beta$
Increase in Net Migration to o	$\sum_j N_j*(1-s_{jo})s_{jo}*\beta$

Table 3.1: Marginal Effects Decomposed By Parts

NEW DEAL GRANTS AND LOANS	
PCPWREL	*Per Capita Relief and Pub Works 1935-1939*
PCAAA	*Per Capita AAA Spending 1935-1939*
IDENTIFYING INSTRUMENTS	
PCTVT28	*Votes in 1928 Election Per by Adult Pop*
AVFARMSIZE	*Average Farm Size*
NGL	*Number of SEAs in State of Origin L*
OTHER COVARIATES	
LNDIST	*Log of Distance Between SEA Centroids*
RLDF3329	*% Change in Per Capita Retail Sales, 1929-1*
RTSAPC29	*Retail Sales Per-Capita, 1929*
PUNPOP30	*Unemployed As % of Pop Aged 10+, 1930*
PLAYOFF	*% Layed off 1930*
POWN30	*% of Families Owning Their Home*
PCTFAIL	*% of Farm Acres with Failures, 1929*
DUSTBOWL	*Dummy for Dustbowl County*
CROPVAL	*Crop Value per Farm Population*
MANUFACT	*% of Adult Pop in Manufacturing Employment*
LPOP3020	*% Change in Population, 1920-1930*
POP30	*Population, 1930*
PCTBLK3	*% Black in 1930*
PCTURB3	*% Urban in 1930*
PRURNF3	*% Rural Non Farm 1930*
PFORB3	*% Foreign Born in 1930*
PCTILL3	*% Illiterate Over Age 10, 1930*
MEAN9628	*Mean of Dem % Vote for President, 1896-192*
PCHURCH26	*% Population Church Members in 1930*

Table 3.2: Variable Descriptions

	OLS		IV Main Equation		First Stage PW Relief		Frist Stage AAA		First Stage Within Group	
	Coeff.	*t-stat.*	*Coeff.*	*t-stat.*	*Coeff.*	*t-stat.*	*Coeff.*	*t-stat.*	*Coeff.*	*t-stat.*
Log of share choosing SEA conditional on migrating out of state	0.213	30.28	0.006	0.17						
AAA per Capita	-0.024	-7.29	-0.024	-3.53						
Public Works and Relief per Capita	0.001	0.68	0.014	2.74						
% voting for president in 1928					-0.044	-0.11	-0.052	-1.99	-0.002	-0.51
Land Area of SEA					0.007	3.06	0.000	-1.24	0.000	-2.7
Average Farm Size, 1929					0.064	1.6	0.100	4.43	0.006	5.45
Number of SEAs					-0.605	-2.09	-0.069	-1.87	-0.202	-27.12
Log (Distance) between SEAs	-1.747	-46.49	-1.405	-21.99	0.894	0.93	-0.228	2.51	1.724	40.42
Retail Sales Growth Rate, 1929-1933	0.099	0.52	0.534	2.15	-22.376	-2.14	-1.304	-0.9	-0.331	-1.21
Retail Sales per Capita, 1929	0.002	3.34	0.002	2.17	-0.034	-0.62	0.001	0.11	-0.001	-1.26
% unemployed, 1930	0.171	3.55	0.117	1.64	-0.979	-0.24	-0.140	-0.42	-0.131	-1.68
% laid off, 1930	-9.267	-1.62	-17.909	-2.99	310.641	0.81	-30.504	-0.88	16.090	1.97
% home owners, 1930	0.007	1.98	-0.008	-1.51	0.572	2.2	-0.100	-2.59	-0.015	-2.97
% acres with farm failures, 1929	0.051	3.52	0.001	0.04	2.130	2.89	-0.056	-0.4	-0.064	-3.06
Dust Bowl	0.310	0.85	0.450	0.7	31.019	1.75	63.714	4.92	0.551	0.98
Crop Value per Acre, 1929	0.004	7.41	0.003	4.21	-0.094	-2.34	0.002	0.2	-0.007	-6.37
% manufacturing, 1930	-0.017	-3.19	0.004	0.4	-1.471	-4.84	-0.034	-1.27	0.013	2.21
Population growth, 1920-1930	1.003	3.4	1.706	3.54	-55.320	-3.34	8.059	4.12	0.331	1.54
Population, 1930	0.000	1.81	0.000	2.06	0.000	-0.82	0.000	0.59	0.000	4.59
% Black, 1930	-0.008	-2.59	-0.001	-0.27	-0.134	-0.8	0.045	1.83	0.007	1.81
% Urban, 1930	-0.002	-0.5	-0.001	-0.13	-0.069	-0.27	-0.024	-1.19	0.005	1.11
% farm population, 1930	-0.056	-0.14	2.123	2.2	-136.212	-3.96	23.717	4.61	1.547	2.95
% foreign born, 1930	0.033	4.37	0.007	0.46	1.690	2.88	0.073	1.19	-0.066	-5.9
% illiterate, 1930	0.046	5.17	0.003	0.18	0.622	0.99	-0.419	-3.49	-0.071	-4.78
% church members, 1926	-0.010	-4.91	0.002	0.45	-0.471	-3.34	0.077	3.4	0.016	5.93
Mean % voting Democrat for President, 1896-1928	0.008	2.53	0.002	0.57	0.251	1.11	-0.047	-1.83	-0.008	-2.16
N	216690		216690		216690		216690		216690	
R^2	0.2754									
Partial R^2 on Instruments					0.0758		0.3956		0.1046	
F on Instruments					5.15		7.53		199.11	
Pvalue of F-statistic					0.0005		0		0	
Hansen J, pvalue			0.973							

Table 3.3: Nlogit Fixed Effects Estimation With Robust Standard Errors

	National	**New England**	**Middle Atlantic**	**East North Central**	**West North Central**	**South Atlantic**	**East South Central**	**West South Central**	**Mountain**	**Pacific**
Effect on Migration Rates										
Migration Retention	0.15%	0.23%	0.22%	0.17%	0.12%	0.18%	0.17%	0.11%	0.05%	0.06%
Increase in Immigration	0.25%	0.99%	0.23%	0.22%	0.12%	0.35%	0.13%	0.13%	0.26%	0.26%
Migration Diversion	0.04%	0.17%	0.04%	0.03%	0.02%	0.06%	0.02%	0.02%	0.03%	0.02%
Migration Creation	0.21%	0.81%	0.19%	0.18%	0.11%	0.29%	0.11%	0.12%	0.23%	0.23%
Total Effect	0.40%	1.22%	0.45%	0.38%	0.24%	0.53%	0.31%	0.24%	0.31%	0.32%
Raw Effect										
Migration Retention	348	736	749	370	252	355	370	219	61	111
Increase in Immigration	410	661	1145	452	221	389	238	226	250	225
Migration Diversion	62	124	180	69	28	65	35	27	29	23
Migration Creation	348	538	965	383	193	323	203	200	221	202
Total Effect	757	1397	1894	822	472	744	608	445	311	336
Decomposition										
Retention	45.91%	52.67%	39.55%	44.99%	53.26%	47.77%	60.82%	49.17%	19.68%	33.11%
Diversion	8.19%	8.86%	9.50%	8.43%	5.93%	8.78%	5.84%	6.01%	9.39%	6.80%
Creation	45.91%	38.47%	50.95%	46.58%	40.80%	43.45%	33.34%	44.82%	70.94%	60.09%

Table 3.4: Marginal Effects for PW and Relief

	National	New England	Middle Atlantic	East North Central	West North Central	South Atlantic	East South Central	West South Central	Mountain	Pacific
Effect on Migration Rates										
Migration Retention	-0.25%	-0.38%	-0.36%	-0.28%	-0.20%	-0.31%	-0.29%	-0.18%	-0.08%	-0.10%
Increase in Immigration	-0.42%	-1.63%	-0.37%	-0.36%	-0.21%	-0.57%	-0.22%	-0.22%	-0.44%	-0.43%
Migration Diversion	-0.06%	-0.29%	-0.07%	-0.05%	-0.03%	-0.10%	-0.03%	-0.03%	-0.05%	-0.04%
Migration Creation	-0.35%	-1.35%	-0.31%	-0.30%	-0.18%	-0.48%	-0.19%	-0.20%	-0.39%	-0.39%
Total Effect	-0.67%	-2.01%	-0.74%	-0.64%	-0.40%	-0.88%	-0.51%	-0.40%	-0.52%	-0.53%
Raw Effect										
Migration Retention	-576	-1219	-1241	-612	-417	-589	-612	-363	-101	-184
Increase in Immigration	-679	-1095	-1896	-749	-366	-644	-394	-375	-414	-372
Migration Diversion	-103	-205	-298	-115	-46	-108	-59	-44	-48	-38
Migration Creation	-576	-890	-1598	-634	-319	-535	-335	-331	-365	-334
Total Effect	-1254	-2314	-3137	-1361	-783	-1232	-1006	-738	-515	-556
Decomposition										
Retention	45.91%	52.67%	39.55%	44.99%	53.26%	47.77%	60.82%	49.17%	19.68%	33.11%
Diversion	8.19%	8.86%	9.50%	8.43%	5.93%	8.78%	5.84%	6.01%	9.39%	6.80%
Creation	45.91%	38.47%	50.95%	46.58%	40.80%	43.45%	33.34%	44.82%	70.94%	60.09%

Table 3.5: Marginal Effects for AAA

	$\overset{o}{M}$	**New Deal Effect** Raw	%
Actual	$9,062,751$		
Estimated with New Deal Counterfactuals	13,720,403		
No PW and REL	10,896,388	2,824,015	20.58%
No AAA	13,742,386	-21,983	-0.16%
No ND	11,440,681	2,279,722	16.62%

Table 3.6: Counterfactuals

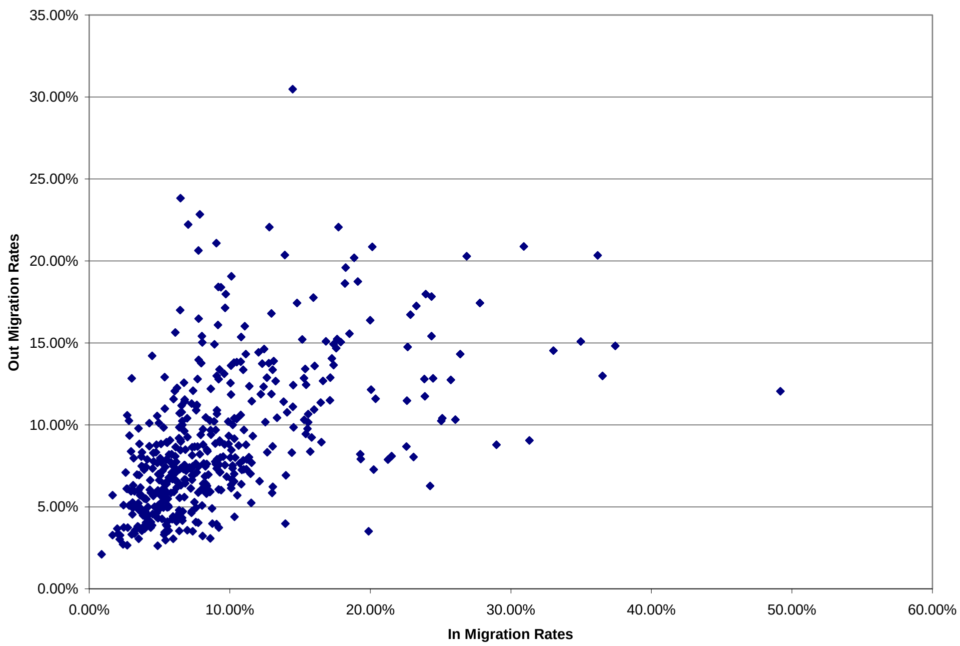

Figure 3.1: In Migration and Out Migration Rates

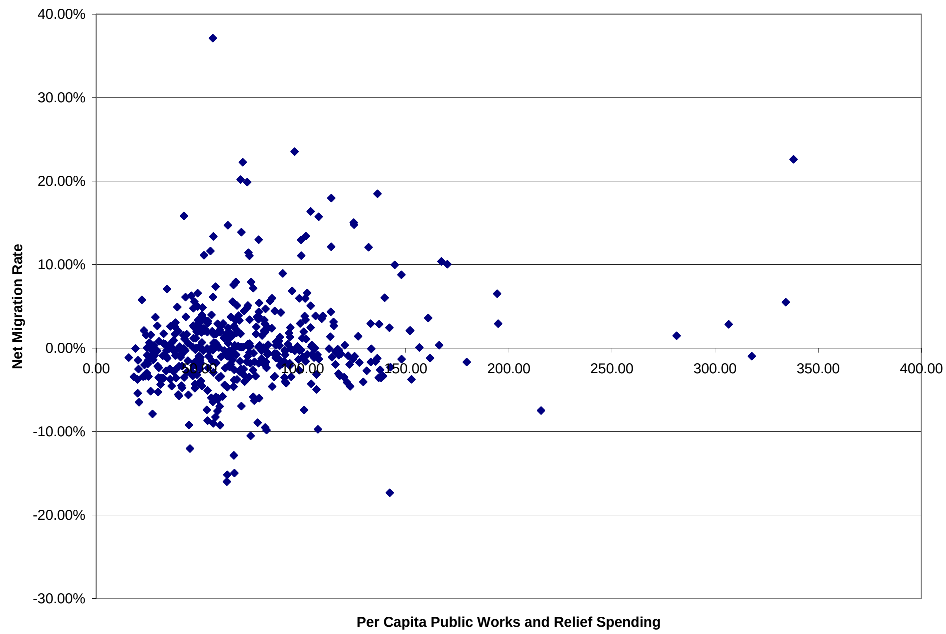

Figure 3.2: Net Migration Rates and Public Works Spending

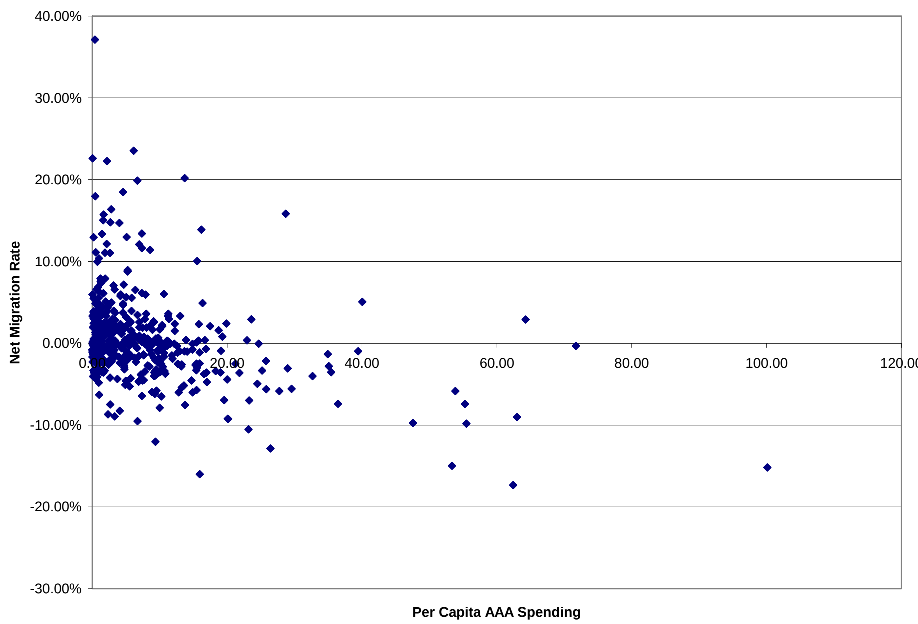

Figure 3.3: AAA Spending and Net Migration Rates

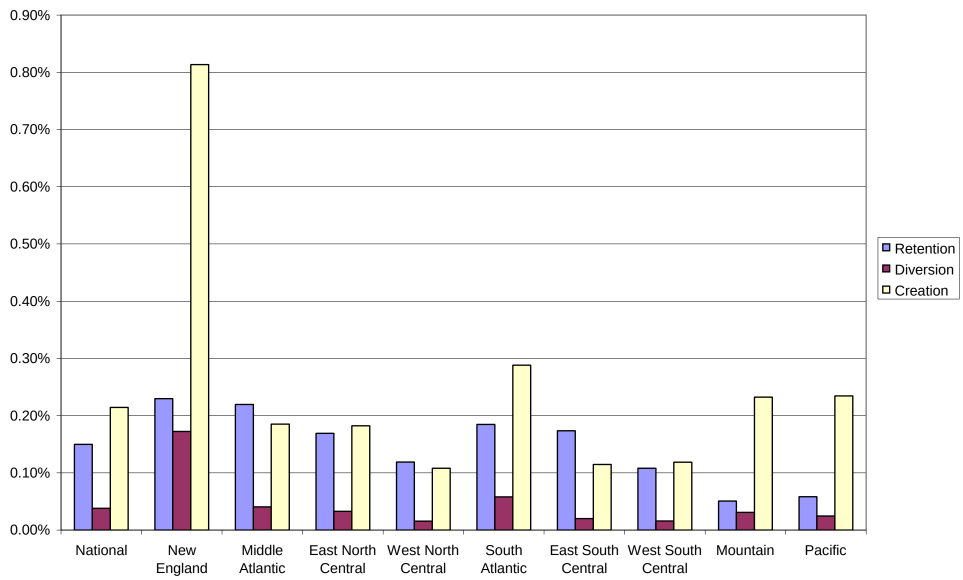

Figure 3.4: Relative Magnitudes of Marginal Effects of Public Works on Migration Rates into SEAs, Average by Region

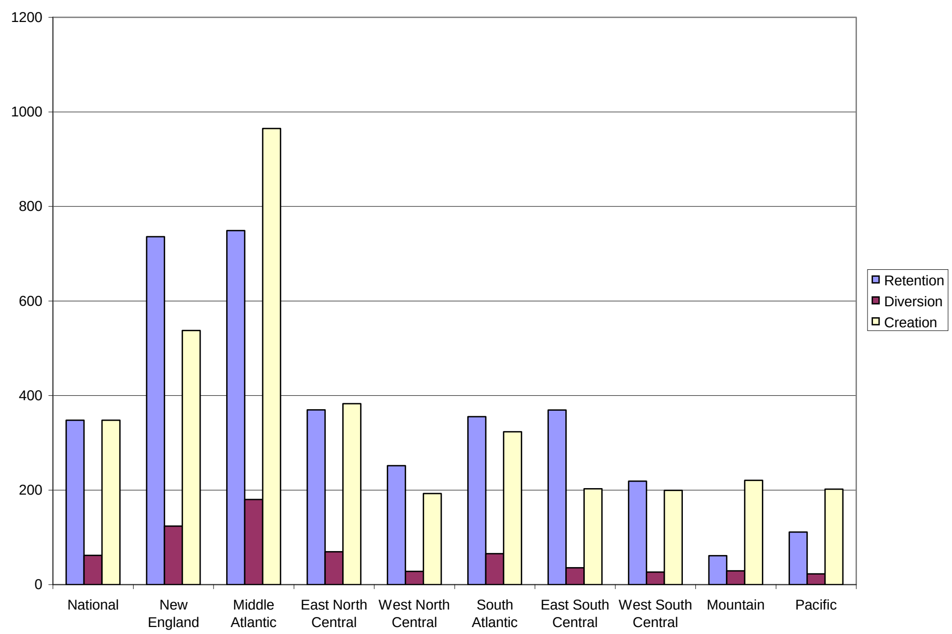

Figure 3.5: Relative Magnitudes of Marginal Effects of Public Works on Migration Flows into SEAs, Average by Region

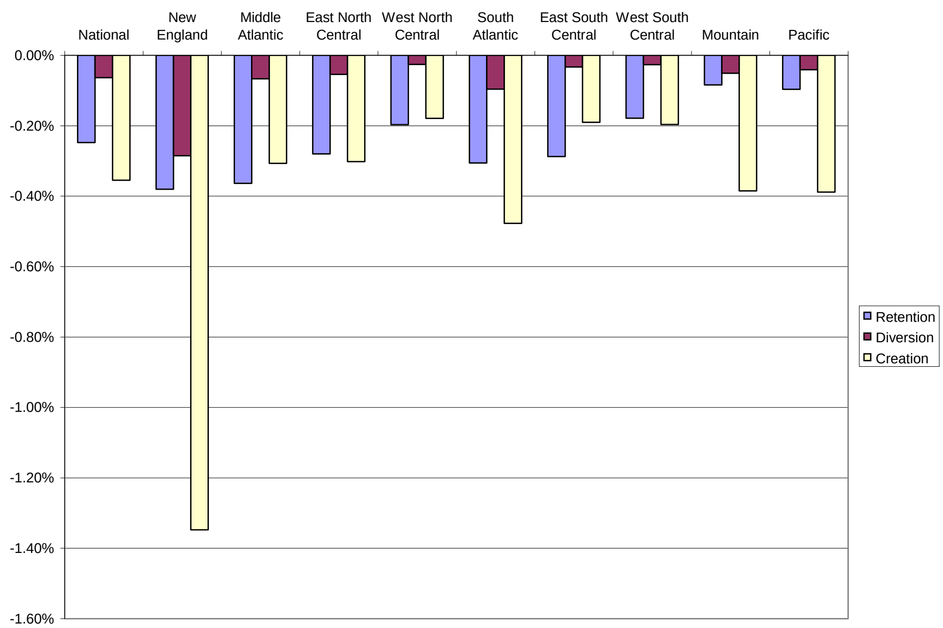

Figure 3.6: Relative Magnitudes of Marginal Effects of AAA on Migration Rates into SEAs, Average by Region

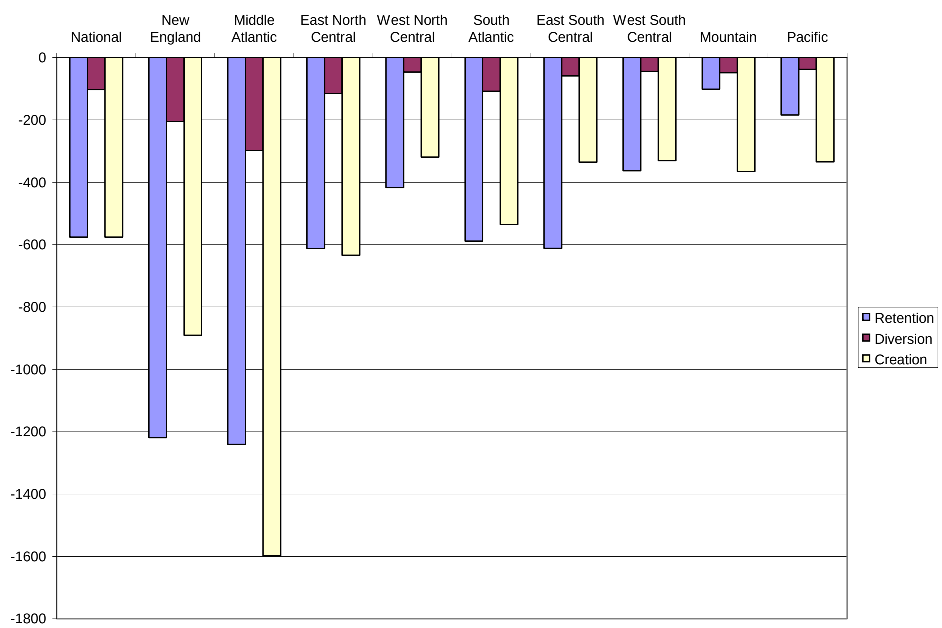

Figure 3.7: Relative Magnitudes of Marginal Effects of AAA on Migration Flows into SEAs, Average by Region

Chapter 4

Do You Receive a Lighter Prison Sentence Because You are a Woman? An Economic Analysis of Federal Criminal Sentencing Guidelines

[1]Gender equity has been one of the major global social issues to emerge out of the 20th century. A major focus of economists in this regard is on disparate labor market outcomes for men and women. Emphasis is placed on human capital explanations for gender wage gaps and though there is some scope for other explanations, such as Becker taste-driven discrimination, statistical discrimination, and market power. There is the potential effect of labor market outcomes on subsequent criminal activities and the effect of criminal activities on subsequent labor market outcomes. The literature on the economics of crime is discussed below. This paper examines the gender equity issue in the criminal justice arena and notes that labor market outcomes and criminal justice outcomes can be jointly determined. A popular perception in the criminal justice system is that female criminal behavior is a less serious problem than male criminal behavior. Detailed statistics compiled by the Bureau of Justice Statistics show that women commit fewer offenses than men and substantially different types of offenses than men. However, the statistics also reveal a rising trend in offenses committed by females and an increase in the incarceration of females in recent years. Beyond the labor market implications of gender equity in the criminal justice system there is also a concern for allocative efficiency regarding resources devoted to deterrence and incarceration.

The Federal Sentencing Guidelines that arose out of the Sentencing Reform Act

[1] Joint work with Supriya Sarnikar and Ronald L. Oaxaca

of 1984 and that were subsequently struck down by the U.S. Supreme Court in 2005 (consolidated cases of United States v. Booker, No. 04-104, and United States v. Fanfan, No. 04-105) required that males and females who commit the same crime and have the same prior criminal record receive equal sentences. Critics of the sentencing guidelines argue that women should be accorded separate treatment because females who are caught in the criminal justice system "enter it due to circumstances that are distinctly different from those of men"[2]. Others argue that gender is not a factor that should in enter into the sentencing decision. The Supreme Court in its split 5 to 4 decision argued that the mandatory guidelines violated the rights of criminal defendants to have a jury rather than a judge decide if defendants had committed all elements of a given crime. Consequently, the guidelines are only advisory to judges who may increase the length of sentences if they determine that the circumstances based on jury determination or admission of the defendant merit a longer prison sentence (Chicago Daily Law Bulletin, 2005). The 2005 decision created some ambiguity in how far judges could stray from the now "advisory" sentencing guidelines. Currently, the Supreme Court is considering a case in which a sentencing judge departed from the guidelines, only to have an appellate court overrule the judge's sentencing decision. The decision the Supreme Court hands down in this case may have as large an impact on the way sentences are given as the 2005 case (New York Times, February 20, 2007).

Whether the circumstances in which a crime is committed should be a consideration in criminal justice is not a question that we propose to answer here. Rather, we address the question of whether or not women do indeed receive more lenient sentences despite the sentencing guidelines. The answer to this question is important to both sides of the debate. Those in the justice system who favor equal treatment but believe that women are let off too lightly may be especially harsh when judging

[2] "Research on Women and Girls in the Justice System." National Institute of Justice Report (September 2000) at page iii. Available at http://www.ncjrs.gov/pdffiles1/nij/180973.pdf. .

a female accused of crime, while those who favor separate treatment of women but believe that they are treated equally may be less stringent. Thus, perceptions of unequal treatment, when they are not based on systematic study and sound facts, may lead to actual inequality in the justice system. A systematic study of whether bias actually exists is therefore not only necessary but timely given the rising trend in offenses committed by women and the increase in female incarceration rates as evidenced by the data compiled by the Bureau of Justice Statistics. Further work can begin to better tie the relationship between gender equity in the criminal justice system with gender equity in the labor market.

An unpublished paper by Oaxaca and Sarnikar (2005) [henceforth OS] uses a rich data set on sentencing outcomes from the United State Sentencing Commission to estimate separate logistic regressions for men and women, where the dependent variable is a binary variable measuring whether or not convicted individuals received federal prison time. While summary statistics from their data set show that females are less likely to receive prison time than males, more sophisticated analysis can take account of covariates that can explain some or all of the gender sentencing differential.

In this paper, we consider outcomes from the sentencing process more broadly for a sample of whites who were convicted while the mandatory sentencing guidelines were still in effect. Specifically, we look beyond the binary Prison/No-Prison outcome to a continuous measure of prison sentence. Ideally, one would want to take into account the fact that defendants must choose whether or not to plea bargain or to take their chances in a trial. Given that we work with a sample of convicted individuals (we do not have data on acquittals), we model the probability of whether the conviction was the result of a trial versus a plea bargain. We treat sentences handed down due to guilty pleas as outcomes from one regime, while sentences given to defendants convicted in a trial are treated as outcomes from a separate regime. This approach allows characteristics to be weighted differently depending on the path to conviction.

One would expect that the average sentences of identical defendants facing identical charges should be lower in the *plea regime*.[3] In our data set, around 25% of all criminal sentences involve no prison time. Because of this considerable mass point at zero, it may be inappropriate to consider the distribution of sentencing outcomes to be continuous. Also, the plea vs. trial regime is a choice variable for the defendant so that we must account for self selection in our model. Accordingly, we treat the outcome variable as a mixed discrete continuous variable. Therefore, our econometric model is a censoring (Tobit) switching regression with endogenous switching, which we estimate by full information maximum likelihood (FIML).

To measure how much of the male/female sentencing differential can be attributed to differences in the characteristics of men and women, compared to how much of the differential can be explained by differences in the weights applied to these characteristics by judges, we develop a new decomposition. This decomposition builds upon Neuman and Oaxaca (2004), which addresses the issue of selectivity in the context of a Heckman sample selection model. We expand this analysis to decompose differentials in the switching regression model with censoring. Our approach takes account of the fact that predicted outcome means will not generally match sample outcome means because of the highly non-linear nature of the model.

Within our data set, the scarcity of observation of females and the preponderance of observations in the plea regime conspire to leave us with an insufficient number of observations of females to properly apply FIML to estimate the female sentence determination model. In the decomposition developed here, we exploit an insight from Oaxaca and Ransom (1994) that allows us to decompose the male-female regime and sentencing differentials without actually estimating the model for females. Rather than comparing weights from a male only and female only model, we instead are able to compare the estimated parameters from the model for males with parameters

[3] If the sentences were lighter in the *trial regime* it would be difficult to believe that defendants would ever do anything but plead not guilty, as this would generate a positive probability of facing no sentence at all.

estimated in a pooled model for males and females.

4.1 Literature

Since the seminal work of Becker (1968), there has been a significant amount of research aimed at understanding the economics of crime. In the basic economic model of crime, a rational individual decides whether or not to allocate his/her time to criminal activity by comparing the expected net return from criminal activity to the expected returns from legitimate activity. The expected net return to criminal activity consists of the potential financial and psychic benefits (B) of committing the crime minus the cost (C) of committing the crime. The cost to the individual of committing the crime is determined as the product of the probability (p) of being caught by law enforcement and the severity of the punishment (S). If the returns to legitimate labor market activity is the wage (W), then a rational, risk-neutral individual will engage in criminal activity only if B-pS > W. This static model therefore predicts that criminal activity can be deterred by either increasing the probability of detection(p), the severity of punishment(S) and/or the wage rate (W) in the labor market.

Economists have since subjected these theoretical predictions to empirical testing using econometric models of varying degrees of sophistication. Ehrlich (1973) and Levitt (1997) estimate the impact of increased law enforcement presence on crime and find that increasing law enforcement efforts have the desired effect of lowering the incidence of crime. Ehrlich (1977) estimated the deterrence effect of capital punishment on crime. Witte (1980) finds that the deterrence effect of higher legal wages was small compared to the deterrence effects of the severity and certainty of state imposed penalties. Johnson, Fishback and Kantor (2006) study the effect of social insurance on crime rates. Block and Gerety (1995) reports on laboratory experiments that examine differences between the criminal population and the general population in the relative responsiveness to the deterrence effects of severity of punishment versus the deterrence effects of the certainty of punishment. The results showed that convicts were more deterred by increases in the certainty of punishment whereas the student subjects were more deterred by increases in the severity of punishment. Kuziemko (2006) uses New York State's reinstatement of the death penalty to identify the effect of capital punishment on plea bargaining outcomes. Freeman (1996), Grogger (1998), and Gould et.al (2002) find that falling real wages were a significant determinant of increasing crime rates during the decades of the 1970s and 1980s.

The link between deterrence efforts and crime rates is an endogenous one. Decisions to increase law enforcement efforts are often made in response to increasing crime rates. Similarly, difficulty in finding legitimate labor market employment might push some individuals into criminal, activity but the fact that an individual has engaged in criminal activity also would lower that individual's probability of finding legitimate employment. Myers (1983) investigates whether poor labor market prospects post-release affect the re-integration of ex-convicts into the mainstream. Using different data sets, Myers finds that better wages post-release significantly reduced recidivism. Witte and Reid (1980) also find that receiving a high wage on the first job after being released from prison decreases recidivism and that the wage rate received by a prison 'releasee' depends mostly on the demand side characteristics such as the industry and occupation rather than on the accumulated human capital of the 'releasee'. Imai and Krishna (2004) estimate a dynamic model of criminal behavior and show that expected future adverse consequences in the labor market prove to be an effective deterrent to crime. Waldfogel (1994) estimated the effects of conviction and imprisonment on post-conviction income and employment probabilities and found that the state-imposed sanctions were much smaller in comparison to the "market sanction", which he estimated as the income lost due to conviction and imprisonment. Also, the "market sanction" was significant only for those offenders who worked at jobs that required much trust. Grogger (1995) used longitudinal data and concluded that the strong negative correlation between arrests and subsequent labor market sanctions that was found in earlier cross-sectional studies was largely due to unobserved characteristics that influence both criminal and labor market behavior. Grogger (1995) however does find that there are significant negative consequences of arrests in the labor market but that they are short-lived.

Consistent with the predictions of the economic model of crime, the Sentencing Reform Act of 1984 (SRA 1984) increased the length of punishment for almost all crimes, eliminated probation and reduced the possibility of parole for good behavior. Kling (2006) estimates the effect of this increased severity of punishment on labor market prospects of criminals post-release. Kling finds that there is no significant adverse effect on employment or earnings of criminals due to longer incarceration lengths and concludes that this may be because prison rehabilitation programs may be offsetting the loss of potential work experience and human capital depreciation while in prison.

The sentencing guidelines formulated pursuant to the SRA 1984 aimed to provide uniform sanctions for the same crime by eliminating gender, age, or racial disparities in sentencing. While economists have studied the deterrence effect of severity of punishment quite extensively, relatively little literature exists on the optimality and desirability of uniform sentencing. Lott (1992) argues against uniform sentencing based on the finding that market sanctions in the form of lost incomes, opportunity costs of imprisonment and the adverse impact of incarceration on labor market prospects are disproportionately higher for individuals with higher incomes. Since the expected total monetary penalty includes the reduction in legitimate earnings capability post release, Lott argues that the state-imposed punishments should be proportionately adjusted. Moreover, since mere conviction can restrict the post-conviction opportunities for higher income individuals more severely than for lower skilled people, Lott argues that rich people should be convicted much less frequently than low-income criminals. The sentencing guidelines however explicitly prohibited sentencing judges from considering factors such as the defendant's socioeconomic status, race, sex, age, and religion. The punishment was to be proportional to the severity of the crime and the defendant's criminal history alone. Judicial discretion to change the sentence based on characteristics of the defendant was thus severely restricted under the guidelines.

Several studies in the criminology literature have examined gender and racial disparities in sentencing prior to the formulation of sentencing guidelines. See Tonry (1996) for a survey of these studies. Whether the guidelines have been successful in reducing the disparity has also been studied extensively both in the criminology and the law and economics literature. Anderson et al (1999), Kempf-Leonard and Sample (2001) study sentencing disparities before and after the federal sentencing guidelines. Mustard (2001) looks at racial and gender disparities in sentencing under the federal guidelines and finds that observed disparities in sentencing are mainly due to the special circumstances when judges are allowed to depart from the guidelines and not due to discriminatory tastes of judges. Schanzenbach (2005) estimates the effect of judicial demographics on sentencing outcomes and finds that increasing the proportion of female judges increases the gender disparity in sentencing and interprets this as evidence that male judges are paternalistic and therefore lenient towards female offenders.

Almost all of the studies mentioned infer gender-based discrimination in sentencing from the statistically significant coefficient on a dummy variable indicating the gender of the criminal offender. Yet, sentencing discrepancies may be observed merely because a judge takes into account extralegal circumstances of the defendant. If the circumstances of male and female criminal defendants are substantially different, as claimed by several authors, then the consideration of circumstances by judges may appear as gender-based bias even when the judge exhibits no such discriminatory tastes. Verdier and Zenou (2004) show that when there is statistical discrimination in the labor market and everyone believes that blacks, for example, are more likely to engage in criminal activity, then such beliefs lead to lower wages for blacks. When the opportunity cost of crime is thus lowered, such beliefs become self-fulfilling and lead to higher crime rates among blacks. It is therefore important to thoroughly investigate whether any bias actually exists in the criminal justice system since perceived bias may itself lead to actual bias. Given the adverse labor market consequences of incarceration, unequal treatment of men and women in the criminal justice system may lead to unequal prospects for men and women in the labor market as well.

Our research design separates the effect of differences in circumstances from the effect of differences in weights attached to circumstances by judges. If a judge attaches different weights to the same circumstances of a male and a female offender, then we may attribute that to a gender-based bias. But if a judge attaches the same weights to circumstances but on average awards different sentences to male and female offenders then that difference in sentencing might be due to differences in circumstances of the two defendants. Oaxaca and Sarnikar (2005) use decomposition analysis to investigate whether there exists any leniency towards women in the binary decision of whether or not to imprison a convicted person. The results of this decomposition show that the differences in characteristics explain more than 100% of the gender sentencing gap. If, when determining whether or not to sentence a woman to prison, judges applied the same weights on characteristics as they use for men, women would actually be slightly less likely to face prison.

4.2 Data

The data used in this study are obtained from the United States Sentencing Commission's data collection efforts and pertain to cases that terminated in convictions over the period 1996-2002. The data set is available from the Federal Justice Resource Statistics Center. In order to abstract from sentencing issues associated with race and ethnicity, we have confined our attention to convicted white males and white females. There were a total of 45,060 sentencing cases in our sample (37,104 cases for males and 7,956 cases for women).

Table 4.1 presents a summary of the share of sentences involving no prison time. Overall, a higher percentage of females receive no prison time upon conviction. This is true for both the trial and guilty plea regimes. For both males and females, conviction by a guilty plea is associated with a larger percentage of sentences involving no prison time.

The variables reported in Table 4.2 are the ones we have constructed for use in our sentence determination model. Both the measure of final offense level and the criminal history variable are set according to a fixed formula. To calculate the offense level, the case is assigned a base level for offense and then adjusted for various aggravating circumstances such as the use of a firearm in the crime or obstruction of justice, or for mitigating circumstances such as acceptance of responsibility. The criminal history measure is a function of both the length of prior imprisonments and how recently these sentences were given.[4]. While men on average are awarded longer prison sentences (42 months) than women (17 months), the severity of their offenses as measured by the final offense level scores are greater on average than those of women. Also, men on average have a higher past criminal history score than women. Convicted men are on average two years older than convicted women and are more likely to have private counsel. A higher percentage of men are college graduates (13% vs. 7%).

In Table 4.3 we present summary statistics pertaining to the average length of sentences imposed on both men and women in each of our sample years. Note that in each year the average male sentence is more than twice that of the average female sentence. If one were to only consider these summary statistics and no covariates, it

[4]For details on their construction of these variables, please see the following documents on the USSC's website:

http://www.ussc.gov/training/sent_ex_rob.pdf
http://www.ussc.gov/training/material.htm

would appear that women receive considerably lighter sentences than do males, and that this difference is considerably greater in the trial regime. Overall and in the trial regime, average male sentences generally declined over the sample period while average female sentences actually rose. Average sentences in the plea regime tended to rise for both males and females.

4.3 Econometric Model

Below we describe the econometric methods used to estimate the necessary parameters to decompose the sentence differentials. First, we describe the model we use to decompose the sentence difference into an explained portion (differences in characteristics) and an unexplained portion (differences in weights).

4.3.1 Sentencing

In our data set, we observe the sentencing outcomes for defendants whose cases reach the sentencing phase. Recall that there are two ways in which a defendant's case can reach the sentencing phase. While a significant number of defendants faced sentencing after being convicted by a jury, the most frequent way a defendant reached the sentencing phase was by pleading guilty. Plea bargains reached with a prosecutor are often the reason for this guilty plea; these defendants are sentenced under what we call the *plea regime.* When a defendant pleads not guilty, but is convicted in a trial, they are sentenced under the *trial regime.* We define y as the months in prison the defendant is sentenced to, X as the vector of the individual's characteristics, and β as the vector of weights on the defendant's characteristics in the respective regimes. Equation (4.1) represents sentencing outcomes when an individual pleads guilty or is convicted by trial:

$$y_i = \begin{cases} X_{Pi}\beta_P + \varepsilon_{P_i} & \text{if defendant is in plea regime} \\ X_{Ti}\beta_T + \varepsilon_{T_i} & \text{if defendant is in trial regime.} \end{cases} \tag{4.1}$$

Although the formal model permits differences in the covariates appearing in each sentencing regime, the empirical specification actually used in this paper restricts covariates to be identical in both sentencing regimes.

The very nature of a plea bargain suggests that the process determining the sentence of the defendant will not be the same in the two regimes. We would then expect the sentences received by two otherwise identical defendants to depend upon the way in which they reached the sentencing phase. Put another way, the weights applied to an individual's characteristics will be different depending on which sentencing regime the defendant is facing. Accordingly, it may be inappropriate to pool observations from individuals in these two regimes into a single sentencing equation. If individuals were exogenously selected into one of the two regimes, we could simply estimate the two models separately.

In order to more formally take account of the regime outcome conditional upon conviction, let π_P represent the probability of a guilty plea, $\pi_{T\&C}$ represent the probability of going to trial and being convicted, and $\pi_{T\&A}$ represent the probability of going to trial and being acquitted. The sum of these probabilities add to 1. Because we do not have observations on those who went to trial and were acquitted, we can only estimate the following conditional probabilities: $\pi_{PC} = \dfrac{\pi_P}{\pi_P + \pi_{T\&C}}$ and $\pi_{TC} = \dfrac{\pi_{T\&C}}{\pi_P + \pi_{T\&C}}$, which sum to 1 and where π_{PC} is the probability that one's conviction was from a guilty plea and π_{TC} is the probability that one's conviction was by trial. Let the variable s^* represent the conditional latent variable corresponding to a defendant's conviction by trial. The variable s takes on a value of 1 if the defendant's conviction is by trial, and a value of 0 if the defendant enters a guilty plea. The vector index variable Z_i is a set of variables affecting this probability.

$$
\begin{aligned}
s_i^* &= Z_i\gamma + u_i & (4.2)\\
s_i &= \begin{cases} 1 \text{ if } s_i^* > 0 \\ 0 \text{ if } s_i^* \leq 0 \end{cases} & (4.3)
\end{aligned}
$$

Correlation between unobservables in the plea decision stage and unobservables in the sentencing stage will create non random selection that will prevent us from obtaining consistent estimates of the parameters if they are estimated by OLS or Tobit. To account for this self-selection, we model the sentence determination process using a switching regression model with endogenous switching. We assume that the error term from each regime's sentence determination equation follows a bivariate normal distribution with the error term from the selection equation. The nature of this model requires that an explicit distributional assumption be made. The structure of the error terms is given in the following variance-covariance matrix, where T denotes the trial regime, P denotes the plea regime, and s denotes the binary selection equation (the variance of which is normalized to 1)[5]:

$$
V = \begin{pmatrix} 1 & \sigma_{Ps} & \sigma_{Ts} \\ \sigma_{Ps} & \sigma_P^2 & \sigma_{PT} \\ \sigma_{Ts} & \sigma_{PT} & \sigma_T^2 \end{pmatrix} \tag{4.4}
$$

The likelihood function of the model is then:

$$
\begin{aligned}
L = & \prod_{i=1}^{N} \left\{ \frac{1}{\sigma_T} \phi\left(\frac{y_i - X_{Ti}\beta_T}{\sigma_T} \right) \Pr(u_i > -Z_i\gamma | \varepsilon_{Ti}) \right\}^{s_i} \\
& \left\{ \frac{1}{\sigma_P} \phi\left(\frac{y_i - X_{Pi}\beta_P}{\sigma P} \right) \Pr(u_i \leq -Z_i\gamma | \varepsilon_{Pi}) \right\}^{1-s_i} \qquad (4.5)
\end{aligned}
$$

This expression is simplified once we take account of the conditional distribution of u on ε :

[5] The errors in the two sentencing regimes could be correlated; however the model neither requires nor provides identification of this parameter.

$$L = \prod_{i=1}^{N} \left\{ \frac{1}{\sigma_T} \phi \left(\frac{y_i - X_{Ti}\beta_T}{\sigma_T} \right) \Phi \left(\frac{Z_i\gamma + \frac{\rho_{Ts}}{\sigma_T}(y_i - X_{Ti}\beta_T)}{1 - \rho_{Ts}} \right) \right\}^{s_i}$$
$$\left\{ \frac{1}{\sigma_P} \phi \left(\frac{y_i - X_{Pi}\beta_P}{\sigma_P} \right) \Phi \left(\frac{-Z_i\gamma - \frac{\rho_{Ps}}{\sigma_P}(y_i - X_{Pi}\beta_P)}{1 - \rho_{Ps}} \right) \right\}^{1-s_i} \tag{4.6}$$

One additional econometric problem we face is the non-continuous distribution of the dependent variable. Because sentence length cannot be negative, and nearly 25% of our sample receives no prison time, it may be necessary to account for this mass point at 0 in order to obtain consistent estimates.[6] In the context of our switching regression model, we treat the dependent variable as a mixed discrete continuous variable, with limit observations at 0. The sentence outcome is now represented as

$$y^*_{Pi} = X_{Pi}\beta_P + \varepsilon_{P_i} \text{ if defendant is in plea regime} \tag{4.7}$$
$$y_{Pi} = \begin{cases} y^*_{Pi} \text{ if } y^*_{Pi} > 0 \text{ and } s_i = 0 \\ 0 \quad \text{if } y^*_{Pi} \leq 0 \text{ and } s_i = 0 \end{cases} \tag{4.8}$$
$$y^*_{Ti} = X_{Ti}\beta_T + \varepsilon_{T_i} \text{ if defendant is in trial regime} \tag{4.9}$$
$$y_{Ti} = \begin{cases} y^*_{Ti} \text{ if } y^*_{Ti} > 0 \text{ and } s_i = 1 \\ 0 \quad \text{if } y^*_{Ti} \leq 0 \text{ and } s_i = 1 \end{cases} \tag{4.10}$$

The likelihood for the switching regression with endogenous switching and censoring allows four different types of entries to the likelihood function: limit and non-limit observations in both of the regimes. The likelihood function is

[6] We also estimate the model without accounting for censoring; the log-likelihood obtained is significantly lower than that obtained in the model where we account for the censoring.

$$
\begin{aligned}
L \;=\; & \prod_{i=1}^{N} \left\{ \Phi_2 \left(\frac{-X_{T_i}\beta_T}{\sigma_T}, -Z_i\gamma, \rho_{Ts} \right) \right\}^{s_i l_i} \left\{ \Phi_2 \left(\frac{-X_{Pi}\beta_P}{\sigma_P}, Z_i\gamma, \rho_{Ps} \right) \right\}^{(1-s_i)l_i} \\
& \cdot \left\{ \frac{1}{\sigma_T} \phi \left(\frac{y_{Ti} - X_{Ti}\beta_T}{\sigma_T} \right) \Phi \left(\frac{Z_i\gamma + \frac{\rho_{Ts}}{\sigma_T}(y_{Ti} - X_{Ti}\beta_T)}{1 - \rho_{Ts}} \right) \right\}^{s_i(1-l_i)} \\
& \cdot \left\{ \frac{1}{\sigma_P} \phi \left(\frac{y_{Pi} - X_{Pi}\beta_P}{\sigma_P} \right) \Phi \left(\frac{-Z_i\gamma - \frac{\rho_{Ps}}{\sigma_P}(y_{Pi} - X_{Pi}\beta_P)}{1 - \rho_{Ps}} \right) \right\}^{(1-s_i)(1-l_i)} \qquad (4.11)
\end{aligned}
$$

where $l = 1$ for limit observations and Φ_2 represents the cumulative bivariate normal distribution.

4.4 Decomposing Sentencing Differentials

To examine how much of the gender difference in sentences is due to leniency toward one sex or the other, we apply empirical methods developed in the labor economics literature to estimate gender bias in criminal sentencing outcomes. These methods have the advantage of decomposing gender differences in sentencing outcomes into two different components – one due to differences in observable circumstances of males and females convicted by the criminal justice system and another due to differences in unobserved circumstances or attitudes of judges towards the sexes. Such decomposition is achieved by a three-step analysis.

The first step typically involves estimation of our empirical model for males and females where the dependent variable is the length of the prison sentence. Here, instead of estimating the empirical model for both males and females, we estimate the model for males only. This approach is consistent with viewing the unexplained gap as a residual. It is also necessary in our case, as the relatively small number of female observations in the trial regime means that we are unable to identify a number of parameters in an estimation of the model for females only. This approach allows us to decompose the differential without estimating the female weights, thus circumventing the problem.

Our analysis departs from previous studies in the second step and adds greater insight into the decision-making process that might lead to gender-based differences in criminal sentencing. In the second step, we predict the average sentence length for females if they faced the male weights. In the third and final step, we use results from the first two steps and decompose the differences in length of sentences for males and females into two components: one attributable to male-female differences in circumstances and a second attributable to unobserved differences in attitudes of judges towards the sexes and unobserved differences in circumstances.

Decomposition methods such as the one described above were first developed in labor market studies of gender and racial wage differences [Oaxaca, 1973] but have not been used in studies of gender or racial bias in criminal sentencing decisions. Such a method of estimating bias is valuable since it not only estimates any gender-based differences in sentencing outcomes but it also identifies whether the observed bias is due to gender differences in circumstances or due to gender-based differences in weights attached to circumstances by judges.

In addition to the problems with identifying the female weights, we face two additional challenges which force us to expand beyond the Oaxaca (1973) decomposition. The issue of selection bias in decompositions is addressed by Neuman and Oaxaca (2004) in the context of a Heckit model. We are able to build off of this work in the decomposition we develop, as the Heckit is essentially a special case of an endogenous switching regression model. Finally, we must account for the existence of the limit observations in our data set.

4.4.1 Decomposing Sentencing Outcomes by Regime

First, consider the sentence determination equation for the trial regime:

$$\begin{aligned}
y_{Ti}^* &= X_{Ti}\beta_T + \varepsilon_{Ti} \text{ if defendant is in the trial regime} && (4.12)\\
y_{Ti} &= \begin{cases} y_{Ti}^* \text{ if } y_{Ti}^* > 0;\ s_i = 1 \\ 0 \ \text{ if } y_{Ti}^* \le 0;\ \ s_i = 1 \end{cases} && (4.13)
\end{aligned}$$

The expected value of a sentence in the trial regime is derived in Appendix C. Define the sample average sentence in the trial regime as $\bar{y}_{Tm}$ for males and $\bar{y}_{Tf}$ for females. The sample is composed of N_{Tm} men and N_{Tf} women. The average predicted value of sentences for males is defined as:

$$\widehat{y}_{Tm} = \frac{1}{N_{Tm}} \sum_{i=1}^{N_{Tm}} \widehat{y}_{Tmi}, \tag{4.14}$$

where $\widehat{y}_{Tmi}$ is the predicted sentence for the ith male in the trial regime. However, in a finite sample the predicted mean and the sample mean terms will not necessarily be equal, i.e.

$$\widehat{y}_{Tm} = \frac{1}{N_{Tm}} \sum_{i=1}^{N_{Tm}} \widehat{y}_{Tmi} \neq \bar{y}_{Tm} = \frac{1}{N_{Tm}} \sum_{i=1}^{N_{Tm}} \ y_{Tmi} \text{ in general.}$$

Assuming that the underlying model can be consistently estimated, we would have

$$\begin{aligned}
\text{plim}(\widehat{y}_{Tm} - \overline{y}_{Tm}) &= 0 && (4.15)\\
\text{plim}(\widehat{y}_{Tf} - \overline{y}_{Tf}) &= 0. && (4.16)
\end{aligned}$$

When the predicted mean outcome does not match the sample mean outcome, we have sample mean prediction error. The proportionate sample mean prediction errors for males and females can be expressed as

$$\widehat{\delta}_{Tm} = \frac{\overline{y}_{Tm}}{\widehat{y}_{Tm}} \tag{4.17}$$

$$\widehat{\delta}_{Tf} = \frac{\overline{y}_{Tf}}{\widehat{y}_{Tf}}. \tag{4.18}$$

It follows from consistency that

$$\text{plim}(\widehat{\delta}) = \text{plim}\left(\frac{\overline{y}}{\widehat{y}}\right) = 1.$$

Appendix D contains a more detailed discussion of the use of sample mean error predictions in the nonlinear decompositions adopted in this paper.

The average value of sentences for females in the trial regime using male weights is defined as:

$$\hat{y}^0_{Tf} = \frac{\sum_{i=1}^{N_f} \hat{y}^0_{Tfi}}{N_{Tf}} \tag{4.19}$$

where $\hat{y}^0_{Tfi}$ is a fitted value of the ith female sentence had they faced the male weights. We decompose the difference in average sentences in the trial regime as follows:

$$\overline{y}_{Tm} - \overline{y}_{Tf} = \widehat{\delta}_{Tm}(\widehat{y}_{Tm} - \widehat{y}^0_{Tf}) + (\widehat{\delta}_{Tm} - \widehat{\delta}_{Tf})\widehat{y}^0_{Tf} + \widehat{\delta}_{Tf}\left(\widehat{y}^0_{Tf} - \widehat{y}_{Tf}\right). \tag{4.20}$$

The first term in eq (4.20) measures the explained sentencing gap while the unexplained gap is the sum of the last two terms. Note that the second term measures the contribution of gender differences in the sample mean prediction error while the last term measures the contribution of gender differences in the estimated parameters of the model.[7] It is therefore possible to separate out the effect of gender differences

[7] Of course there are many instances in which there is no discrepancy between sample means and predicted sample means, e.g. the linear regression model with a constant term, the logit model with a constant term, and the second stage regression of a heckit sample selection model.

in $\hat{\delta}_T$ if the econometrician estimates both $\widehat{\delta}_{Tm}$ and $\widehat{\delta}_{Tf}$. While we are able to decompose the difference in outcomes into the portion caused by differences in weights and differences in characteristics, we will be unable to isolate the difference caused by weights into a portion caused by different $\hat{\delta}_T$ terms. However, if it is the case that $\widehat{\delta}_{Tm} - \widehat{\delta}_{Tf} \approx 0$, the unexplained gap is totally captured by $\widehat{\delta}_{Tf}\left(\widehat{y}^0_{Tf} - \widehat{y}_{Tf}\right) \approx \widehat{\delta}_{Tm}\left(\widehat{y}^0_{Tf} - \widehat{y}_{Tf}\right)$. Under these circumstances one could identify the predicted mean outcome for females as $\widehat{y}_{Tf} \approx \widehat{y}^0_{Tf} - \left(\frac{1}{\widehat{\delta}_{Tm}}\right)\left[\left(\overline{y}_{Tm} - \overline{y}_{Tf}\right) - \widehat{\delta}_{Tm}\left(\widehat{y}_{Tm} - \widehat{y}^0_{Tf}\right)\right]$.

The decomposition of sentences in the plea regime follows closely that of the trial regime. Now using male weights from the plea regime, the fitted value of the length of sentence in the regime becomes $\hat{y}_P$, which differs slightly in form from $\hat{y}_T$.[8]

4.4.2 Decomposing Regime Choice

Now consider a decomposition of regime choice. Consider the regime determination model given in (4.2) and (4.3) where a positive outcome indicates conviction by trial. The observed proportion of females and males going to trial are, respectively

$$\bar{p}_{Tf} = \frac{\sum_{i=1}^{N_f} s_{fi}}{N_f} \tag{4.21}$$

$$\bar{p}_{Tm} = \frac{\sum_{i=1}^{N_m} s_{mi}}{N_m} \tag{4.22}$$

We define the difference in outcomes for males and females as the observed differences in proportions of males and females in the trial regime, $\bar{p}_{Tm} - \bar{p}_{Tf}$.

Recall that we do not estimate the model separately for females. However, we are still able to decompose the difference in male and female outcomes into the por-

[8] The fitted value is now for individuals who are "selected in" in the plea equation, rather than the "selected out" observations in the conviction by trial equation. The form of the selectivity term will differ slightly. See Appendix C for the expressions governing the calculations of the mean outcomes.

tion caused by differences in characteristics and the portion caused by differences in weights. We go about these *single model decompositions* by decomposing differentials using only the estimated weights for males.

Here, we decompose the difference in the propensity of males and females to be convicted by trial regime using only male weights. Consider the regime determination model estimated for males:

$$s_{mi}^* = Z_{mi}\gamma_m + u_i \tag{4.23}$$

$$s_{mi} = \begin{cases} 1 \text{ if } s_{mi}^* > 0 \\ 0 \text{ if } s_{mi}^* \leq 0 \end{cases} \tag{4.24}$$

The estimated weights in this model allow us to obtain a predicted probability of conviction by trial for each individual in the sample:

$$\hat{p}_{Tmi} = \Phi(Z_{mi}\hat{\gamma}_m) \tag{4.25}$$

We compute the average predicted probability by averaging the individual predicted probabilities:

$$\hat{p}_{Tm} = \sum_{i=1}^{N_m} \frac{\Phi(Z_{mi}\hat{\gamma}_m)}{N_m} \tag{4.26}$$

Note that in the probit model, unlike the logit model, the average predicted probability of entering the trial regime will not necessarily equal the proportion of the sample who do in fact enter the regime (for further work on the decomposition of differentials in the context of a probit model, see Fairlie(2005) and Yun(1999)).

In practice the difference is typically negligible. However, the selection probability parameters in our model are obtained from FIML applied to the joint estimation of the selection probability and sentencing equations. Hence, there is a need to scale the mean predicted probabilities when conducting a decomposition of gender differences in the propensity to be convicted via the trial regime. As above for the sentencing outcomes, the sample mean (probability) prediction errors for males can be expressed as follows:

$$\widehat{\delta}_{sm} = \frac{\bar{p}_{Tm}}{\hat{p}_{Tm}} \tag{4.27}$$

The same consistency argument applies here as in the case of sentencing outcomes.

We estimate the average predicted probability of females being in the trial regime had they faced the same weights as the males:

$$\hat{p}^0_{Tf} = \sum_{i=1}^{N_f} \frac{\Phi(Z_{fi}\hat{\gamma}_m)}{N_f} = \sum_{i=1}^{N_f} \frac{\hat{p}^0_{Tfi}}{N_f} \tag{4.28}$$

The difference in the average probability of conviction via the trial regime can then be decomposed as follows:

$$\bar{p}_{Tm} - \bar{p}_{Tf} = (\bar{p}_{Tm} - \widehat{\delta}_{sm}\hat{p}^0_{Tf}) + (\widehat{\delta}_{sm}\hat{p}^0_{Tf} - \bar{p}_{Tf}) \tag{4.29}$$

where the first term on the right hand side represents the difference in probabilities that can be attributed to differences in characteristics, and the second term represents the part of the difference that can be attributed to differences in weights.

4.4.3 Total Decomposition

Consider an algebraic decomposition of sentencing differences by regime. Define $\bar{y}_m$ as the average sentence for males in our sample, and $\bar{y}_f$ as the average sentence for females. Each gender's average sentence will be a weighted average of the average sentence in the two regimes:

$$\bar{y}_m = \bar{y}_{Tm}\bar{p}_{Tm} + \bar{y}_{Pm}\,(1 - \bar{p}_{Tm}) \tag{4.30}$$

$$\bar{y}_f = \bar{y}_{Tf}\,\bar{p}_{Tf} + \bar{y}_{Pf}\,(1 - \bar{p}_{Tf}) \tag{4.31}$$

The difference in average sentences can then be expressed as

$$= \bar{y}_{Tm}\ \bar{p}_{Tm} + \bar{y}_{Pm}\ (1 - \bar{p}_{Tm}) - \bar{y}_{Tf}\ \bar{p}_{Tf} - \bar{y}_{Pf}\ (1 - \bar{p}_{Tf})$$

Adding and subtracting the terms $\bar{y}_{Tf}\ \bar{p}_{Tm}$ and $\bar{y}_{Pf}\ (1 - \bar{p}_{Tm})$, and collecting terms appropriately yields

$$\begin{aligned} \bar{y}_m - \bar{y}_f &= (\bar{y}_{Tm} - \bar{y}_{Tf})\ \bar{p}_{Tm} + (\bar{y}_{Pm} - \bar{y}_{Pf})\ (1 - \bar{p}_{Tm}) \\ &\quad + (\bar{y}_{Tf} - \bar{y}_{Pf})\ (\bar{p}_{Tm} - \bar{p}_{Tf}). \qquad (4.32) \end{aligned}$$

The first two terms in (4.32) can be interpreted as a weighted average of the differences in mean sentence outcomes for men and women (weighted by the probability of being in each of the two regimes). The final term can be interpreted as the difference in mean sentence outcomes that can be attributed to gender differences in the propensities of being in the trial regime (weighted by the differences in mean outcomes among females in the two regimes).

Recall how we decomposed each of the single decomposition terms. Denote the portion of the difference attributed to differences in characteristics (the *explained* portion) as E. The portion of the differences attributed to differences in the characteristics (the *unexplained* portion) is denoted as U. Each portion also contains a subscript denoting the part of the estimation from which it originates:

$$\begin{aligned} \bar{y}_{Tm} - \bar{y}_{Tf} &= \left[\widehat{\delta}_{Tm}(\widehat{y}_{Tm} - \widehat{y}^0_{Tf})\right] + \left[(\widehat{\delta}_{Tm} - \widehat{\delta}_{Tf})\widehat{y}^0_{Tf} + \widehat{\delta}_{Tf}\left(\widehat{y}^0_{Tf} - \widehat{y}_{Tf}\right)\right] \\ &= E_T + U_T \qquad (4.33) \\ \bar{y}_{Pm} - \bar{y}_{Pf} &= \left[\widehat{\delta}_{Pm}(\widehat{y}_{Pm} - \widehat{y}^0_{Pf})\right] + \left[(\widehat{\delta}_{Pm} - \widehat{\delta}_{Pf})\widehat{y}^0_{Pf} + \widehat{\delta}_{Pf}\left(\widehat{y}^0_{Pf} - \widehat{y}_{Pf}\right)\right] \\ &= E_P + U_P \qquad (4.34) \\ \bar{p}_{Tm} - \bar{p}_{Tf} &= (\bar{p}_{Tm} - \widehat{\delta}_{sm}\widehat{p}^0_{Tf}) + (\widehat{\delta}_{sm}\widehat{p}^0_{Tf} - \bar{p}_{Tf}) \\ &= E_s + U_s \qquad (4.35) \end{aligned}$$

The decomposition of the overall gender sentencing gap can then be expressed as

$$\begin{aligned}\bar{y}_m - \bar{y}_f &= \left[(E_T + U_T)\ \bar{p}_{Tm} + (E_P + U_P)\ (1 - \bar{p}_{Tm})\right] \\ &\quad +(\bar{y}_{Tf} - \bar{y}_{Pf})\ (E_s + U_s) \\ &= \underbrace{E_T\ \bar{p}_{Tm} + E_P\ (1 - \bar{p}_{Tm}) + E_s\ (\bar{y}_{Tf} - \bar{y}_{Pf})}_{E} \\ &\quad + \underbrace{U_T\ \bar{p}_{Tm} + U_P\ (1 - \bar{p}_{Tm}) + U_s\ (\bar{y}_{Tf} - \bar{y}_{Pf})}_{U},\end{aligned} \tag{4.36}$$

where E is the total amount of the overall gender sentencing gap that is explained by differences in characteristics, and U is the total unexplained gap associated with differences in weights.

We note that a more straight forward total decomposition of the mean sentencing differences between men and women can be calculated as

$$\bar{y}_m - \bar{y}_f = \left(\bar{y}_m - \hat{\delta}_m \hat{y}_f^0\right) + (\hat{\delta}_m \hat{y}_f^0 - \bar{y}_m) \tag{4.37}$$

where

$$\hat{y}_f^0 = \frac{\sum_i \left[\hat{p}_{Tfi}^0 \hat{y}_{Tfi}^0 + \left(1 - \hat{p}_{Tfi}^0\right) \hat{y}_{Pfi}^0\right]}{N_f}$$

and

$$\hat{\delta}_m = \left\{ \frac{\sum_i \left[\hat{p}_{Tmi} \hat{y}_{Tmi} + (1 - \hat{p}_{Tmi})\, \hat{y}_{Pmi}\right]}{N_m} \right\} \left\{ \frac{1}{\bar{y}_m} \right\}.$$

In this decomposition $\hat{y}_f^0$ is the mean fitted overall sentence for females using the male weights. Empirically, it turns out that both (4.36) and (4.37) yield virtually identical values of the total explained and unexplained portions of the overall gender sentencing gap. However, a shortcoming of the decomposition given by (4.37) is that it obscures the sources of the overall gender sentencing gap revealed by the more detailed decomposition given in (4.36).

4.5 Results

Formal theory does not offer very much guidance on the actual specification of the regime selection and sentencing equations. The sentencing guidelines largely confined federal court judges to considering only current offense level and criminal history when passing sentence. Specifically, the guidelines exclude race, sex, national origin, creed, religion, and socioeconomic status. Furthermore, employment and family ties and responsibilities are also not to be considered in awarding criminal sentences. With only limited exception, age and education are not supposed to be relevant for sentencing decisions. Judges are permitted to award lighter prison sentences to elderly defendants. Since we have data on these various potential factors, we are able to empirically determine the extent to which they turn out to influence sentences because of, or despite, the guidelines. The variables that appear jointly in the regime selection and sentencing equations are indicators for females (in the pooled sample), education, marital status, the circuit court district, and year while the continuous variables appearing jointly pertain to prior criminal history, number of dependents, and age. An indicator for U.S. citizenship appears in the regime selection equation but not in the sentencing equations. While judges should not take into account the nationality of a defendant when determining her sentence, citizenship should serve as a proxy for this defendant's knowledge of and experience with the U.S. criminal justice system; we would expect risk averse individuals with less knowledge of how this process works to be less likely to take their chances in a trial rather than striking a plea bargain deal. An indicator for a defendant's fine being waived appears in the sentencing equation but not in the regime selection equation. This variable serves as a crude proxy for income. Also, a cubic polynomial function of the severity of the final offense level appears in the sentencing equations but are excluded from the regime selection equation. Both the fine variable and the final criminal offense level variables are not determined at the time that the individual makes the decision about going to trial. Given that defendants do not have perfect foresight, these variables should determine the final sentence given but not affect the plea decision.

Although our data span both cases and years, it is not treated as a panel. The data are available as separate cross-sections by case for each year. Each case corresponds to all prosecutions ending in convictions of an individual in the given year and the total prison time awarded. While it is theoretically possible for an individual to appear in more than one year's cross-section, we suspect that this is not very common. Among males the average prison sentence is 3.5 years over a period of 7 years. This does not leave much time for multiple year convictions unless offenses are committed while the individual is in prison. In the case of females the average prison sentence is 1.4 years over the period of our study. This would allow for multiple year convictions except that the crime rate is still much lower for females. Female cases account for just under 18% of the total number of cases in our data set.

To get a sense of whether or not there may be favoritism towards women, we first estimate our model on a pooled sample of males and females, including an indicator variable for whether the observation is that of a female offender. In Table 4.4 we present parameter estimates from this pooled sample of males and females The estimated coefficient on the female indicator variable is negative and significant in the selection equation, indicating that women are less likely to obtain their convictions via the trial regime, where average sentences are higher. More educated and married individuals are more likely to obtain their convictions through trial rather than through guilty pleas. Being a U.S. citizen is associated with a lower probability of obtaining one's conviction via trial as opposed to a guilty plea. The chances that one would obtain their conviction via trial rather than via a guilty plea rise with age until around 73 years after which the trial regime probability declines. The circuit court district in which the conviction took place does affect the probability of conviction via trial vs. guilty plea. The year indicators (where 2002 is the omitted reference group) suggest that the probability of obtaining conviction via trial relative to guilty plea steadily declined over time. A more extensive past criminal history was positively associated with conviction by trial vs. a guilty plea. Having a private defense counsel has a statistically significant negative impact on the probability of conviction by trial.

The estimated coefficients on the female gender indicator are negative and statistically significant in both sentencing regimes, but they are of a greater magnitude (in absolute value) in the trial regime. Even before we allow all weights to differ by gender, this indicates that women may receive lighter sentences than men. This would seemingly violate the sentencing guidelines. Contrary to the guidelines, marital status and number of dependents do affect prison sentences, but only in the plea regime. Married defendants receive shorter sentences in the plea regime. Having more dependents leads to shorter sentences in the plea regime. Age and education exhibit some effect on sentences though ordinarily these are not considered relevant by the guidelines. Sentence length rises with age and peaks at 69 years if one is convicted in the trial regime and peaks at 29 years in the plea regime. Although the guidelines permit lighter sentences for the elderly, a peak of 29 years in the plea regime and the strong significance of the age terms in the trial regime would not seem to be entirely consistent with the guidelines. Education appears to lower sentences in the plea regime and raise them in the trial regime. Those who have been convicted and had fines waived receive longer sentences in the plea regime. If this variable adequately proxies incomes of the defendants, then it would seem that poorer defendants receive longer sentences in the plea regime. As expected the extent of a defendant's criminal history and severity of current final criminal offense contribute to longer prison sentences in both regimes. The signs and magnitudes of the linear, quadratic, and cubic terms jointly imply that, for all relevant values of the variable, as the severity of the crime for which one is convicted increases, sentence length increases at an increasing rate. Having a private defense counsel lowered prison sentences in both conviction regimes.[9]

[9] If the choice of defense counsel and the conviction regime are jointly determined, then the choice of defense counsel would be endogenous in the model. Accordingly, we estimate a model to determine

Similar to the case with conviction regime selection, the circuit court district in which the conviction took place does affect sentence lengths. The estimated coefficients on the time indicator variables reveal that, ceteris paribus, sentence length had been declining over time in the trial regime while rising in the plea regime. Estimates of the correlations between the conviction regime error and the sentencing regime errors suggest that unobservables in the selection equation are negatively correlated with unobservables in the trial sentencing equation and positively correlated with unobservables in the plea regime. Roughly speaking, this means that those who are more likely to select into the conviction by trial regime can expect shorter sentences in the trial regime and longer sentences in the plea regime. While this is a sensible result, one potential problem is that the estimated correlation coefficient between the regime selection equation error term and the plea regime sentencing error term is close to the boundary value of 1. It is probably the case that this extreme estimate of the correlation coefficient is caused by the fact that ninety five of the sample represent convictions via guilty pleas.

In Table 4.5 we report the FIML estimates based on just the male sample. Since the results for males are qualitatively the same as those for the pooled sample, we do not separately discuss these estimates. The major purpose behind estimating the model separately for males is to provide us with the necessary parameter estimates to compute the decomposition of gender differences in prison sentences.

Decomposition results are reported in Tables 4.6 through 4.8. We begin with Table 4.6 which presents mean sentencing outcomes by regime and regime selection differences as well as predicted outcomes using estimated male weights. On average men are awarded nearly 25 more months of prison than women. This varies by sen-

if the decision to be represented by a private attorney is jointly determined with regime choice. By estimating the model with a bivariate probit, we can test for this possible correlation in the two error terms related to these decisions. Our estimation finds the error term correlation coefficient to be insignificant, suggesting that the coefficient on the defense counsel variable in the main model is consistently estimated.

tencing regime. For those convicted by trial, men received an average of 69 more months of prison than women. Among those who plead guilty, men received an average of almost 22 more months of prison time than women. A higher percentage of men than women received their convictions via trial vs. a guilty plea, 5.5% vs. 3.5%. From the fitted (predicted mean) sentences for males, we are able to calculate the proportionate mean sample prediction errors. The most accurate prediction corresponds to the plea regime which is the one into which the vast majority of the cases fall. The last column of Table 4.6 reports the predicted outcomes for females using the FIML estimated weights for men and are comparable to the calculated fitted values for men reported in the next to the last column in Table 4.6. For the actual decompositions, the proportionate mean prediction errors for men are applied to the predicted outcomes for women obtained using the estimated male weights. The figures in Table 4.6 clearly imply that if females had faced the same sentence determination process as men, they would have experienced longer prison sentences in each regime, though still less than those of men, and would have had a higher propensity to have received their convictions from the trial regime as opposed to the plea regime.

Our decompositions of gender sentencing differences in each regime and gender differences in conviction regime probabilities are reported in Table 4.7. Differences in the female mean characteristics explain 46% of the gender sentencing differential in the trial regime and 66% of the sentencing differential in the plea regime. We observe that of the 69 month sentencing gap that favors women in the trial regime, nearly 38 months of the gap cannot be accounted for by gender differences in circumstances. Of the 22 month sentencing gap that favors women in the plea regime, 7 months of the gap cannot be accounted for by gender differences in circumstances. Only about 21% of the 2.1 percentage point gender gap in the propensity to obtain conviction in the trial regime can be explained by gender differences in characteristics. Females are also less likely to be sentenced in the trial regime, though their characteristics suggest they would actually be more likely to be sentenced in this regime if they were to face the male weights (though still less likely than males).

In Table 4.8 we parse out the components that add to the overall gender sentencing difference across both conviction regimes. These components weight the explained and unexplained portions of the sentencing gaps in each regime by the probabilities of being in each regime and gender differences in these probabilities. Of the nearly 25 month overall gender sentencing gap favoring women, 3.8 months (15.4%) arises from gender sentencing differences in the trial regime. Gender sentencing differences in the plea regime account for a little over 20 months (81.6%) of the overall gap. The remainder of less than one month (3.0%) is accounted for by gender differences in conviction regime probabilities. Overall, the explained portion of the gap accounts for about 15.4 months (62.7%) of the total gender sentencing difference. This leaves about 9.5 months (38.3%) that cannot be explained by gender differences in circumstances. Table 4.8 disaggregates the explained and unexplained portions of the overall sentencing gap by contributions from each sentencing regime and sentencing regime probabilities. The plea regime accounts for the largest contribution to the overall explained gap (13.5 months or 87.6%) and to the overall unexplained gap (6.8 months or 72.0%). In fact the largest single component of the constituent parts of the overall gender sentencing gap is the 13.5 month explained gap from the plea regime which accounts for 54.0% of the overall advantage of women in awarded sentences.

4.6 Conclusion

Unlike any studies in the literature so far, our study separates observed gender differences in sentencing into two different components – one attributable to differences in circumstances of male and female criminal defendants, and the second attributable to differences in attitudes of sentencing judges towards male and female defendants and the differences due to unobservable characteristics of the male and female defendants. Our model takes account of the joint determination of sentences by regime and conviction regime selection as well as censoring occasioned by sentences that do not involve prison time. We are able to determine the role of gender differences in selection regime probabilities. Such decomposition provides a better insight into the decision-making process of sentencing judges. Knowing whether judges consider extralegal circumstances in their decision making is important, but knowing how they consider extralegal circumstances is useful to policy makers in deciding how to reform sentencing guidelines to ensure equal treatment. This study not only examines whether judges consider extralegal circumstances but if they do, it asks whether they attach the same weight to circumstances of males and females. Even in light of the Supreme Court's decision in 2005 to strike down the Federal Sentencing Guidelines, our results may offer some guidance as to what to expect now that judges are less constrained in imposing sentences.

We find that women receive prison sentences that average a little over 2 years less than those awarded to men. Even after controlling for circumstances such as the severity of the offense and past criminal history, women receive more lenient sentences. Approximately 9.5 months of the female advantage cannot be explained by gender differences in individual circumstances. In other words if women faced the same sentencing structure as men, women would on average receive 15.4 months less prison time than men rather than 24.9 months less prison time. Most of the gender gap arises from convictions via guilty pleas, which account for the vast majority of the convictions observed in our data. Besides gender, we find evidence that judges took into account factors such as family circumstances which are expressly prohibited from consideration when awarding sentences.

One should bear in mind that our data permit us to examine only the end stage of the criminal justice system. A more comprehensive treatment would take account of the fact that before arriving at the judge for sentencing, a defendant must also pass through a jury or possible plea bargain with a prosecutor. Before reaching this stage, other groups, such as the police and the prosecution, have the potential to create bias in the criminal justice system. Future work will focus on separating out differential outcomes layer by layer, as well as making explicit the impact of gender bias in the criminal justice system on gender differences in labor market outcomes.

Year	Males			Females		
	Total (%)	Trial (%)	Plea (%)	Total (%)	Trial (%)	Plea (%)
1996	25.26	6.45	27.00	44.41	12.73	46.34
1997	25.25	4.83	26.99	41.85	21.74	42.80
1998	21.63	4.56	22.94	37.67	18.60	38.42
1999	21.98	9.76	22.74	39.97	34.88	40.15
2000	23.21	4.78	24.13	38.03	10.26	39.00
2001	21.67	9.90	22.10	35.84	14.81	36.32
2002	22.42	5.63	22.86	42.57	17.39	43.04

Table 4.1: Percentage of Sentences Involving No Prison Time

Variable Definitions and Summary Statistics

Variable	Description	Overall		Males		Females	
		Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
TOTALMONTHS	Length of prison sentence in months	37.27	73.12	41.67	78.03	16.76	37.07
REGIME	Indicator for trial regime	0.05	0.22	0.06	0.23	0.03	0.18
FEMALE	Indicator for female	0.18	0.38	0.00	0.00	1.00	0.00
FINEWAIV	Indicator of fine being waived	0.84	0.37	0.83	0.38	0.87	0.34
HISCHOOL	Indicator for high school education	0.24	0.43	0.23	0.42	0.28	0.45
GED	Indicator for general equivalency diploma	0.13	0.34	0.14	0.34	0.11	0.32
SOMECOLL	Indicator for some college attended	0.26	0.44	0.25	0.43	0.29	0.45
COLLGRAD	Indicator for a college degree or higher	0.12	0.32	0.13	0.33	0.07	0.26
NUMDEPEN	Number of dependents	1.13	1.41	1.13	1.44	1.09	1.29
MARRD	Indicator for married or cohabiting	0.26	0.44	0.26	0.44	0.25	0.43
CITIZN	Indicator for US citizen	0.95	0.21	0.95	0.22	0.97	0.16
DEFENSEP	Indicator for private counsel	0.36	0.48	0.37	0.48	0.30	0.46
XCRHISSR	Final criminal history category	2.11	1.60	2.23	1.66	1.57	1.16
CRIMHIS1	Final criminal history category =1	0.58	0.49	0.54	0.50	0.74	0.44
CRIMHIS2	Final criminal history category =2	0.12	0.32	0.12	0.33	0.09	0.29
CRIMHIS3	Final criminal history category=3	0.13	0.33	0.13	0.34	0.10	0.29
CRIMHIS4	Final criminal history category =4	0.06	0.24	0.07	0.25	0.03	0.17
CRIMHIS5	Final criminal history category =5	0.04	0.19	0.04	0.20	0.02	0.13
CRIMHIS6	Final criminal history category =6	0.08	0.27	0.09	0.29	0.03	0.16
XFOLSOR	Final offense level	16.80	8.30	17.37	8.35	14.11	7.51
XFOLSOR2	Final offense level squared	351.02	325.02	371.49	332.24	255.55	269.16
XFOLSOR3	Final offense level cubed	8507.77	11499.56	9140.23	11899.77	5558.17	8832.43
AGE	Age of defendant	37.74	11.14	38.08	11.22	36.16	10.59
AGE2	Age of defendant squared	1548.66	888.58	1576.29	900.04	1419.79	820.88
CIRC1	Circuit indicators	0.03	0.16	0.03	0.17	0.02	0.14
CIRC2	Circuit indicators	0.11	0.31	0.12	0.32	0.08	0.27
CIRC3	Circuit indicators	0.04	0.21	0.05	0.21	0.04	0.19
CIRC4	Circuit indicators	0.05	0.21	0.05	0.21	0.05	0.21
CIRC5	Circuit indicators	0.10	0.30	0.10	0.30	0.12	0.32
CIRC6	Circuit indicators	0.08	0.28	0.08	0.27	0.09	0.29
CIRC7	Circuit indicators	0.04	0.20	0.04	0.20	0.04	0.20
CIRC8	Circuit indicators	0.08	0.27	0.08	0.27	0.09	0.29
CIRC9	Circuit indicators	0.27	0.44	0.27	0.44	0.28	0.45
CIRC10	Circuit indicators	0.05	0.23	0.05	0.22	0.06	0.24
CIRC11	Circuit indicators	0.14	0.35	0.14	0.35	0.13	0.34
1996	Year indicators	0.13	0.33	0.13	0.33	0.12	0.33
1997	Year indicators	0.13	0.34	0.14	0.34	0.13	0.34
1998	Year indicators	0.13	0.34	0.13	0.34	0.14	0.35
1999	Year indicators	0.14	0.35	0.14	0.34	0.15	0.36
2000	Year indicators	0.14	0.35	0.14	0.35	0.15	0.35
2001	Year indicators	0.15	0.36	0.15	0.36	0.15	0.36
2002	Year indicators	0.17	0.37	0.17	0.38	0.16	0.36

Table 4.2: Variable Definitions and Summary Statistics

	Males			Females		
Year	Total	Trial	Plea	Total	Trial	Plea
1996	43.26	133.44	34.92	15.00	44.37	13.21
1997	43.81	136.47	35.93	19.27	94.50	15.73
1998	42.00	124.29	35.66	16.07	36.17	15.28
1999	42.15	112.62	37.79	15.08	30.66	14.51
2000	40.12	108.01	36.75	15.93	44.85	14.92
2001	41.20	94.50	39.24	18.27	59.19	17.34
2002	39.87	111.85	38.00	17.58	55.26	16.88

Table 4.3: Mean Sentences in Months

	Regime Selection		Trial Regime		Plea Regime	
Variable	Parameter	Asmp Z	Parameter	Asmp Z	Parameter	Asmp Z
Constant	-2.685	-17.17	-768.983	-17.66	-128.508	-30.91
FEMALE	-0.198	-5.90	-48.700	-5.67	-9.223	-10.29
FINEWAIV			3.944	0.80	4.695	5.73
HISCHOOL	0.001	0.03	-3.030	-0.41	-1.105	-1.50
GED	0.020	0.57	4.163	0.49	1.192	1.53
SOMECOLL	0.017	0.55	0.882	0.12	-1.519	-2.20
COLLGRAD	0.186	4.88	30.596	3.31	-1.836	-1.73
CITIZN	-0.043	-1.65				
MARRD	0.053	1.67	10.319	1.38	-2.715	-4.02
NUMDEPEN	-0.008	-0.97	-2.081	-1.12	-0.873	-4.55
DEFENSEP	-0.039	-1.72	-9.220	-1.66	-5.525	-9.48
CRIMHIS2	0.038	1.12	22.950	2.82	8.531	9.04
CRIMHIS3	0.030	0.87	29.251	3.50	18.795	24.54
CRIMHIS4	0.019	0.39	39.424	3.44	30.676	29.92
CRIMHIS5	0.107	1.93	76.201	5.84	38.444	27.52
CRIMHIS6	0.077	2.19	74.597	9.53	53.103	69.13
XFOLSOR			13.632	5.16	15.799	51.80
XFOLSOR2			-0.480	-4.73	-0.723	-53.66
XFOLSOR3			0.009	7.43	0.014	79.57
AGE	0.035	4.91	8.396	4.93	0.249	1.45
AGE2x10^{-2}	0.024	-2.79	-0.061	-2.99	-0.004	-2.01
CIRC2	-0.385	-5.92	-98.366	-6.30	-6.391	-3.24
CIRC3	-0.252	-3.37	-50.159	-2.75	1.783	0.79
CIRC4	-0.160	-2.14	-30.082	-1.70	12.412	5.70
CIRC5	-0.266	-3.98	-60.598	-3.79	10.880	5.49
CIRC6	-0.039	-0.59	-3.004	-0.19	8.319	4.00
CIRC7	-0.023	-0.31	6.603	0.36	17.144	7.76
CIRC8	-0.217	-3.30	-36.302	-2.29	-1.485	-0.76
CIRC9	-0.242	-4.04	-53.839	-3.68	2.105	1.13
CIRC10	-0.184	-2.58	-34.254	-2.02	3.417	1.57
CIRC11	0.003	0.05	7.136	0.49	10.045	5.30
1996	0.535	12.04	123.182	11.53	-0.603	-0.56
1997	0.506	11.45	115.919	11.12	-0.030	-0.03
1998	0.464	10.28	104.472	9.68	0.529	0.47
1999	0.348	8.05	71.825	7.04	1.511	1.49
2000	0.258	5.80	56.937	5.38	1.800	1.78
2001	0.128	2.82	25.239	2.29	1.759	1.80
Sigma 0	47.109	958.51				
Rho 0u	-0.663	-67.95				
Sigma 1	215.842	102.38				
Rho 1u	0.994	1692.69				
N	45060		2333		42727	
Log-Likelihood	-189907.6					

Table 4.4: Censored Switching Regression with Endogenous Switching: Pooled Sample

	Regime Selection		Trial Regime		Plea Regime	
Variable	Parameter	Asmp Z	Parameter	Asmp Z	Parameter	Asmp Z
Constant	-2.638	-15.64	-767.458	-15.67	-128.041	-26.57
FINEWAIV			3.340	0.60	5.149	5.52
HISCHOOL	0.017	0.51	-1.423	-0.17	-1.611	-1.88
GED	0.036	0.94	7.070	0.75	1.172	1.32
SOMECOLL	0.025	0.77	1.778	0.21	-1.885	-2.37
COLLGRAD	0.171	4.18	26.557	2.58	-1.521	-1.28
CITIZN	-0.042	-1.52				
MARRD	0.055	1.60	11.390	1.35	-3.354	-4.30
NUMDEPEN	-0.006	-0.70	-1.706	-0.83	-0.948	-4.33
DEFENSEP	-0.049	-1.96	-11.229	-1.80	-6.532	-9.74
CRIMHIS2	0.033	0.89	22.167	2.42	8.394	7.75
CRIMHIS3	0.041	1.13	32.000	3.49	19.239	21.88
CRIMHIS4	0.035	0.71	43.492	3.50	31.494	27.17
CRIMHIS5	0.113	1.94	78.039	5.39	38.609	24.59
CRIMHIS6	0.077	2.11	76.024	9.01	53.490	61.44
XFOLSOR			11.693	3.94	15.581	43.19
XFOLSOR2			-0.404	-3.54	-0.699	-44.14
XFOLSOR3			0.008	5.97	0.014	65.83
AGE	0.034	4.46	8.507	4.43	0.185	0.94
AGE2x10^{-2}	0.024	-2.64	-0.063	-2.76	-0.003	-1.40
CIRC2	-0.397	-5.80	-103.768	-6.08	-6.478	-2.89
CIRC3	-0.258	-3.26	-52.720	-2.63	1.005	0.39
CIRC4	-0.166	-2.10	-33.349	-1.71	12.327	4.94
CIRC5	-0.278	-3.89	-61.521	-3.49	9.425	4.14
CIRC6	-0.074	-1.06	-9.518	-0.53	7.088	2.98
CIRC7	-0.034	-0.43	4.737	0.24	15.882	6.31
CIRC8	-0.246	-3.52	-42.872	-2.44	-1.560	-0.69
CIRC9	-0.259	-4.08	-57.704	-3.59	1.188	0.56
CIRC10	-0.176	-2.33	-30.395	-1.63	2.539	1.01
CIRC11	-0.013	-0.21	4.682	0.29	9.672	4.48
1996	0.538	11.14	126.383	10.40	-1.064	-0.85
1997	0.514	10.70	120.490	10.17	-0.825	-0.68
1998	0.481	9.72	111.016	8.99	-0.272	-0.21
1999	0.357	7.65	75.419	6.58	1.409	1.19
2000	0.257	5.34	55.756	4.69	1.362	1.16
2001	0.128	2.62	24.521	1.99	1.539	1.36
Sigma 0	49.543	858.88				
Rho 0u	-0.692	-72.20				
Sigma 1	224.546	93.54				
Rho 1u	0.994	1535.52				
N	37104		2057		35047	
Log-Likelihood	-163499.9					

Table 4.5: Censored Switching Regression with Endogenous Switching: Males

Variable	Males	Females	Difference	Male Fitted	Females Fitted (Male Weights)
$\bar{y}$	41.673	16.757	24.916	43.320	25.737
$\bar{y}_T$	120.845	51.736	69.109	141.770	89.281
$\bar{y}_P$	37.027	15.500	21.526	37.320	22.776
$\bar{p}_T$	0.055	0.035	0.021	0.062	0.051
$\hat{\delta}_T$	0.852				
$\hat{\delta}_P$	0.992				
$\hat{\delta}_s$	0.894				
$\hat{\delta}$	0.962				

Table 4.6: Mean Sentences and Conviction-by-Trial Probabilities

Variable	Explained	Unexplained	Total Gap
$\bar{y}_{Tm} - \bar{y}_{Tf}$	31.564	37.545	*69.109*
$\bar{y}_{Pm} - \bar{y}_{Pf}$	14.251	7.275	*21.527*
$\bar{p}_{Tm} - \bar{p}_{Tf}$	0.004	0.016	*0.021*
$\bar{p}_T$	0.055	0.035	0.021

Table 4.7: Decomposition by Part

Explained		Unexplained		Total Gap	
$E_T\, \bar{p}_{Tm}$	1.750	$U_T\, \bar{p}_{Tm}$	2.081	$E_T\, \bar{p}_{Tm} + U_T\, \bar{p}_{Tm}$	3.831
$E_P\, (1 - \bar{p}_{Tm})$	13.461	$U_P\, (1 - \bar{p}_{Tm})$	6.872	$E_P\, (1 - \bar{p}_{Tm}) + U_P\, (1 - \bar{p}_{Tm})$	20.333
$E_s\, (\bar{y}_{Tf} - \bar{y}_{Pf})$	0.159	$U_s\, (\bar{y}_{Tf} - \bar{y}_{Pf})$	0.593	$E_s\, (\bar{y}_{Tf} - \bar{y}_{Pf}) + U_s\, (\bar{y}_{Tf} - \bar{y}_{Pf})$	0.752
E	15.370	U	9.546	$E + U$	24.916

Table 4.8: Contribution to Total

Appendix A: Inversion of Nested Logit in Our Model

Berry derives the inversion of the nested logit model as follows:

$$s_{oj} = s_{oj|g_j} * s_{g_j} \tag{4.38}$$

$$s_{oj|g_j} = \frac{\exp(\frac{\delta_{oj}}{1-\sigma})}{D_{g_j}} \tag{4.39}$$

$$s_{g_j} = \frac{D_{g_j}^{1-\sigma}}{[\sum_{j \in G_g} D_{g_j}^{1-\sigma}]} \tag{4.40}$$

$$D_{g_j} = \sum_{j \in G_g} \exp(^{\frac{\delta_j}{1-\sigma}}) \tag{4.41}$$

There are two key simplifying assumptions Berry makes to derive the inversion of the nested logit model:

a) The utility of the outside good is normalized to 0, which does not hold for our model

b) The outside good is the sole member of its group, which does not hold for our model

These assumptions imply:

$$s_{oo} = s_{g0} * 1 = \frac{1}{[\sum_{j \in G_g} D_{g_j}^{1-\sigma}]} \tag{4.42}$$

The share ratios can then be expressed as:

$$\frac{s_j}{s_o} = \frac{\exp(\frac{\delta_j}{1-\sigma})}{D_{g_j}} \frac{D_{g_j}^{1-\sigma}}{[\sum_{j \in G_g} D_{g_j}^{1-\sigma}} / [\frac{1}{[\sum_{j \in G_g} D_{g_j}^{1-\sigma}]}] = \frac{\exp(\frac{\delta_{oj}}{1-\sigma})}{D_g} \tag{4.43}$$

$$\ln(s_{oj}) - \ln(s_{oo}) = \frac{\delta_{oj}}{1-\sigma} - \ln(D_{g_j}) \tag{4.44}$$

Going back to equation 4.40, we see

$$s_{g_j} = \frac{D_{g_j}^{1-\sigma}}{[\sum_{j \in G_g} D_{g_j}^{1-\sigma}} = D_{g_j}^{1-\sigma} * s_{oo} \Rightarrow \ln(s_{g_j}) = (1-\sigma) * \ln(D_{g_j}) + \ln(s_{oo}) \tag{4.45}$$

Solving for $\ln(D_g)$ and plugging into equation (7)

$$\ln(s_{oj}) - \ln(s_{oo}) = \delta_j + \sigma * \ln(s_{j|g}) \tag{4.46}$$

Because there is more than one choice in the outside good's group (also we do not normalize the utility of the outside good to 0, but that is not the key difference between Berry and the migration problem):

$$s_{g_o} = \frac{D_o^{1-\sigma}}{[\sum_{j \in G_g} D_{g_j}^{1-\sigma}]} \tag{4.47}$$

$$s_{oo} = s_{g_o} * s_{o|g_0} = \frac{\exp(\frac{\delta_{o0}}{1-\sigma})}{D_0} \frac{D_0^{1-\sigma}}{[\sum_{j \in G_g} D_g^{1-\sigma}]} \tag{4.48}$$

So equation 4.43 now becomes:

$$\frac{s_{oj}}{s_{oo}} = \frac{\exp(\frac{\delta_{oj}}{1-\sigma})}{D_g} \frac{D_g^{1-\sigma}}{[\sum_{j \in G_g} D_g^{1-\sigma}]} / [\frac{\exp(\frac{\delta_{oo}}{1-\sigma})}{D_o} \frac{D_o^{1-\sigma}}{[\sum_{j \in G_g} D_g^{1-\sigma}]}] = \frac{\exp(\frac{\delta_{oj}}{1-\sigma})}{\exp(\frac{d_{oo}}{1-\sigma})} * (\frac{D_{g_j}}{D_o})^{-\sigma} \tag{4.49}$$

Taking logs,

$$\ln(s_{oj}) - \ln(s_{oo}) = \frac{\delta_{oj} - \delta_{oo}}{1-\sigma} - \sigma * [\ln(D_{g_j}) - \ln(D_o)] \tag{4.50}$$

Taking the log of equation 4.41 and 10, respectively, we now have

$$\ln(D_{g_j}) = \frac{\ln([\sum_{j \in G_g} D_g^{1-\sigma}]) + \ln(s_{oj})}{1-\sigma} \tag{4.51}$$

$$\ln(D_o) = \frac{\ln([\sum_{j \in G_g} D_{g_j}^{1-\sigma}]) + \ln(s_{oo})}{1-\sigma} \tag{4.52}$$

Subtracting these two terms,

$$\ln(D_{g_j}) - \ln(D_o) = \frac{\ln(s_{g_j}) - \ln(s_{oo})}{1-\sigma} \tag{4.53}$$

Which gives us (as the differencing of the deltas is equivalent to normalizing the utility of the outside good to 0) the same results at Berry:

$$\ln(s_{oj}) - \ln(s_{oo}) = \delta_{oj} - \delta_{oo} + \sigma * (\ln_{j|g_j}) \tag{4.54}$$

Appendix B: Symmetry of Creation and Retention

One complication we face in reporting marginal effects is the choice of which marginal effect to report. As the s variable varies across observations, a different estimation of the marginal effect can be computed at each observation One natural solution to this problem would be to report the mean of all estimated marginal effects. However, as we show below, this approach also results in a serious complication.

Consider again the migration retention term in Table 3.1. Table 3.1 expresses the number of migrants retained in location o as the number of individuals in the o who substitute to location o when a characteristic of o changes (own effects):

$$retention_o = N_o(1 - s_{oo}) * s_{oo} * \beta \tag{4.55}$$

Another way in which we can think of this term is as the number of people in o who substitute away from every other possible choice when a characteristic of o changes (summing over cross effects)

$$retention_o = \sum_{j \neq o} N_o s_{oo} s_{oj} \beta \tag{4.56}$$

Now consider the measure of creation in Table 3.1.

$$creation_o = \sum_{j \neq o} N_j * s_{jj} s_{jo} * \beta \tag{4.57}$$

Suppose that we want to find the average amount of retention. This would entail computing $retention_o$ for each o. To do so, we simply sum over o and divide by the number of choices in the set:

$$\overline{retention} = \frac{\sum_o \sum_{j \neq o} N_0 s_{oo} s_{oj} \beta}{J} \tag{4.58}$$

To find the average amount of creation, we do the same operation to $creation_o$:

$$\overline{creation} = \frac{\sum_o \sum_{j \neq o} N_j * s_{jj} s_{jo} * \beta}{J} \tag{4.59}$$

Note this symmetry: the set containing the $N_j * s_{jj} s_{jo}$ term for every possible j and o will be identical to the set containing $N_0 s_{oo} s_{oj}$ for every possible value of o and j. Thus, the average effects of retention and creation will be identical if we sum over all observations. To obtain more intuitive results, we instead estimate averages of the marginal effects for each of the 9 census regions.

Appendix C: Expected Value of Dependent Variable with Censoring

The expected value of a censored dependent variable is simply the product of the probability of observing a non-limit observation and the expected value of the dependent variable given that it is a non-limit observation, plus the probability of observing a limit observation times the expected value of the dependent variable given that it is a limit observation. Because the censoring point is at zero, the expected value of limit observations is 0, causing the second term to drop from the expression. We first consider the trial regime:

$$
\begin{aligned}
E[y_{Ti}|s_i = 1] = \Pr(y^*_{Ti} > 0|s_i = 1) \cdot E[y_{Ti}|y^*_{Ti} > 0 \cap s_i = 1] \\
+ \Pr(y^*_{Ti} \leq 0|s_i = 1) \cdot E[y_{Ti}|y^*_{Ti} \leq 0 \cap s_i = 1] \\
= \Pr(y^*_{Ti} > 0|s_i = 1) \cdot E[y_{Ti}|y^*_{Ti} > 0 \cap s_i = 1] \\
+ \Pr(y^*_{Ti} \leq 0|s_i = 1) \cdot 0 \\
= \Pr(y^*_{Ti} > 0|s_i = 1) \cdot E[y_{Ti}|y^*_{Ti} > 0 \cap s_i = 1] \qquad (4.60)
\end{aligned}
$$

Consider each of the two right hand size terms separately. First, consider the probability of observing a non-limit observation, conditional upon selection. From our specification of the data generating process for y^* and s, we can express this as the function of two random variables, ε and u.

$$\Pr(y^*_{Ti} > 0|s_i = 1) = \Pr(\varepsilon_{Ti} < X_{Ti}\beta_T|u_i < Z_i\gamma) \qquad (4.61)$$

By Bayes' rule we can express this as the joint probability that a non-limit observation is selected into the trial regime, divided by the probability of that observation being in the trial regime. This term can then be expressed using values from the cumulative normal and cumulative bivariate normal distributions.

$$
\begin{aligned}
\Pr(y^*_{Ti} &> 0|s_i = 1) = \frac{\Pr(\frac{\varepsilon_{Ti}}{\sigma_T} < \frac{X_{Ti}\beta_T}{\sigma_T} \cap u_i < Z_i\gamma)}{\Pr(u_i < Z_i\gamma)} \\
&= \frac{\Phi_2(\frac{X_{Ti}\beta_T}{\sigma_T}, Z_i\gamma, \rho_{sT})}{\Phi(Z_i\gamma)} \ . \qquad (4.62)
\end{aligned}
$$

Finally, we must consider the expected value of the dependent variable, given that it is a non-limit observation in the trial regime. Recall that non-limit observations take on the value

$$
\begin{aligned}
E[y_{Ti}|y^*_{Ti} &> 0 \cap s_i = 1] = E[y^*_{Ti}|y^*_{Ti} > 0 \cap s_i = 1] \\
&= E[y^*_{Ti}|y^*_{Ti} > 0 \cap s^*_i > 0] \\
&= E[y^*_{Ti}|\frac{\varepsilon_{Ti}}{\sigma_T} < \frac{X_{Ti}\beta_T}{\sigma_T} \cap u_i < Z_i\gamma] \ . \qquad (4.63)
\end{aligned}
$$

This expected value appears similar to the expected value of the dependent variable in the main equation of the Heckit model: it is truncated by the draw for the error term in the selection equation. It also appears similar to the expected value of the dependent variable in the Tobit model: it is truncated by the draw for the error term in the main equation. This incidence of "double truncation" however, is substantially more complex than the single truncation in either the Tobit or the Heckit. We derive it for our model based on page 72 of Johnson and Kotz (1972):

$$\begin{aligned} E[y_{Ti}|y_{Ti}^* &> 0 \cap s_i = 1] = \frac{X_{Ti}\beta_T}{\Phi_2(\frac{X_{Ti}\beta_T}{\sigma_T}, Z_i\gamma, \rho_{sT})} \\ &\cdot \left\{ \sigma_T\{\phi(\frac{-X_{Ti}\beta_T}{\sigma_T})\Phi(\frac{-1}{\sqrt{1-\rho_{sT}^2}}[-Z_i\gamma - \rho\frac{-X_{Ti}\beta_T}{\sigma_T}]) \right. \\ &\left. + \rho_{sT}\phi(-Z_i\gamma)\Phi(\frac{-1}{\sqrt{1-\rho_{sT}^2}}[-X_{Ti}\beta_T - \rho(-Z_i\gamma)]) \right\} \end{aligned} \tag{4.64}$$

The resulting expected value of the length of sentence in the trial regime is:

$$E[y_{Ti}|s_i = 1] = \frac{\Phi_2(\frac{X_{Ti}\beta_T}{\sigma_T}, Z_i\gamma, \rho_{sT})}{\Phi(Z_i\gamma)} * E[y_{Ti}^*] \tag{4.65}$$

We can then define the $\hat{y}_T(X, Z, \hat{\theta}_m) = E[y_{Ti}^*]$ as given above.

Appendix D: A Note on Sample Mean Prediction Error in Decompositions

In decomposition analysis, the standard term to decompose is the difference between the sample mean of the dependent variable for two groups. Define the sample mean values for groups m and f as $\overline{y}_m$ and $\overline{y}_f$, where each group has N_m and N_f members, respectively. After estimating an econometric equation for both of the groups, we can then calculate fitted values $\widehat{y}_{mi}$ and $\widehat{y}_{fi}$ for each individual in groups m and f, respectively. The average fitted value for members of these groups is:

$$\widehat{y}_m = \frac{1}{N_m}\sum_{i=1}^{N_m}\widehat{y}_{mi} \tag{4.66}$$

$$\widehat{y}_f = \frac{1}{N_f}\sum_{i=1}^{N_f}\widehat{y}_{fi} \tag{4.67}$$

Define $\widehat{y}^o_{fi}$ as the fitted value of an observation in group f, had that individual faced the group m estimated parameters. The mean of this variable for group f is then:

$$\widehat{y}^0_f = \frac{1}{N_f}\sum_{i=1}^{N_f}\widehat{y}^0_{fi} \tag{4.68}$$

By adding and subtracting the $\widehat{y}^0_f$ term, the decomposition is then expressed as:

$$\overline{y}_m - \overline{y}_f = (\overline{y}_m - \widehat{y}^0_f) + (\widehat{y}^0_f - \overline{y}_f) \tag{4.69}$$

where the first term expresses the difference in the left hand side variable which can be attributed to differences in the characteristics of the two groups, and the second term expresses the difference caused by differences in the parameters the two groups face.

Assuming that the underlying model can be consistently estimated, we would have

$$\begin{aligned} \text{plim}(\widehat{y}_m - \overline{y}_m) &= 0 & (4.70) \\ \text{plim}(\widehat{y}_f - \overline{y}_f) &= 0 & (4.71) \end{aligned}$$

However, in a finite sample, the $\widehat{y}$ and $\overline{y}$ terms will not necessarily be equal. We can express the sample mean prediction error in the model as follows:

$$\begin{aligned} \overline{y}_m &= \widehat{\delta}_m \widehat{y}_m & (4.72) \\ \overline{y}_f &= \widehat{\delta}_f \widehat{y}_f & (4.73) \end{aligned}$$

It follows from consistency that

$$\text{plim}(\widehat{\delta}) = \text{plim}\left(\frac{\overline{y}}{\widehat{y}}\right) = 1$$

The decomposition can now be expressed as:

$$\overline{y}_m - \overline{y}_f = (\widehat{\delta}_m \widehat{y}_m - \widehat{y}_f^o) + (\widehat{y}_f^o - \widehat{\delta}_f \widehat{y}_f) \quad (4.74)$$

The impact of the estimation error becomes more clear if, instead of adding and subtracting $\widehat{y}_f^o$, we instead add and subtract $\delta_m \widehat{y}_f^o$

$$\begin{aligned} \overline{y}_m - \overline{y}_f &= (\widehat{\delta}_m \widehat{y}_m - \widehat{\delta}_m \widehat{y}_f^0) + (\widehat{\delta}_m \widehat{y}_f^0 - \widehat{\delta}_f \widehat{y}_f) \\ &= (\widehat{\delta}_m \widehat{y}_m - \widehat{\delta}_m \widehat{y}_f^0) + \left(\widehat{\delta}_m - \widehat{\delta}_f\right) \widehat{y}_f^0 + \widehat{\delta}_f \left(\widehat{y}_f^0 - \widehat{y}_f\right) & (4.75) \\ &= (\overline{y}_m - \widehat{\delta}_m \widehat{y}_f^0) + (\widehat{\delta}_m \widehat{y}_f^0 - \overline{y}_f) & (4.76) \end{aligned}$$

Thus, the $\widehat{\delta}$ terms contribute to both the explained and unexplained portions of the mean decomposition.

In principle it is possible to separate out the effect of gender differences in the $\widehat{\delta}$

parameter from the effect of differences in other parameters eq (4.75). However, this is only feasible if the econometrician estimates both the $\widehat{\delta}_m$ and $\widehat{\delta}_f$ terms. In our case, we lack sufficient data to identify the weights in the model for females. Consequently, we only are able to decompose the difference in mean outcomes into the portion caused by differences in weights and differences in characteristics according to eq (4.76).

References

[1] Allard, Scott; Danziger, Sheldon (2000). "Welfare magnets: myth or reality?", *The Journal of Politics,* 62: 350-368.

[2] Alston, Lee (1981). "Tenure choice in Southern agriculture", *Explorations in Economic History,* 18: 211-232.

[3] Anderson, James M.; Kling, Jeffrey R.; Stith, Kate (1999). "Measuring inter-judge sentencing disparity: before and after the Federal sentencing guidelines." *Journal of Law & Economics*, 42: 271 - 298.

[4] Anderson, Gary M.; Robert D. Tollison (1991). "Congressional influence and patterns of New Deal spending, 1933-1939", *Journal of Law and Economics*, 34: 161-75.

[5] Angelucci, Manuela (2004). "U.S. Border enforcement and the net inflow of illegal migration", *UCL Working Paper.*

[6] Arizona Daily Star (2006). "Sealing our border: it won't work", *Arizona Daily Star*, September 24th 2006.

[7] Arizona Daily Star (2006). "Illegal border crossers now favoring California." *Arizona Daily Star*, August 8th 2006.

[8] Bartel, Ann P (1989). "Where do the new U.S. immigrants live?", *Journal of Labor Economics*, 7: 371-391.

[9] Becker, G. (1968). "Crime and punishment: an economic approach", *Journal of Political Economy*, 76: 169-217.

[10] Berry, Steven T (1994). "Estimating discrete-choice models of product differentiation", *The RAND Journal of Economics*, 25: 242-262.

[11] Berry, Steven; Levinsohn, James; Pakes, Ariel (1995). "Automobile prices in market equilibrium", *Econometrica,* 63: 841–889.

[12] ______ (2004). "Differentiated products demand system from a combination of micro and macro data: the new car market", *Journal of Political Economy,* 112: 68-105.

[13] Berry, Steven T.; Waldfogel, Joel (2001). "Do mergers increase product variety? Evidence from radio broadcasting", *The Quarterly Journal of Economics*, 30: 1009-1025.

[14] Bean, Frank D; Chanove, Roland; Cushing, Robert G; de la Garza, Rodolfo; Freeman, Gary P.; Haynes, Charles W.; Spencer, David (1994). "Illegal Mexican migration and the United States/Mexico border: the effects of Operation Hold the Line on El Paso/Juarez", *U.S. Commission on Immigration Reform Research Paper.*

[15] Biles, Roger (1994). The South and the New Deal. University of Kentucky Press, Lexington, KY.

[16] Blank, Rebecca (1998). "The effect of welfare and wage levels on the location decisions of female-headed households", *Journal of Urban Economics*, 24: 186-211.

[17] Block, Michael K.; Gerety, Vernon E (1995). "Some experimental evidence on differences between student and prisoner reactions to monetary penalties and risk", *Journal of Legal Studies*, 24: 123 -138.

[18] Bogue, Donald (1951). State economic areas: A description of the procedure used in making a functional grouping of the counties of the United States. U.S. Bureau of the Census. Washington, D.C.: Government Printing Office.

[19] Bogue, Donald ; Beale, Calvin (1953). "Economic Subregions of the United States", Farm Population. Series Census-BAE, No. 19. Washington, D.C. Government Printing Office.

[20] Bogue, Donald.J.; Shyrock Jr., H.S.; Hoermann, S.A; (1957) Subregional migration in the United States, 1935-40. Volume I, Streams of Migration Between Subregions. Scripps Foundation, Oxford, OH.

[21] Borjas, George J.; Bronars, Stephen G.; Trejo, Stephen J (1992). "Self-Selection and Internal Migration in the United States." *Journal of Urban Economics*, 32: 159-185.

[22] Blundell, Richard W.; Smith, Richard J (1989). "Estimation in a class of simultaneous equation limited dependent variable models", *The Review of Economics Studies*, 56: 37-57.

[23] Brown, Josephine Chapin (1940). Public Relief, 1929-1939 Henry Holt and Company, New York, NY.

[24] Bureau of Justice Statistics, Federal Justice Statistics Program (2001). "Defendants sentenced under the guidelines during fiscal year 2001: SC01OUT". Washington D.C.:The Urban Institute (distributor).

[25] Cardell, Scott (1997). "Variance components structures for extreme-value and logistic distributions with application to models of heterogeneity", *Econometric Theory*, 13: 185-213.

[26] Carrión-Flores, Carmen E. (2005). "What makes you go back home? Determinants of the duration of migration of Mexican immigrants in the U.S." Mimeo.

[27] Chicago Daily Law Bulletin (2005),.December 27, Law Bulletin Publishing Company.

[28] Congressional Quarterly Almanac: 1975-1997.

[29] Congressional District Almanac: 1977, 1983, 1993.

[30] Cornelius WA (2001). "Death at the border: efficacy and unintended consequences of U.S. immigration control policy", *Population and Development Review*, 27: 661-685.

[31] Couch, Jim, Atkinson, Keith E.; Wells, William H. (1998). "New Deal agricultural appropriations: a political influence", *Eastern Economic Journal*, 24: 137-48.

[32] Couch, Jim; Shughart II, William (1998). The political economy of the New Deal. Edward Elgar, New York, NY.

[33] Couch Jim; Williams, Peter M. (1999). "New Deal or same old shuffle? The distribution of New Deal dollars across Alabama", *Economics and Politics*, 11: 213-223.

[34] Cragg, Michael; Kahn, Matthew (1997). "New estimates of climate demand: evidence from location choice", *Journal of Urban Economics*, 42: 261-284.

[35] Ehrlich, I (1973). "Participation in illegitimate activities: a theoretical and empirical investigation", *Journal of Political Economy*, 81: 521-565.

[36] _________ (1977) "Capital punishment and deterrence: some further thoughts and additional evidence", *Journal of Political Economy*, 85: 741-788.

[37] Fairly, Robert W (2005). "An extension of the Blinder-Oaxaca Decomposition technique to logit and probit models", *Journal of Economic and Social Measurement*, 30: 305-316.

[38] Federal Works Agency, Work Projects Administration. Final Statistical Report of the Federal Emergency Relief Administration (Washington, DC: GPO, 1942).

[39] Fishback, Price V., Haines, Michael R.; Kantor, Shawn (2001). "The impact of New Deal programs on black and white infant mortality in the South", *Explorations in Economic History*, 38: 93-122.

[40] Fishback, Price V.; Horrace, William C.; Kantor, Shawn (2006). "Do federal programs affect internal migration? The impact of New Deal expenditures on mobility during the Great Depression", *Explorations in Economic History*, 43: 179-222.

[41] Fishback, Price V.; Horrace, William C.; Kantor, Shawn (2005a). "Do federal programs affect internal migration? The impact of New Deal expenditures on mobility during the Great Depression", NBER Working paper No. w8283. Jan 2005.

[42] Fishback, Price V.; Horrace William C.; Kantor, Shawn (2005b). "The impact of New Deal expenditures on local economic activity: an examination of retail sales, 1929-1939", *Journal of Economic History*, 65: 36-71

[43] Fishback, Price V., Kantor, Shawn; Wallis, John Joseph (2003) "Can the New Deal's Three Rs be rehabilitated? A program-by-program, county-by-county analysis", *Explorations in Economic History*, 40: 278-307.

[44] Fleck, Robert Kenneth (1994). "Essays on the political economy of the New Deal", unpublished Ph.D. dissertation, Stanford University.

[45] _____(2001a). "Inter-party competition, intra-party competition, and distributive policy: a model and test using New Deal data", *Public Choice,* 108 : 77-100.

[46] _____(2001b). "Population, land, economic conditions, and the allocation of New Deal spending", *Explorations in Economic History*, 38: 296-304.

[47] _____(1999a). "The marginal effect of New Deal relief work on county-level unemployment statistics", *Journal of Economic History*, 59: 659-87.

[48] _____(1999b). "The value of the vote: a model and test of the effects of turnout on distributive policy", *Economic Inquiry*, 37: 609-23.

[49] Freeman, Richard B (1996). "Why do so many young American men commit crimes and what might we do about it?", *Journal of Economic Perspectives*, 10: 25-42.

[50] General Accounting Office (2006) "Illegal immigration: border-crossing deaths have doubled since 1995; Border Patrol's efforts to prevent deaths have not been fully evaluated" August 2006.

[51] Gramlich, Edward; Laren, Deborah (1984). "Migration and income redistribution responsibilities", *Journal of Human Resources*, 19: 489-511.

[52] Gould, Eric D.; Weinberg, Bruce A.; Mustard, David B (2002). "Crime rates and local labor market opportunities in the United States: 1979–1997", *Review of Economics and Statistics*, 84: 45-61.

[53] Greene, William H (2003). Econometric analysis, fifth edition: Prentice Hall, Upper Saddle River NJ.

[54] Grogger, Jeffrey (1995). "The effect of arrests on the employment and earnings of young men.," *Quarterly Journal of Economics*, 110: 51-71.

[55] __________ (1998). "Market wages and youth crime", *Journal of Labor Economics*, 16: 756-791.

[56] Hanson, Gordon (2006). "Illegal migration from Mexico to the United States." *Journal of Economic Literature*, forthcoming.

[57] Hanson, Gordon H.; Spilimbergo, Antonio (1999). "Illegal immigration, border enforcement, and relative wages: evidence from apprehensions at the U.S.-Mexico border", *The American Economic Review,* 89: 1337-1357.

[58] Hanson, Gordon H.; Spilimbergo, Antonio (2001). "Political economy, sectoral shocks, and border enforcement", *The Canadian Journal of Economics*, 34: 612-638.

[59] Hanson, Gordon H.; Robertson, Raymond; Spilimbergo, Antonio (2002). "Does border enforcement protect U.S. workers from illegal immigration", *Review of Economics & Statistics*, 84: 73-92.

[60] Herzog, Henry W., Jr.; Schlottmann, Alan M.(1986). "The metro-rating game: what can be learned from recent migrants?" *Growth and Change*, 17: 37-50.

[61] Holley, William C.; Winston, Ellen; Woofter, Thomas Jackson (1971). The Plantation South, 1934-1937. Freeport, NY, Books for Library Press.

[62] Howard, Donald S (1943). The WPA and Federal Relief Policy. Russell Sage Foundation, New York, NY.

[63] Imai, Susumu; Krishna, Kala (2004). "Employment, deterrence and crime in a dynamic model", *International Economic Review* 45: 845-872.

[64] Jaeger, D.A. (2000). "Local Labor Markets, Admission Categories, and Immigrant Location Choice" Mimeo, Hunter College and Graduate School of CUNY, New York and IZA, Bonn.

[65] Johnson, Norman L.; Kotz, Samuel (1972). Distributions in statistics: continuous multivariate distributions. John Wiley and Sons, New York.

[66] Johnson, Ryan S.; Kantor, Shawn; Fishback, Price V. (2007) "Striking at the roots of crime: the impact of social welfare spending on crime during the Great Depression", NBER Working Paper No.12825.

[67] Kauffman, Kyle; Kiesling, Lynne (1997). "Was there a nineteenth entury welfare magnet in the United States? Preliminary results from New York City and Brooklyn.", *The Quarterly Review of Economics and Finance,* 37: 439-448.

[68] Kempf-Leonard, Kimberly; Sample, Lisa L. (2001). "Have federal sentencing guidelines reduced severity? An examination of one circuit", *Journal of Quantitative Criminology*, 17: 111-144.

[69] Kling, Jeffrey (2006). "Incarceration length, employment and earnings", *The American Economic Review*, 96: 863-876.

[70] Kuziemko, Ilyana (2006). "Does the threat of the death penalty affect plea bargaining in murder cases? Evidence from New York's 1995 reinstatement of capital punishment." *American Law and Economics Review* 8: 116-142.

[71] Levine, Phillip; Zimmerman, David (1999). "An empirical analysis of the welfare magnet debate using the NLSY", *Journal of Population Economics,* 12: 391-409.

[72] Levitt, Steven D (1997). "Using electoral cycles in police hiring to estimate the effect of police on crime", *American Economic Review*, 87: 270-290.

[73] Lott, John R. Jr. (1992). "An attempt at measuring the total monetary penalty from drug convictions: the importance of an individual's reputation", *Journal of Legal Studies*, 21: 159-187.

[74] McFadden, Daniel (1973). "Conditional logit analysis of qualitative choice behavior", in P.Zarembka, ed., Frontiers in Econometrics, Academic Press, New York.

[75] Mertz, Paul E (1978). New Deal policy and Southern rural poverty. Louisiana State University Press, Baton Rouge, LA.

[76] Moffitt, Robert (1992). "Incentive effects of the U.S. welfare system: a review", *Journal of Economic Literature*, 30: 1-61.

[77] Mustard, D.B (2001). "Racial, ethnic, and gender disparities in sentencing: evidence from the U.S. Federal Courts", *Journal of Law and Economics*, 44: 285 -314.

[78] Myers, Samuel L Jr (1983). "Estimating the economic model of crime: employment versus punishment effects", *Quarterly Journal of Economics*, 98: 157-166.

[79] Neuman, Shoshana; Oaxaca, Ronald L. (2004). "Wage decompositions with selectivity-corrected wage equations: a methodological note", *Journal of Economic Inequality*, 2: 3-10.

[80] New York Times (2007). "Justices to revisit thorny issue of sentencing guidelines in first case after recess". Section A15: Feb 20.

[81] Oaxaca, Ronald L (1973). "Male-female wage differentials in urban labor markets", *International Economic Review*, 14: 693-709.

[82] Oaxaca, Ronald L.; Ransom, Michael R (1994). "On discrimination and the decomposition of wage differentials", *Journal of Econometrics*, 61: 5-21.

[83] Oaxaca, Ronald L.; Sarnikar, Supriya (2005). "Do females receive lenient sentences despite the federal sentencing guidelines?". Mimeo.

[84] Orrenius, P.M.; Zayodnym M. (2003). "Do amnesty programs reduce undocumented immigration? Evidence from IRCA", *Demography*, 40: 347-450.

[85] Petrin, A; Train, K. (2002). "Omitted product attributes in discrete choice models.", Working Paper, Department of Economics, University of California, Berkeley.

[86] Reading, Don C. (1973). "New Deal activity and the states, 1933 to 1939", *Journal of Economic History*, 1973: 792-810.

[87] Roy, A. D. (1951). "Some thoughts on the distribution of earnings." *Oxford Economic Papers*, 3: 135-146.

[88] Rhode, Paul; Strumpf, Koleman (2003). "Assessing the importance of Tiebout sorting: local heterogeneity from 1850 to 1990", *American Economic Review*, 93: 1648-77.

[89] Saloutos, Theodore (1974). "New Deal agricultural policy: an evaluation", *The Journal of American History*, 61: 394-416.

[90] Schanzenbach, M.(2005). "Racial and sex disparities in prison sentences: the effect of district-level judicial demographics", *Journal of Legal Studies*, 34: 57-92.

[91] Schaefer, Donald F (1989). "Location Choice in the Antebellum South", *The Journal of Economic History,* 49: 145-165.

[92] Sjaastad, Larry A (1962). "The costs and returns of human migration", *The Journal of Political Economy*, 70: 80-93.

[93] Tonry, Michael (1996). Sentencing Matters, Oxford University Press, Oxford.

[94] Urrea, Luis.Alberto (2004). The Devil's Highway. Back Bay Books: New York, 2004.

[95] Verdier, Thierry; Zenou, Yves (2004). "Racial beliefs, location, and the causes of crime." *International Economic Review* 45: 731-760.

[96] Waldfogel, Joel (1994). "The effect of criminal conviction on income and the trust 'reposed in the workmen", *The Journal of Human Resources*, 29: 62-81.

[97] Wallis, John; Fishback, Price; Kantor, Shawn (2006). "Politics, relief, and reform: Roosevelt's efforts to control corruption and political manipulation during the New Deal." in Corruption and Reform Edward Glaeser and Claudia Goldin eds. NBER Volume. Chicago: University of Chicago Press: 343-372.

[98] Wallis, John Joseph; Benjamin, Daniel K (1981). "Public relief and private employment in the Great Depression", *Journal of Economic History*, 41: 97-102.

[99] Wallis, John (1987). "Employment, politics, and economic recovery during the Great Depression", *Review of Economics and Statistics*, 69: 516-20.

[100] _____(1998). "The political economy of New Deal spending revisited, again: with and without Nevada", *Explorations in Economic History*, 35: 140-70.

[101] ____(2001). "The political economy of New Deal spending, yet again: a reply", *Explorations in Economic History*, 38: 305-14.

[102] Webb, John N (1936). "The transient unemployed", *Monthly Report of the Federal Emergency Relief Administration* January 1 through January 31, 1936 (Washington, DC: GPO, 1936), 1-25.

[103] Whatley, Warren C (1983). "Labor for the picking: the New Deal in the South", *The Journal of Economic History*, 45: 905-929.

[104] Williams, Edward Ainsworth (1968). Federal Aid for Relief. AMS Press, New York.

[105] Witte, Ann Dryden (1980). "Estimating the economic model of crime with individual data." *The Quarterly Journal of Economics*, 94: 57-84.

[106] Witte, Ann D.; Reid, Pamela A. (1980). "An exploration of the determinants of labor market performance for prison releasees", *Journal of Urban Economics*, 8: 313-329.

[107] Wooldridge, Jeffery M (2002). Econometric analysis of cross section and panel data. MIT Press,Cambridge, Massachussetts.

[108] Wright, Gavin (1974). "The political economy of New Deal spending: an econometric analysis", *Review of Economics and Statistics*, 56: 30-38.

[109] Yun, Myeong-Su (1999)."Generalized selection bias and the decomposition of wage differentials", IZA Discussion Paper #69.