Two Essays in Applied Microeconomics

Author: Francis Stanley Georges

Boston College Electronic Thesis or Dissertation, 2015

Boston College

The Graduate School of Arts and Sciences

Department of Economics

"TWO ESSAYS IN APPLIED MICROECONOMICS"

a dissertation

by

FRANCIS GEORGES

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

May 2015

Two Essays in Applied Microeconomics

Francis Georges

Advisor: Peter Ireland

Abstract

This dissertation consists of two chapters. The first chapter: Does going to prison increase the chance that one eventually applies for U.S. disability insurance (DI)? Since the 1980's, there have been substantial increases in both the number of people who have been incarcerated and the number of people applying for DI. Both increases have caused higher costs to taxpayers. While several studies have explored the causes of the increased DI applications and several others have looked at the labor outcomes of ex-inmates, no study has yet asked whether prison itself has any effect on the DI application process. Prison, with its harsh conditions, could cause physical and mental disabilities that increase the chance of a DI application. Properly measuring this, however, requires considering any endogeneity that predisposes ex-inmates to a DI application prior to incarceration. To do this, I use the instruments of states' minimum wages and legal high school drop-out ages to explore the effect of increasing incarceration numbers on state-level DI applications. I find that prison does have a significant effect on DI applications; a 1.0% increase in incarceration causes approximately a 0.5% increase in DI applications six years after the initial increase in incarceration numbers. I find that prison's effect is especially strong for a means-tested group who also concurrently applies to Supplemental Security Income (SSI); here a 1.0% increase in prison leads to a 0.9%

increase in people who apply for both DI and SSI after a six year lag. This suggests lower income groups are more sensitive to incarceration. Also, the cost of imprisonment should take into account the cost of subsequent DI applications and awards.

The second chapter: This paper assesses the specific case of when a monopolist manufacturer producing two types of goods is allowed to bundle the goods when selling to retailers who are allowed to resell the goods individually, have territorial market power and have heterogeneity in the resale demand functions. While the literature covers bundling in a variety of forms, no paper has considered the effect that the presence of multiple retailers may have on an upstream manufacturer who bundles and how benefits to bundling may accrue to consumers, retailers, and manufacturer in the presence of retailer heterogeneity. It is shown that under plausible circumstances, the ability of a retailer to retain profit in the face of bundling may prevent consumers in other markets from realizing greater welfare-enhancing effects although bundling in these cases at least weakly improves consumer welfare and never diminishes it. It is also shown by example, that in the case of three retailers, some retailers may actually profit more when the upstream manufacturer bundles while other retailers may profit less. This suggests that in certain cases some retailers may even favor upstream bundling as their interests align with that of the manufacturer.

Acknowledgements

I would like to thank the members of my committee for their unwavering support during the time it took me to write this. My chair, Peter Ireland, deserves huge thanks for steering me towards the topic of disability insurance. He also encouraged me and helped me to find data related to my question on this topic when it first seemed to me that not much data was available that would lend any new insights to this issue. Although he may not realize it until he reads this, Peter also had a hand in the bundling theory paper as it was he who taught me my first lesson in constrained optimization during the math for economists course from my first semester in graduate school.

Hideo Konishi deserves credit for helping me to hone my micro theory skills and also for providing comments that clarified the exposition in the bundling paper. Frank Gollop, now retired, also deserves credit for inspiring me to wonder about when restrictions on the market, or lack thereof, can affect the interests of manufacturers, retailers and consumers. It was along these lines that I decided to try to look at removing the restriction of bundling in a certain limited framework and to see what happened.

Don Cox provided invaluable assistance with the econometric issues. I learned much about the importance of making logical assumptions when trying to use instrumental variables from him. He also guided me towards literature and tests that I feel eventually bolstered the case for the methodology I used in the paper on disability insurance. He also continues to encourage me as I become less prolix in my presentations.

Dick Tresch was extremely helpful in various ways on both papers in this dissertation. His editing suggestions were a huge help as I worked to improve the exposition of the papers. I believe the incorporation of his suggestions will eventually make this material available to a wider audience. Dick also served as a good sounding board when I was pondering issues both theoretical and empirical and could often point me in the right direction of where to get answers or whom to get answers from.

Also, many people made other helpful comments on these papers over the years. I thank Arthur Lewbel for early comments and encouragement on my bundling paper. Sara Parker, of Rutgers University, explained to me the usefulness of the Sargan test at a chance meeting at a party one night and I eventually used this in the disability insurance paper. Scott Fulford also later provided helpful comments in this regard.

I am also indebted to the Center for Retirement Research at Boston College for providing me with much of my data in my empirical paper. Without that, I don't know how that paper gets written.

Finally, my biggest thank you is to my mom and dad who were unbelievably patient and supportive as I found my way through graduate school and wrote this. My gift to them is finishing this piece of writing, which would not have happened without them.

Contents

Chapter 1

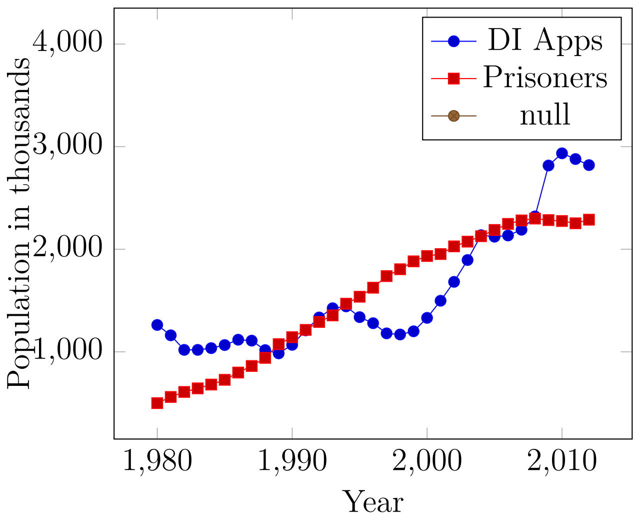
The Effect of Incarceration on Disability Insurance Applications

1.1 Introduction

The mid-1980s saw the beginnings of significant U.S. policy changes in the seemingly distinct realms of disability insurance and criminal justice. These policy changes led to a substantial increase in both the number of people incarcerated in prisons and the number of people applying for and receiving Disability Insurance (DI) awards. Numbers for DI increased from about 1,066,000 applications and 416,000 awards in 1985 to 2,935,000 applications and 1,049,000 awards in 2012. These numbers greatly outstripped any increases in the working-age population over this time frame; awards per 1000 insured workers increased from 3.7 in 1988 to 7.0 in 2010. Overall, there are about 8.8 million people collecting disability insurance today compared with about 2.5 million in 1985. (Social Security Administration data.) Meanwhile, the overall number of prisoners in the U.S. has changed from approximately 600,000 in 1985 to just under 2.5 million in 2010. The magnitude of the change is comparable to the rise in DI applications over the same time-period. From 1984 to 1999 alone, the ratio of jailed persons to the population changed from 96 per 100,000 to 468 per 100,000 (Pastore and Maguire, 2000).

Theories abound for the occurence of such changes. With respect to DI, the federal government reduced the stringency with which applications for awards were screened. The loosening of the criteria of who was eligible for DI benefits led to increases in both the number of DI applications and awards in the subsequent decades. Autor and Duggan (2003) also cite the increased attractiveness of benefits to low-skill workers, who found that the replacement rates offered by DI awards rose relative to the falling real wages these workers faced. The changes in the prison numbers have been explained by evolving attitudes towards criminal punishment, with tough sentences for drug-related offenses being a large factor.

Graph 1: Disability Applications and Prisoners by Year



The literature is still considering all the causes leading to increased DI awards and its resultant effects on the social welfare system and labor force participation. Another literature exists for exploring the effect of prison on the workforce and labor outcomes for inmates. Little to no work exists, however, on the relationship between the two. Open questions include not only whether prisoners are more likely to apply for DI, but also whether *prison itself*, with its hard conditions, can make someone suffering an incarceration spell *more* likely to become disabled and apply for a DI award. In this paper I seek to tie together these questions. First, I seek to determine some of the important factors driving DI applications. Second, I also ask if the increase in prison spells are themselves responsible for some of the rise in DI applications.

These are important policy questions. As the number of DI awards has grown by millions, the size of the welfare state has also grown and become increasingly costly to taxpayers. If the increase in prison numbers is responsible for a sizable fraction of the DI award increase then one can say that the cost of incarceration needs to be adjusted upward as well. A study of the factors that have lead to greater DI applications (and subsequent DI awards) will help shape future policy discussions about both the prison system and disability. In addition, these increases accompany changing trends in the labor force participation rate and are a small part of a larger empirical puzzle in the literature: the labor force participation rate, on the rise for decades because of women's increasing entry into the labor force, has now come back down in recent years to approximately where it was in the late 1970's. Labor force participation was 62.8 % in 1978 and had risen to 66.2, 67.2, and 66.2 % in 1988, 1998, and 2008 respectively; in 2014 it was back to its 1978 level. Since eligibility for DI requires removal from the labor force, one may wonder to what extent the two phenomenon are linked. If prison has some disabling effect on people to the point that they remove themselves from the workforce, then prison policy is relevant for issues involving labor force participation.

However, the effect of prison spells on disability applications is difficult to determine as some prisoners likely have conditions that make them also likely to apply for DI before any incarceration spell ever begins. For instance, prisoners could have mental disorders that both make it difficult to obtain paid employment and increase the tendency to commit crime. To disentangle the effect of prison itself on disability from the pre-dispositions towards DI that prisoners already have requires careful analysis. In this paper I will use the instruments of minimum wage and minimum school drop-out ages to address any endogenous factors causing disability that prisoners already have prior to going to jail. This is explained later in the paper but to get a sense of how DI may be interconnected with the prison experience, it is first useful to consider some demographic data showing how the DI and prison populations overlap.

The groups most likely to apply for disability share some demographic characterisitics. Autor and Duggan find that 47 percent of all DI applications come from high school dropouts. Racial minority status, poverty, poor health, and lower education also increase the probability of applying to DI, even when controlling for other factors (Coe, Haverstick, Webb, and Munnell, 2011.) Interestingly, some of these same racial, educational, and income characteristics also increase the probability of suffering spells of incarceration. From survey and panel data, Raphael (2011) finds that 19 percent of blacks who were high-school dropouts were imprisoned versus only 5.0%, 4.1% and 2.3% for whites, Hispanics, and Asians, respectively. For college graduates, the numbers drop: 1.1 % of blacks were incarcerated versus 0.2%, 0.1% and 0.4% for whites, Hispanics, and Asians, respectively. Raphael also finds that the proportion of people who have spent time in prison at any point has also increased in past decades. In 1974, 1.4% of whites, 8.7%

of blacks, and 2.3% of Hispanics had prior spells in jail. By 2001, those numbers had risen to 2.6%, 16.6% and 7.7%, respectively. Over ten million people in the U.S. have had prior incarceration at some point in their lives. Raphael points out that young men in certain demographics have been especially hard hit. More than 20 percent of black men in both the 25-34 and 35-44 age groups have served time, with the 35-44 group having the highest proportion of ex-inmates for every ethnicity. (Raphael, 2011) Beck suggests the problem is acute for black and Hispanic males, with 11% and 4% of each group in prison at a single moment in time. (Beck, 2000) Western and Pettit report that 33 % of black high school dropouts were in jail in 2000. (Western and Pettit, 2000)

Lack of skills is a problem for many inmates upon their release. Even with the advent of prison training programs, Raphael reports that 54% of released inmates have less than a high school education or G.E.D. (Raphael, 2011) If one supposes that lack of skills and job experience compound the stigma of a criminal record in attempts to procure employment, it is not unreasonable to think that many ex-prisoners are future candidates to disappear from the labor force. (Holzer, Raphael and Stoll 2004) Since labor force non-participation is a prerequisite for DI application, if many ex-prisoners can't find jobs then they could have little to lose by also applying for disability claims, whether such applications are successful or not. The possibilty exists that prison leads to further DI applications through this channel even if prison itself is not disabling.

Autor and Duggan (2006) mention that there are three primary reasons for the sharp increase of DI awards in recent decades: reduced stringency of eligibility criteria, higher replacement rates for some insured workers, and female participation in the labor force. Autor and Duggan point out that eligibility criteria were liberalized in the 1980s, perhaps as a result of politics or more sympathetic policy managers at the state level. However, they indicate that not only did the number of people deemed disabled increase within each type of disabling category, but also that the relative proportions of awards across the categories began to significantly vary. In 1983, 16.8% and 21.9% of all DI awards were given for neoplasms and circulatory disorders, respectively; only 13.4% of awards were for musculo-skeletal disorders and 16.3% were for mental disorders. By 1999, the same breakdown looked like this: 10.6% for neoplasms, 12.1% for circulatory disorders, 23.7% for musculo-skeletal disorders, and 22.5% for mental disorders. The new emphasis switched the prior DI awards for more-lethal conditions to less-lethal ones: neoplasms and circulatory disorders had mortality rates exceeding 16% while mortality rates for musculo-skeletal conditions and mental disorders were on the order of 5% per year over the same period. The trend of giving DI awards to less lethal conditions has continued to the present day and does much to explain the increasing rolls. Not only are more people applying and receiving awards, but they are also staying on the rolls longer because the disabling condition is not fatal.

Of relevance for the formerly incarcerated population is the increase in a mental disorder diagnosis as a reason to receive DI. She and Stapleton (2006) report that "based on data from three Department of Justice (DOJ) surveys, about 37 percent of jail inmates, 31 percent of state prison inmates, and 23 percent of Federal prison inmates report a disability of some sort." They claim that this is about evenly split between three types of self-reported disabilities: physical, mental, and learning. Even if approximately 10 percent of jail inmates have mental disorders, and these disorders persist when inmates become ex-inmates, then the number of ex-inmates with mental disorders is in the millions. This is a potentially sizable fraction of all DI awardees. Considering the preponderance of ex-inmates in their 30s and 40s, the mental disorder disability issue becomes more acute when looking at who, under 50 years old, is collecting DI. About 33% of male DI recipients are "young" - or under 50, about 1.5 million people. For women, about 28% of DI recipients are under age 50 and this amounts to just over 1 million people. 50% of all DI recipients under age 50 receive their DI award as a result of a mental disorder. This amounts to 1.25 million people. (Social Security Administration.) If being in jail has an exacerbating effect on causing mental disorders or interferes with proper psychological treatment, one can imagine then it could affect the DI statistics.

Other factors may also be important for increased DI rolls. The "replacement rate" used to determine DI awards is based on the past income of insurees. As low-skilled workers have seen their real median incomes drop over the past few decades relative to higher skilled workers, the incentive to apply to DI has increased. Also, more females are eligible for DI simply because of their increased participation in the workforce. Autor and Duggan (2003) cite higher mortality rates, lower median wages of dropouts, and lower proportions of immigrants in the population (who potentially tend to apply for DI awards less than natives) as factors that can increase DI applications. They find the length of unemployment insurance benefits to be insignificant as is the number of people incarcerated per state, although they do not explicitly consider a time lag between the time when jail ends and DI applications begin, treating those two phenonmenoa as simultaneous.

Interestingly, in their later 2006 paper, Autor and Duggan find that, when looking at the lowest 10th percentile income group, replacement rates increased more for 30-39 year-old men (60.6% to 85.7 %) than for 50-61 year-old men (67.8% to 86.0%) over the 1984-2002 time period. The percentage change in DI awards has also been larger for 25-39 and 40-54 year old groups than for the 55-64 year old group. Even though the older group still leads the DI awards in levels, the younger groups have had the largest proportional change. The factors causing these younger groups to apply to DI more frequently are still largely undetermined.

Coe, Haverstick, Munnell and Webb (2011) examine whether there are differences between those who apply only for DI versus those who apply concurrently for DI and Supplemental Security Income (SSI). (SSI is a means-tested supplemental income program which does not require a concurrent application to DI.) They look at state-level application data and find that factors such as race, labor force participation rate, length of unemployment benefits, health insurance regulation, and the prevalence of smoking matter for the DI-only group. For applying to the means-tested SSI and non-means tested DI concurrently, health measures, education-level, gender, mandated disability insurance through state law, and the presence of a Republican governor matter. This study considers lots of variables, however, and it is sometimes difficult to figure out which factors may be driving disability claims, especially for the young population; the idea that labor force participation and unemployment numbers, for instance, are related to disability application may skew the results, as they are both conditional for application in the first place. Including these variables in a regression for DI applications is likely a case of controlling for too many factors if we are to assess the other causes of applying for DI.

Putting aside other causes for changes in DI applications, the question of incarceration's effect on DI remains open. Most of the literature on labor patterns of ex-inmates focuses on whether there is an employment participation effect or earnings effect as a result of jail. There is virtually nothing on DI. The literature that does exist reveals mixed results on the employment and earning patterns. Some studies actually suggest a positive effect of incarceration on both employment and earnings, but the bulk of the literature suggests a negative one. Kling (2006) uses panel data on prisoners to assess the effect of longer prison sentences on earnings. He uses randomly assigned judges, with different sentencing behavior, as an instrument and finds that longer incarceration spells are initially accompanied by increases in employment and earnings for one to two years after release. However, he says these findings are also "largely explained by differences in offender characteristics and incarceration conditions, such as participation in work-release programs." He finds that any effect of longer prison sentences on employment or earnings vanishes after seven to nine years. He also constrains his study to compare only prisoners with shorter sentences with prisoners with longer sentences; there is no non-prisoner control group. Cho and LaLonde (2008) find that women who go to prison experience declining employment rates prior to going to prison and experience a bump of up to 5% in employment and earnings immediately after leaving prison, although again this generally dissipates to pre-prison levels after a few years. They claim that prison has no long-term effect on subsequent earnings or employment.

Most studies indicate that jail has a subsequent negative employment effect, beyond a positive or merely neutral one. Grogger (1996) uses data from the adult criminal justice system in California and merges it with unemployment insurance data from the early 1980's. He looks up to six quarters after incarceration for the employment and earnings effects. He finds negative effects on employment and earnings, which are dampened somewhat by state fixed effects. Grogger, however, has limitations with his data as some of the men might have returned to jail or still be in jail in the "after" period. Geller, Garfinkel and Western (2006) do a careful study in which they look at a panel of families, some of whom include ex-inmates and some of whom do not. They use propensity score matching to analyze the treatment of jail on similar groups. They find that employment rates of formerly incarcerated men are about 6 percentage points lower than for similar men who have not been incarcerated. They claim imprisonment is also associated with a 14 to 26 percent decline in hourly wages. Western (2000) shows that the age-earning profile of inmates is flat compared with men who have never experienced jail. This gap grows with age. (The inference is that this contributes to differences in replacement rates for any future DI applications.) Freeman (1992) similarly finds that there is a 15-25% lower employment rate for those jailed when looking at the 1979 National Longitudinal Survey of Youth. Waldfogel (1994) finds that first-time conviction reduces employment probabities by 5 points and has a significant negative effect on income - as much as 30 percent. He uses panel data on federal offenders and estimates pre- and post- conviction income on age and other covariates. However, Waldfogel admits unobserved heterogeneity can be a problem and that there is a need for better data. Lott (1992a) shows that those convicted of embezzlement and large-scale fraud do much worse after prison while, in a different paper, Lott (1992b) shows drug offenders experience little change in their income after a prison spell. He rationalizes this by saying those who did the best financially prior to prison have the most to lose and may suffer the worst effects. Needels (1996), in a separate study, finds that there is no effect on employment from going to jail but did find a negative earnings effect of 12 percent. However, Needels primarily looks at people who were inmates in the 1970s and 1980s and their subsequent labor force participation in the 1980's and 1990s. One wonders if the same results would hold up today given the changes in the justice system in the last 30 years.

Given the different findings of the effect of incarceration on earnings and employment, one would expect that there would be even more theories to explain the direction of the effect. Indeed there are and some of these theories also suggest reasons for possible bias in jail's effect on employment. To begin, the theories that proclaim a positive effect on jail on earnings or employment often point to things like training programs or education that inmates receive in prison. Tyler and Kling (2007) find that attendance in a G.E.D. program leads to about a 15 percent increase in earnings for inmates upon immediate release but that the effect dissipates completely after several quarters. Using data on federal offenders, Nagin and Waldfogel (1998) claim that ex-prisoners have less access to career jobs offering stable, long-term employment. As a result, a criminal record often "relegates offenders to spot-market jobs, which may have higher pay at the outset of the career but do not offer...rising wages." This suggests that any gains in the labor market from prior prison experience are short-lived. On the other hand, much has been written about reasons for the potentially detrimental effects of a prison spell on employment. Waldfogel (1994) mentions the possibility of eroding skills from incapcitation. Irwin and Austin (1997) mention the possibility of encouraging behaviors that are inconsistent with positive workplace routines. Granovetter (1995) discusses how a jail spell can cause social capital to be diminshed; with fewer connections to the working world it can be hard to find a new job as many jobs are found through social contacts. Haney (2002) mentions other factors such as hypervigilance, interpersonal distrust, suspicion, emotional over-control, alienation, psychological distancing, dependence on institutional structure, diminshed sense of self-worth, physical trauma, and retraumatization of childhood/past experiences can all have an effect on post-prison employment prospects. Liat Ben-Moshe (2013) also notes that prison has "hard labor, toxic conditions, circulation of drugs, unsanitary needles, poor air quality and lack of appropriate medical care." This latter description moves perhaps closer to the realm of creating a nascent environment for directly disabling conditions, although clearly there is a continuum of reasons why incarceration might not be helpful for worker productivity.

Some of the effects of prison on the labor force, and the resultant effect on DI applications, may be even more nuanced. Western and McLanahan (2000) found a low likelihood of marriage or cohabitation for ex-prisoners. Laub, Nagin and Sampson (1998) find a higher rate of divorce for prior inmates. If spouses pool resources, then it is possible a lower family income could affect the incentive to apply for disability insurance and other entitlements. Garfinkel (2001) posits that incarcerated men with children are often hopelessly in arrears with child support payments by the time they exit prison. This may diminish the incentive to work as prisoners are able to keep less of their earnings and thereby induce them to apply for DI. Using employer survey data, Holzer, Raphael and Stoll (2007) make the point that the stigma of having a criminal record or "rap sheet" alone might be enough to dissuade potential employers from hiring. Pager (2003) found similar results when sending out fake c.v.'s; those with c.v.'s that listed prison experience rarely got callbacks for interviews. Becket and Western (1999) notice that many inmates tend to come from the same neighborhoods in many states. For instance, Baltimore is the origin of 50% of all state prisoners in Maryland and 7 New York City neighborhoods account for 75% of state prisoners in New York. They claim that this can often lead to stereotyping and that the stigma of coming from a neighborhood where prisoners agglomerate could be used in place of a revealed criminal record when making employment decisions. Such stereotyping would presumably bias the direct effect of prison on employment downward, but still result in a greater indirect effect on employment for the larger population. This could have increased ramifications for labor force non-participation and therefore for disability insurance applications. Another source of potential bias when measuring jail's impact on DI is pointed out by Haney (2002.) He mentions that prisons have become more overcrowded over time and conditions have potentially become worse. This raises the possibility that the effect of prison on DI applications has become greater in later years, if prison life itself generates disabilities. Of note also is a paper by Conly (2005) which describes state agencies in New York, Pennsylvania and Texas whose explicit purpose is to help ex-inmates apply for and obtain DI benefits. The existence of such agencies suggests that a substantial portion of the ex-inmate population may be disabled.

Haney (2002) and She and Stapleton (2006) indicate that prisons may be filling the role of now nearly-extinct previous institutions like sanitariums in housing large numbers of the mentally ill. This conjecture is consistent with the separate trends of a large and growing number of people collecting disability insurance and a large and growing number of people who have been incarcerated. Given that many incarcerated individuals report potentially disabling conditions and that prison itself may contribute to diminishing the ability to work, the time has come for a more formal investigation of whether the two things are related.

This paper hopes to expand the understanding of factors driving changes in applications for DI while also exploring the issues regarding the costs and benefits of criminal justice policy in the spirit of Becker's (1968) seminal paper on crime and punishment. The remainder of the paper proceeds as follows: section II describes the regression approach used in assessing the effect of incarceration effect on disability insurance applications, section III describes the data used by the model, section IV describes the results, and section V concludes with some policy implications.

1.2 Model

The principal equation to be estimated is:

$$DIApps_{it} = \beta_0 + \beta_1 Prisoners_{it-q} + \beta_i X_{it} + f_i + u_{it}$$

Here, $DIApps_{it}$ is the number of federal disability insurance applications as a proportion of the eligible working age population in state i in year t. $Prisoners_{it-q}$ is the number of prisoners per capita in state prisons in state i in year $t-q$. X_{it} is a vector of covariates at the state level that also explain changes in disability insurance applications. f_i is a fixed effect at the state level which can explain persistent differences in the screening and approval of disability insurance applications that exist across states and thus explain some of the variation in the number of applications. u_{it} represents unobserved variables that affect disability applications.

Estimation of β_1, the effect of the lagged number of prisoners per capita on the number of disability applications, is the focus of this paper. One would expect that OLS estimation of the above equation to lead to biased estimates if some of the unobserved conditions, such as mental disorders, that lead a person to a prison spell also are likely to lead to eventual application for disability insurance. If prison has simply become the latest place to house more people with prior mental disorders then OLS estimates of β_1 would be expected to show a very small effect on disability claims.

To determine whether attendance in prison itself leads to additional (mental or physical) disabilities that increase the chance of a disability application, the underlying disablities that lead a person to a prison sentence must be disentangled from the additional effect that incarceration itself has on disablity. To do this, I adopt an instrumental variable approach in which the instruments show little correlation with such underlying disabilities but still are correlated with additional prison sentences. The following section explains my choice of instruments.

1.2.1 Instruments

To estimate the effect of prisoners per capita on disability insurance, I use two instruments that I assume satisfy the criteria of being uncorrelated with disabilities in the population but correlated with incarceration numbers in each state. These two instruments are minimum wage changes and minimum school drop out ages. In addition to satisfying these two conditions for use as appropriate instruments, I will also argue that these variables are also best used as instruments and not as independent variables in the main regression.

To begin the argument that the minimum wage variable[1] has a relationship with incarceration, preliminary data suggests that there is a 0.2265 correlation with minimum wage and prisoners per capita. Also, research by Beauchamp and Chan (2012) shows that minimum wage increases have a significant effect on crime. While the labor economics literature on the effects of minimum wage on labor force participation and employment is inconclusive and a source of much academic debate, Beauchamp and Chan suggest that a unemployment effect may be pronounced for a sector of unskilled workers who have a proclivity to commit crime.[2] They find that a 1 percent increase in the minimum wage is associated with a 2.1% to 2.4% increase in violent crime[3], a 1.8% to 2.3% increase in property crime and a 1.4% to 2.8% increase in juvenile drug crimes. As minimum wage increases occur in jumps of more than one percent, they could generate

[1] State minimum wages are considered in this model in place of federal minimum wages, even if the state levels fall below federal levels. Sub-federal minimum wages are applicable to sizable groups such as tipped employees, young workers under 20 years of age with less than 90 days job tenure, student-learners, workers in small businesses with little interstate commerce, and (perhaps relevantly) workers with disabilities.

[2] Ahn, Arcidiacono and Wessels (2011) posit that minimum wage increases affect unskilled workers employment prospects more than skilled workers. The thought is that unemployed, unskilled workers could remain idle and commit more crimes.

[3] Violent crime increases were found concentrated among those crimes with a clear monetary reward.

large increases in crime. These results are consistent with an older literature (e.g. Hashimoto (1987)) that finds a positive effect of minimum wage on certain categories of teenage crime.

In a different paper related to predicting juvenile crime, Chan (2012) also posits that an “incapacitation effect" is responsible for the decrease in crime due to exposing late adolescents to stricter high school drop-out laws.[4] She uses variation in state minimum high-school drop-out ages to determine this. She finds that increases in drop-out ages do have some effect on decreasing crime for those close to high school graduation age although this effect dissipates through the mid-to-late 20’s. My own data indicate a small negative correlation of 0.0257 between minimum high-school drop-out age and incarceration.

These arguments pertain to the strength of the instruments and are bolstered by Cragg-Donald statistics which are included in the results. Typically, a Cragg-Donald F-statistic of 10 indicates sufficient strength of the instruments in the first stage regressions. Values of 11 to 20 are obtained for most specifications in this paper.

Besides instrument strength, instrument validity, which is lack of correlation with the error term in the main regression, must be assessed. One argument here is that minimum wage increases on their own should have no bearing on the the amount of disability in the population. I assume that the minimum wage paid to workers is completely independent of the level of disability such workers face. Its argument for inclusion as an instrument is that wage increases can force an

[4]The “incapacitation effect" theory is that students and young adults do not commit crimes close to the period when they are still in school but may still offend several years later upon emancipation from stay-in-school laws. Forcing the students to physically remain in school “incapacitates" them and removes them from the opportuntity to commit crimes outside of school.

employer to lay workers off in the face of rising costs. This, in turn can lead to unemployment, idle time and increased risk of prison. Any resultant prison sentence is more likely to have an effect on disability claims several years down the road than the original minimum wage increase would.

One possible objection to this logic may be that any policy resulting in disemployment (such as potential minimum wage increases) will increase the hazard of at least applying for disability insurance, because unemployment is a precondition for such application. This would indicate that minimum wage belongs in the main regression and not as an instrument. To address this concern, I will argue that while, in fairness, minimum wage changes may certainly have an immediate effect on unemployment (and DI applications), this effect should not last for more than a year or two and should not survive periods on the order of half a business cycle (or approximately five years.) By the time four to six years have passed, any direct effect minimum wage changes had on individual employment probablilties should dissipate. Its effect on disability insurance applications, however, will still show up several years later through the channel of prison. This is especially true if prison spells last longer than the direct effect of minimum wage on employment and also have themselves a potentially stronger effect on disability.

An argument can also be made for the validity of the mandatory school drop-out age instrument. I will assume that increased mandatory school drop-out ages are not done as a response to changes in the underlying distributions of disabilities in the population but are due more to policies designed to meet the changing educational needs of a workforce. The additional argument for proper identification of mandatory school drop-out ages as an instrument for prison and not as a direct contributor itself to DI applications relies again on arguing that this policy has little effect on the factors that lead to employment (and resultant DI-ineligiblity) several years down the road. Ostensibly, the primary factor here that has the most direct effect on both employment and DI application tendency is population educational level. While it is fair to wonder whether mandatory school drop-out ages have any effect on education level, two ideas suggest that it does not directly matter this way in my regression. First, my data indicate a nearly zero but still slightly *negative* correlation of -0.0011 between the mandatory minimum drop-out age and the percent of the population with a high school degree; thus, forcing students to stay in school longer has no obvious effect on educational outcomes in this data set. Second, I will directly control for the percent of the population with a high school degree in the other covariates in my regression. This last part should help isolate any "incapacitation" effect that keeping students in school for longer periods has on prison numbers apart from the actual educational outcome of such a policy.

Sargan tests are performed to test the validity of the instruments. The Sargan test works only with overidentification and assumes that at least some of the instruments are valid before testing whether the additional instruments have any correlation with the error terms in the main regression. Accordingly, the above arguments for identification are necessary. Still, the Sargan test can be used as a robustness check on instrument validity given the conditional assumption that at least some of the variables have proper identification as instruments. Starting with a null hypothesis of valid instruments, a chi-square statistic is reported. p-values for these Sargan numbers much larger than 0.1, and often very close to 1, were routinely found in a variety of specifications and are reported in this paper. This gives further support to the suitability of the two variables as suitable instruments. The rest of this section discusses the suitability of the other variables in the model.

1.2.2 Dependent Variables

The main dependent variable in the paper involves applications for disability insurance at the state level. Through use of a special data set, I consider different kinds of categories here. It is possible to consider the number of people who apply only for DI separately from the number of people who concurrently apply for both DI and the means-tested Supplemental Security Income (SSI) program.[5] Using this convenient variation in data, it is possible to determine the effects of certain variables on overall DI applications as well as the effect of the variables on the two subgroups. This is useful in assessing whether incarceration or other variables have differing effects on the two populations, as those without a low income cannot apply for the means-tested SSI and may potentially have other different underlying characteristics as well.

1.2.3 Use of Fixed Effects

There is considerable variation in the number of prisoners per capita in each state, from 0.069% to 0.873%. Clearly, states may have different attitudes about

[5]I am grateful to Paul Davies of the Social Security Administration for providing the Title 2 (DI) only, Title 16 (SSI) only, and concurrent Title 2 and Title 16 receipts by state for FY1993-FY2010. The FY1993-FY2000 receipts data came from paper records from SSA's State Agency Operations Reports (SAOR) system. The FY2001-FY2010 receipts data are from SSA's Payment Management System (PMS). I am also grateful for Boston College's Center for Retirement Research for providing me with a table form of these data at the state level.

crime and punishment. States may also have different attitudes about awarding disability insurance claims. If successful awards also lead to a preponderance of additional applications, some of the large variation in DI application numbers across states may be a result of different within-state cultures, some of which may have a more permissive attitude towards entitlements than others. The actual awarding of DI benefits is largely determined within-state so much of the variation in application numbers could be endogenous. To account for built-in cultural differences across states in both justice system outcomes and disability insurance administration, I propose a fixed effects approach and use binary variables to control for each state's individual history and attitudes towards the factors leading to DI applications. I then regress the DI applications against the fixed effects, incarceration measures and other relevant variables. The goal is to see whether changes in prison policy or other variables lead to within-state changes in disability applications.

1.2.4 Other Covariates

Since the literature suggests that other variables are also a factor in disability applications, I run my regressions on different specifications. To control for health variables, I include mortality rates and the proportion of people in each state with a BMI at an overweight level or higher. I control for age by including the proportion of the working population (aged 18-64) over age 50. I consider educational status through the proportion of workers who have obtained at least a high school education. To assess the impact of race and ethnicity, I include the percentage of the population classified as white and I also include a measure of the number of foreign-born people in each state to test Autor and Duggan's claim that immigrants apply for DI less than natives do. I include the percent married in each state as a variable to partly control for family dynamics and possible self-insurance through marriage. I also include an indicator variable if there is a Republican governor in the state to assess changing cultural norms and bureaucratic changes that affect the chances of a DI application, which may itself be affected by the chance of an application being successful. I chose not to include variables with little explanatory power or those that cause collinearity issues. Labor force participation, for instance, was left out as non-participation is highly correlated with DI application and some of the other variables. As it is a necessary condition for any DI award, I found labor force participation absorbed much of the explanatory power of the other variables when included. The goal of my regression is to find which variables cause an increase in DI applications overall and not simply find the effect of each variable on DI applications conditioned on non-participation in the labor force.

1.2.5 Use of Lags for Prisoners Per Capita

It is reasonable to wonder how soon after leaving prison that ex-inmates begin applying for disability benefits. It is expected that most inmates try to enter the workforce upon release, although it is unclear how long most inmates expect to remain in the work force if they are repeatedly unsuccessful at procuring employment. Clearly, the application for disability benefits may not be instantaneous upon release in most cases. To account for a possible lag between release from prison and application for disability benefits, I regress DI applications on state incarceration numbers on anywhere from two years prior to nine years prior. The identification arguments made earlier suggest that some of the instruments may have an effect on the main equation during some of the early lags. Once this effect has dissipated, the principal effect of the instruments on the DI channel should come through the prison channel. Raphael (2011) indicates average prison spell lengths for various categories of offense for both 1984 and 2002. Robbery, for instance, had an average sentence of 3.13 years in 1984 and 3.80 years in 2002. Other statistics for average prison sentence length in the two periods include averages for assault (2.01 and 2.86 years), burglary (1.99 and 2.48 years), drugs (1.63 and 2.11 years), motor vehicle violations (1.42 and 1.87 years), rape (2.98 years and 5.30 years), and murder (6.49 years and 8.13 years.) The more violent crimes have longer sentences but tend to be less frequent. The median sentence is currently on the order of a little over three years, although this also reflects a several month increase from earlier decades. Given these numbers and the time it may take to remove oneself from the workforce after a prison spell, I would expect prison to have the largest effect on DI applications approximately four to six years after the beginning of incarceration. The extent to which prison spells affect DI applications in even longer lags is unclear but I explore them in accordance with the reasonable limits that the data allow.

1.2.6 Other notes

In this paper, the instruments themselves are concurrent with the lag of the prison variable which they are used to predict. I explored the possibility that prison should be instrumented using lags of the instruments since some time may pass before any changes in minimum wage or school drop-out ages result in new crimes. However, I found that the concurrent instruments were the strongest and most valid according to the Cragg-Donald and Sargan numbers respectively; thus I use these instruments instead of lagged ones.

As an additional robustness check, logs of both the dependent and independent variables are used in the main regressions. This allows more direct calculation of elasticities. I consider the appropriateness of the log model using the instrumental variables approach in the results section.

Errors are assumed to be i.i.d. This allows calculation of the p-values for the Sargan tests and allows the Cragg-Donald statistics to be related to critical values calculated by Stock and Yogo (2006). Heteroskedastic robust errors did not change the estimated coefficients and variances appreciably.

1.3 Data

My model can be estimated through the use of a combination of data sets. Information on DI-only and DI-SSI concurrent applications at the state level came from a data set provided by the Social Security Administration and obtained through the Center for Retirement Research (CRR) at Boston College. The CRR also aggregated and provided many relevant state-level demographic variables. The main variables of interest are disability application rates by state (from 1994 to 2007), expressed as a percentage of the state's working-age population (ages 18-64) not receiving DI benefits.[6] Similar to Coe, Haverstick, Munnell and Webb

[6] The denominator is the number of residents aged 18-64 in a state as of July 1st from the U.S. Census Bureau. From this I subtract out the number of beneficiaries of each program,

(2011), I consider what factors explain the variation in application rates to DI and also the two components of this overall application rate - the DI-only application rate and the concurrent DI-SSI application rate.

Data on incarceration numbers come from the Bureau of Justice Statistics' website. I was able to obtain the numbers of incarcerated individuals in each state prison system from 1988-2012. This allowed me to regress disability insurance applications on prisoners from several years prior. I did not consider the number of federal prisoners in each state as these prisoners may come from different states. Approximately 10 percent of prisoners are federal prisoners. I did not have data on which states federal prisoners come from, but I assume that the states are somewhat uniform in the percentage of the population each contributes to the federal system, so I do not expect exclusion of the federal prison numbers to bias the results. Also, some states house prisoners from other states in their state prisons or in out-of-state for-profit prisons. Data on interstate prison transfers is scarce but one study by Kirby (2013) found that interstate prisoners currently number about 10,500. Since this is a small number compared to the total number of prisoners currently held in state prisons or jails (approximately 2,500,000), I also do not expect the lack of data on interstate prison transfers to bias the results. In addition, Kirby suggests that interstate prison transfer is a more recent phenonomenon of the past few years. Therefore, the lack of this interstate prison transfer data should not affect much of the 1988-2009 time period of my study.

The District of Columbia is excluded from the regression as D.C. began placing its prison population in other states (mainly Virginia and North Carolina) in

obtained from the Social Security Administration Statistical Bulletins (SSA 1994-2009) since current beneficiaries are not at risk of applying.

1998 and completed the process in 2001. Interestingly, the District of Columbia is also mostly urban. It has a much higher prisoner per capita measure prior to 1999 than all other states, and almost twice as high as the next closest state, Texas. Whether the results would be different by including this urban outlier is an open question, but the incomplete data precludes estimation.

To assess BMI levels in each state - and the percent of the population deemed as overweight - I use the Center for Disease Control's Behavioral Risk Factor Surveillance Survey (BRFSS). This a telephone-based survey that has been in place since 1984 and interviews about 350,000 people per year. To control for declining land-line usage that may accompany telephone surveys, I use the CRR's weights for land-line usage in each state. The proportion of residents in each state who are in each age group, who are white and not Hispanic, who are married, and who have at least a high school diploma comes from the Annual Social and Economic Supplement found in the Current Population Survey (March CPS) and the tables that the CRR aggregated using this survey. Political variables regarding the presence of a Republican governor come from the National Governors Association's *Governors Database* and the Council of State Governments' *The Book of the States* and also were aggregated in tables used by the CRR.

In addition, separate information on state-level mortality rates was compiled using the Center for Disease Control's website. Statistics on immigration were obtained through the U.S. Census Bureau and the Department of Homeland Security. To construct a measure of the number of foreign-born people in each state, I used the Census data in 1990 and 2000 to obtain the number of foreign-born people living in each state at the end of each decade. I then obtained the number of persons obtaining "legal permanent resident status" by state of residence from the Department of Homeland Security's website in each year from 1990 to 2009. Then, for each year in the sample I added the cumulative total of legal residents entering the state since the start of the relevant decade (either 1990s or 2000's) to the number of foreign-born immigrants listed in the census for that state at the start of the decade. This allowed me to construct a measure of foreign-born people in each state for *each year*. The cumulative numbers of legal residents entering each state throughout a decade closely matched the overall increase in census-reported foreign-born population across the decades in most states, so I presume this is a reasonable way of constructing the number of foreign-born people in each state for all years. Also, unless there is evidence of significant movement of immigrants obtaining residence in one state and then moving to another state, I do not expect much mismeasurement or bias here. In any case, such evidence is hard to come by.

The instruments used to predict incarcerated individuals per capita in each state were obtained as follows: state-level minimum wage data comes directly from the Bureau of Labor Statistics (BLS) website. Data on minimum legal school drop-out ages for the various years and states is from Oreopoulous (2009).

1.4 Results

1.4.1 Overall Results

The main dependent and independent variables (DI applications and prisoners) in the linear regressions are expressed as percentages of the working age population. A dummy variable is used to indicate a Republican governor. The other demographic variables which are percentages are percentages of the entire population and not just the working age population due to limitations on data. With the other variables, a one unit change represents a change from no one in the population having the indicated characteristic to everyone having said characteristic. Accordingly, the coefficients mainly represent the proportional change in disability insurance application if everyone undertook the listed characteristics as opposed to no one. It is not expected that these effects would be linear for the entire population but the coefficients serve to give us a sense of changes at the margin. A coefficient of "1" approximately means a one percent increase in the X variable is accompanied by a one percent change in disability applications. Besides this model where the percentages were inserted into the regression linearly, logs of the percentages are also considered to determine if changing the specification matters in the determination of the elasticities.

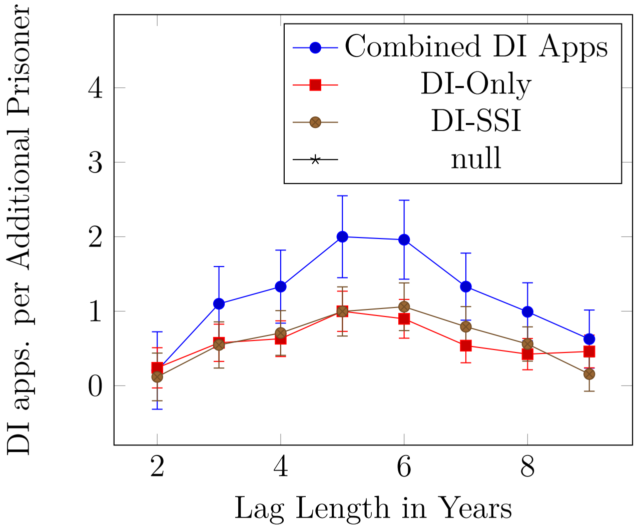
The main coefficient of interest is the one measuring the effect of lagged prisoners per capita on disability applications. When included without instrumental variables (but with the other listed covariates) against the combined (both SSI & DI as well as DI-only) categories of DI applications in a regression, the effect of the prison variable is found to range from small but negative and significant values (-0.2 for the earliest lag) to small but increasingly positive and significant values for the later lags (such as 0.3 for the nine year lag.) It seems that, even when uninstrumented, prison spells are correlated with some effect on disability insurance 7 to 9 years down the road.

When the instruments are used, the effect increases positively for all lags and is substantial in both magnitude and significance. The coefficient on prison is maximized at a lag of approximately five to six years and suggests that a given increase in the number of prisoners per capita accounts for twice as many disability applications six years later. The effect is still strong nine years out (0.6 applications per prisoner) but weaker and insignificant two years out (0.3 applications per prisoner.)

The instruments appear to be correcting for some downward bias. On the one hand, if prison is simply correlated with people who already exhibit disability and many of these people have been shifted to prisons over the past few decades, the rise in prison numbers may not be responsible for a drastic change in DI applications. On the other hand, prison is itself disabling, then the instruments, which are chosen to be independent of any pre-prison disabling features in the population, should show the disabling effect of prison. The fact that disability insurance applications increase with prison spells, while controlling for prior levels of population disability, suggests that prison itself elevates the hazard of eventually applying for DI several years after the initial incarceration.

If the effect is strong and significant, one can still wonder if the numbers listed here make sense. How can one additional prisoner lead to more than one additional disability application? It may be that prison affects not only the incarcerated individual but also his or her acquaintances. Prison splits up families and forces those who are not in prison to make ends meet on one less income and to raise children with one less spouse. It causes stress for other people who know the prisoner and it's conceivable that this could lead to further disabling conditions. There could also be neighborhood effects a la Becket and Western in which people who merely live in high crime districts have difficulty finding work because of the bad reputations of those areas; they may drop out of the labor force as a result. Such effects could be persistent and have labor effects down the road. That said, the coefficient of "2" on the five year lag does still seem somewhat inexplicably high but nevertheless points to a strong effect of prison on overall DI-applications.

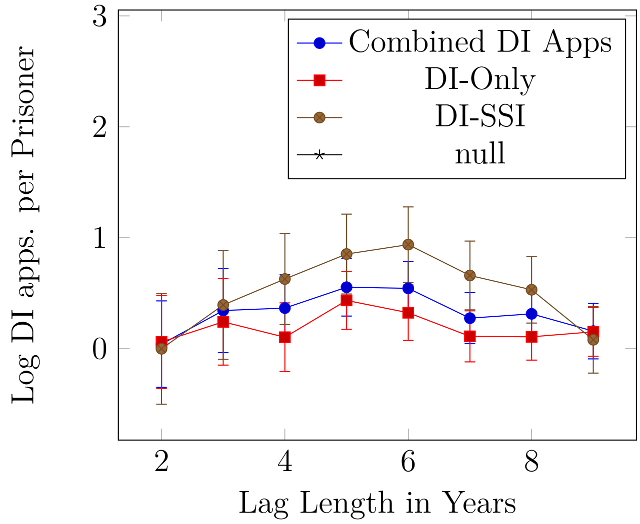
Graph 2: Effect on DI Apps by Prisoners per Capita sorted by lag time



It may be that a log specification does better at prediciting the results than the linear model. One problem with the log model is that the instruments are somewhat weak for the early lag specifications (suggesting a misspecified model perhaps due to lack of variation with the log variables or better correlation between instruments and prison in later periods) but recover enough to be useful for lags of five years or more. If one focuses on the five year and six year lag period, where the instruments are suitably strong and valid, the effect of prison on DI is positive and significant. An elasticity of approximately 0.5 is reported here, which is more realistic than the linear coefficients. Again, a downward bias seems to be corrected from the uninstrumented case, which actually had prison showing a small but negative and significant elasticity of -0.09 in the five year lag specification.

It is interesting to consider whether DI application is more sensitive to prior prison spells conditioned on whether people concurrently apply for SSI when applying for DI. The number of applications is approximately evenly split between the two categories. The summary statistics reveal that the mean application rate for DI-only was about 0.42% of the age 18-64 working population and was 0.39% for the concurrent category's application rate in the same age group. (The standard deviations, maximums, and minimums are also comparable across categories.) Accordingly, the coefficients on overall DI applications are close to the sum of the comparable coefficients in the regressions where either DI-only applications or combined DI with SSI applications are the dependent variables. Looking at these coefficients reveals an effect of 0.9 DI-only applications per prisoner and 1.0 DI-SSI applications per prisoner in the linear models at a lag of six years. Thus, prison has approximately the same effect on the group applying for means-tested SSI relief as the group which does not apply for such relief. This is also true in mostly all the lags. At the longest (8 and 9 year) lags for the SSI-DI group, the Sargan test starts to break down but this may be due to reductions in the amount of the data at the latest lags. (Nine year lags require mapping onto disability applications beginning no earlier than 1997.)

Graph 3: Effect on Log DI Apps by Log Pris. per Cap. sorted by lag time



While there does not seem to be much difference in prison's effect on the two groups in the linear model, the result is somewhat different for the log model. In the log model, prisoners per capita results in a positive and significant elasticity of 0.9 on the DI-SSI group when six years out. With the DI-only group, prison has an elasticity of 0.3 here and is not significant. Thus, it may be that people who do not fall into the post-prison means-tested category find that prison had less of a disabling effect. The rationale here is that the population that satisfies the means-test may be at higher risk for disability insurance with an additional prison spell; they may lack support services that a higher income group has to either cope with disability or be able to survive without public assistance. These numbers must be taken with caution however. The Sargan p-value numbers for DI-only log model are somewhat low (0.12) whereas they are quite strong in the DI-SSI log model. Although the Sargan value for the log DI-only case is not enough on its own to reject the validity of the instruments, it is possible that misspecification with the log model could be responsble for its difference with the DI-SSI case.

The response of the different classes of DI application to the various lags of prison for both the linear and log models is shown in graphs 2 and 3, respectively. The peak of the prison effect on DI applications is approximately four to six years out. For early lags, many inmates might still be incarcerated and there could also be effects of the instruments directly on the tendency to apply for DI. For the later lags, the effect of prison on additional applications is probably no longer as great a factor, since any increased risk for DI because of prison probably has already resulted in prior applications to the program at that point.

The other variables in the regression reveal some interesting results as well. Most of the coefficients in the linear specifications are small (on the order of suggesting that a 1% change in the variable is responsible for only .01% to .1 % of the change in the category DI applications), but many are significant when looking at the regressions with the six-year lagged prison variable. For instance, the percentage of the working population over age 50 matters for explaining within-state variation in DI applications and is significant but only slightly positive (coefficient of 0.0685) in the linear specifications. Age is insignificant with the log specifications although the magnitude increases (to a coefficient of 0.139). The percentage of people who are overweight has a significant effect in all specifications but is, interestingly enough, negative. This suggests, perhaps, that disabled people eat less or working people eat more. This is a small effect (coefficient of -.016) for the linear model but is much larger (-.958) for the log model. The Republican governor variable suggests Republican administrations have an insignificant effect on overall DI applications in the linear model but is significant but rather small when considering DI-SSI applications. This indicates that something happens during GOP governorships to dissuade the means-tested group from applying to

SSI concurrently with DI. However, the effect is insignificant in the log model for all classes of DI applications.

Other variables have inconclusive effects. In the linear model, mortality has a strong positive effect on DI-only applications but a negative effect for the DI-SSI group. This may mean poor people facing a terminal illness do not bother to go through the DI application process as frequently as the group that is not means-tested. The sign of the coefficient also is sensitive to whether a log model or linear model is specified. Marriage has little effect on DI-application rates in all categories and specifications. The percentage of white people in each state has some positive predicitive power in the log specification for the DI-SSI application group and the combined group (elasticities of 0.675 and 0.450 respectively) but is insignificant in predicting concurrent DI & SSI applications. However, when the linear model is used, the proportion of white people has more predictive power with the DI-only group, although the magnitude of the coefficient drops substantially.

The percentage of people who have a high school diploma has no effect on overall DI applications in the uninstrumented log regression but is negative and significant (with an elasticity of -1.4) in the instrumented one. A higher percentage of high school graduates also matters for reducing concurrent DI-SSI applications (with a -2.4 elasticity in the instrumented regression.) Of all the other variables besides prisoners per capita, the high school coefficient seemed to move around the most with the inclusion of the instruments. It was thought there may be some correlation between the education policy variable of minimum legal drop-out age and high school graduation rates, but this was only about -0.0011. However, the percent of high school graduates was positively correlated with the minimum wage in each state; this correlation was approximately 0.32. The positive effect of minimum wage on DI application combined with its positive correlation with the proportion of high school graduates in the population suggests a positive bias on the high school coefficient without instruments. It also might be that a high school diploma is intrinsically more valuable for remaining in the work force in areas with higher minimum wages. Similarly, it is conceivable that a high school diploma might serve as a stronger signal to employers if one is actually obtained when forced to stay in school through a minimum drop-out age.

The foreign-born variable predicts a small but negative (-.0297) effect on applying for DI in the linear model and this effect increases with the instruments. This difference in the coefficient could occur as a result of foreign-born individuals being attracted to states with higher minimum wages. The 0.31 correlation of the foreign-born and minimum wage variables presumably results in a positive bias in the uninstrumented version. The negative coefficient suggests foreigners are less likely to apply for DI, all else equal. However, in the log version of the regression this effect is not significant.

1.4.2 Robustness and Sensitivity Checks

A few additional adjustments to the model revealed how the coefficients were sensitive. I had disability application data from 1993 to 2009 but chose to drop the last two years as this made some of the Sargan numbers problematic. The 2008 and 2009 period contained many changes in the economy and the number of people who became eligible for disability applications increased as labor force participation diminshed more than at any time in the last 50 years. I argue that, due to the Great Recession, this period is systematically different than the 1993-2007 period and can be explored in a future study when more data about subsequent years becomes available. For now, it is hard to say how some of the variables may have changed in their effect on DI post-recession.

I also found that the instruments were somewhat weaker for the eight largest states but worked very well for the remainder of the states. The low Cragg-Donald statistic for the large states might be because the largest state, California, did not vary its minimum school drop-out age during the entire period of the sample.

I considered other specifications where a time trend was used. This didn't change the results much although it usually made the coefficient on prison slightly stronger for all classes of the DI variable. However, the standard errors also increased to the point that the coefficient was insignificant in many lags. This may be due to high collinearity in the data between minimum wage and time, which was 0.84. Other specifications in which the prison variable was treated as a flow rather than as stock were discarded as the instruments lost much of their strength, perhaps due to the noise of the flows.

To check for collinearity between the explanatory variables, variance inflation factors are given for the uncentered (non-fixed effects) regression of overall disability applications in Table 12. The general "rule of thumb" is that VIF's lower than five are acceptable. All of the variables in the regression fit this criteria.

1.5 Conclusion

This paper uses a novel approach to assess the effect of prison on subsequent disability insurance applications. Thinking that some prisoners may be inherently disabled, either mentally or physically, prior to going to prison, one can conclude that a simple regression of prisoners on disability applications may not reveal much. However, if life in prison generates disabilities for some inmates then one might expect a relationship between the two. I use the policies of minimum wage increases and minimum legal school drop-out ages as instrumental variables to exogenously change the rate at which people are sent to prison. This shows that there is a significant effect on DI applications a few years later. The 0.54 positive elasticity of prison on DI applications at a lag of six years in the log model reflects an additional cost accompanying incarceration. As of 2012, according to Social Security Administration data, approximately 34.9% of all DI applications were successful. While the number of DI applications has increased drastically since 1980, the success rate of these applications has changed through the years: rising somewhat through the late 1990s and falling back to early 1980's levels since then. In 1982, the success rate was 33%. It hit a high of 52.0% in 1998 before it began a steady decrease. To put numbers into perspective, a 0.54 per prisoner application increase corresponds to a 0.18 per prisoner award increase today and would correspond to a 0.28 per prisoner award increase in the late 1990's, a period covered by this paper. One can consider that the cost of incarcerating additional prisoners now must include a figure that corresponds to paying 18% to 28% of these prisoners their disability replacement wages. This is perhaps an unforeseen consequence of excessive incarceration, but potentially a real one. It remains to be seen whether prison is getting harsher over time as Haney suggested. If he is correct in this supposition, then one can only expect the percentage of ex-inmates seeking DI benefits to continue to increase.

The effect of prison on DI insurance is roughly the same whether one applies to the means-tested SSI program concurrently or not. This is true at least with the linear specification although the log specification suggests that the means-tested group applies for DI more often as a result of an incarceration spell. Explanations for why this may be are tenuous at best but it could be due to a lack of resources for the poor group to recover from the harsh, disabling conditions of prison. It also might be that the effects of prison are harder on the acquaintances of poor prisoners than richer ones. Poorer people related to someone incarcerated are likely stretched even thinner and may come to rely more on public assistance and also experience more relative hardship than richer people. The poor may also experience increased "neighborhood effects" and suffer the stigma of living in an area disproportionately populated by prisoners; this reduces employment possibilities and subsequently increases the hazard for DI.

In any event, there does seem to be an effect of prison on DI applications and subsequent awards. This has policy relevance as the number of people on DI has grown since the days when Bound and Burkhauser published their seminal article about it in 1999. Even though the prospect of successful applications have dropped from about 50% to 34% since that time, application numbers have increased 250% and the diagnoses accompanying successful awards have shifted more to non-lethal conditions, causing longer spells on the DI program. These cases include mental disorders, many of which may be the result of a prison sentence. Given that many people are applying in their 30's and 40's (especially including ex-inmates) some of the longer DI spells have not even been concluded in the fifteen year time frame studied here, even if began prior to 1994. If a person is on DI until "retirement" at age 65, a spell by a younger worker is significantly more costly than that of an older worker, despite smaller replacement wages. As a person at 35 would be on disability for six times as long as one at 60, the younger DI population deserves substantial policy attention. Since the population of ex-inmates is growing fastest in the youngest age groups (age under 40), the question of when ex-inmates apply for DI is crucial for determining the cost of incarceration and its effect on entitlement programs. If these inmates are removing themselves from the labor force as a result of prison-induced disability, prison may also be having an unforseen role on the dwindling labor force participation issue. The results of this paper suggest that prison, employment, and entitlements may tie together in unexpected ways. More study with different data and models, such as panel data (if available), should be considered.

1.6 References

- 1. Ahn, Tom, Peter Arcidiacono, and Walter Wessels, "The Distributional Impacts of Minimum Wage Increases When Both Labor Supply and Labor Demand Are Endogenous," *Journal of Business & Economic Statistics*, 2011, 29 (1), 12-23.
- 2. Autor, David H. and Mark G. Duggan, "The Rise in Disabilty Rolls and the Decline in Unemployment,"*Quarterly Journal of Economics,* CXVIII (2003), 157-205.
- 3. Autor, David H. and Mark G. Duggan, "The Growth in the Social Secu-

rity Disability Rolls: A Fiscal Crisis Unfolding", *NBER Working Paper No. 12436*, August 2006

- 4. Beauchamp, Andrew and Stacey Chan, "Crime and the Minimum Wage", *Ph.D. Thesis*, Boston College 2012.
- 5. Beck, A.J. (2000) "Prison and Jail Inmates at Midyear 1999". *Bureau of Justice Statistics Bulletin.* NCJ 181643
- 6. Becker, Gary "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76 (no. 2): 169-217, March/April, l968.
- 7. Ben-Moshe, Liat. "The Institution yet to Come: Analyzing Incarceration Through a Disability Lens" in *The Disability Studies Reader.* Edited by Lennard J. Davis. Hardcover: 600 pages. Publisher: Routledge; 4th edition (February 8, 2013)
- 8. Bonczar, T.P. & A.J. Beck (1997). "Lifetime Likelihood of Going to State or Federal Prison." *Bureau of Justice Statistics Special Report.* NCJ 160092.
- 9. Bound, John and Richard V. Burkhauser "Economic Analysis of Transfer Programs Targeted on People with Disabilities," Center for Policy Research, Syracuse University, 1999.
- 10. Coe, Norma, Kelly Haverstick, Alicia Munnell, and Anthony Webb "What Explains State Variation in SSDI Application Rates?", *Working Paper, Center for Retirement Research*, Boston College, 2011.
- 11. Chan, Stacey, "Not Until You're Older: The Minimum Dropout Age and the Crime-Age Profile," *Ph.D. Thesis*, Boston College 2012.
- 12. Cho, Rosa and Robert LaLonde. "The Impact of Incarceration in State Prison on the Employment Prospects of Women," *Journal of Quantitative Criminology,* September 2008, Volume 24, Issue 3, pp 243-265
- 13. Conly, Catherine H. "Helping Inmates Obtain Federal Disability Benefits: Serious Medical and Mental Illness, Incarceration, and Federal Disability Entitlement Programs," Abt Associates, Inc., United States of America, Date Published: 06/2005. Sponsoring Agency: National Institute of Justice.
- 14. Freeman, R.B. "Crime and the Employment of Disadvantaged Youth." Pp. 201-37 in George Peterson & Wayne Vronman (Eds.), *Urban Labor Markets and Job Opportunity.* (1992), Washington DC: Urban Institute Press.
- 15. Garfinkel, Irwin. 2001. "Child Support in the New World of Welfare." In R. Blank and R. Haskins eds. The New World of Welfare. Washington D.C.: Brookings Institution, 442-60., 2001.
- 16. Geller, Amanda, Irwin Garfinkel, and Bruce Western. "The Effects of Incarceration on Employment and Wages: An Analysis of the Fragile Families Survey." Center for Research on Child Wellbeing. *Working Paper #2006-01-FF.* (2006)

- 17. Granovetter, Mark. "Getting a Job: A study of Contracts and Careers." (2nd ed.). Chicago: University of Chicago Press. (1995)
- 18. Grogger, J., "The Effect of Arrests on the Employment and Earnings of Young Men." *Quarterly Journal of Economics,* (1996), 110:51-71
- 19. Haney, Craig. "The Psychological Impact of Incarceration: Implications for Post-Prison Adjustment." *Working Paper for Conference.* U.S. Department of Health and Human Services. The Urban Institute. 2002.
- 20. Hashimoto, Masanori, "The Minimum Wage and Youth Crimes: Time-Series Evidence," *Journal of Law Economics,* 1987, 30 (2), 443-64
- 21. Holzer, Harry J. "Collateral costs: the effects of incarceration on the employment and earnings of young workers", *IZA Discussion Papers*, No. 3118, 2007.
- 22. Holzer, Harry J., Steven Raphael and Michael Stoll. "Will Employers Hire Former Offenders? Employer Preference, Background Checks and their Determinants." In M. Pattillo, D. Weiman and B. Western eds. Imprisoning America: The Social Effects of Mass Incarceration. New York: Russell Sage Foundation. 2004.
- 23. Holzer, Harry J., Steven Raphael and Michael Stoll. "The Effect of an Applicant's Criminal History on Employer Hiring Decisions and Screening Practices: New Evidence from Los Angeles. In S. Bushway, M. Stoll and D. Weiman eds." Barriers to Reentry? The Labor Market for Released Prisoners in Post-Industrial America. 2007, New York: Russell Sage Foundation.
- 24. Irwin, J. & J. Austin, "It's About Time: America's Imprisonment Binge." Second Edition. Belmont, CA: Wadsworth. (1997)
- 25. Kirby, H., "Locked Up and Shipped Away: Interstate Prison Transfers and The Private Prison Industry." Grassroots Leadership. Charlotte, NC. (2013)
- 26. Kling, Jeffrey R. "Incarceration Length, Employment and Earnings." *American Economic Review,* 96:3 (June 2006), 863-876
- 27. Laub, John H., Daniel S. Nagin, and Robert J. Sampson "Trajectories of Change in Criminal Offending: Good Marriages and Desistance Process." *American Sociological Review,* (1998), 63:225-38.
- 28. Lott, John R., "Do We Punish High Income Criminals Too Heavily?" *Economic Inquiry.* (1992a), 30:583-608.
- 29. Lott, John R. "An Attempt at Measuring The Total Monetary Penalty From Drug Convictions: The Importance of an Individual's Reputation." "*Journal of Legal Studies,* (1992b), 21:159-187
- 30. Lyons, Christopher J. and Becky Pettit "Compounded Disadvantage: Race, Incarceration, and Wage Growth" *Social Problems* Vol. 58, No. 2 (May 2011), pp. 257-280

- 31. Nagin, Daniel and Joel Waldfogel. "The Effect of Convicton on Income Through the Life Cycle" *International Review of Law & Economics,* Vol. 18, no. 1 (March 1998), pp. 25-40.
- 32. Needels, Karen E. "Go Directly to Jail and Do Not Collect? A Long-Term Study of Recidivism, Employment, and Earnings Patterns among Prison Releases." *Journal of Research in Crime and Delinquency,* (1996), 33:471-96
- 33. Oreopoulos, Philip, "Would More Compulsory Schooling Help Disadvantaged Youth? Evidence from Recent Changes to School-Leaving Laws," in "The Problems of Disadvantaged Youth: An Economic Perspective" NBER Chapters, National Bureau of Economic Research, Inc., 2009, pp. 85-112.
- 34. Pager, Devah "The Mark of a Criminal Record." *American Journal of Sociology,* (2003,) 108(5):937-975.
- 35. Pastore, Ann L. & Kathleen Maguire. (2000) "Sourcebook of Criminal Justice Statistics" [Online] Available: http://www.albany.edu/sourcebook
- 36. Raphael, Steven "Incarceration and Prisoner Reentry in the United States." *The ANNALS of the American Academy of Political and Social Science,* May 2011, Vol. 635, No. 1, 192-215
- 37. She, Peiyun and David C. Stapleton. "A Review of Disability Data for the Institutional Population: Research Brief." Cornell University Institute for Policy Research. Rehabilitation Research and Training Center on Disability Demographics and Statistics, May 2006.
- 38. Social Security Administration Data. Available: http://www.ssa.gov/policy/docs/statcomps/di_asr/2013/
- 39. Stock, James & Motohiro Yogo. "Testing for Weak Instruments in Linear IV Regression" In Identification and Inference for Econometric Models edited by Donald W. K. Andrews. New York: Cambridge University Press; 2005. pp. 80-108.
- 40. Tyler, John H. and Jeffrey R. Kling. "Prison-Based Education and Re-Entry into the Mainstream Labor Market." In *Barriers to Reentry? The Labor Market for Released Prisoners in Post-Industrial America.* Edited by Shawn Bushway, Michael Stoll, and David Weiman. New York: Russell Sage Foundation Press, 2007. 227-256
- 41. Waldfogel J., "Does Conviction Have a Persistent Effect on Income and Employment?," *International Review of Law and Economics,* March 1994.
- 42. Western, Bruce & Katherine Beckett. "How Unregulated is the U.S. Labor Market: The Penal System as a Labor Market Institution." *American Journal of Sociology,* (1999), 104:1030-60.
- 43. Western, Bruce, Jeffrey R. Kling and David F. Weiman. "The Labor Market Consequences of Incarceration." *Crime and Delinquency* 47:3 (July 2001), 410-427

- 44. Western, Bruce & Sara McLanahan. "Fathers in Fragile Families: The Impact of Incarceration on Family Formation." In Families and Crime edited by Greer Litton Fox. London: Oxford University Press. (2000).

- 45. Western, Bruce & Becky Pettit. "Incarceration and Racial Inequality in Men's Employment." *Industrial and Labor Relations Review*, (2000), 54:3-16.

1.7 Tables

Table 1.1: Variable Key

Variable	Label	Description
Prisoner Per Capita	PPCL6	Prisoners Per Capita Lagged Six Years in Each State
Mortality Rate	Mort. Rate	Mortality Rate Per Capita
Percent HS Plus	% HSPlus	Percent in Each State with a HS education or more
Percent Age 50+	% Age+50	Percent of Adult Pop 18-64 Aged over 50 in Each State
Percent Overweight	% Overweight	Percent Identified as Overweight (by BMI) in Each State
Republican Governor	Rep. Gov.	Binary Variable for Republican Governor in State
Percent Married	% Married	Percent Married in Each State
Percent White	% White	Percent Identified as White in Each State
Percent Foreign-Born	% For. Born	Percent Identified as Foreign-Born in Each State

The Prisoners Per Capita variable is a percentage of the working age population in each state. The other percentage variables are based on total state populations.

Table 1.2: Response of Combined DI-Only and DI-SSI Applications to Prison Variables†

	Specifications			
Variable	2 yr. lag	3 yr. lag	4 yr. lag	5 yr. lag
Pris. Per. Cap.	-2.12 x 10^{-1}**	-1.41 x 10^{-1}	-4.25x 10^{-2}	7.31 x 10^{-3}
	(1.00 x 10^{-2})	(9.00 x 10^{-2})	(9.54 x 10^{-2})	(8.98 x 10^{-2})
Pris. Per. Cap. (with IV)	3.23 x 10^{-1}	1.10 x 10^{-0}**	1.33 x 10^{-0}***	2.00 x 10^{-0}***
	(5.18 x 10^{-1})	(4.95 x 10^{-1})	(4.88 x 10^{-1})	(5.56 x 10^{-1})
Sargan p-value	0.2075	0.9897	0.9998	0.8141
Cragg-Donald Statistic	12.82	15.28	15.42	14.79
R-squared	0.843	0.842	0.841	0.841
R-squared (IV)	0.821	0.828	0.804	0.800
Variable	6 yr. lag	7 yr. lag	8 yr. lag	9 yr. lag
Pris. Per. Cap.	1.41 x 10^{-1}	2.36 x 10^{-1}**	2.67x 10^{-1}***	3.55 x 10^{-1}***
	(9.44 x 10^{-2})	(9.74 x 10^{-2})	(9.93 x 10^{-2})	(9.94 x 10^{-2})
Pris. Per. Cap. (with IV)	1.96 x 10^{-0}***	1.33 x 10^{-0}***	9.92 x 10^{-1}***	6.26 x 10^{-1}**
	(5.28 x 10^{-1})	(4.47 x 10^{-1})	(3.93 x 10^{-1})	(3.88 x 10^{-1})
Sargan p-value	0.9224	0.8604	0.2793	0.1794
Cragg-Donald Statistic	15.48	16.30	18.20	15.47
R-squared	0.835	0.857	0.870	0.890
R-squared (IV)	0.830	0.855	0.889	0.880
State Level Co-Variates	Yes.	Yes.	Yes.	Yes.
State FE	Yes.	Yes.	Yes.	Yes.

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

† Dependent variable is combined rate of DI-only and concurrent SSI-DI applications per adult population aged 18-64. State-level covariates include the variables for mortality rate, percent white, percent with at least a HS diploma, percent at least 50 years old, percent foreign born, percent married, percent overweight, and an indicator for Republican governor.

Table 1.3: Response of DI-Only Applications to Prison†

	Specifications			
Variable	2 yr. lag	3 yr. lag	4 yr. lag	5 yr. lag
Pris. Per. Cap.	3.24×10^{-2}	6.93×10^{-2}	1.09×10^{-1}**	1.08×10^{-1}**
	(5.29×10^{-2})	(5.01×10^{-2})	(4.80×10^{-2})	(4.70×10^{-2})
Pris. Per. Cap. (with IV)	2.40×10^{-1}	5.76×10^{-1}**	6.31×10^{-1}***	1.00×10^{-0}***
	(2.71×10^{-1})	(2.50×10^{-1})	(2.40×10^{-1})	(2.74×10^{-1})
Sargan p-value	0.2566	0.8573	0.4559	0.6846
Cragg-Donald Statistic	12.82	15.28	15.42	14.78
R-squared	0.787	0.789	0.789	0.789
R-squared (IV)	0.771	0.779	0.756	0.755
Variable	6 yr. lag	7 yr. lag	8 yr. lag	9 yr. lag
Pris. Per. Cap.	1.30×10^{-1}**	1.14×10^{-1}**	7.78×10^{-2}	9.89×10^{-2}*
	(5.02×10^{-2})	(5.37×10^{-2})	(5.47×10^{-2})	(5.39×10^{-2})
Pris. Per. Cap. (with IV)	8.98×10^{-1}***	5.38×10^{-1}**	4.24×10^{-1}**	4.60×10^{-1}**
	(2.61×10^{-1})	(2.34×10^{-1})	(2.14×10^{-1})	(2.18×10^{-1})
Sargan p-value	0.4019	0.2992	0.7889	0.7242
Cragg-Donald Statistic	15.49	16.30	18.20	15.48
R-squared	0.790	0.791	0.807	0.835
R-squared (IV)	0.774	0.790	0.819	0.771
State Level Co-Variates	Yes.	Yes.	Yes.	Yes.
State FE	Yes.	Yes.	Yes.	Yes.

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

† Dependent variable is rate of DI-only applications per adult population aged 18-64. State-level covariates include the variables for mortality rate, percent white, percent with at least a HS diploma, percent at least 50 years old, percent foreign born, percent married, percent overweight, and an indicator for Republican governor.

Table 1.4: Response of DI-SSI Applications to Prison†

	Specifications			
Variable	2 yr. lag	3 yr. lag	4 yr. lag	5 yr. lag
Pris. Per. Cap.	-2.40×10^{-1}***	-2.05×10^{-1}***	-1.44×10^{-1}**	-9.16×10^{-2}
	(6.29×10^{-2})	(5.97×10^{-2})	(5.75×10^{-2})	(5.63×10^{-2})
Pris. Per. Cap. (with IV)	1.18×10^{-1}	5.47×10^{-1}*	7.08×10^{-1}**	9.97×10^{-1}***
	(3.25×10^{-1})	(3.06×10^{-1})	(3.03×10^{-1})	(3.29×10^{-1})
Sargan p-value	0.3025	0.8802	0.5517	0.4676
Cragg-Donald Statistic	12.91	15.53	15.72	14.93
R-squared	0.843	0.841	0.841	0.840
R-squared (IV)	0.821	0.827	0.812	0.802
Variable	6 yr. lag	7 yr. lag	8 yr. lag	9 yr. lag
Pris. Per. Cap.	2.13×10^{-2}	1.22×10^{-1}**	1.88×10^{-1}***	2.53×10^{-1}***
	(5.87×10^{-2})	(5.95×10^{-2})	(5.90×10^{-2})	(5.86×10^{-2})
Pris. Per. Cap. (with IV)	1.06×10^{-0}***	7.92×10^{-1}***	5.60×10^{-1}**	1.56×10^{-1}
	(3.17×10^{-1})	(2.72×10^{-1})	(2.31×10^{-1})	(2.28×10^{-1})
Sargan p-value	0.3922	0.2301	0.0372	0.0526
Cragg-Donald Statistic	15.66	17.79	16.31	15.48
R-squared	0.834	0.861	0.878	0.898
R-squared (IV)	0.835	0.867	0.899	0.902
State Level Co-Variates	Yes.	Yes.	Yes.	Yes.
State FE	Yes.	Yes.	Yes.	Yes.

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

† Dependent variable is rate of DI-SSI applications per adult population aged 18-64. State-level covariates include the variables for mortality rate, percent white, percent with at least a HS diploma, percent at least 50 years old, percent foreign born, percent married, percent overweight, and an indicator for Republican governor.

Table 1.5: Response of Log Combined DI-Only and DI-SSI Applications to Selected Variables†

	Specifications			
Variable	2 yr. lag	3 yr. lag	4 yr. lag	5 yr. lag
Pris. Per. Cap.	-1.22 x 10^{-2}**	-1.05 x 10^{-2}**	-7.80x 10^{-2}*	-9.04 x 10^{-2}**
	(4.92 x 10^{-2})	(4.81 x 10^{-2})	(4.69 x 10^{-2})	(3.95 x 10^{-2})
Pris. Per. Cap. (with IV)	-4.07 x 10^{-2}	3.44 x 10^{-1}	3.66 x 10^{-1}	5.54 x 10^{-1}**
	(3.86 x 10^{-1})	(3.75 x 10^{-1})	(3.00 x 10^{-1})	(2.56 x 10^{-1})
Sargan p-value	0.9447	0.6393	0.7958	0.3447
Cragg-Donald Statistic	5.29	5.96	8.97	13.46
R-squared	0.764	0.763	0.764	0.764
R-squared (IV)	0.726	0.746	0.723	0.708
Variable	6 yr. lag	7 yr. lag	8 yr. lag	9 yr. lag
Pris. Per. Cap.	-4.26 x 10^{-2}	-1.85 x 10^{-2}	3.48x 10^{-2}	1.01 x 10^{-1}
	(5.00 x 10^{-2})	(5.39 x 10^{-2})	(5.82 x 10^{-2})	(6.21 x 10^{-2})
Pris. Per. Cap. (with IV)	5.43 x 10^{-1}**	2.75 x 10^{-1}	3.13 x 10^{-1}	1.58 x 10^{-1}
	(2.43 x 10^{-1})	(2.33 x 10^{-1})	(2.32 x 10^{-1})	(2.50 x 10^{-1})
Sargan p-value	0.2865	0.4750	0.7324	0.4377
Cragg-Donald Statistic	15.63	15.88	16.93	14.26
R-squared	0.767	0.770	0.791	0.812
R-squared (IV)	0.754	0.774	0.809	0.796
State Level Co-Variates	Yes.	Yes.	Yes.	Yes.
State FE	Yes.	Yes.	Yes.	Yes.

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

† Dependent variable is combined rate of DI-only and concurrent SSI-DI applications per adult population aged 18-64. State-level covariates include the variables for mortality rate, percent white, percent with at least a HS diploma, percent at least 50 years old, percent foreign born, percent married, percent overweight, and an indicator for Republican governor.

Table 1.6: Response of Log DI Apps. to Log Six Year Prison Lags and Other Variables†

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Specifications					
PPCL6	-.043	.543**	.044	.324	-.142	.938***
	(.050)	(.244)	(.055)	(.248)	(.065)	(.344)
Mor. Rate	-.039	.402	.646***	.856***	-.891***	.083
	(.201)	(.278)	(.222)	(.283)	(.262)	(.392)
% HSPlus	-.243	-1.38**	-.151	-.694	-.275	-2.406***
	(.317)	(.570)	(.348)	(.579)	(.410)	(.812)
% Age+50	.336***	.139	.229**	.135	.521***	.160
	(.085)	(.120)	(.094)	(.122)	(.110)	(.169)
% Overwt.	-.839***	-.958***	-.566***	-.622***	-1.209***	-1.434***
	(.163)	(.179)	(.180)	(.182)	(.212)	(.253)
Rep Gov.	-.037***	-.012	-.020	-.008	-.063***	-.016
	(.014)	(.018)	(.01)	(.016)	(.019)	(.026)
% Married	-.397	-.175	-.448	-.341	-.404	-.019
	(.356)	(.386)	(.390)	(.393)	(.461)	(.543)
% White	-.024	.450*	.165	.391	-.181	.675**
	(.157)	(.253)	(.172)	(.258)	(.203)	(.353)
% For.	.196***	.057	.218***	.151**	.192***	-.061
	(.043)	(.072)	(.047)	(.074)	(.055)	(.102)
I.V.	No.	Yes.	No.	Yes.	No.	Yes.
Cragg-Donald stat.	–	15.63	–	15.63	–	15.56
Sargan p-value	–	0.2865	–	0.1296	–	0.9795
R-Squared	0.767	0.754	0.669	0.662	0.760	0.779
State FE	Yes.	Yes.	Yes.	Yes.	Yes.	Yes.
Application Type	Combined	Combined	DI-Only	DI-Only	DI-SSI	DI-SSI

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

† Dependent variable is combined rate of DI-only and concurrent SSI-DI applications per adult population aged 18-64. Prisoner per capita is not instrumented.

Table 1.7: Response of DI Apps. to Six Year Prison Lags and Other Variables[†]

Variable	Specifications (1)	(2)
PPCL6	$1.42\text{x}10^{-1}$	$1.96\text{x}10^{-0}$***
	($9.44\text{x}10^{-2}$)	($5.27\text{x}10^{-1}$)
Mor. Rate	$-3.02\text{x}10^{-1}$*	$-3.34\text{x}10^{-2}$
	($1.81\text{x}10^{-1}$)	($2.31\text{x}10^{-1}$)
% HSPlus	$2.45\text{x}10^{-3}$	$-1.10\text{x}10^{-2}$**
	($2.39\text{x}10^{-3}$)	($4.77\text{x}10^{-3}$)
% Age+50	$1.19\text{x}10^{-2}$ ***	$6.85\text{x}10^{-2}$**
	($2.12\text{x}10^{-3}$)	($2.94\text{x}10^{-3}$)
% Overwt.	$-1.56\text{x}10^{-2}$ ***	$-1.60\text{x}10^{-2}$***
	($3.08\text{x}10^{-3}$)	($3.72\text{x}10^{-3}$)
Rep Gov.	$-2.78\text{x}10^{-4}$ ***	$-1.42\text{x}10^{-4}$
	($9.37\text{x}10^{-5}$)	($1.20\text{x}10^{-4}$)
% Married	$-3.07\text{x}10^{-3}$	$-7.68\text{x}10^{-4}$
	($2.36\text{x}10^{-3}$)	($2.92\text{x}10^{-3}$)
% White	$-1.54\text{x}10^{-3}$	$4.80\text{x}10^{-3}$
	($2.07\text{x}10^{-3}$)	($3.07\text{x}10^{-3}$)
% For.	$-3.42\text{x}10^{-3}$	$-2.97\text{x}10^{-2}$***
	($5.38\text{x}10^{-3}$)	($9.86\text{x}10^{-3}$)
I.V.	No.	Yes.
Cragg-Donald stat.	-	15.49
Sargan p-value	-	0.9224
R-Squared	0.847	0.797
State FE	Yes.	Yes.
Application Type	Combined	Combined

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

[†] Dependent variable is combined rate of DI-only and concurrent SSI-DI applications per adult population aged 18-64. Prisoner per capita is not instrumented.

Table 1.8: Response of DI Apps. to Six Year Prison Lags and Other Variables†

	Specifications			
Variable	(3)	(4)	(5)	(6)
PPCL6	1.29x10^{-1}**	8.98x10^{-1}***	2.13x10^{-3}	1.06x10^{-0}***
	(5.02x10^{-2})	(2.61x10^{-1})	(5.87x10^{-2})	(3.17x10^{-1})
Mor. Rate	1.81x10^{-1}*	2.95 x 10^{-1}**	-4.85x10^{-1}***	-3.26x10^{-1}**
	(9.62x10^{-2})	(1.14x10^{-1})	(1.13x10^{-1})	(1.40x10^{-1})
% HSPlus	2.27x10^{-3}**	-3.40x10^{-3}	4.37x10^{-4}	-7.40x10^{-3}**
	(1.27x10^{-3})	(2.36x10^{-3})	(1.49x10^{-3})	(2.91x10^{-3})
% Age+50	5.18x10^{-3}***	3.03x10^{-3}**	6.70x10^{-3}***	3.78x10^{-3}***
	(1.13x10^{-3})	(1.45x10^{-3})	(1.33x10^{-3})	(1.78x10^{-3})
% Overwt.	-4.95x10^{-3}***	-5.09x10^{-3}***	-1.05x10^{-2}***	-1.08x10^{-2}***
	(1.64x10^{-3})	(1.83x10^{-3})	(1.93x10^{-3})	(2.26x10^{-3})
Rep Gov.	-5.21x10^{-5}	4.42x10^{-6}	- -2.27x10^{-4}***	-1.44x10^{-4}**
	(4.98x10^{-5})	(5.91x10^{-5})	(5.85x10^{-5})	(7.28x10^{-5})
% Married	-1.59x10^{-3}	-6.10x10^{-4}	-1.37x10^{-3}	-1.34x10^{-4}
	(1.26x10^{-3})	(1.44x10^{-3})	(1.47x10^{-3})	(1.76x10^{-3})
% White	1.43x10^{-3}	4.11x10^{-3}***	-2.89x10^{-3}**	-6.38x10^{-4}
	(1.10x10^{-3})	(1.52x10^{-3})	(1.23x10^{-3})	(1.84x10^{-3})
% For.	1.43x10^{-3}	-9.64x10^{-3}**	5.30x10^{-3}	-2.00x10^{-2}***
	(2.86x10^{-3})	(4.87x10^{-3})	(3.35x10^{-3})	(5.86x10^{-3})
I.V.	No.	Yes.	No.	Yes.
Cragg-Donald stat.	-	15.49	-	15.66
Sargan p-value	-	0.4019	-	0.3922
R-Squared	0.791	0.745	0.846	0.805
State FE	Yes.	Yes.	Yes.	Yes.
Application Type	DI-Only	DI-Only	DI-SSI	DI-SSI

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

† Dependent variable is combined rate of DI-only and concurrent SSI-DI applications per adult population aged 18-64. Prisoner per capita is not instrumented.

Table 1.9: Response of Combined DI-Only & DI-SSI Apps. to Selected Variables†

	Specifications			
Variable	(1)	(2)	(3)	(4)
PPCL6	5.38×10^{-1}***	1.69×10^{-1}**	1.42×10^{-1}	1.42×10^{-1}
	(6.55×10^{-2})	(8.61×10^{-2})	(8.67×10^{-2})	(9.44×10^{-2})
Mor. Rate	-	-2.69×10^{-1}*	-2.73×10^{-1}*	-3.02×10^{-1}*
		(1.58×10^{-1})	(1.57×10^{-1})	(1.81×10^{-1})
% HSPlus	-	3.29×10^{-3}	2.31×10^{-3}	2.45×10^{-3}
		(2.37×10^{-3})	(2.38×10^{-3})	(2.39×10^{-3})
% Age+50	-	1.25×10^{-2}***	1.21×10^{-2}***	1.19×10^{-2}***
		(2.03×10^{-3})	(2.04×10^{-3})	(2.12×10^{-3})
% Overwt.	-	-1.49×10^{-2}***	-1.55×10^{-2}***	-1.56×10^{-2}***
		(3.06×10^{-3})	(3.05×10^{-3})	(3.08×10^{-3})
Rep Gov.	-	-	-2.77×10^{-4}***	-2.78×10^{-4}***
			(9.28×10^{-5})	(9.37×10^{-5})
% Married	-	-	-3.44×10^{-3}	-3.07×10^{-3}
			(2.25×10^{-3})	(2.36×10^{-3})
% White	-	-	-	-1.54×10^{-3}
				(2.07×10^{-3})
% For.	-	-	-	-3.42×10^{-3}
				(5.38×10^{-3})
I.V.	No.	No.	No.	No.
State FE	Yes.	Yes.	Yes.	Yes.
R-Squared	0.812	0.842	0.847	0.847

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

† Dependent variable is combined rate of DI-only and concurrent SSI-DI applications per adult population aged 18-64. Prisoner per capita is not instrumented.

Table 1.10: Response of Combined DI-Only and DI-SSI Applications With Instruments†

	Specifications			
Variable	(1)	(2)	(3)	(4)
PPCL6	9.86×10^{-1}***	1.14×10^{-0}***	1.17×10^{-0}***	1.96×10^{-0}***
	(9.37×10^{-2})	(2.60×10^{-1})	(2.82×10^{-1})	(5.27×10^{-1})
Mor. Rate	-	3.10×10^{-1}	3.28×10^{-1}	-3.34×10^{-2}
		(2.21×10^{-1})	(2.27×10^{-1})	(2.31×10^{-1})
% HSPlus	-	-5.67×10^{-3}*	-6.18×10^{-3}*	-1.10×10^{-2}**
		(3.35×10^{-3})	(3.35×10^{-3})	(4.77×10^{-3})
% Age+50	-	6.03×10^{-2}**	5.82×10^{-3}**	6.85×10^{-2}**
		(2.70×10^{-3})	(2.70×10^{-3})	(2.94×10^{-3})
% Overwt.	-	-1.75×10^{-2}***	-1.76×10^{-2}***	-1.60×10^{-2}***
		(3.29×10^{-3})	(3.29×10^{-3})	(3.72×10^{-3})
Rep. Gov.	-	-	-2.45×10^{-4}**	-1.42×10^{-4}
			(9.89×10^{-5})	(1.20×10^{-4})
% Married	-	-	-9.37×10^{-4}	-7.68×10^{-4}
			(2.63×10^{-3})	(2.92×10^{-3})
% White	-	-	-	4.80×10^{-3}
				(3.07×10^{-3})
% For.	-	-	-	-2.97×10^{-2}***
				(9.86×10^{-3})
I.V.	Yes.	Yes.	Yes.	Yes.
State FE	Yes.	Yes.	Yes.	Yes.
R-Squared	0.795	0.825	0.827	0.797
Cragg-Donald stat.	307.00	44.06	37.71	15.49
Sargan p-value	0.1622	0.2026	0.3755	0.9224

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

† Dependent variable is combined rate of DI-only and concurrent SSI-DI applications per adult population aged 18-64. Prisoner per capita is instrumented using state minimum wage and minimum legal high school drop-out age.

Table 1.11: Response of DI-Only Applications to Selected Variables†

	Specifications			
Variable	(1)	(2)	(3)	(4)
PPCL6	2.43×10^{-1}***	1.26×10^{-1}***	1.17×10^{-1}**	1.29×10^{-1}**
	(3.40×10^{-2})	(4.55×10^{-2})	(4.61×10^{-2})	(5.02×10^{-2})
Mor. Rate	-	1.87×10^{-1}**	1.86×10^{-1}**	1.81×10^{-1}*
		(8.35×10^{-2})	(8.35×10^{-2})	(9.62×10^{-2})
% HSPlus	-	2.68×10^{-3}**	2.40×10^{-3}*	2.27×10^{-3}**
		(1.25×10^{-3})	(1.26×10^{-3})	(1.27×10^{-3})
% Age+50	-	5.07×10^{-3}***	4.91×10^{-3}***	5.18×10^{-3}***
		(1.08×10^{-3})	(1.08×10^{-3})	(1.13×10^{-3})
% Overwt.	-	-4.93×10^{-3}***	-5.12×10^{-3}***	-4.95×10^{-3}***
		(1.62×10^{-3})	(1.63×10^{-3})	(1.64×10^{-3})
Rep. Gov.	-	-	-5.73×10^{-5}	-5.21×10^{-5}
			(4.94×10^{-5})	(4.98×10^{-5})
% Married	-	-	-1.13×10^{-3}	-1.59×10^{-3}
			(1.20×10^{-3})	(1.26×10^{-3})
% White	-	-	-	1.43×10^{-3}
				(1.10×10^{-3})
% For.	-	-	-	1.43×10^{-3}
				(2.86×10^{-3})
I.V.	No.	No.	No.	No.
State FE	Yes.	Yes.	Yes.	Yes.
R-Squared	0.766	0.789	0.790	0.791

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

† Dependent variable is rate of DI-only application per adult population aged 18-64. Adults who applied to SSI concurrently are not in the dependent variable here. Prisoner per capita is not instrumented.

Table 1.12: Response of DI-Only Applications With Instruments†

Variable	Specifications (1)	(2)	(3)	(4)
PPCL6	3.99×10^{-1}***	4.48×10^{-1}***	4.60×10^{-1}***	8.98×10^{-1}***
	(4.77×10^{-2})	(1.30×10^{-1})	(1.42×10^{-1})	(2.61×10^{-1})
Mor. Rate	-	3.80×10^{-1}***	3.86×10^{-1}***	2.95×10^{-1}**
		(1.10×10^{-1})	(1.14×10^{-1})	(1.14×10^{-1})
% HSPlus	-	-2.91×10^{-3}	-4.12×10^{-4}	-3.40×10^{-3}
		(1.68×10^{-3})	(1.68×10^{-3})	(2.36×10^{-3})
% Age+50	-	2.90×10^{-3}***	2.84×10^{-3}**	3.03×10^{-3}**
		(1.35×10^{-3})	(1.35×10^{-3})	(1.45×10^{-3})
% Overwt.	-	-5.78×10^{-3}***	-5.79×10^{-3}***	-5.09×10^{-3}***
		(1.65×10^{-3})	(1.65×10^{-3})	(1.83×10^{-3})
Rep. Gov.	-	-	-4.80×10^{-5}	4.42×10^{-6}
			(4.96×10^{-5})	(5.91×10^{-5})
% Married	-	-	3.29×10^{-4}	-6.10×10^{-4}
			(1.32×10^{-3})	(1.44×10^{-3})
% White	-	-	-	4.11×10^{-3}***
				(1.52×10^{-3})
% For.	-	-	-	-9.64×10^{-3}**
				(4.87×10^{-3})
I.V.	Yes.	Yes.	Yes.	Yes.
State FE	Yes.	Yes.	Yes.	Yes.
R-Squared	0.757	0.781	0.782	0.745
Cragg-Donald stat.	307.00	44.06	37.71	15.49
Sargan p-value	0.0595	0.0339	0.0478	0.4019

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

† Dependent variable is rate of DI-only application per adult population aged 18-64. Adults who applied to SSI concurrently are not in the dependent variable here. Prisoner per capita is instrumented using state minimum wage and minimum legal high school drop-out age.

Table 1.13: Response of Concurrent SSI-DI Applications to Selected Variables†

	Specifications			
Variable	(1)	(2)	(3)	(4)
PPCL6	2.99×10^{-1}***	4.63×10^{-2}	2.80×10^{-2}	2.13×10^{-3}
	(4.14×10^{-2})	(5.42×10^{-2})	(5.44×10^{-2})	(5.87×10^{-2})
Mor. Rate	-	-4.47×10^{-1}***	-4.51×10^{-1}***	-4.85×10^{-1}***
		(9.93×10^{-2})	(9.82×10^{-2})	(1.13×10^{-1})
% HSPlus	-	8.53×10^{-4}	1.66×10^{-4}	4.37×10^{-4}
		(1.49×10^{-3})	(1.41×10^{-3})	(1.49×10^{-3})
% Age+50	-	7.35×10^{-3}***	7.01×10^{-3}***	6.70×10^{-3}***
		(1.28×10^{-3})	(1.28×10^{-3})	(1.33×10^{-3})
% Overwt.	-	-9.82×10^{-3}***	-1.03×10^{-2}***	-1.05×10^{-2}***
		(1.93×10^{-3})	(1.92×10^{-3})	(1.93×10^{-3})
Rep Gov.	-	-	-2.21×10^{-4}***	-2.27×10^{-4}***
			(5.81×10^{-5})	(5.85×10^{-5})
% Married	-	-	-2.14×10^{-3}	-1.37×10^{-3}
			(1.40×10^{-3})	(1.47×10^{-3})
% White	-	-	-	-2.89×10^{-3}**
				(1.23×10^{-3})
% For.	-	-	-	5.30×10^{-3}
				(3.35×10^{-3})
I.V.	No.	No.	No.	No.
State FE	Yes.	Yes.	Yes.	Yes.
R-Squared	0.806	0.840	0.845	0.846

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

† Dependent variable is rate of concurrent SSI and DI application per adult population aged 18-64. Adults who applied only to DI are not in the dependent variable here. Prisoner per capita is not instrumented.

Table 1.14: Response of Concurrent SSI-DI Applications With Instruments†

Variable	Specifications (1)	(2)	(3)	(4)
PPCL6	5.88×10^{-1}***	6.86×10^{-1}***	7.10×10^{-1}***	1.06×10^{-0}***
	(5.93×10^{-2})	(1.66×10^{-1})	(1.79×10^{-1})	(3.17×10^{-1})
Mor. Rate	-	-6.84×10^{-2}	-5.61×10^{-2}	-3.26×10^{-1}**
		(1.40×10^{-3})	(1.44×10^{-1})	(1.40×10^{-1})
% HSPlus	-	-5.05×10^{-3}**	-5.47×10^{-3}**	-7.40×10^{-3}**
		(2.13×10^{-3})	(2.13×10^{-3})	(2.91×10^{-3})
% Age+50	-	3.06×10^{-3}*	2.91×10^{-3}*	3.78×10^{-3}***
		(1.71×10^{-3})	(1.71×10^{-3})	(1.78×10^{-3})
% Overwt.	-	-1.16×10^{-2}***	-1.17×10^{-2}***	-1.08×10^{-2}***
		(2.09×10^{-3})	(2.08×10^{-3})	(2.26×10^{-3})
Rep. Gov.	-	-	-1.98×10^{-4}***	-1.44×10^{-4}**
			(6.25×10^{-5})	(7.28×10^{-5})
% Married	-	-	7.10×10^{-3}	-1.34×10^{-4}
			(1.66×10^{-3})	(1.76×10^{-3})
% White	-	-	-	-6.38×10^{-4}
				(1.84×10^{-3})
% For.	-	-	-	-2.00×10^{-2}***
				(5.86×10^{-3})
I.V.	Yes.	Yes.	Yes.	Yes.
State FE	Yes.	Yes.	Yes.	Yes.
R-Squared	0.786	0.816	0.820	0.805
Cragg-Donald stat.	306.61	43.81	37.50	15.66
Sargan p-value	0.5030	0.7344	0.8609	0.3922

*** = p-value <0.01; ** = p-value < 0.05; * = p-value <0.1

† Dependent variable is rate of concurrent SSI and DI application per adult population aged 18-64. Adults who applied only to DI are not in the dependent variable here. Prisoner per capita is instrumented using state minimum wage and minimum legal high school drop-out age.

Table 1.15: Summary Statistics of Selected Variables

Variable	Mean	Standard Deviation	Minimum	Maximum
DI-Only Apps. Per Adult Pop.	0.00428	0.00106	0.00037	0.01017
SSI-DI Apps Per Adult Pop.	0.00395	0.00148	0.00027	0.00920
Combined Apps Per Adult Pop.	0.00828	0.00235	0.00065	0.01648
Prisoner Per Cap	0.00329	0.00136	0.00069	0.00873
Mortality Rate	0.00855	0.00130	0.00380	0.01190
Percent HSPlus	0.851	0.046	0.690	0.944
Percent Age+50	0.279	0.037	0.138	0.377
Percent Overweight	0.349	0.015	0.271	0.425
Republican Governor	0.549	0.497	0	1
Percent Married	0.551	0.031	0.471	0.649
Percent White	0.766	0.152	0.156	0.989
Percent Foreign Born	0.0670	0.0578	0.0080	0.300
Relevant Instruments				
Minimum Wage	5.20	1.42	1.60	9.04
Minimum Dropout Age	16.67	0.86	16	18
Dependent Variables				
DI-Only Apps Per Adult Pop.	0.00428	0.00106	0.000373	0.01017
Concurrent DI-SSI Apps Per Adult Pop.	0.00395	0.00148	0.00027	0.00920

n =1250. Observations at state level 1988-2012

Table 1.16: Collinearity Matrix of Selected Variables

	PPC	Mort.	%HSPlus	%Age+50	% Ovwt.	RepGov.	Married	%White	%For.
PPC	1.00								
Mort Rate.	-0.06	1.00							
% HSPlus	-0.20	-0.42	1.00						
% Age+50	-0.02	0.50	0.13	1.00					
% Ovwt.	-0.03	0.06	0.14	0.10	1.00				
Rep. Gov.	0.02	-0.03	-0.07	-0.12	0.04	1.00			
% Marr.	-0.27	0.01	0.02	-0.30	0.02	0.00	1.00		
% White	-0.47	0.33	0.28	0.07	0.17	-0.04	0.47	1.00	
% Foreign	0.18	-0.40	0.06	0.02	-0.13	0.06	-0.48	-0.61	1.00

Collinearity Matrix

Table 1.17: Collinearity Matrix of Prisoners Per Capita with DI Applications Variables (per Adult Pop. 18-64)

	PPC	Combined DI-SSI	DI Only	DI-SSI
PPC	1.00			
Combined DI-SSI	0.13	1.00		
DI Only	0.07	0.88	1.00	
DI-SSI	0.14	0.94	0.66	1.00

Collinearity Matrix

Table 1.18: Collinearity Matrix of Selected Variables with Instruments (per Adult Pop. 18-64)

	% HS Plus	% Foreign	Min. Wage	Min. Dropout Age
% HS Plus	1.00			
% Foreign	0.0487	1.00		
Min. Wage	0.2558	0.3167	1.00	
Min. Dropout Age	-0.0011	0.1536	0.0284	1.00

Collinearity Matrix

Table 1.19: Collinearity Matrix of Prison Per Capita with Instruments (per Adult Pop. 18-64)

	% Prison Per. Capita	Min. Wage	Min. Dropout Age
% Prison Per. Capita	1.00		
Min. Wage	0.2265	1.00	
Min. Dropout Age	-0.0257	0.1010	1.00

Collinearity Matrix

Table 1.20: Uncentered VIF's of Independent Variables in Regression of Combined DI-only and DI-SSI applications (per Adult Pop. 18-64)

Variable	VIF
Mortality Rate	3.24
% White	2.96
% HSPlus	2.27
% Age+50	2.10
% Foreign Born	2.03
% Married	1.67
Prisoner Per Capita lagged 3 years	1.41
% Overweight	1.06
Republican Governor	1.03
Mean VIF	1.97

Chapter 2

Manufacturer Bundling and Price Discrimination With Heterogeneous Retail Markets

2.1 Introduction

Manufacturers often discriminate among consumers by offering different quantities of a good at nonlinear prices. Large fries versus small fries, a season ticket versus a single game ticket, and 900 cell phone minutes per month versus 450 minutes per month are all examples of setting different quantities and prices in a package so that excess surplus can be extracted from buyers to maximize profit. The idea that a manufacturer can package his products into different quantities to increase his profit through price discrimination is long established. Also, manufacturers may combine different types of goods in a package to price discriminate. Such examples may include offering a hotel room with an airline seat; a movie ticket with a voucher for a burger; or a free photography lesson with the purchase of a camera. The idea that a multi-product manufacturer can include different types of goods in a package to facilitate discrimination is also nothing new.

The literature has given much attention to what can happen when a monopolist manufacturing one good seeks to optimize profit non-linearly while selling different quantities of the good to heterogeneous buyers. The literature also considers in detail the ramifications of nonlinear pricing by a multi-product monopolist who offers bundles of different type products to heterogeneous buyers, but often only when the bundles are constrained to those in which the number of each type of a good in a bundle has a maximum of one. There has been scant research done, however, on a manufacturer's non-linear pricing strategy when he may offer different quantities of different types of goods in a package to heterogeneous buyers. Nevertheless, such deals are often common when a manufacturer deals with retailers, who not only may buy many of each good but many of different types of goods together. For example, a supermarket could buy a package with a given quantity of ketchup bottles and a given quantity of mustard bottles from the same producer; or a sporting goods store could buy a package of many hockey sticks and many basketballs from an athletics supplier; or an auto-dealer could buy various numbers of certain types of cars from an upstream manufacturer.

Accordingly, there has also been a lack of analysis about how downstream parties in these cases may be affected when the upstream manufacturer is allowed to sell different types of goods in the same multiple-quantities package. This paper seeks to analyze the extent to which downstream parties, such as retailers and consumers are harmed, or even helped, by permitting a manufacturer to bundle different types of goods. While the fact that the upstream parties' profit is weakly improved by bundling is not a stunning result, this paper takes the focus off the manufacturer and looks at the impact downstream on welfare when the parties buying from the manufacturer have heterogeneity. The added value here is assessing which retailers and consumers are harmed or helped by bundling and how that harm or help relates to the heterogeneity of demand for the manufacturer's goods. The model explores how demand heterogeneity may lead to asymmetry to benefits and harms accrued downstream with bundling. While it is nowhere near a complete analysis, this paper will lay out some conditions under which retailers earn a profit or not and specify conditions under which some retailers may be harmed more than others. It will also analyze some relationships between the ability of the retailers to earn profit after bundling and the ability of consumers to benefit from bundling. Also, a specific example will specify conditions under which some retailer revenues improve with bundling thereby aligning these retailers' interest with that of the manufacturer but against other retailers. All of these issues could have implications for the regulation of multi-product firms or mergers in which merged firms may opt to bundle their products and resell downstream to retailers.

This paper can also contrast the effect that bundling has on welfare when an intermediate, such as a retailer, is added to the good supply chain with the case of a manufacturer selling directly to customers. Typically, the literature has examples in these latter cases in which total surplus improves as a result of bundling but customers suffer the effect of price discrimination and see welfare diminish. It is shown here that the retailer can act as a buffer between the consumer and the welfare-diminishing effects of price discrimination of bundling. Bundling weakly allows total production to go up; and since retailers seek to sell the extra goods by pricing at or below the value of the marginal consumer, some gains are passed onto the consumer even if taken from the retailer. The manufacturer and consumers' interests become aligned while that of the retailer may often suffer. Such issues could be quite important either in pre-merger investigations of mergers between firms that subsequently are likely to engage in bundling; or in rule of reason applications balancing potential welfare improvements, especially to final consumers, and the anticompetitive effects of a firm's unilateral transition to bundling.

2.2 Literature Review

The idea that a multi-product manufacturer can bundle his products to increase his profit is long established in a variety of ways. Bundling is taken here as the ability to sell multiple types of goods in potentially different quantities in a single package. Adams and Yellen (1976) analyze bundles with one of each type of good in a bundle and show that bundling can lead to welfare-enhancing or welfare-diminishing distortions provided other distortions (such as monopoly producers) are present. Schmalensee (1982) and Lewbel (1985) show specifically that mixed bundling, when the number of each type of a good in a bundle has a maximum of one, can effectively act as a mechanism for price discrimination amongst multiple buyers; this holds in the cases with independent goods and complements or substitutes, respectively. Shaffer and O'Brien (2005) take a somewhat different approach. They show that in the market for intermediate goods, where a manufacturer may sell to only a *single* retailer, bundling still allows manufacturer profit to improve when contracts are non-linear and various prices can be charged for various quantities in a package. In this case, the manufacturer gains ability to extract surplus through bundling by removing the retailer's "threat" to buy a package of just one type of good or the other. The goods in Shaffer's paper are modeled as substitutes. Without bundling, the manufacturer has an incentive to lower the quantity of each good sold to the retailer below vertically-integrated levels in order to minimize a retailer's incremental profit from purchase of each good. It is then willing to pay the two fixed fees for packages of the two separate goods that maximize profit for the manufacturer. With bundling allowed, quantities may be increased to vertically-integrated levels and fixed fees can be raised since the retailer's threat is now reduced to completely passing on a manufacturer's offer. Whinston (1990) analyzes bundling and tying contracts as a means to foreclose competition. He shows that a manufacturer's profits may increase through exclusion of rivals who produce the same goods competitively.

The focus here will not be the use of bundling to exclude rivals and maximize profit in the strand of Whinston. The approach taken in this paper will eventually perhaps be best described as a mix between the worlds of Shaffer and O'Brien and that of Schmalensee or Lewbel; this paper explores what happens when manufacturers can sell to multiple retailers and can use bundling to price discriminate between them. Unlike Schmalensee & Lewbel, however, it is not assumed that bundles are constrained to those in which the number of each type of a good in a bundle has a maximum of one. Manufacturers are able to sell packages containing any quantity of any type good they produce; if bundling is allowed the different type goods may be sold together. Thus, this paper allows for more general contracts of quantities and prices between retailer and manufacturer then they consider. Also, unlike Shaffer and O'Brien this paper does not consider the goods produced by a manufacturer to be substitutes. In addition, Shaffer and O'Brien allows for there to be some competition between the manufacturers producing the upstream substitute goods but allow the middleman retailer to be a monopolist.

This paper's focus will instead be the impact on bundling among different sets of retailers with heterogenous demands for the various goods to assess the effect of bundling combined with the heterogeneity; also assessed will be the interplay of the relative demand strength with benefits that may accrue to all three parties: monopolist manufacturer, territorial retailers with market power, and consumers.

The results sought here will add insight to the extent to which downstream parties, such as retailers and consumers are harmed, or even helped, by permitting a manufacturer to bundle his goods. This paper takes the focus off the manufacturer and looks at the impact downstream on welfare when the parties buying from the manufacturer have heterogeneity. Focusing on the manufacturer is not the interesting case as permitting bundling simply allows manufacturers to weakly increase profit, which is standard in the literature. The added value here is assessing which retailers and consumers are harmed or helped by bundling and how that harm or help relates to the heterogeneity of demand for the manufacturer's goods. Demand heterogeneity may lead to asymmetry to benefits and harms accrued dowstream with bundling; this is especially so when requiring that manufacturers must offer the same terms to all retailers as this paper does.

So, while papers have been written about bundling's ability to enhance a manufacturers's means to price discriminate there is scant literature on how downstream parties are affected by this practice. Schmalenesee and Lewbel seem to look at only certain examples for single buyers of limited bundles. As there is also scant literature on bundling in the intermediate goods market *and* bundling when non-linear contracts are allowed (with Shaffer (2005) being the only notable exception), there has been next to none formal analysis of how downstream parties are affected by bundling in these contexts. Shaffer and Kolay (2003) consider a manufacturer's ability to price discriminate among different consumers with non-linear arrangements but this paper only looks at packages of one type of good and has manufacturers directly deal with consumers. Horn and Wolinsky (1988) only look at linear pricing by upstream parties who sell directly to downstream parties and are not concerned with retailers who resell to consumers. Dobson and Waterson (1997) look only at linear pricing from manufacturer to retailer and do not consider bundling or multiple goods. They also allow the retailers to compete for the same consumers where as this paper allows retailers to have market power in a territory. Milliou and Petrakis (2007) discuss upstream mergers and multiple products but each upstream product can only be sold specifically to one downstream firm. There, a manufactured product is exclusive to a downstream firm unlike in this paper where the same downstream firms can buy the same goods. The main result in Milliou and Petrakis is upstream firms prefer to remain unmerged and act independently if bargaining is constrained to two-part tariffs rather than wholesale prices. The mergers are shown to be welfare-decreasing under these assumptions. Rennhoff and Serfes (2009) only consider bundles in which there is a maximum of one good of each type in each bundle. In industries such as cable T.V., they also look at downstream bundling directly to consumers, who buy the final good, and do not consider upstream bundling by content providers or manufacturers. They find that requirements that cable T.V. packages be "unbundled" actually enhance welfare. In contrast, this paper will look at circumstances where there can be multiple numbers of the goods in each bundle and bundling can be welfare-enhancing. Armstrong (1996) also discusses a monopolist who can produce bundles of different type goods and can price non-linearly but still constrains the bundles to ones in which the number of each type of good in a bundle has a maximum of one. Armstrong finds that the monopolist may exclude some buyers through certain pricing strategies to increase profit. However, there is no intermediate “retailer” in this paper and, accordingly, Armstrong also doesn’t consider the interplay of consumer welfare with retailer profit and doesn’t consider the various characterizations of retailer profit with and without bundling.

The goal in this paper is to create a realistic model which describes how a multi-product manufacturer may sell their goods to downstream parties, such as retailers, who may buy multiple quantities of multiple types of goods and then resell the goods to consumers. However, there is no paper in the literature which examines what happens if 1) there is an incremental change of allowing a manufacturer to include multiple types of goods in a package (or, that is, to go from prohibiting bundling to allowing it); 2) the manufacturer is free to include more than one of each type of good in a package; 3) the manufacturer may sell to heterogeneous buyers; 4) these buyers are retailers and they will take the items in their package and sell them individually to final consumers; 5) the manufacturer is able to price the packages non-linearly; 6) the manufacturer is a monopolist and the retailers have monopolistic power in a territory. While some papers look at a subset of these six conditions, I have found no paper that looks at all six. Given that these conditions seem realisitic and that there are clearly multi-product manufacturers with some market power reselling to retailers with at least some market power in a territory, it is hoped that this paper will contribute to the analysis of bundling’s effect on a market that consists of manufacturers, retailers and consumers. The focus is the interplay of some of the characteristics of each with the effects of bundling that may accrue to each.

2.3 Model

2.3.1 Modeling Assumptions

The purpose of this paper is to examine the effects of the bundling of goods by an upstream, monopolist manufacturer who sells items to downstream retailers with differences in the demand functions they face when selling the items to final consumers. The retailers can then resell the goods to consumers. The model is kept parsimonious to best isolate these effects. To model legal realities, the manufacturer, M must offer the same terms to all retailers. There are two goods in the model, a and b. M may offer "packages" containing (q_a, q_b), which are quantities of the a and b goods respectively. These packages may be sold at transfer prices, $T(q_a, q_b)$, to any retailer. In this setting, the manufacturer can not only adjust the types of goods included in a package but also the volume of each type. This gives the manufacturer additional flexibility to price discriminate, even though the same terms must be offered to all retailers. These packages can be sold at any price and the pricing format can be non-linear. In a world where bundling is illegal, packages may contain positive quantities of only one of the two goods. Multiple items of the same good can be included in the packages. For example, in the non-bundling world, packages such as (0,152), (45,0), and (1,0) are allowed. Packages such as (22, 33) or (1,1) are not allowed if bundling is illegal.

The demand functions faced by a retailer i are as follows:

$$p_{ai}(q_{ai}) = \alpha_{ai} - q_{ai} \tag{2.1}$$

$$p_{bi}(q_{bi}) = \alpha_{bi} - q_{bi} \tag{2.2}$$

where p_{ai} and p_{bi} are the prices that a retailer i can sell a good for when they are sold in quantities q_{ai} or q_{bi}, respectively. In general, the values α_{ai} and α_{bi} can take on any non-negative value but will usually not be the same. For many of the ensuing examples, notation such as α_h and α_l will be used instead where 'h' indicates a relatively high value to a lower value denoted by l.

In this paper, these demand functions are kept simple since one of the main focuses of the paper is to see how the relative strength of the demand functions for the different goods facing the different retailers affects the terms which M will offer. Heterogeneity in retailer profits and revenues can be modeled by simply adjusting the α terms. Allowing different slopes for the demand functions adds little to the model if the profits and revenues of the retailers relative to one another is the main concern. Accordingly, the above demand functions are chosen as they enable one to model sufficient heterogeneity in the demand functions for this purpose. To keep the focus on overall revenue and profits, the cost of manufacturing the goods is kept at zero. There are no shelf space or capacity issues for the retailer or manufacturer. The goods are kept independent. Whether substitutability or complementarity between the goods could affect the eventual terms offered by M is an open question, but one which might be difficult to find tractable answers to under this set-up.

To be clear, there are two types of prices here: the transfer prices $T(q_a, q_b)$ for the packages which are sold from the manufacturer to the retailer and the final prices p_{ai} and p_{bi} at which retailer i resells the goods to the final consumers. The manufacturer may sell the goods together in multiple quantities but the retailer is permitted to sell each good in individual units. Cournot pricing is assumed. The retailer is a territorial monopolist with market power in its own territory.

The order of events is assumed as follows: The manufacturer chooses the quantities of a and b in the packages and the prices of packages. He may effectively make certain packages unavailable by charging a price ∞ and can limit the retailers to only buying certain packages at certain prices. The monopolist manufacturer has most of the bargaining power here. Retailers, however, are free to buy any package which is available or buy nothing and accept a zero profit.

As a matter of semantics, it is worth emphasizing again that "packages" may include various numbers of one or both types of goods but may only have one type of good if bundling is per se illegal. "Bundled packages" or "bundles" are packages made available by M which include both types of goods. The "vertically-integrated" package is the revenue-maximizing amount. It is what the monopolists would sell if it were managing the territories rather than retailers. As the paper proceeds, it is good to keep in mind that the "vertically-integrated" quantities may actually be welfare-enhancing compared to the cases where less or even nothing is produced. One will see that the manufacturer often has an incentive to sell less than these vertically-integrated quantities to retailers in cases where bundling both is and is not allowed.

Manufacturer and retailers alike seek to maximize profit. For the manufacturer, M, this occurs by obtaining the most value for packages sold. For retailers, this occurs by maximizing the difference between product revenue and transfer prices paid to M. Retailers are allowed to dump any unsold quantity of any good at no cost. Consumer welfare is measured in the typical way which is by considering the net excess of consumer valuation over price.

2.3.2 The Single Good Case

Before proceeding further it is useful to consider what happens in this set up when only one good is produced. The results in this section are not new and have been found in previous papers like O'Brien and Shaffer (2005) but will serve to help frame the results in future sections. It is included as instructive for the reader who may not be familiar with some specific results in the price discrimination literature.

Assume there are two retailers, $R1$ and $R2$. They face the following respective demand functions for the good a:

$$p_{a1}(q_{a1}) = \alpha_h - q_{a1} \tag{2.3}$$

$$p_{a2}(q_{a2}) = \alpha_l - q_{a2} \tag{2.4}$$

where $\alpha_h \geq \alpha_l$.

When deciding what packages to offer at what prices, M has a choice. Keeping in mind it must offer the same terms to both retailers, it can offer either one package or two. If it offers only one package, then both retailers are free to buy the package although it is possible not all will do so. If the manufacturer only sells to one of the retailers, it would make the most sense if it were $R1$ since its potential revenue is higher. M could maximize the revenue extracted from $R1$ by making available a package of $\frac{\alpha_h}{2}$ of good a at a price of $\frac{\alpha_h^2}{4}$, which is the revenue $R1$ could obtain by selling this quantity.[1] This is the vertically-integrated quantity and maximizes consumer surplus for the a good as well. $R1$ would agree to buy this package as it assures a non-negative profit. $R2$ would not agree to do so as it could not sell enough items of a in its market to recoup the initial transfer price of $\frac{\alpha_h^2}{4}$.

However, M could do better by selling to both retailers. $R2$ is willing to buy q_{a2} at a price of $q_{a2}\alpha_l - q_{a2}^2$, the revenue it can make by selling this quantity. If $R1$ buys this package it can earn a profit of $(\alpha_h - \alpha_l)(q_{a2})$. $R1$ can profit off any package sold to $R2$ as long as $\alpha_h \geq \alpha_l$, which is assumed.[2]

The monopolist could either try to sell the same identical package of q_a to both or construct two different packages appealing to the different retailers in order to improve its profit. If it sells the same package to both, M will not allow $R2$ to earn any profit as M can always otherwise do better by charging $R2$, the retailer with "weaker demand", whatever revenue it can earn from selling a fixed quantity of the a good. Accordingly, in this case where only one package is offered, M will sell $R2$ it's vertically-integrated quantity at a price equal to the revenue generated by that quantity. $R1$ can then buy the same package at the same price

[1] In choosing the quantity here, M is maximizing $R1$'s revenue function which is $(\alpha_h - q_{a1})(q_{a1})$

[2] $R1$'s profit is its revenue from selling R2's package minus the transfer cost of purchasing this package. This is $(\alpha_h - q_{a2})(q_{a2}) - (\alpha_l - q_{a2})(q_{a2}) = (\alpha_h - \alpha_l)(q_{a2})$

and earn a small profit. M accordingly seeks to choose q_{a2} to maximize:

$$2(\alpha_l - q_{a2})(q_{a2}) \tag{2.5}$$

which is 2 times the revenue $R2$ can earn by selling good a. M will set this quantity to equal $\frac{\alpha_l}{2}$ and can sell this to $R2$ at $\frac{\alpha_l^2}{4}$. Since $R1$ finds it profitable to buy this package it also does so and M earns $\frac{\alpha_l^2}{2}$ in profit. Note that $\frac{\alpha_l^2}{2} \leq \frac{\alpha_h^2}{4}$ whenever $\alpha_l \leq \frac{\sqrt{2}\alpha_h}{2}$.

We can assume that when this condition holds that if M only sells one package then it will only sell to $R1$ as described previously rather than sell a single package to both $R1$ and $R2$.

However, this is a potentially unimportant point as M has the option to sell separate packages to the different retailers, can entice each to do so and can earn more profit by doing so. However, before one sees this, one needs to keep in mind that $R1$ can always earn a profit of $(\alpha_h - \alpha_l)(q_{a2})$ off any quantity q_{a2} sold to $R2$ and always has the option of buying $R2$'s package to acheive this. If M is to sell any package to $R1$, then its best option is to set a quantity which creates $R1$'s maximum revenue in its own territory but still assures $R1$ the profit it can earn by buying and reselling $R2$'s package. This means M does best by offering $R1$ a quantity of $a = \frac{\alpha_h}{2}$ at a price equal to $\frac{\alpha_h^2}{4} - (\alpha_h - \alpha_l)(q_{a2})$. However, M can still earn more profit by selling $R2$ a package of q_{a2} at a price of $(\alpha_l - q_{a2})q_{a2}$ as well. Importantly, M may reduce q_{a2} if it sufficiently reduces the profit it must assure $R1$ in buying any package. Specifically, to balance these competing concerns, M is going to choose q_{a2} to maximize it's profit, π_m:

$$\pi_m = (\alpha_l - q_{a2})q_{a2} + \frac{\alpha_h^2}{4} - (\alpha_h - \alpha_l)(q_{a2}) \tag{2.6}$$

where the first term is the revenue generated from $R2$'s package, the second term is the revenue from $R1$'s vertically-integrated quantity and the third-term is $R1$'s profit from buying $R2$'s vertically-integrated package, which must be assured to $R1$ in any deal. It is a given that q_{a1} will be $R1$'s vertically-integrated quantity of $\frac{\alpha_h}{2}$ but solving the first order condition for q_{a2} yields:

$$q_{a2} = \frac{2\alpha_l - \alpha_h}{2} \tag{2.7}$$

where it is seen that M actually finds it profitable to reduce $R2$'s quantity from its vertically-integrated level. In fact, the only time $R2$ will receive its vertically-integrated quantity here is if $\alpha_l \geq \alpha_h$, which violates a previous condition. For $R2$ to earn any positive quantity at all, it must be that $\alpha_l \geq \frac{\alpha_h}{2}$. Otherwise, M finds it better to sell only $R1$'s vertically-integrated quantity to $R1$ and extract all revenue here. If α_l is between $\frac{\alpha_h}{2}$ and α_h then putting $q_{a2} = \frac{2\alpha_l - \alpha_h}{2}$ into the profit equation reveals that M earns a profit of $\frac{\alpha_h^2 + 2\alpha_l^2 - 2\alpha_h\alpha_l}{2}$ by selling to both retailers. For any value of $\frac{\alpha_h}{2} \leq \alpha_l \leq \alpha_h$, this profit is always higher than $\frac{\alpha_h^2}{4}$, the maximum M can get by selling a just to $R1$, or $\frac{\alpha_l^2}{2}$, the maximum he can get by selling the same package of a to both retailers. Thus, provided $\alpha_l \geq \frac{\alpha_h}{2}$, M will always opt to sell two packages of a single good, each aimed at separate retailers.

As an aside, note that M can not simply offer a vertically-integrated package of the single good to $R2$ at a price equal to the revenue generated by that package and do better off than by only selling $R1$ it's vertically-integrated package. To see this, note that if M offered $R2$ a package of $\frac{\alpha_l}{2}$ of the a good at a price of $\frac{\alpha_l^2}{4}$ its maximum profit would be:

$$\pi_m = \frac{\alpha_l^2}{4} + \frac{\alpha_h^2}{4} - (\alpha_h - \alpha_l)\frac{\alpha_l}{2} \tag{2.8}$$

where the first two terms are the vertically-integrated revenues by each retailer and the last is the profit which M must assure $R1$ in order to prevent it from buying $R2$'s vertically integrated package in any deal. In total, this is equal to $\frac{\alpha_h^2 - 2\alpha_h\alpha_l + 2\alpha_l^2}{4}$. This is always less than $\frac{\alpha_h^2}{4}$, the profit simply gained by offering $R1$ its own vertically-integrated quantity of a. This shows that in order for selling the single good to both retailers to be profitable, M must reduce the quantity sold to $R2$ below vertically-integrated levels.

2.3.3 Introducing multiple goods and bundling

Now let us consider what happens when another good is introduced and the prospect of bundling is allowed. It's easy to show consumer welfare can improve with bundling. As a simple example consider that there are symmetric but opposite demand functions for two goods, a and b Here, $R1$ has a relatively higher valuation of good b and $R2$ has the relatively high valuation of good a:

$$p_{a1}(q_{a1}) = \alpha_l - q_{a1} \tag{2.9}$$

$$p_{b1}(q_{b1}) = \alpha_h - q_{b1} \tag{2.10}$$

$$p_{a2}(q_{a2}) = \alpha_h - q_{a2} \tag{2.11}$$

$$p_{b2}(q_{b2}) = \alpha_l - q_{b2} \tag{2.12}$$

where q_{ji} is the quantity of the j good in retailer i's market and is sold at price p_{ji} and where $\alpha_h \geq \alpha_l$. Prior to bundling, the markets behave as in the previous section. The retailer with the lower valuation of the good never earns a profit and purchases a quantity of that good equivalent to $\frac{2\alpha_l - \alpha_h}{2}$ provided that $\alpha_l > \frac{\alpha_h}{2}$. This is less than that retailer's vertically-integrated quantity. The retailer with a higher valuation of the good can earn a profit but will still purchase a package of the good which consists of its vertically-integrated quantity, provided that he is still assured the profit which can be earned by purchasing his counterpart retailer's package and reselling it. If $\alpha_l \leq \frac{\alpha_h}{2}$, nothing is sold to the retailer with the low valuation; the retailer with the high valuation earns no profit as he has no option to buy another package and will accept a package of his own vertically-integrated quantity at minimal profit to him.

Once bundling is introduced, M can sell each retailer their vertically-integrated quantities for both goods. M can offer packages of $(q_{a1} = \frac{\alpha_l}{2}, q_{a2} = \frac{\alpha_h}{2})$ and $(q_{b1} = \frac{\alpha_h}{2}, q_{b2} = \frac{\alpha_l}{2})$ to $R1$ and $R2$ respectively at a price equal to the revenue generated by their vertically-integrated quantities. This price is $\frac{\alpha_l^2}{4} + \frac{\alpha_h^2}{4}$. If these are the only packages made available, neither retailer will ever buy the other's package and earn a positive profit. If $R1$ buys $R2$'s package, for instance, he will sell no more than $\frac{\alpha_l}{2}$ of the a good and earn a revenue of $\frac{\alpha_l^2}{4}$ since selling more beyond this diminishes profit (as, beyond the vertically-integrated level, elasticity goes above one) and will choose to dump any excess quantity. When $R1$ sells the b good from $R2$'s package, he earns $(\frac{\alpha_l}{2})\frac{2\alpha_h - \alpha_l}{2} = \frac{2\alpha_h\alpha_l - \alpha_l^2}{4}$. The total revenue of $\frac{2\alpha_h\alpha_l}{4}$ never is larger than the cost $\frac{\alpha_h^2 + \alpha_l^2}{4}$ of buying $R2$'s vertically-integrated package. A similar argument can be made for $R2$. Each will buy their own vertically-integrated quantities at a price equivalent to the revenues generated.

Thus, bundling in this case of symmetric demand functions means a complete elimination in the profits of the retailers, if any were earned prior to bundling. However, there is some benefit to consumers in the markets with the “weaker” valuation of the goods. Prior to bundling, only below-vertically-integrated levels of these goods (a for $R1$ and b for $R2$) were available for purchase. After bundling, the higher vertically-integrated quantities are sold. There are improvements to consumer welfare as well as overall welfare. The monopolist M sees profits goes up as he can extract the complete vertically-integrated revenues in both markets with bundling.

The reduction of retailer profits by way of bundling was mentioned in Shaffer and O’Brien as was the use of bundling to achieve vertically-integrated quantities. However, in that paper it was a single monopolist retailer buying substitute goods from many upstream manufacturers. There, the retailer’s profit diminished when two manufacturers merged and could offer their goods in a bundle. The retailer lost his “threat” of buying one of the other substitute goods with bundling and suffered as a result. In the example above, bundling also eliminates the “threat” of buying some other package to earn a profit, but in this case, it is the package of another retailer with the same classes of goods not a package containing somewhat differentiated types of goods. In some ways, bundling in the above example differs from Shaffer and O’Brien in that it seems to make a greater type of goods available to more retailers even if it is diminishing their profits. So while profits diminish for retailers and consumer welfare improves through bundling in both instances, with Shaffer and O'Brien it is done by removing a single retailer's threat to buy alternative goods while here it is done by removing a threat to buy the same goods sold to other retailers. Perhaps this is a subtle difference but it shows that the power and effects of bundling could be linked just as much to the presence of other retailers as much as the presence of the different goods.

2.3.4 Retailer has threat of earning profit

The next natural question involves what happens when the demand functions are not symmetric. How is consumer welfare impacted? What happens to retailer profits? Can retailers still earn positive profit even with bundling? To start answering these questions, the demand functions are adjusted for a new example:

$$p_{a1}(q_{a1}) = \alpha_l - q_{a1} \tag{2.13}$$

$$p_{b1}(q_{b1}) = \alpha_s - q_{b1} \tag{2.14}$$

$$p_{a2}(q_{a2}) = \alpha_h - q_{a2} \tag{2.15}$$

$$p_{b2}(q_{b2}) = \alpha_l - q_{b2} \tag{2.16}$$

There are obviously numerous possibilities for strength of the parameters. Initially, let us choose $\alpha_l = \frac{\alpha_h}{2}$ and $\alpha_s > \alpha_h$. The first condition allows neither retailer to earn a profit prior to bundling as the "weaker" market is too weak for M to find it worth selling anything to and will thus choose to simply extract the maximum revenue from the stronger market. The second condition gives $R1$ a larger overall surplus than $R2$. $R1$ could in theory potentially profit from buying and reselling $R2$'s package from bundling if the strength of its strong good, evidenced by α_s is strong enough. The following example assesses this and looks at whether such a threat of $R1$ earning a profit has any effect on consumer welfare or output as well.

As with the single good case, $R2$ should never earn a profit. To see this, realize the manufacturer could offer each retailer a package at a price equivalent to the consumer revenue generated in their market by that package. A retailer's ability to earn a profit comes form threatening to buy the other's package. If both retailers are earning a profit, then M continues to increase the price at least until one of the retailers doesn't profit anymore. If $R2$ is the one earning a profit by buying the package intended for $R1$, the manufacturer can always turn around and create a case of symmetrically opposite packages and prices where $R1$ is the one who earns the profit by threatening to buy $R2$'s package. M can then do better by offering $R1$ a package containing its own vertically-integrated quantities at the same price and extract the additional surplus, created by $R1$'s relatively high demand for the b good creates; this is all done while assuring $R1$ the same profit from threatening to buy $R2$'s package.

How strong does $R1$'s demand for b need to be for $R1$ to earn a profit? At the minimum, $R1$ should be be able to profit off a package containing $R2$'s vertically-integrated quantities even if it's purchased at the price $R2$ could resell this for. (Otherwise, M could just do best by offering each retailer their vertically-integrated quantities at transfer prices equivalent to the revenue these generate in each retailer's market, and extract maximum revenue.) For this condition to hold it must be:

$$(\alpha_s - \frac{\alpha_l}{2})(\frac{\alpha_l}{2}) + (\alpha_l - \frac{\alpha_l}{2})(\frac{\alpha_l}{2}) - \frac{\alpha_l^2}{4} - \frac{\alpha_h^2}{4} \geq 0 \qquad (2.17)$$

where the first term is what $R1$ earns in revenue by selling $R2$'s vertically-integrated quantity of b, the second term is what $R1$ earns in revenue by selling $R2$'s vertically-integrated quantity of a (while dumping any excess goods past $\frac{\alpha_l}{2}$), and where the last two terms represent the cost of buying $R2$'s package, which is the revenue generated by $R2$'s vertically-integrated quantity. Remembering that $\alpha_l = \frac{\alpha_h}{2}$, the above equation is binding when $\alpha_s = \frac{5\alpha_h}{4}$.

To ensure that $R1$ can earn at least some profit when buying $R2$'s vertically-integrated bundle at the level of revenue this bundle generates, let us suppose that $\alpha_s = \frac{4\alpha_h}{3}$. Prior to bundling b is only sold to $R1$ and a is only sold to $R2$. This generates a surplus of $\frac{\alpha_s^2}{4} + \frac{\alpha_h^2}{4} = \frac{25\alpha_h^2}{36}$, which is entirely extracted by M. In any arrangement with bundling, M must earn a higher profit than this (and this is shown below.)

Before proceeding much further, it is instructive to point out that when M can bundle it has the option of selling either one type of package (which targets only the retailer with the higher surplus) or two types of packages (targeting the retailers differently), much as it did in the one-good case. If it sells only one package to one retailer, it will choose to do so with $R1$ and charge it the value of its vertically-integrated level for those same quantities. This package has a value of $\frac{1}{4}(\alpha_s^2 + \alpha_l^2) = \frac{73\alpha_h^2}{144} \leq \frac{25\alpha_h^2}{36}$ so clearly M will not offer only one package even with bundling as such an outcome can never exceed the profit from selling packages to multiple retailers without bundling. Something will be available for both $R1$ and $R2$ to purchase if bundling is to happen.

What should the two packages to the retailers look like? $R1$ should and will be assured the same profit from each package as otherwise M can continue to raise the price until $R1$ is indifferent between them. As mentioned before, M can charge $R2$ the price at least up to the revenue its package generates, as to do otherwise gives profit unnecessarily to $R2$...as well as $R1$, whose own hand would be strengthed by a more affordable package for the other retailer. If $R2$'s package will be q_{a2}, q_{b2} then M should let $q_{a2} = \frac{\alpha_h}{2}$ as this extracts maximum revenue from $R2$ but also forces $R1$ to carry the greatest excess amount of the a good while still paying for it. This latter point makes $R2$'s package less attractive to $R1$ and diminishes any profit which M must assure him, thereby increasing M's own profit. When it comes to q_{b2}, however, M may actually find it profitable to reduce this quantity below $R2$'s vertically-integrated level in order to reduce $R1$'s margin on $R2$'s package in order to entice $R1$ to buy it's own vertically-integrated package. A partial recoupment here of the larger surplus by M may offset any losses endured by reducing the quantity of q_{b2} sold to $R2$.

Accordingly, M seeks to maximize it's profit from selling the above-described package to $R2$ and a package containing $R1$'s vertically-integrated quantities at a price which still assures $R1$ the same profit it gets from buying and reslling $R2$'s package. The only choice variable of interest here is q_{b2} as the others (q_{a2}, q_{a1} and q_{b1}) should all be the vertically-integrated ones. M seeks to maximize its profit expressed as:

$$(\frac{\alpha_h^2}{4}) + (\alpha_l - q_{b2})(q_{b2}) + (\frac{\alpha_s^2}{4}) + (\frac{\alpha_l^2}{4}) - ((\frac{\alpha_l^2}{4}) + (\alpha_s - q_{b2})(q_{b2}) - (\frac{\alpha_h^2}{4}) - (\alpha_l - q_{b2})(q_{b2})) \tag{2.18}$$

Here the first two terms represent the profit sold from selling $R2$ its package at a level equal to the revenue it generates, the second two terms represent the value of $R1$'s vertically-integrated package, and the final term in parentheses (which is subtracted) represents the assured profit $R1$ gets by threatening to turn around and buy $R2$'s package, thus discounting the profit which M can make. Simple solving of the first-order condition reveals $q_{b2} = \frac{2\alpha_l - \alpha_s}{2}$, which is the same as the result in the one-good case when M limited the quantity sold to one retailer to get the other retailer to buy more. Inspection of this equation when the parameters above are inserted indicates $q_{b2} = 0$. However, unlike the one-good case, this is not acceptable or realistic as it means $R1$'s profit from buying and reselling the $R2$ package of $(0, \frac{\alpha_h}{2})$ is now negative itself.[3]

Clearly it is not reasonable to expect that $R1$ agree to a negative profit so a non-negativity profit condition must be imposed. M now seeks to maximize its profit, π_m, subject to the condition that $R1$ earns a non-negative profit. This is expressed as:

$$\pi_m = (\frac{\alpha_h^2}{4}) + (\alpha_l - q_{b2})(q_{b2}) + (\frac{\alpha_s^2}{4}) + (\frac{\alpha_l^2}{4}) - ((\frac{\alpha_l^2}{4}) + (\alpha_s - q_{b2})(q_{b2}) - (\frac{\alpha_h^2}{4}) - (\alpha_l - q_{b2})(q_{b2})) + \lambda((\alpha_s - q_{b2})q_{b2} + \frac{\alpha_l^2}{4} - \frac{\alpha_h^2}{4} - (\alpha_l - q_{b2})q_{b2}) \quad (2.19)$$

where the term following λ is the non-negative profit condition, which is the difference between the revenue $R1$ receives from buying $R2$'s package and the cost of that package to $R2$. This constraint will bind in this example and we have:

[3]In fact, it is equal to $\frac{\alpha_l^2}{4} - \frac{\alpha_h^2}{4} = \frac{-3\alpha_h^2}{4}$ as $R1$ dumps most of the a good despite paying for it.

$$q_{b2} = (\frac{\alpha_h^2}{4} - \frac{\alpha_l^2}{4})(\frac{1}{\alpha_s - \alpha_l}) \tag{2.20}$$

substituting in $\alpha_l = \frac{\alpha_h}{2}$ and $\alpha_s = \frac{4\alpha_h}{3}$, we see that $q_{b2} = \frac{9\alpha_h}{40}$. This is under $\frac{\alpha_h}{4}$ (or $\frac{\alpha_l}{2}$), which is $R2$'s vertically-integrated level of q_{b2}. $R1$ earns zero profit but its *threat* to earn a profit off the b good causes M to reduce the quantity offered from vertically-integrated levels in the *other* market. Consumer welfare in $R2$'s market improves with bundling as some of this good is now available but it doesn't improve all the way to vertically-integrated levels because of the strength of $R1$'s demand for this good.

One last thing to show in this example is that M prefers selling bundles of $(q_{a2} = \frac{\alpha_h}{2}, q_{b2} = \frac{9\alpha_h}{40})$ to $R2$ and $(q_{a1} = \frac{\alpha_l}{2}, q_{a2} = \frac{\alpha_s}{2})$ to $R1$ versus remaining in a world without bundles. M can charge $R2$ its complete revenue from selling its items. In terms of α_h this is equal to:

$$\frac{\alpha_h^2}{4} + (\frac{\alpha_h}{2} - \frac{9\alpha_h}{40})(\frac{9\alpha_h}{40}) = \frac{499\alpha_h^2}{1600} \tag{2.21}$$

From $R1$, M can receive the value of $R1$'s vertically-integrated package minus the profit it must assure $R1$ to not buy $R2$'s package:

$$\frac{\alpha_s^2}{4} + \frac{\alpha_l^2}{4} - ((\alpha_s - q_{b2})(q_{b2}) + \frac{\alpha_l^2}{4} - \frac{499\alpha_h^2}{1600}) \tag{2.22}$$

which in terms of α_h is equal to:

$$\frac{4\alpha_h^2}{9} + \frac{\alpha_h^2}{16} - ((\frac{4\alpha_h}{3} - \frac{9\alpha_h}{40})(\frac{9\alpha_h}{40}) + \frac{\alpha_h^2}{16} - \frac{499\alpha_h^2}{1600}) \tag{2.23}$$

However, unsurprisingly, this simplifies to $(\frac{4}{9} + \frac{1}{16})\alpha_h^2 = \frac{\alpha_s^2}{4} + \frac{\alpha_l^2}{4}$, which is just the revenue from $R1$'s vertically-integrated package as no profit is given to $R1$ when the constraint binds. The total profit to M from both retailers is thus:

$$\alpha_h^2(\frac{499}{1600} + \frac{4}{9} + \frac{1}{16}) = \frac{11791\alpha_h^2}{14440} \tag{2.24}$$

This is larger than the $\frac{25\alpha_h^2}{36}$ that M earned from selling the goods separately so M will choose to bundle. Accordingly the following observation can be made:

Observation 1. *When bundling under the conditions in this section, more goods will be sold in each market but the two retailers still do not earn profits. $R2$'s consumers suffer sub-vertically-integrated levels of b as a result of $R1$'s threat to earn a profit from sales of the b good to its own consumers after $R1$ has bought $R2$'s vertically-integrated package from M.*

2.3.5 Implications when a retailer can earn profit even with bundling

Let us consider some more general demand forms to assess if it is possible if one of the retailers can earn a profit with bundling and what the implications may be for consumers, manufacturer and total welfare in this case. Let the demand functions faced by the retailers now be:

$$p_{a1}(q_{a1}) = \alpha_l - q_{a1} \tag{2.25}$$

$$p_{b1}(q_{b1}) = \alpha_h - q_{b1} \tag{2.26}$$

$$p_{a2}(q_{a2}) = \alpha_m - q_{a2} \quad (2.27)$$

$$p_{b2}(q_{b2}) = \alpha_n - q_{b2} \quad (2.28)$$

where $\alpha_m > \alpha_l$ and $\alpha_h > \alpha_n$. Let the second inequality be larger so $R1$ has the larger vertically-integrated surplus and can potentially profit by buying $R2$'s vertically-integrated package. M will reduce the quantity of q_{b2} sold to $R2$ to maximize :

$$\pi_m = (\frac{\alpha_m^2}{4}) + (\alpha_n - q_{b2})(q_{b2}) + (\frac{\alpha_h^2}{4}) + (\frac{\alpha_l^2}{4}) - ((\frac{\alpha_l^2}{4}) + (\alpha_h - q_{b2})(q_{b2}) - (\frac{\alpha_m^2}{4}) - (\alpha_n - q_{b2})(q_{b2})) + \lambda((\alpha_h - q_{b2})q_{b2} + \frac{\alpha_l^2}{4} - \frac{\alpha_m^2}{4} - (\alpha_n - q_{b2})q_{b2}) \quad (2.29)$$

where the term following λ is the non-negative profit condition for $R1$. The first two terms represent the revenue from $R2$'s package, the next two terms represent the revenue from $R1$'s vertically-integrated package and the following subtracted term in parentheses represents the profit which M must assure $R1$ to get it to buy its own vertically-integrated package rather than the package sold to $R2$.

If the constraint binds then:

$$q_{b2} = (\frac{\alpha_m^2 - \alpha_l^2}{4})(\frac{1}{\alpha_h - \alpha_n}) \quad (2.30)$$

If the constraint above fails to bind, then M will continue to decrease q_{b2} until either $q_{b2} = \frac{2\alpha_n - \alpha_h}{2}$ or until no positive amount of the b good is sold to $R1$. (The latter case occurs if $\alpha_n \leq \frac{\alpha_h}{2}$.) Even if positive amount of b are sold to $R2$, the quantity $\frac{2\alpha_n - \alpha_h}{2}$ is still less than the vertically-integrated quantity of $\frac{\alpha_n}{2}$.

For $R1$ to earn a profit, the term in (28) must be larger than $\frac{2\alpha_n - \alpha_h}{2}$. The top term $\frac{\alpha_m^2 - \alpha_l^2}{4}$ is a measure of $R1$'s relative weakness in the market for its weaker good, a. The larger this difference, the more q_{b2} that M can produce and sell in a package to $R2$ as $R1$'s higher cost of carrying the a good also serves as an incentive for $R1$ to not buy $R2$'s package and dampens the need to restrict the quantity of b sold to the other market. The bottom term in (28), $\frac{1}{\alpha_h - \alpha_n}$, is an inverse measure of $R1$'s relative strength in the market for its stronger good. A stronger $R1$ which more easily earns a higher profit requires M to produce less q_{b2} in order to not make $R2$'s package so attractive that $R1$ is not enticed to buy its own vertically-integrated bundle.

For $R1$ to earn a profit and for the constraint not to bind it must be that:

$$\frac{2\alpha_n - \alpha_h}{2} > \frac{(\alpha_m^2 - \alpha_l^2)}{4(\alpha_h - \alpha_n)} \tag{2.31}$$

This occurs when:

$$6\alpha_n\alpha_h - 2\alpha_h^2 - 4\alpha_n^2 > \alpha_m^2 - \alpha_l^2. \tag{2.32}$$

or when there is a sufficiently-high α_h to α_n spread relative to a sufficiently-low α_m to α_l spread.

Assuming that $\alpha_n = x\alpha_h$, for $0 < x < 1$, the term on the left above in (30) is positive whenever $\frac{1}{2} < x < 1$. Basically for $R1$ to earn a profit, $R1$'s demand for b must be strong but not so strong that M decides it's better to sell only to $R1$ and not to $R2$. Thus, there is sort of a "Goldilocks-zone" of the parameters for which $R1$ can profit. Also, α_l must be sufficiently large as to be close to α_m. Under these conditions, $R1$ can still earn a profit with bundling. While $q_{b2} = \frac{2\alpha_n - \alpha_h}{2}$ is given to $R2$ in a package, this is no different than the quantity that would be given to it prior to bundling. $R1$'s ability to earn any positive profit even after bundling here completely wipes out any gains to consumers that may occur in the *other* market as a result of bundling.

Of course, this only happens if M can itself earn a better profit by bundling than without bundling and chooses to bundle accordingly. The below example shows that M earns a better profit with bundling for the case where $\alpha_m = \alpha_h$, $\alpha_n = x\alpha_h$, and $\alpha_l = y\alpha_h$ for $\frac{1}{2} < x < y < 1$. (Eventually, the parameters $x = \frac{3}{5}$ and $y = \frac{19}{20}$ will be used.)

Before bundling, using the parameters chosen above, M will sell positive quantities of goods to both consumers. The results are obtained from the one-good case. For good a, M earns $\frac{\alpha_m^2}{2} + \alpha_l^2 - \alpha_m\alpha_l$. For good b, M earns $\frac{\alpha_h^2}{2} + \alpha_n^2 - \alpha_h\alpha_n$.

With bundling and a quantity of $q_{b2} = \frac{2\alpha_n - \alpha_h}{2}$ sold in $R2$'s package, insertion of q_{b2} into M's profit function reveals a profit of $\frac{\alpha_m^2}{2} + \frac{\alpha_h^2}{2} - \alpha_n\alpha_h + \alpha_n^2$ for M.[4]

Comparison of the two cases reveals that M's profit with bundling is larger by $\alpha_m\alpha_l - \alpha_l^2$ than it is without bundling. Thus, M chooses to bundle. The increase in M's profit comes only from the increased sale of the a good to $R1$. This explains the presence of parameters only from the a demand functions here. Letting $\alpha_m = \alpha_h$, $\alpha_n = \frac{3\alpha_h}{5}$, and $\alpha_l = \frac{19\alpha_h}{20}$, we would see that the increase in M's profit is equivalent to $\frac{19\alpha_h^2}{400}$.

After bundling $R1$ earns a profit equivalent to the value at which in can sell $R2$'s package less the cost of purchasing this package (which is the value at which

[4]Terms with α_l drop out completely indicating the relative strength of $R1$'s "weak" good has no effect on M's profit with bundling although this term does affect the profit without bundling.

$R2$ can sell it.) This is:

$$\frac{\alpha_l^2}{4}+(\frac{2\alpha_n-\alpha_h}{2})(\alpha_h-\frac{2\alpha_n-\alpha_h}{2})-\frac{\alpha_m^2}{4}-(\frac{2\alpha_n-\alpha_h}{2})(\alpha_n-\frac{2\alpha_n-\alpha_h}{2}) \quad (2.33)$$

where the first two terms represent $R1$'s revenue from selling $R2$'s package and the second two terms, which are subtracted, represents $R2$'s revenue from the same package. This simplifies to:

$$\frac{6\alpha_n\alpha_h-2\alpha_h^2-4\alpha_n^2-\alpha_m^2+\alpha_l^2}{4} \quad (2.34)$$

which is reminisicent of the profit condition in (30). Again, letting $\alpha_m=\alpha_h$, $\alpha_n=\frac{3\alpha_h}{5}$, and $\alpha_l=\frac{19\alpha_h}{20}$, this leads to a profit of $\frac{\alpha_h^2}{64}$ for $R1$. Prior to bundling, $R1$ earned profits from the sale of good b of $(\frac{2\alpha_h-\alpha_n}{2})(\alpha_h-\alpha_n)=\frac{3\alpha_n\alpha_h-\alpha_h^2-2\alpha_n^2}{2}$, which would be $\frac{11\alpha_h^2}{50}$ for the same parameters. Thus, we see that while $R1$ earns a profit after bundling it is substantially reduced from what it was before bundling. Specifically, it decreases by $\frac{\alpha_l^2-\alpha_h^2}{4}$. $R2$'s profit of $\frac{3\alpha_l\alpha_m-\alpha_m^2-2\alpha_l^2}{2}$ from selling a prior to bundling is wiped out completely. With the chosen parameters, this was only a more modest level of $\frac{9\alpha_h^2}{400}$ considering that $R1$'s demand for a was allowed to be close to the same demand $R2$ faced; this allowed M to only have to restrict $R1$'s quantity slightly to entice $R2$ to buy its own vertically-integrated quantity of the single good a.

More generally, we could also consider the case where $R1$ has superior demand for both goods. Now, $\alpha_l>\alpha_m$ and $\alpha_h>\alpha_n$. Let $R2$'s demand parameters be at least 1/2 of those of $R1$ ensuring that some good is sold in each market prior to bundling. With bundling, M will again try to maximize its profit and do so by charging $R2$ whatever value it can get from the revenue $R2$'s package sells for in its own market but also by offering $R1$ a vertically-integrated package that assures it gets the same profit received from buying $R2$'s package. M will choose the quantities q_{a2} and q_{b2} which are sold in both markets to maximize the profit:

$$((\alpha_m - q_{a2})(q_{a2}) + (\alpha_n - q_{b2})(q_{b2}) + (\frac{\alpha_l^2}{4} + \frac{\alpha_h^2}{4}) - ((\alpha_l - \alpha_m)(q_{a2}) + (\alpha_h - \alpha_n)(q_{b2})) \quad (2.35)$$

where the first two terms are the value of $R2$'s package to $R2$, the second two terms are the value of $R1$'s vertically-integrated package, and the final terms in the subtracted set of parentheses represent the profit $R1$ can make by buying $R2$'s package. Solving of the first order conditions simply reveal that $q_{a2} = \frac{2\alpha_l - \alpha_m}{2}$ and $q_{b2} = \frac{2\alpha_h - \alpha_n}{2}$. These quantities are the same as when goods are sold separately so bundling allows no increase in the final quantities or consumer welfare if the retailer has a stronger demand for both goods. This is not surprising as any increase in quantities by way of bundling came through M's increased power to turn a retailer away from the threat of buying another retailer's goods by forcing the retailer earning a profit to buy excess undesirable goods of another type in a "bundle". Since $R1$ was already happily buying any quantities of either good that $R2$ wished to buy, this power is not gained here and so the equilibrium quantities remain unchanged.

Accordingly, to sum up this section with an observation:

Observation 2. *Improvements in consumer welfare by way of bundling (in this world with two territorial monopolistic retailers buying nonlinear packages from one monopolistic manufacturer) seem tied to retailer profit. Specifically, these* *improvements are related to the ability of the retailer in the other market to profit by buying his counterpart retailer's goods. The stronger the threat of profiting off another retailer's vertically-integrated package of either good, the less the benefit derived by consumers from bundling in the other market. If the retailer's threat of profiting off one of the goods is so strong that he can still earn an overall profit in the case of bundling, consumer welfare and quantity gains in that good are eliminated in the other market. If the retailer has stronger demand for both goods, consumers in all markets generally do no better with bundling than without although it is still possible they can purchase some level of sub-vertical quantities in the weaker market. This shows there is a relative relationship between the demand functions retailers may face and benefits which may accrue to consumers. Strength in one market can cause repercussions elsewhere.*

2.3.6 Three retailer example

One can wonder what happens if there are more territories with more retailers. Figuring out the different bargains which M may offer the various retailers with different demand functions can be complex as M must take into account whether retailers will buy other retailers' packages. There are also many possible forms of the heterogeneity in the demand functions and establishing general or, even tractable, results may be difficult under these circumstances. However, to show some possibility of what may happen with three retailers, I establish the following example:

I consider what happens if there are three retailers. I still assume retailers are local monopolies who face idiosyncratic demand functions for two different goods a and b. The goods are neither complements nor substitutes at this point. Consider them again to be two unrelated goods. The goods are manufactured by a single monopolist, who sells to all retailers. (Index the three retailers R_i with $i = 1, 2, 3$.) The manufacturer can sell to retailer i packages of the goods (q_{ai}, q_{bi}) at a price $T(q_{ai}, q_{bi})$. The manufacturer M can set the price T of the packages at any value; there is no linear or other pricing constraint. If there is no "bundling" allowed, the manufacturer is constrained to offer only packages of types $(q_{ai}, 0)$ or $(0, q_{bi})$. That is, any positive quantity of a type of good offered in one package must have zero of the other type of good in the same package. When bundling is introduced, the quantities of both goods in the package can take on non-zero numbers.

M will choose packages and prices that maximize total profit. M can make certain packages unavailable by effectively charging $T = \infty$ for them. Assume cost of production is zero for now. M has knowledge of the demand functions that retailers face for the two goods. They are:

$$p_{a1}(q_{a1}) = \alpha_l - q_{a1} \tag{2.36}$$

$$p_{b1}(q_{b1}) = \alpha_h - q_{b1} \tag{2.37}$$

$$p_{a2}(q_{a2}) = \alpha_h - q_{a2} \tag{2.38}$$

$$p_{b2}(q_{b2}) = \alpha_l - q_{b2} \tag{2.39}$$

$$p_{a3}(q_{a3}) = \alpha_m - q_{a3} \tag{2.40}$$

$$p_{b3}(q_{b3}) = \alpha_m - q_{b3} \tag{2.41}$$

In this example, let $\alpha_l = \alpha_h/2$. (The demand functions above allow for some convenient symmetry.) Let α_m, which is less than α_h, be large enough so that R_3 would opt to buy the vertically-integrated bundle (or the bundle which extracts the maximum of R_1 or R_2's profits) that is sold to either R_1 or R_2. Such a bundle would contain $\alpha_h/2$ of one good and $\alpha_l/2$ of the other. The value to R_1 or R_2 of this bundle would be $\frac{\alpha_h^2+\alpha_l^2}{4}$. If the manufacturer seeks to maximize surplus extracted from R_1 or R_2, then it charges this price for the bundle. R_3's valuation[5] of either of the others' bundles is $\frac{\alpha_m^2+2\alpha_m\alpha_l-\alpha_l^2}{4}$. $R3$ will buy the vertically-integrated package of the other two retailers if its valuation of their package exceeds their valuation of it or if:

$$\alpha_m^2 + \alpha_m\alpha_h - 3/2\alpha_h^2 > 0 \tag{2.42}$$

This is true if $\alpha_m \geq \frac{\alpha_h(-1+\sqrt{7})}{2} \approx 0.823\alpha_h$. Assume this condition holds.

Here, R_3 would like to buy R_1 or R_2's bundle unless the manufacturer restricts the quantities in these bundles to make it less favorable to R_3. The manufacturer may find it worthwhile to do this if the lost revenue from restricting the quantites in packages sold to R_1 or R_2 can be gained back by having to assure R_3 less of a profit when it presents R_3 with its vertically integrated bundle of $(\alpha_m/2, \alpha_m/2)$. The manufacturer can offer R_3 a package $(\alpha_m/2, \alpha_m/2)$ at a price equal to the revenue generated by R_3's sales of the items in this package minus the profit R_3 receives from buying one of the other two bundles. (R_3 would be willing to buy this package.) In choosing the quantities of the low and high valuation goods to

[5]These revenues are obtained by putting in quantities of $\frac{\alpha_m}{2}$ for one good and $\frac{\alpha_l}{2}$ for the other into the revenue functions. Remember $R3$ dumps any quantity past $\frac{\alpha_m}{2}$ so even though it buys a package which includes $\frac{\alpha_h}{2}$ of a other good, it will not sell this quantity.

sell to R_1 and R_2, M seeks to maximize:

$$2(q_h(\alpha_h - q_h) + q_l(\alpha_l - q_l)) + (\alpha_m^2/4 + \alpha_m^2/4) - (\alpha_m^2/4 + q_l(\alpha_m - q_l) - (q_h(\alpha_h - q_h) + q_l(\alpha_l - q_l))) \quad (2.43)$$

with respect to q_h and q_l. Here q_h is the quantity of the high valued good sold to each retailer and q_l is the quantity of the low valued good sold to each retailer. The first term in the parentheses (following the "2") is revenue that could be generated from a bundle sold to either R_1 or R_2. M will charge these retailers this amount for this bundle since it has nothing to lose by not doing so. The second term in parentheses is the revenue generated by R_3's vertically integrated bundle. The third term in parentheses is profit R_3 can generate by purchasing and reselling either R_1 or R_2's bundle. The above equation makes use of the reverse symmetry of the demand functions for R_1 and R_2. The manufacturer has an incentive to lower q_l simultaneously on each bundle for R_1 and R_2. If this were not the case, R_3 could simply opt to buy the package with the higher q_l to earn the largest profit. The manufacturer must transform each of R_1 and R_2's bundles into an equally less profitable option in order to have any deterrence on R_3's "threat" to buy the other retailers' bundles. Above, the manufacturer will earn 2 times the revenue from one of the symmetric bundles sold to R_1 or R_2 plus additional profit gained by selling R_3 his producer surplus maximizing bundle $(\alpha_m/2, \alpha_m/2)$ at a price that assures him the profit from buying one of the other bundles. The term $\alpha_m^2/4$ is present as R_3 never sells more than its producer surplus maximizing quantity.

The f.o.c.'s here yield $q_h = \frac{\alpha_h}{2}$ and $q_l = \frac{3\alpha_l - \alpha_m}{4}$. Note that the quantity of q_l is higher than in the two retailer case. R_3 is able to play off the two other retailers better than if just one other retailer were present; this is because the manufacturer can earn more profit from two retailers with low valuations of the different goods than if there were just one such retailer. This lessens the manufacturer's necessity to limit the amounts of the low valuation goods it sells in the packages to R_1 or R_2 in order to reduce the profit it must assure R_3. Here R_3's revenue[6] from buying either the package of R_1 or R_2 is $\frac{18\alpha_m\alpha_l - 9\alpha_l^2 - \alpha_m^2}{16}$. The revenue to R_1 or R_2 is $\frac{4\alpha_H^2 + 3\alpha_L^2 + 2\alpha_m\alpha_l - \alpha_m^2}{16}$. The former exceeds the latter if $\alpha_m \geq \frac{7\alpha_h}{8}$. This is the condition for R_3 making a positive profit on buying the bundle of either R_1 or R_2. Assume this holds.

Before bundling was allowed, the manufacturer had to sell the goods separately. Consider how good b should be priced in this case. M could choose to sell only to R_1, who had the highest value. He could have chosen to sell to R_1 and R_3, who had the two highest valuations. Or he could have sold to all of R_1, R_2 and R_3. If he sold to R_2, who had the lowest value, he would seek to need to assure the other retailers the same profit as if they bought R_2's package in order to get them to buy any other quantity. Because the other retailers have such a large potential producer surplus relative to R_2 it can be shown he chooses not to sell to R_2 as it would cut into his extraction of these other surpluses (since R_1 and R_3 would just go ahead and buy R_2's package instead). Specifically, if M opted to sell to R_2, he would seek to maximize:

$$3(q_{b2}(\alpha_l - q_{b2}) + (\alpha_h^2/4 + \alpha_m^2/4) - (q_{b2}(\alpha_h - q_{b2}) + q_{b2}(\alpha_m - q_{b2})) \qquad (2.44)$$

[6] These revenues are obtained by putting in quantities of $\frac{\alpha_m}{2}$ for one good and $\frac{3\alpha_l - \alpha_m}{4}$ for the other into the revenue functions. Remember that $R3$ dumps any quantity past $\frac{\alpha_m}{2}$ again.

with respect to q_{b2}. The first term in parentheses (after the "3")is revenue that can be generated and extracted from R_2. The second term in parentheses reflects revenue generated by selling vertically integrated quantities to R_1 and R_3. The last term in parentheses is revenue that R_1 and R_3 can get by selling R_2's quantity. (The "3" simplifies the fact that $R1$ and $R3$ also want to buy $R2$'s bundle and are willing to do so.) The equation takes into account the revenue that can be gained by selling to R_2 versus assured profit to the other retailers. First order conditions show that $q_{b2} = \frac{3\alpha_l - \alpha_h - \alpha_m}{2}$. If $\alpha_m + \alpha_h$ is greater than $3\alpha_l$, then M does best by not choosing to sell any positive quantity to R_2. This condition would hold for the cases above in which R_3 earns a profit in the event of bundling, where $\alpha_m \geq \frac{7\alpha_h}{8}$.

At this point, with $R2$ receiving nothing, we are back to the two-retailer case without bundling. Since $\alpha_m > \frac{\alpha_h}{2}$, M won't sell only to $R1$ but rather to both $R1$ and $R3$. To maximize profit for the market for good b, M only sells to R_1 and R_3. Now the manufacturer must consider what happens if R_1 chooses to buy R_3's quantity of b. As in the previous cases, the manufacturer will seek to sell R_3 some quantity q_{b3} of b which is less than R_3's vertically integrated quantity to limit the profit R_1 can earn. Specifically, M seeks to maximize:

$$2(q_{b3}(\alpha_m - q_{b3})) + (\alpha_h^2/4) - q_{b3}(\alpha_h - q_{b3}) \tag{2.45}$$

with respect to q_{b3}, the quantity of b offered to R_3. Here, $q_{b3} = \frac{2\alpha_m - \alpha_h}{2}$, which is a familiar quantity. It is sold to R_3 at the price equal to the revenue that quantity generates for him. M eventually offers R_1 his vertically integrated quantity of $\frac{\alpha_h}{2}$ for b at a price equal to the revenue that quantity generates for him minus the profit that he can earn by buying R_3's package of b.

Since $\alpha_m > \alpha_h/2$ a positive quantity will still be sold to R_3 and some profit assured to R_1. However, R_3's revenue is completely extracted by M in the b market. M simply charges R_3 the revenue that his quantity of b generates. Since the same story can be told for R_3 and R_2 in the a market, we conclude that R_3 earns no profit prior to bundling. Yet, after bundling is introduced, R_3 is the *only* retailer that earns a profit. In this example, the retailers which earned a profit prior to bundling no longer do so afterwards; whereas the retailer who earns nothing before bundling, earns something after bundling.[7]

This section can be summarized with the following observation:

Observation 3. *In this example with three retailers, the retailers which earn or do not earn profit can flip with bundling. This is another example of how bundling can not just affect manufacturer, product output, and consumer welfare but can also have direct effects on the profitability of the retailer. A single retailers' interest in bundling could be aligned with that of the manufacturer and consumers but against those of other retailers in certain cases. Consumer welfare improves (weakly) through bundling, as usual with the two retailer case. (This can be seen by simply noting that consumers in the low valuation markets are now able to buy some positive amount of these goods with bundling and consumers in R_3's market also get to buy more of both of these goods with bundling.) Vertically-integrated quantities still aren't obtained for the lower valued goods in R_1 and R_2 due to the desire of the manufacturer to protect itself against R_3's threat to buy these retailer's packages yet the quantities in these markets are higher nonetheless than before bundling. Overall, social welfare increases through bundling here; consumers, manufacturers*

[7] This all assumes of course that M earns more bundling than not bundling and thus willingly bundles. This is shown in the appendix and is actually true for all positive α_m in the range of $(0, \alpha_h)$

and R_3 benefit. R_1 and R_2 see profits diminish.

2.4 Conclusion

Given all that has been said in this paper what are the main things to take away about manufacturer bundling in the face of retailer heterogeneity? First, it seems that the presence of the retailers and the ability of the manufacturer to offer flexible non-linear packages when bundling weakly allows more product to be produced and sold. Bundling allows manufacturers to have more control in extracting surplus and offering packages which approach vertically-integrated quantities by making certain packages tailored to one retailer less attractive to another. The reduction in alternatives for the retailers gives the manufacturer greater bargaining power. In Shaffer and O'Brien, this was done by forcing the single retailer to buy extra goods. Here, the addition of extra retailers seems to strengthen this power. As vertically-integrated quantities are approached, consumers do benefit as, prior to bundling, manufacturers may have been restricting the quantities offered to one retailer in order to assure less profit to another, especially in a world where all packages are available to all retailers. The introduction of bundling of an unprofitable good allows another tool apart from quantity restriction to extract surplus from retailers by forcing them to buy other packages.

Interestingly, here a retailer's ability to earn a profit in a two-good, two-retailer case seems tied to the ability of bundling to improve consumer welfare in a different market when the same terms of price and quantity must be offered to all retailers. If a retailer can still profit after being forced to buy some other, less-profitable goods when attempting to buy another retailer's package, the manufacturer will revert to restricting the quantity of the profitable good as it did with the one-good case prior to bundling. This is because the higher margin with which this retailer can sell the profitable good in a retailer's package causes him to need an assured profit in any deal and, as this cuts into the producer's own profit, this may lead to quantity restrictions in packages sold to the other retailer. The parameters for which this is possible seems to require that the "unprofitable" good be not too unprofitable and the "profitable" good have a moderate margin. It was shown that too high a margin on the "profitable" good will prevent it from being offered to a competitor. This eliminates a retailer's threat of making a profit by buying the other's package. However, too low a margin may not be enough to overcome the unprofitability of a bundled good. The interplay of the markets for the two goods was examined using crude demand functions but it is believed that more complex demand functions would not model the relative strengths of the demands much differently; this seems especially true since the monopolist manufacturer is pricing non-linearly and can pick the optimal quantity for itself with any demand function.

The case with three retailers shows, somewhat surprisingly perhaps, that some retailers may do better with bundling even when the bundle contains excess goods which are unprofitable and may be dumped. This seems to happen for a certain set of parameters where one retailer finds it again not-too unprofitable to carry the undesireable good (where $\alpha_m = \frac{7\alpha_h}{8}$) and substantially profitable to carry the desirable one (where $\alpha_m = \frac{7\alpha_l}{4}$). The fact that this effect is so strong that some retailers could go from having profits before bundling to losing them with bundling and that another could earn no profit without bundling but earn one with it was unexpected and surprising. This may have anti-trust implications. If retailers in certain territories can profit with bundling or with mergers that result in bundling, they may have reason to pool their legal resources with upstream parties in insisting on such practices.

The way in which bundling may affect the fortunes of retailers and consumers is complex and this paper only looks at a subset of cases to determine some possible effects. Including more goods or more retailers, allowing the retailers to directly compete, enabling manufacturers to have competition, looking at firm entry and exit, and including the possibility of goods as substitutes or complements would enhance the model but also potentially greatly complicate it. In offering deals to the retailers, the manufacturer in this paper needed to consider how other retailers would respond under the realisitic legal assumption that the same terms must be offered to each party. Complicating the model further makes consideration of these terms perhaps untractable. This paper at least ponders the case of bundling when there is some retailer heterogeneity without the added complexities that may alter, enhance or muddle the results. While it is not the last word on bundling, it can add to what is already understood about it.

2.5 References

- 1. Adams, W.J. and Yellen, J.L., 1976, “Commodity Bundling and the Burden of Monopoly. "*Quarterly Journal of Economics* 90 (Aug), 475-498.
- 2. Armstrong, M., 1996,“Multiproduct nonlinear pricing." *Econometrica* vol. 64, No. 1, pp. 51-75

- 3. Dobson, P. and Waterson, M. “Countervailing Power and Consumer Prices." *Economic Journal*, Vol. 107 (1997), pp. 418-430.
- 4. Horn, H. and Wolinsky, A. “Bilateral Monopolies and Incentives for Mergers," *Rand Journal of Economics*, Vol. 19 (1988), pp. 408-419.
- 5. Lewbel, A., 1985, “Bundling of Substitutes or Complements" *International Journal of Industrial Organization* 3 (1985), 101-107.
- 6. Kolay, S. and Shaffer, G. 2003, “Bundling and menus of two-part tariffs." *The Journal of Industrial Economics* 51, September 383.
- 7. Milliou C. and Petrakis, E. 2007 “Upstream horizontal mergers, vertical contracts, and bargaining" *International Journal of Industrial Organization* Volume 25, Issue 5, October, Pages 963-987
- 8. O’Brien, DP and Shaffer, G., 2005. “Bargaining, Bundling and Clout: the Portfolio Effects of Horizontal Mergers" *RAND Journal of Economics* 36 (3), 573-595
- 9. Rennhoff, A. and Serfes, K., 2009, “The role of upstream-downstream competition on bundling decisions: Should regulators force firms to unbundle?", *Journal of Economics and Management Strategy* 18, 547-588
- 10. Schmalensee, R., 1982, “Commodity bundling by single product monopolies." *Journal of Law and Economics* 25, April 67-71.
- 11. Shaffer, G. 1991,“Capturing Strategic Rent: Full-Line Forcing, Brand Discounts, Aggregate Rebates, and Maximum Retail Price Maintenance." *The Journal of Industrial Economics* 39, September 557.
- 12. Werden, G. and Froeb, L. 2007, “Unilateral Competitive Effects on Horizontal Mergers." *Handbook of Antitrust Economics* 43-104.
- 13. Whinston, M. “Tying, Foreclosure, and Exclusion." *American Economic Review*, Vol. 80 (1990), pp. 837-859.

2.6 Appendix

2.6.1 Proof Manufacturer Earns More Profit with Bundling in the Three-Retailer example

Without bundling, M sells each good to the two retailers with the two highest valuations of the product. For good b, these are $R1$ and $R3$. Using results from section 3.2 when there is just a single good sold to two retailers, we see that M’s profit from selling separate packages of b to $R1$ and $R3$ is:

$$\frac{\alpha_h^2 - 2\alpha_h\alpha_m + 2\alpha_m^2}{2} \tag{2.46}$$

Since there is also the a good, and as $R2$'s demand for a is the same as $R1$'s demand for b, M can earn an identical profit as above by selling a to both $R2$ and $R3$. Thus, without bundling, M's total profit is double that of the previous equation:

$$\frac{2\alpha_h^2 - 4\alpha_h\alpha_m + 4\alpha_m^2}{2} \tag{2.47}$$

Using the value $\alpha_m = \frac{7\alpha_h}{8}$, we have a total profit for M of $\frac{25\alpha_h^2}{32}$ when there is no bundling.

To determine M's profit with bundling, we use equation (2.43) and enter the profit maximizing quantities found in section 3.6. These are $q_h = \frac{\alpha_h}{2}$ and $q_l = \frac{3\alpha_l - \alpha_m}{4}$. Inserting these into (41) gives M a profit of:

$$\frac{3(4\alpha_h^2 + 3\alpha_l^2 + 2\alpha_m\alpha_l - \alpha_m^2)}{16} + \frac{\alpha_m^2}{4} - \frac{18\alpha_m\alpha_l - 5\alpha_m^2 - 9\alpha_l^2}{16} \tag{2.48}$$

Remembering again that $\alpha_l = \frac{\alpha_h}{2}$ and $\alpha_m = \frac{7\alpha_h}{8}$, this simplifies to

$$\frac{66\alpha_h^2 - 24\alpha_m\alpha_h + 24\alpha_m^2}{64} = \frac{4056\alpha_h^2}{4096} = \frac{507\alpha_h^2}{512} \tag{2.49}$$

As this exceeds $\frac{25\alpha_h^2}{32}$, M chooses to bundle in the example in section 3.6 as it is more profitable.