

Manoli, Dayanand; Weber, Andrea

Working Paper

Intertemporal Substitution in Labor Force Participation: Evidence from Policy Discontinuities

NRN Working Paper, NRN: The Austrian Center for Labor Economics and the Analysis of the Welfare State, No. 1011

Provided in Cooperation with:

NRN: The Austrian Center for Labor Economics and the Analysis of the Welfare State, Johannes Kepler University Linz

Suggested Citation: Manoli, Dayanand; Weber, Andrea (2010) : Intertemporal Substitution in Labor Force Participation: Evidence from Policy Discontinuities, NRN Working Paper, NRN: The Austrian Center for Labor Economics and the Analysis of the Welfare State, No. 1011, Johannes Kepler University Linz, NRN - The Austrian Center for Labor Economics and the Analysis of the Welfare State, Linz

This Version is available at:

<https://hdl.handle.net/10419/115053>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**Intertemporal Substitution in Labor Force Participation:
Evidence from Policy Discontinuities**

by

Dayanand MANOLI
Andrea WEBER

Working Paper No. 1011

October 2010

Supported by the
Austrian Science Funds

FWF

**The Austrian Center for Labor
Economics and the Analysis of
the Welfare State**

JKU Linz
Department of Economics
Altenberger Strasse 69
4040 Linz, Austria
www.laborrn.at

Corresponding author: a.weber@uni-mannheim.de
phone +49 (0621) 181-1928

Intertemporal Substitution in Labor Force Participation: Evidence from Policy Discontinuities*

Dayanand Manoli
UCLA & NBER

Andrea Weber
University of Mannheim

October 11, 2010

Abstract

This paper presents new empirical evidence on intertemporal labor supply elasticities. We use administrative data on the census of private sector employees in Austria and variation from mandated discontinuous changes in retirement benefits from the Austrian pension system. We first present graphical evidence documenting delays in retirement in response to the policy discontinuities. Next, based on the empirical evidence, we develop a model of career length decisions. Using an estimator that exploits the graphical evidence, we estimate an intertemporal labor supply elasticity of 0.30; this relatively low estimate reflects that the disutility of labor supply rises relatively quickly with additional years of work.

*We are grateful to Joe Altonji, David Card, Raj Chetty, Julie Cullen, Patrick Kline, Kathleen Mullen, Jesse Rothstein and numerous seminar and conference participants for insightful comments and suggestions. We are also grateful for financial support from the Steven H. Sandell Grant Program. Andrea Weber acknowledges research funding by the Leibniz Association (Pakt für Forschung und Innovation) and the Austrian Science Fund (NRN Labor Economics and the Welfare State).

1 Introduction

The theory of intertemporal labor supply is the workhorse theory of dynamic labor supply decisions in economics. In numerous applications of this theory in macroeconomics, labor economics and public economics, the intertemporal labor supply elasticity plays a central role in understanding business cycle fluctuations, lifetime devoted to market work, and responses to income tax and transfer programs. Despite its importance in macroeconomic and microeconomic models, there is a wide-spread debate regarding the magnitude of the intertemporal labor supply elasticity, with higher and lower elasticities having vastly different policy implications.

In this study, we provide new empirical evidence on intertemporal labor supply elasticities in a life-cycle setting using responses to policy discontinuities in retirement benefits in Austria. We first present nonparametric graphical evidence documenting individuals' labor supply responses to the policy discontinuities. Next, we develop a strategy to estimate the intertemporal labor supply elasticity. The strategy exploits the observed labor supply responses. Based on the observed patterns in individuals' retirement decisions, we estimate an intertemporal labor supply elasticity of 0.30; this estimate reflects that the disutility of labor supply increases relatively quickly with more years of work.

There has been significant research on intertemporal labor supply elasticities yielding a wide range of values. Specifically, macroeconomic models explaining aggregate labor supply responses assume relatively high elasticities, while estimates based on micro data typically find small labor supply elasticities.¹ Recent efforts to reconcile higher and lower elasticities have emphasized the importance of distinguishing between the intensive and extensive margins in labor supply decisions.² Intuitively, small labor supply responses on the intensive margin reflecting hours of work decisions may well be compatible with large responses at the extensive margin reflecting career length or participation decisions. As most previous studies examining individual-level labor supply have focused on intensive margin decisions, the responsiveness in labor supply along the extensive margin in micro data has been identified as a key issue.³ In this study, we are able to estimate an intertemporal labor

¹For microeconomic evidence on intertemporal substitution in labor supply, see MaCurdy (1981), Browning, Deaton and Irish (1985), Altonji (1986), Card (1994) and the survey discussions in Blundell and MaCurdy (1999) and Browning, Hansen, and Heckman (1999). For macroeconomic evidence, see Mulligan (1999), Ljungqvist et al (2006), Ohanian et al (2008), Rogerson and Wallenius (2009), Ljungqvist and Sargent (2010) and the survey discussions in Prescott (2006) and Keane and Rogerson (2010).

²Other efforts to reconcile higher and lower elasticities have focused on human capital (see Imai and Keane (2009)) and adjustment costs (see Chetty (2009)).

³Heckman and MaCurdy (1980), Heckman (1993), Blundell and MaCurdy (1999) and Browning,

supply elasticity while focusing explicitly on an extensive margin decision. In particular, we estimate the extensive margin Frisch elasticity, or more intuitively, the elasticity of career length with respect to anticipated wages.

The policy discontinuities exploited in this study arise because a lump-sum component of retirement benefits in Austria increases discontinuously once individuals complete specific threshold amounts of tenure prior to their retirements. These benefits are fully anticipated by the workers and thus incorporated into lifetime wealth. This allows us to focus on marginal-utility-of-wealth-constant labor supply responses. While the lump-sum benefits increase discontinuously by a considerable amount (about 30% of an annual salary), they are small relative to lifetime income. We therefore assume that income effects from the discontinuous benefit increases are negligible so that the labor supply responses to the severance payments are only delays in individuals' retirement decisions.

We examine behavior before and after multiple tenure thresholds to determine if individuals extend their careers in response to the anticipated discontinuous increases in benefits. Graphical evidence based on a large sample of individual retirements from administrative records indicates reduced numbers of retirements just prior to the thresholds and excess numbers of retirements just after the thresholds. The empirical analysis provides clear evidence on the nature of labor supply decisions in the face of retirement benefits. Specifically, we can identify how long individuals are willing to delay retirement to become eligible for benefits. Further, heterogeneity analysis allows us to distinguish between individuals who are able to respond to the benefit incentive and others who are constrained by health or job related problems.

Motivated by the empirical analysis, we build a labor supply model that generates the retirement patterns observed in the data. In particular, the model captures the reduced retirements prior to the tenure thresholds and the excess retirements just after the tenure thresholds. We then develop a strategy to estimate the extensive margin intertemporal labor supply elasticity based on relating the observed retirement patterns to the model's predicted patterns. More specifically, the estimation strategy is based on matching the observed maximum length of time that individuals delay their retirements in response to the severance payments to the model's corresponding predicted maximum delay time. While this estimator relies on discontinuities in individuals' budget constraints, it is similar in spirit to previous bunching estimators that exploit kinks in individuals' budget constraints

Hansen, and Heckman (1999) also have emphasized the distinction between the intensive and extensive margins in labor supply decisions.

(see Saez (1999, 2009) and Chetty et al (2010)). Furthermore, we highlight that the estimation strategy exploits exogenous variation in individuals' budget sets coming from the policy discontinuities and allows for estimation of a policy-relevant structural parameter without requiring ad hoc distributional assumptions.⁴

Overall, the empirical analysis yields the following results. The graphical analysis indicates that individuals do not delay their retirements by much time in response to the severance payments. We estimate a maximum delay time of 1.25 years, and this drives a relatively low estimate for the intertemporal labor supply elasticity of 0.30. These results highlight that the disutility of labor supply rises relatively quickly with additional years of work. Thus, it is difficult to reconcile larger macro intertemporal labor supply elasticities with smaller microeconomic estimates based solely on the differences between intensive and extensive margin labor supply decisions. Larger macro elasticities may be driven by other factors such as adjustment costs or search frictions (see Chetty et al (2010) and Hall (2006) respectively).

This paper is organized as follows. Section 2 discusses both the institutional background regarding the Austrian pension system and the administrative data from the Austrian Social Security Database. Section 3 presents a nonparametric graphical analysis of the data. Section 4 develops an intertemporal labor supply model based on the empirical evidence presented in section 3. Section 5 develops the elasticity estimation strategy and then presents the estimation results and sensitivity analysis. Section 6 concludes.

2 Institutional Background & Data

2.1 Retirement Benefits in Austria

There are two forms of government-mandated retirement benefits in Austria: (1) government-provided pension benefits and (2) employer-provided severance payments. We start with the description of severance payments since these payments are the primary focus of the current study. The employer-provided severance payments are made to private sector employees who have accumulated sufficient years of tenure by the time of their retirement. Tenure is defined as uninterrupted employment time with a given employer and retirement is based on claiming a government-provided pension. The payments must be made within

⁴It is possible to estimate alternative dynamic structural models, though we lack data on individuals' consumption and savings decisions. We leave these considerations for future work.

4 weeks of claiming a pension according to the following schedule. If an employee has accumulated at least 10 years of tenure with her employer by the time of retirement, the employer must pay one third of the worker's last year's salary. This fraction increases from one third to one half, three quarters and one at 15, 20 and 25 years of tenure respectively. This schedule for the severance payments is illustrated in Figure 1. The payments are made in lump-sum and, since payments are based on an employee's salary, overtime compensation and other non-salary payments are not included when determining the amounts of the payments. Provisions to make these payments come from funds that employers are mandated to hold based on the total number of employees. Severance payments are also made to individuals who are involuntarily separated (i.e. laid off) from their firms if the individuals have accumulated sufficient years of tenure prior to the separation. The only voluntary separation that leads to a severance payment, however, is retirement. Employment protection rules hinder firms from strategically laying off workers to avoid severance payments and there is no evidence on an increased frequency of layoffs before the severance pay thresholds.⁵ In general, older workers approaching retirement age enjoy the highest level of job protection in Austria.

The Austrian income tax system, which is based on individual taxation, applies particular rules to tax income from severance payments. Specifically, all mandated severance payments are exempt from social security contributions and subject to a tax rate of 6%. The income taxation of the severance payments differs from the general income tax rules. Generally, gross monthly earnings net of social security contributions⁶ are subject to the income tax with marginal tax rates in the different tax brackets of 0%, 21%, 31% 41% and 50%.^{7 8}

Because the timing of the severance payments relates to pension claiming, eligibility for government-provided retirement pensions interacts with the severance payment system. Austria has a public pension system that automatically enrolls every person employed in the private sector. Fixed pension contributions are withheld from each individual's wage

⁵For more details regarding the severance payments at times of unemployment, see Card, Chetty and Weber (2007).

⁶Contributions for pension, health, unemployment, and accident insurance of 39% are split in half between employer and employee and the employee's share is withheld from gross annual earnings up to a contribution cap.

⁷These tax brackets are based on legislation in 2002; there have subsequently been relatively small changes due to several small tax reforms.

⁸Additionally, Austrian employees are typically paid 13th and 14th monthly wage payments in June and December. These payments, up to an amount of one sixth of annual wage income, are also subject to a 6% tax rate; amounts in excess of one sixth of annual income are subject to the regular income tax rates.

and annuitized benefits during retirement are then based on prior contributions (earnings histories). Replacement rates from the annual payments are roughly 75% of pre-retirement earnings and there are no actuarial adjustments for delaying retirement to a later age. Individuals can retire by claiming Disability pensions, Early Retirement pensions and Old Age pensions. Eligibility for each of these pensions depends on an individual's age and gender, as well as having a sufficient number of contribution years. Beginning at age 55, private sector male and female employees can retire by claiming Disability pensions, where disability is based on reduced working capacity of 50% relative to someone of a similar educational background. At age 55, women also become eligible to claim Early Retirement pensions, but the Early Retirement Age is age 60 for men. Lastly, men and women become eligible for Old Age pensions at age 65 and 60 respectively.⁹ Figure 2 illustrates survival functions for entry into the pension system for the sample of private sector employees. The graphs are presented separately for men and women given the different eligibility ages. The survival functions illustrate sharp declines at ages 60 and 65 highlighting a significant amount of entry into the pension system once individuals become eligible for the Early Retirement and Old Age pensions. Additionally, the figure demonstrates that, for both men and women, most retirements occur between ages 55 and 60. Further, the graph shows that roughly 25% of the male sample retire by claiming disability pensions prior to age 60.

2.2 Administrative Data & Sample Restrictions

Our empirical analysis is based on administrative registers from the Austrian Social Security Database (ASSD, see Zweimüller et al (2009)), which is collected with the principle aim of verifying individual pension claims. The data provide longitudinal information for the universe of private sector workers in Austria throughout their working lives. Specifically, information on employment and earnings as well as other labor market states relevant for computing insurance years such as military service, unemployment, and maternity leave is collected. Detailed electronic records with employer identifiers that allow the measurement of tenure are recorded in the period from 1972 onwards; here we use information up to 2006. For the years prior to 1972 retrospective information on insurance relevant states is available for all individuals who have retired by the end of the observation period.

⁹Benefits from disability and early retirement are entirely withdrawn if an individual earns more than about 300 Euros per month; therefore we see very few individuals returning to the labor force once they are retired.

Combining the administrative data from 1972 onwards and the retrospective data prior to 1972 yields information on complete earnings and employment careers of retirees. Because firm identifiers are available only from 1972 onwards, uncensored tenure can be measured for jobs starting after January 1, 1972.

To investigate the effect of severance pay eligibility on retirement decisions we consider all individuals born between 1930 and 1945. For these individuals we observe sufficiently long uncensored tenure at retirement.¹⁰ We focus on workers who are still employed after their 55th birthday and follow them until entry into retirement or up to the age of 70. We make several restrictions to the original sample of about 650,000 workers, which are summarized in the top panel of Table 1. Most importantly, we exclude individuals who worked as civil servants or whose last job was in construction, because they are subject to different pension and severance pay rules. As we are interested in tenure at retirement, we further exclude workers with left censored tenure at retirement and we only consider retirement entries which occur within 6 months of the worker's last job. Individuals with longer gaps between employment and retirement are only followed until the end of the last employment. With these restrictions, we have a final sample of 269,411 retiring individuals.

Table 2 presents summary statistics separately for the full retirement sample and for the sub-sample of individuals with more than 10 years of tenure at retirement, who are eligible for severance pay. The median retirement age is at 59 years in both groups, which reflects that most individuals retire through disability or early retirement (28% and 38% in the full sample, respectively).¹¹ Years of employment and annual earnings in the last year before retirement are slightly higher for workers with longer tenure and these workers also appear to be of better health given their average time spent in sick leave. Overall the differences between both groups are minor. Earnings relevant for the calculation of retirement benefits and therefore recorded by the ASSD are top coded; roughly 14% of the sample has censored earnings at retirement.

¹⁰In addition, these individuals retire after a pension reform in 1985 which changed the assessment basis for benefit calculation and thereby the type of information recorded.

¹¹The actual share of retirements through early retirement is higher than the presented number, as separate insurance categories for early retirement are only recorded as of 07/1993 and individuals retiring before the statutory pension age before that are coded as old age pension entries.

3 Nonparametric Graphical Analysis

In this section we present graphical evidence on the individual labor supply responses to the severance payment thresholds at retirement. We start with a discussion of patterns in the distribution of tenure at retirement that are observed in the raw data. In particular, the observed patterns highlight that individuals delay their retirements in response to the severance payment thresholds. As we discuss in the estimation strategy below, a key element for the elasticity estimation is detecting the point at which the delays begin to occur because the maximum delay time will drive the elasticity estimation. The graphical evidence in this section indicates that the maximum delay times are relatively short since delays can only be detected amongst individuals who are relatively close to the thresholds.

To confirm that the observed patterns correspond to reactions to the severance payment rule, we present three pieces of additional empirical evidence in this section. First, we investigate the variation of other observables around the tenure thresholds and examine whether or not this variation in other observables can explain the observed patterns in the distribution of tenure at retirement. Second, we examine whether decisions earlier in life such as job changes at particular ages are responsible for the retirement patterns. Finally, we investigate how the patterns in tenure at retirement vary across various subgroups within the sample. We confirm that there is heterogeneity in the retirement patterns such that there are less (more) distinctive patterns amongst groups that we expect to be less (more) responsive to the severance payments. While this section focuses on highlighting the empirical evidence on labor supply responses to the severance payments, the next section presents a model of retirement decisions motivated by the empirical evidence.

3.1 Distribution of Tenure at Retirement

Figure 3 presents the distribution of tenure at retirement for the full sample with the number of individuals on the vertical axis and years of tenure at retirement on the horizontal axis; tenure at retirement is measured at a monthly frequency. Several features are immediately evident from this graph. First, the plot shows discontinuous spikes in the number of retirements at the tenure thresholds. Second, there are dips in the number of retirements just before the tenure thresholds, which are generally concentrated within 1 year before the threshold. These patterns are regularly repeated at each tenure threshold but are not apparent at any other point in the tenure distribution. This evidence suggests that individuals who would have retired just before the thresholds in the absence of the

severance pay discontinuities end up delaying their retirements until they just qualify for the (larger) severance payments. The plot also indicates a seasonal pattern illustrated by small spikes in the number of retirement at each integer value of years of tenure at retirement. The seasonality can be explained by a relatively large fraction of job starts in January and corresponding retirement exits in December.

Some noteworthy features are indicated by the pattern in Figure 3. First, the dips and spikes around the tenure thresholds are clearly separated from each other. This suggests that labor supply responses to each tenure threshold occur in a relatively narrow time window around the threshold. An impact of the severance pay schedule on intertemporal labor supply decisions beyond a five-year horizon is therefore not supported by the data. Second, we do not detect any evidence for income effects based on the observed patterns. In the presence of detectable income effects, individuals who qualify for the severance payments would end up retiring earlier than they would have in the absence of the severance payments; the observed patterns, however, seem to indicate that the only responses to the severance payments are delays in retirement decisions. Third, even though there are decreases prior to the thresholds, the frequency of retirements never goes to zero just prior to the thresholds. This means there appears to be a substantial number of individuals who are unresponsive to the severance pay system at retirement. Our analysis of heterogeneity in labor supply responses will concentrate on identifying the unresponsive groups; we will examine how health, earnings, firm size and job rigidity relate to responsiveness to the severance pay thresholds.

3.2 Accounting for Covariates

We exploit panel variation in the probability of retirement to examine whether or not other observable characteristics change around the tenure thresholds. In particular, we estimate the following regression

$$r_{it} = \sum_{\tau=0}^{34} \gamma_{\tau} d_{\tau} + X_{it}\beta + \epsilon_{it}$$

where r_{it} is an indicator equal to 1 if individual i retires within time period t . The set of observations per individual covers all quarters from age 55 to retirement or age 70. The sample used for estimation includes all 380,737 individuals left at the last step of sample selection in Table 1, not only those observed retiring within 6 month of their last job.

Including all job exits allows us to examine whether or not regularities in general job exits (as opposed to just retirements) after 5, 10, 15, ... year intervals are responsible for the observed retirement patterns in Figure 3. For computational reasons, time is measured at a quarterly frequency instead of the monthly frequency presented in Figure 3.

The regressors in the estimated equation are a set of indicators d_τ equal to 1 if the individual's quarterly tenure at time t equals τ . Further, we include a large set of time-varying control variables X_{it} relating to age, calendar years, industry, region, seasonality, earnings histories, firm characteristics, health and experience.¹²

Figure 4 plots the coefficients on the quarterly tenure dummies from the estimated regressions. The graph shows a pattern of dips before and large spikes at the thresholds that is very similar to Figure 3. The yearly seasonality pattern is now removed by controls for quarter of the year. Overall, Figure 4 confirms that incentives in the severance pay system are driving the retirement pattern around the tenure thresholds rather than other observable characteristics or regularities in job-leaving behavior.

3.3 Job Starts

After highlighting individuals' responsiveness to the severance payments at retirement in the figures above, we now turn to the question whether these payments affect individuals' decisions to begin new jobs. Specifically, we investigate whether individuals time the beginning of new jobs so that they can retire at the Early Retirement Ages (ERAs, respectively 55 and 60 for women and men) and also claim severance payments at the time of their retirements. To explore this idea, Figure 5 plots the number of individuals starting new jobs (vertical axis) against age measured at a quarterly frequency (horizontal axis). If individuals are timing the beginning of their new jobs so that they can just complete 10, 15, or 20 years of tenure at the ERAs, then we would expect to see sharp increases in the number of individuals starting new jobs at ages 50, 45, 40 etc. The evidence in Figure 5 shows no discernible change in job starts at any age prior to the ERAs. This smoothness across age emphasizes that, while there is evidence that some individuals delay their

¹²Firm size is grouped into the following categories: ≤ 5 , 6–10, 11–25, 26–99, 100–499, 500–999, ≥ 1000 . Health status through age 54 is based on the following categories of sick leave through age 54: ≤ 0.5 years, 0.5 – 1 years, 1 – 2 years, and ≥ 2 years. Health in the current quarter is based on the following categories for sick leave in the current quarter: 0 days, 1 – 30 days, 31 – 60 days, and ≥ 61 days. Earnings growth dummies are based on positive, negative, or zero growth relative to earnings in the corresponding quarter. Quarterly earnings for individuals with continuous employment during a calendar year are equal to total annual earnings divided by 4. Earnings for individuals retiring at the beginning of a quarter are set equal to earnings from the previous quarter. For women, the base controls also include a dummy for having kids.

retirements to qualify for (larger) severance payments at retirement, there is no evidence that individuals reallocate their labor supply (or participation) at earlier ages in response to the sizeable anticipated incentives from the severance payments.

3.4 Heterogeneity

Above we have seen that, while there is a clear pattern in the frequencies of retirement around the tenure threshold, there are also retirements occurring in the months directly before a tenure threshold. This means that a substantial fraction of the sample seems to be unresponsive to the incentives created by the severance system. Here we examine differences in responsiveness along observable individual and job characteristics. In particular, we consider heterogeneity by health status, position in the earnings distribution, firm size, and job rigidity.

We start by investigating heterogeneity related to health status. We measure ill health based on the fraction of time between age 54 and retirement spent on sick leave.¹³ We define an individual as unhealthy if the fraction of time between age 54 and retirement spent on sick leave is above the median fraction of time for individuals with positive sick leave days. Figure 6 presents frequency plots for unhealthy and healthy individuals, respectively. As expected, unhealthy individuals are not very flexible in the timing of their retirements. We basically see no response to the thresholds among retirees with health problems. Thus, some of the pre-threshold retirement is likely to be driven by negative health shocks and also more permanently poor health status.

Next we turn to heterogeneity related to earnings. We group individuals by the calendar year when they turn 55 and by tenure at the end of age 54; within each group, we compute percentiles of the distribution of average real earnings between ages 50 through 54. We condition on tenure at the end of age 54 because we want to account for returns to tenure and compare higher and lower earnings individuals with similar tenure levels at retirement. Figure 7 presents the distributions of tenure at retirement for different earnings percentiles. Because of the relatively small sample sizes, this graph shows frequencies for pooled observations in the two years before and after each of the tenure thresholds. The plots illustrate less pooling amongst individuals at higher earnings percentiles. These

¹³Roughly 35% of individuals in our sample have no sick leave days over their entire careers and 68% have no sick leave between ages 54 and retirement. Health status is highly correlated with the likelihood of claiming disability pension; about 64% of individuals with some sick leave between age 54 and retirement claim disability pensions as opposed to 15% of those with no sick leave between age 54 and retirement.

high-earning individuals are most likely affected by the social security earnings cap and therefore have other savings and private pensions. This makes changes in their budget sets due to the severance payments relatively small.

Job and employer characteristics are also likely to influence a worker's flexibility in timing his retirement date. Therefore we next examine retirement patterns by firm size. Intuitively, individuals employed in smaller firms may be more restricted in choosing their retirement dates around the tenure thresholds. Small employers may put more pressure on their employees to retire prior to qualifying for a (larger) severance payment. Additionally, employees at smaller firms may have less ability to leave their firms just after reaching a tenure threshold since their employers may rely on them to complete their projects since there are fewer substitutable employees available to do so. The evidence presented in Figure 8 is consistent with these intuitions as the plots indicate that the pre-threshold dips and post-threshold spikes increase monotonically with firm size.

As firm size plays a considerable role for individual retirement decisions, we examine also other rigidities that may be imposed by an individual's job situation. In particular, we use firm level information on job exits and retirements to infer the restrictions an individual may face in the choice of their retirement date. To summarize different impacts we create a job rigidity index based on three components. First, we measure the rate of exits from the firm in the year of retirement by the number of job spells with the employer ending during the year divided by the number of employees at the beginning of the year. We then rank jobs according to the firm level exit rates and define high exit rate jobs as the top decile. Second, the Austrian labor market is highly seasonal and we observe that many firms hire and let go workers only in certain months of the year. This seasonal demand pattern may also restrict the choice of retirement dates. Therefore we exploit the distribution of exits from the firm over the calendar year and compute the level of exit concentration by the share of all exits that occur the calendar month with the highest exit rate. Jobs in the top decile of the exit concentration distribution are defined as jobs in firms with highly concentrated exits. Third, we investigate retirement behavior of coworkers at the firm around the tenure thresholds. Specifically, from all retirements at the firm in the past 5 years, we compute the share of retirements that occurred at a tenure level in the year after a threshold. The bottom decile of jobs in firms with the lowest shares of post-threshold retirements are defined as jobs in low post-threshold retirement firms. The rigidity index takes the values from 0 to 2 if the job hits none, one, or at least two of the three rigidity components (job in firm with high exit rate, with highly concentrated exits, or in firm with

low level of post-threshold retirements). Figure 9 clearly shows that responsiveness to the severance pay thresholds decreases as the level of job rigidity increases.

3.5 Restricted Sample

Figures 7 through 9 demonstrate that there are considerable differences in responsiveness to the severance pay thresholds across the population. In principle, there are two ways to interpret these heterogeneous responses shown in the graphs. First, they could be the result of heterogeneous preferences and thus heterogeneity in labor supply elasticities throughout the population. Second, the difference in responsiveness could be due to constraints and adjustment costs. Following the second interpretation, we assume homogeneous preferences so that differences in constraints lead to different responses to the incentives. The reasoning behind this assumption is as follows. The nonresponsive groups identified in the graphical analysis have patterns that are mostly smooth through the tenure thresholds rather than having smaller pre-threshold dips and post-threshold spikes. In addition, heterogeneity in the graphs occurs along dimensions that could constitute obstacles to responsiveness such as ill health and job rigidities. We do not detect heterogeneity along more neutral characteristics such as gender, birth cohorts, skill types, or number of kids (not shown here). Our interpretation of the heterogeneity patterns thus follows Chetty (2009) who considers fixed costs of adjusting labor supply and their implications for estimating labor supply elasticities.

We base the main empirical analysis on a restricted sample excluding the most constrained groups of individuals. The bottom panel in Table 1 summarizes the decreases in sample size resulting from excluding the least responsive individuals along each dimension of heterogeneity that we have examined. The total number of individuals in the restricted sample is 154,484 individuals. Figure 10 illustrates the distribution of retirements by distance to the nearest threshold after pooling across all thresholds; this figure is based on 76,400 individuals in the restricted sample with at least 7 years of tenure at retirement. The basic patterns are the same as for the full sample. Even in the restricted sample, we still observe several individuals retiring just prior to the severance pay thresholds. However, eliminating the unresponsive groups does reduce the probability of retirement shortly before the thresholds. In the full sample, the probability of retiring in a quarter within 1 year before a threshold is 22% lower than the probability of retiring in any other quarter. Each sample cut further lowers this probability so that in the restricted sample, the probability

of retiring in a quarter within 1 year before a threshold is 29% lower than the probability of retiring in any other quarter.¹⁴

4 Theoretical Background

4.1 Preliminaries

We use the empirical evidence from the previous section to guide us in modelling labor supply responses to the severance payments. The empirical evidence is summarized as follows. First, the empirical evidence indicates that individuals do not time their job starts earlier in their careers to be eligible for severance pay at the minimum retirement ages mandated by the government pension system. Therefore, we focus only on the effects of the severance payments on retirement decisions. Second, we will abstract from income effects in our theoretical model. Intuitively, the severance payments are small relative to lifetime income, and individuals may be unlikely to respond to such small changes in lifetime income. In addition, we find no clear evidence for income effects in the retirement patterns. The absence of income effects implies that the only effect severance pay thresholds can have on retirement decisions in our model is to provide incentives to delay retirement relative to a counterfactual date without severance pay. Third, given the lack of long-term planning in relation to the severance payments and the relatively short time-space over which retirements take place in Austria, the empirical evidence suggests that individuals' retirement decisions take into account at most one tenure threshold when deciding when to retire. We therefore model the decision to delay retirement based on the nearest, upcoming tenure threshold. Lastly, the empirical evidence from the previous section indicates that controlling for age and other observable covariates does not alter the observed retirement patterns. Therefore, we assume that the age or date until which each individual would need to work to reach the next severance pay threshold is varying exogenously across individuals.

The empirical evidence suggests a basic decision process for retirement decisions and responses to the financial incentives from the severance payments. First, individuals ignore the severance payments and select an optimal retirement age while taking all other financial incentives for retirement into account. We refer to these retirement ages as counterfactual retirement ages. Second, individuals examine the age at which they would reach

¹⁴Results from linear probability models of the retirement indicator by quarter on a pre-threshold dummy and basic controls for gender, age, season, and a set of threshold indicators for different subsamples are available on request.

their (next) tenure thresholds to qualify for (larger) severance payments. In particular, individuals compute their delay time based on the difference between their threshold ages and their counterfactual retirement ages. Finally, individuals compare their delay times against a critical value capturing the maximum amount of time they are willing to delay their retirements. This comparison determines individuals' ultimate retirement ages; if their delay time is less than the critical value, then they will retire at their threshold ages, otherwise they will retire at their counterfactual ages.

We specify the formal optimization problems behind this decision process in the next subsection. We highlight that the critical value is determined based on an indifference condition that expresses indifference between retiring at an early date without severance pay and retiring at a threshold with severance pay. The estimation strategy that we discuss in the next section focuses on this indifference condition to estimate the structural parameter of interest, the intertemporal labor supply elasticity.

4.2 Model

In this subsection, we formalize the economic model of individuals' retirement decisions that we discussed in the previous subsection. We consider a population of workers who are employed at age 55 and who decide when to retire in the remaining T years of their lifetime. Each individual behaves according to a life-cycle labor supply model in which preferences are defined over consumption c_t in each period $t \in \{0, 1, \dots, T\}$, and years of work R beyond age 55. We assume that there is no uncertainty or time discounting and that each individual lives for T periods. Especially individuals are not subject to the risk of layoffs, job changes, or major health shocks.¹⁵ To concentrate on the effect of severance pay on delays in retirement we further assume a quasi-linear utility function, which eliminates income effects.¹⁶ In this setting without discounting or uncertainty, an individual will consume a constant fraction of her total income in each period.

Following the sequential decision procedure outlined in the previous subsection, we start with the model for counterfactual retirement choices. In the absence of any severance

¹⁵The assumption of no uncertainty is a useful approximation to describe the environment in which retirement decisions of higher-tenured workers in Austria take place for multiple reasons. First, layoffs are concentrated amongst lower-tenured, younger workers. Second, collective bargaining agreements determine a significant portion of earnings based on age, experience, tenure and other observable employee characteristics.

¹⁶To analyze the sensitivity of our results to this assumption, we will experiment with more general utility functions in section 5.3.

payment thresholds, an individual chooses his retirement date by solving the following optimization problem

$$\max_R yR + x - \frac{\theta}{1 + \frac{1}{e}} \left(\frac{R}{\theta} \right)^{1 + \frac{1}{e}}. \quad (1)$$

where y denotes the individual's annual earnings, x denotes unearned income, and θ denotes an individual's taste for work. We assume that tastes for work are distributed across individuals according to the distribution function $F(\theta)$ defined on $(0, \theta_{max}]$ and a higher value of θ corresponds to higher taste for work as opposed to leisure. The parameter e captures the convexity in the disutility of work; we discuss this parameter in more detail below. Solving the optimization problem, counterfactual retirement choices are given by

$$R(\theta) = \theta (\lambda y)^e \quad (2)$$

where $\lambda = 1$ is the marginal utility of income. We highlight that these counterfactual retirement choices capture retirement decisions when ignoring the severance payments but still taking all other financial incentives for retirement into account.

After selecting a counterfactual retirement date while ignoring the severance payments, individuals follow the second step of the decision process and select an optimal retirement date while taking the financial incentives from the severance payments into account. We assume that each individual has a threshold retirement age, $\bar{R} \geq R(\theta)$, such that, when retiring after the threshold age, the individual will have accumulated sufficient years of tenure to qualify for a lump-sum payment dx . Since individuals differ in their years of tenure at age 55, the threshold retirement ages are heterogeneously distributed across individuals. Formally, each individual solves the following optimization problem to select his optimal retirement date when taking his severance payment into account,

$$\max_{R > R(\theta)} yR + x + dx \mathbf{1}(R \geq \bar{R}) - \frac{\theta}{1 + \frac{1}{e}} \left(\frac{R}{\theta} \right)^{1 + \frac{1}{e}}. \quad (3)$$

Optimal retirement choices for individuals with counterfactual retirement choice $R(\theta) = R$ and different values of the severance payment threshold at \bar{R} are illustrated in Figure 11. As shown, the severance payments create discontinuous increases in individuals' budget sets at the threshold retirement ages. For some individuals, the thresholds are sufficiently early such that these individuals would prefer to retire at the tenure thresholds with the severance payments rather than retire at the counterfactual date without the severance payments.

For other individuals, the thresholds are sufficiently late such that these individuals would prefer to retire earlier, at the counterfactual date, without the severance payments rather than retire later with the severance payments.

Because the threshold ages vary across individuals with constant taste for work θ , the amount of time between each individuals' tenure threshold and her counterfactual retirement age, denoted by $dR = \bar{R} - R(\theta)$, also varies across individuals. Intuitively, dR captures the amount of delay from one's counterfactual retirement age that is necessary to qualify for the severance pay. An individual's optimal labor supply strategy incorporating the severance payment threshold is therefore to set a critical value $\Delta(\theta)$ that captures the maximum amount of time he is willing to delay retirement from the counterfactual retirement age to qualify for his severance payment. If $dR \leq \Delta(\theta)$, then the individual will delay retirement and retire at his threshold age \bar{R} ; otherwise the individual will retire at his counterfactual retirement age $R(\theta)$. The critical value $\Delta(\theta)$ is determined by solving for the length of time that makes the individual indifferent between retiring early without the severance payment and retiring at the tenure threshold with the severance payment. Formally, the critical value is characterized by solving the following indifference equation for $\Delta(\theta)$,

$$\underbrace{y(\bar{R} - \Delta(\theta)) + x - \frac{\theta}{1 + \frac{1}{e}} \left(\frac{\bar{R} - \Delta(\theta)}{\theta} \right)^{1 + \frac{1}{e}}}_{\text{utility when retiring early without sev pay}} = \underbrace{y\bar{R} + x + dx - \frac{\theta}{1 + \frac{1}{e}} \left(\frac{\bar{R}}{\theta} \right)^{1 + \frac{1}{e}}}_{\text{utility when retiring at threshold with sev pay}}. \quad (4)$$

As indicated by the notation, the critical value or maximum length of time an individual is willing to delay retirement to qualify for severance pay varies with the taste for work θ . Specifically, $\Delta(\theta)$ is longer for individuals with a higher value of θ . All individuals of type $\theta = \theta_{max}$ are willing to delay retirement up to the maximum length of delay $\Delta(\theta_{max})$. But for individuals with smaller values of $\theta < \theta_{max}$ the willingness to delay retirement is lower. In the aggregate, the model implies that the frequency of retirements by tenure follow a declining pattern before the tenure threshold. The number of retirements relative to the counterfactual retirement date drops when the highest types θ_{max} start delaying retirement and as we move closer to the threshold lower types join which further lowers the observed number of retirements relative to the counterfactual.¹⁷

Figure 12 illustrates the declining pattern in the number of retirements prior to the sev-

¹⁷Whether the frequency of retirements goes all the way to zero prior the threshold, or whether there are individuals retiring just before a tenure threshold, depends on the distribution of θ .

erance pay threshold obtained from a simulation of the model.¹⁸ The simulated model also takes into account that there are excess retirements just after the government-mandated tenure thresholds in addition to excess retirements exactly at the government-mandated tenure thresholds in Figure 10. To capture this observed pattern, we assume that some employers impose tenure thresholds beyond the government-mandated tenure thresholds for some employees. Intuitively, some employers require some employees to complete specific tasks or projects prior to paying these employees their severance payments. Using \bar{r} to denote the retirement age corresponding to the government-mandated tenure threshold, we assume that an individual's threshold retirement age is given by

$$\bar{R} = \bar{r} + z$$

where z represents the amount of additional time required for project completion. We assume that z is distributed on $[0, \infty)$ according to density $p(z)$.

4.3 The Intertemporal Labor Supply Elasticity

The elasticity of intertemporal substitution in labor supply is defined to capture how labor supply responds to anticipated variation in earnings per unit of time (i.e., anticipated wage variation). Intuitively, when a wage increase is anticipated, it is already factored into lifetime income so that the marginal utility of lifetime income can be assumed to be held constant. Thus, using λ to denote the marginal utility of lifetime income (the multiplier on the individual's budget constraint), the intertemporal labor supply elasticity is defined by $\frac{d \ln R}{d \ln y} |_{\lambda}$. Solving the individual's optimization problem, the intertemporal labor supply elasticity in our model is given by

$$\frac{d \ln R}{d \ln y} |_{\lambda} = e.$$

Intuitively, when the marginal disutility from additional labor supply rises very rapidly, an individual will not adjust his labor supply very much in response to an anticipated wage increase. In the life-cycle labor supply model above, we specifically refer to e as an intertemporal (λ -constant) elasticity because this parameter governs responsiveness to severance payments that can be fully anticipated and hence factored into the marginal utility of income. The intertemporal labor supply elasticity is reflected in the curvature

¹⁸The parameter values and distributions used in the simulation are described in the Sensitivity Analysis section.

of an individual's indifference curves, or equivalently in the critical value Δ that captures an individual's willingness to delay retirement for a severance payment. In particular, a larger elasticity corresponds to a willingness to delay retirement for a longer amount of time, which increases the critical value.

4.4 Intensive & Extensive Margin Elasticities

While we have defined the above intertemporal labor supply elasticity in terms of career length, this extensive margin elasticity differs from the more traditional intensive margin intertemporal labor supply elasticity which is usually defined in terms of hours of work. To highlight the distinction between these elasticities, we relate the model above to a more traditional life-cycle labor supply model. Specifically, in the setting above with no discounting and no uncertainty, a more traditional life cycle labor supply model could define preferences in terms of hours of work (h) so that an individual's optimization problem would be

$$\max_{\{h_t\}} \int_0^T w_t h_t dt + x - \int_0^T \theta(t) v(h_t) dt$$

where w_t captures the hourly wage at time t , $\theta(t)$ captures how the disutility of work evolves with time (age), and $v(\cdot)$ captures the disutility over hours of work. In this problem, the individual chooses a profile of hours of work at each point in time $\{h_t\}$. The intensive margin intertemporal labor supply elasticity captures the curvature in the disutility over hours of work,

$$e^{intensive} = \frac{h_t v''(h_t)}{v'(h_t)}.$$

To make the relationship with the career-length model above more explicit, suppose that an individual is restricted such that he can choose to either work \bar{h} hours or none at all, i.e. $h_t \in \{0, \bar{h}\}$ for all t and $v(0) = 0 < v(\bar{h})$. In this case, if $\theta(t)$ is monotonically increasing (i.e. $\theta(t') > \theta(t)$ for $t' > t$), then an individual's optimization problem would simplify to choosing an optimal retirement date R ; in particular, using $h^*(t)$ to denote an individual's optimal profile of hours of work, the optimal retirement date would be defined by

$$R \in [0, T] \quad \text{s.t.} \quad h^*(t) = \begin{cases} \bar{h} & \text{for } t \leq R \\ 0 & \text{for } t > R \end{cases}.$$

Intuitively, if an individual can only work a fixed number of hours in every period and working is becoming more difficult over time, the individual's optimal strategy is to simply

pick a time at which he will stop working. Furthermore, after re-normalizing such that $\tilde{\theta}(t) = \theta(t)v(\bar{h})$ and $y_t = w_t\bar{h}$, the individual's optimization problem can be re-formulated in terms of the choice of retirement date

$$\max_R \int_0^R y_t dt + x - \int_0^R \tilde{\theta}(t) dt$$

Comparing the disutility of work in this model $\left(\int_0^R \tilde{\theta}(t) dt\right)$ to the disutility from work in the career length model $\left(\frac{1}{1+\frac{1}{e}} (R)^{1+\frac{1}{e}}\right)$ highlights that the extensive margin, or career length, intertemporal elasticity e reflects how quickly the marginal disutility of work increases with age. In other words, while $e^{intensive}$ is determined by the curvature of $v(\cdot)$, the extensive margin elasticity e depends on $\theta(t)$. This implies that the extensive margin elasticity can be completely independent of the intensive margin elasticity. While both intensive and extensive margin elasticities are important to estimate, we focus only on estimating the extensive margin elasticity since we lack data on hours of work and since the severance payment incentives apply only to career length decisions.

An alternative, but closely related, view of the distinction between labor supply elasticities at the extensive and intensive margins is the focus of recent work in the macroeconomics literature. In particular, Rogerson and Wallenius (2009) develop a life-cycle model of labor supply in which both intensive margin and extensive margin labor supply decisions arise due to of fixed costs of labor force participation. The intensive margin decision corresponds to the choice of how many hours to work in a given period while the extensive margin decision corresponds to the choice of which, or how many, periods to work. Using this model in a general equilibrium setup, Rogerson and Wallenius examine how aggregate (lifetime) hours of work respond to taxes. They emphasize that, even when the intensive margin elasticity is small, the aggregate responses may be large because the extensive margin elasticity can be large. Moreover, they show that there is an important interaction between the intensive and extensive margins: less adjustment on the extensive margin necessarily implies more adjustment at the intensive margin. Intuitively, the less an individual likes to change hours over the life cycle, the higher is the fixed cost required to induce a retirement and thus the larger the response at the extensive margin.

In our empirical analysis and in the model above, we are able to focus explicitly on the extensive (career length) margin highlighted by Rogerson and Wallenius. Importantly, the intertemporal labor supply elasticity in our model, i.e. the structural parameter e , exactly captures the extensive margin elasticity that Rogerson and Wallenius emphasize. The

identification of this parameter comes from variation in the nonlinearities in budget sets due to severance payments, which corresponds to variation in the fixed costs of participation in the Rogerson and Wallenius framework.

Additionally, similar to Rogerson and Wallenius, this extensive margin elasticity captures the elasticity of aggregate (lifetime) hours of work when the intensive margin elasticity is small. To demonstrate this, we note that with a small intensive margin elasticity, an individual must essentially work some set number of hours \bar{h} in each period that he works, but he can choose the number of periods that he works, R . In this case, the elasticity of aggregate (lifetime) hours of work with respect to the wage rate is given by $\frac{d \ln(\bar{h}R)}{d \ln y} = \frac{d \ln(R)}{d \ln y} = e$.

5 Elasticity Estimation

5.1 Estimation Strategy

The estimation the intertemporal labor supply elasticity is based on the counterfactual labor supply equation and the indifference condition that determines the critical value Δ , equations (2) and (4) respectively, which we repeat here for clarity:

$$\begin{aligned} \bar{R} - \Delta &= \theta(\lambda y)^e \\ y(\bar{R} - \Delta) - \frac{\theta}{1 + \frac{1}{e}} \left(\frac{\bar{R} - \Delta}{\theta} \right)^{1 + \frac{1}{e}} &= y\bar{R} + dx - \frac{\theta}{1 + \frac{1}{e}} \left(\frac{\bar{R}}{\theta} \right)^{1 + \frac{1}{e}}. \end{aligned}$$

Our estimation strategy is to solve the two equations for θ and e using estimated information on y , \bar{R} , dx and Δ from the data. The crucial moment we derive from the data to make this procedure work is the length of delay Δ that makes the individual indifferent between retiring at the counterfactual retirement date and delaying retirement to the tenure threshold. Essentially, this moment is estimated from the observed distribution of retirements by tenure (see Figure 10, presented in the previous section).

Our model assumes that individuals are heterogeneous in their taste for work θ and that types are randomly distributed in the population. As we have discussed, each type has a different indifference condition and consequently there is variation in $\Delta(\theta)$. The identification of e requires, however, that we pin down a specific type. Our strategy is to focus on the type of individuals with the highest taste for work θ_{max} . For those individuals the maximum length of delay can be identified from the observed retirement patterns in Figure 10 as the point where number of retirements starts declining, relative to the

counterfactual frequencies, prior to the tenure threshold. As the individuals with the highest taste for work are willing to delay retirement the longest (i.e. they have the highest value of Δ), we observe them relocating their retirement earliest and can thus identify the responses of the type with highest value of $\theta = \theta_{max}$. Note that the assumption that there is a fixed type with maximum taste for work θ_{max} in the population is crucial for identification of the intertemporal labor supply elasticity. To examine the sensitivity of the elasticity estimate e with respect to the point of indifference we will hold the type θ_{max} fixed and vary Δ .

In the next subsection we describe in detail the steps taken to estimate Δ and introduce the calibrated values for the other parameters. Subsequently we present estimation results and sensitivity analysis.

5.2 Estimation Procedures

The steps of the procedure to estimate Δ are illustrated in Figures 13 through 15 and summarized as follows:

1. Estimate the spike in retirements and a continuous approximation of the observed frequencies (Figure 13).
2. Estimate the counterfactual retirement frequencies using an interpolation of actual frequencies away from the threshold (Figure 14).
3. Estimate the excess mass of retirements after the tenure threshold based on the cumulative differences between the actual and counterfactual retirement frequencies (Figure 15).
4. Estimate Δ by equating the pre-threshold delayed retirements with the post-threshold excess retirements (Figure 15).

To estimate the critical value Δ , we start by estimating a continuous approximation of the *observed retirement frequencies* so that we can accurately characterize the amount of excess mass at the tenure thresholds. Specifically, in each interval between two tenure thresholds, we regress the observed frequencies on a continuous 4th order polynomial in tenure at retirement. We then predict the fitted values and set the values at the tenure thresholds equal to the observed values. We re-scale the predicted values so that the total number of individual retirements based on the fitted values is equal to the number of

observed retirements. We refer to the frequencies based on this continuous approximation as the *actual frequencies*. Figure 13 plots the observed frequencies against the actual frequencies when pooling across all of the tenure thresholds.

In the second step, we estimate *counterfactual frequencies* of tenure at retirement that would occur in the absence of the severance payments. For that purpose, we regress the actual frequencies in each interval between two thresholds on a continuous 4th order polynomial in tenure at retirement and a set of threshold dummies which are equal to 1 if the level of tenure is within 1 year before, at or within 1 year after a tenure threshold.¹⁹ We then set the threshold dummies equal to 0 and obtain the fitted values. Again, we re-scale the fitted values so that the total number of actual retirements is equal to the total number of counterfactual retirements. The identifying assumption implied by using dummies to capture retirement behavior around the tenure thresholds is the following: in the absence of the severance payment incentives, individuals with tenure around the thresholds would behave similarly to individuals away from the tenure thresholds. Thus, the counterfactual frequencies are identified based on individuals away from the tenure thresholds. Figure 14 plots the actual frequencies against the counterfactual frequencies when pooling across all of the tenure thresholds. The plot highlights that, in the absence of the severance payments, the counterfactual frequencies would be smooth through the tenure thresholds. While the first two steps treat each tenure threshold separately, the remaining steps in estimating the point of indifference are based on the frequencies when pooling across all of the tenure thresholds.

In the third step, we estimate the number of excess retirements at the tenure thresholds based on the cumulative differences between the actual and counterfactual frequencies at and just after the tenure thresholds. We select a post-threshold cutoff, R_H , to capture all of the excess retirements because the graphical evidence indicates that some of the excess retirements come from individuals who retiring just after the tenure thresholds rather than exactly on the thresholds. We use an iterative procedure to select R_H . We choose an initial R_H just above \bar{R} and compute the excess retirements based on the sum of the differences between the actual and counterfactual frequencies at \bar{R} through R_H . Next, we increase R_H and compute the number of additional excess retirements that are added to the previous measure of excess retirements. We continue to increase R_H until the number of additional excess retirements that is added to the previous measure of excess retirements is sufficiently

¹⁹The results are robust to increasing the window around the tenure thresholds to more than 1 year before and after the thresholds and also to using higher order polynomials.

small (i.e., less 3% of the measured excess mass). The determination of R_H and the excess retirements are illustrated in Figure 15.

In the final step, the critical value Δ is estimated based on equating the number of individuals delaying their retirements with the number of excess retirements. Analogous to the estimation of the number of excess retirements, we estimate the number of individuals delaying their retirements using the sum of the differences between the actual and counterfactual frequencies just prior to the tenure threshold. In particular, Δ is determined by the pre-threshold value such that the number of individuals delaying their retirements is equal to the number of excess retirements. Intuitively, the point of indifference is therefore estimated as the maximum possible length of delay amongst individuals who delayed their retirements to qualify for a (larger) severance payment. This strategy for determining Δ is illustrated in Figure 15.²⁰

Standard errors for the estimated value of Δ are generated by a bootstrap procedure that repeats the above steps 1 – 4 over 100 random samples of the data. We subsequently transform these standard errors into standard errors for the estimate of the elasticity e .

For the remaining parameters in equations (2) and (4) we use the following calibrated values. For annual earnings we use the mean annual earnings in the sample $y = 20,000$. The amount of severance pay is given by the severance pay rule with $dx = y/3$. To select a plausible value for the tenure threshold we note that we model the retirement decision for an individual with the highest taste for work, who should consequently choose a late retirement age. In Austria almost all retirements occur within 10 years from age 55 (see Figure 2), so we select $\bar{R} = 10$.

5.3 Estimation Results

Elasticity Estimate

Table 3 presents the estimation results. We estimate $\hat{R}_H = 0.75$ indicating that the excess retirements occur within roughly 9 months after the tenure thresholds. We estimate that the total number of excess retirements to be roughly 5,200 individuals. To put this

²⁰We note that Δ is measured as the distance to the threshold \bar{R} when there is no task completion ($z = 0$). This is because otherwise we would expect to see delays in retirements further before the threshold at $\Delta + z$ with $z > 0$ since task completions and tastes for work are independently distributed in the population. For example, suppose that we observe delays beginning at 1.25 years prior to \bar{R} and a maximum task completion of $z = 9$ months. It would be incorrect to conclude that individuals are willing to delay their retirements by $1.25 + 0.75 = 2.00$ years; if this were the case, we would observe delays beginning at 2.00 years prior to the threshold \bar{R} .

figure in perspective, we also report the excess fraction which computes the excess retirements as a fraction of the total number of counterfactual retirements between \bar{R} and \hat{R}_H . We estimate the excess fraction to be roughly 0.40; this indicates that the number of excess retirements is less than half of the total number of individuals that one would expect to retire just after the tenure thresholds in the absence of the severance payments.

Next, we estimate the point of indifference to be $\hat{\Delta} = 1.25$, which indicates that the maximum length of delay amongst individuals retiring just after the tenure thresholds is roughly 1 year and 3 months. Additionally, we report the number of delayed retirements based on the pre-threshold differences and check to make sure that the difference between the estimated number of delayed retirements and the estimated number of excess retirements is close to 0. Finally, based on the estimated point of indifference, we estimate the intertemporal labor supply elasticity to be $\hat{e} = 0.30$. The standard error for this estimate is roughly 0.05 indicating that the elasticity is estimated with relatively high statistical precision. Indeed, with this relatively small standard error, elasticities greater than one are easily outside of the 95% confidence interval of the estimated elasticity. Overall, the low elasticity reflects that the disutility of work rises relatively quickly with additional years of work.

Sensitivity Analysis

While the estimation results highlight a relatively low elasticity, we now examine the sensitivity of this result to alternative assumptions to gauge the plausibility of higher labor supply elasticities given the available empirical evidence. First, since the point of indifference Δ plays an important role, and since the quasi-linear utility assumption implies a relatively high valuation of the severance payments, we examine the implied elasticities at alternative points of indifference and under different utility specifications in Figure 16. In particular, the elasticities in this figure are computed using the following indifference condition,

$$Tu\left(\frac{y(\bar{R} - \Delta) + x}{T}\right) - \frac{\theta}{1 + \frac{1}{e}} \left(\frac{\bar{R} - \Delta}{\theta}\right)^{1 + \frac{1}{e}} = Tu\left(\frac{y\bar{R} + x + dx}{T}\right) - \frac{\theta}{1 + \frac{1}{e}} \left(\frac{\bar{R}}{\theta}\right)^{1 + \frac{1}{e}}$$

where we take advantage of the result that, with no uncertainty and no time discounting, an individual will consume a constant fraction of his total income in each period. We consider constant relative risk aversion utility functions $u(c) = \frac{c^{1-g}}{1-g}$ with $g \geq 0$, which nest

the baseline quasi-linear utility function used in the estimation at $g = 0$.²¹ The following table gives the remaining parameter values,

$$\begin{aligned}
 T &= 30 \\
 y &= 20,000 \\
 x &= 250,000 \\
 dx &= 0.333 * y \\
 \bar{R} &= 10.
 \end{aligned}$$

The value for annual earnings y is chosen to roughly correspond to the mean annual earnings reported in the summary statistics in Table 2. The value for unearned income x is chosen so that the individual consumes 15,000 euros in each period, consistent with a 75% replacement rate from government-provided pension benefits.

Figure 16 plots the elasticity e on the vertical axis against the point of indifference Δ on the horizontal axis given different values of g . We focus first on the black solid line which captures the relationship between e and Δ given the baseline utility specification at $g = 0$. The plot highlights that the elasticity increases as the maximum amount of time that individuals are willing to delay their retirements for the severance payments increases. Intuitively, if individuals are willing to delay their retirements by a longer time, this would imply that they are more responsive to the anticipated benefits. The figure illustrates that, under the baseline utility function, estimating an elasticity greater than 1.0 would require estimating a point of indifference more than 2 years prior to the threshold. Similarly, the alternative utility functions would also require a larger point of indifference to estimate an elasticity greater than 1.0. A larger point of indifference is inconsistent with the estimated excess mass and the graphical patterns observed in the data.

Continuing with Figure 16, we next consider the impacts of different utility specifications on the implied elasticities. The figure highlights that as g increases, the implied elasticities at each point of indifference also increase. The intuition for this result is as follows. As g increases, utility over consumption becomes more curved so that the marginal utility of consumption decreases faster. A lower marginal utility of consumption implies a lower valuation of the severance payment. Holding the point of indifference fixed while increasing g therefore amounts to holding the change in labor supply fixed while reducing

²¹For $g = 1$, we use $u(c) = \log(c)$.

the magnitude of the incentives causing the change in labor supply, and this implies a higher labor supply elasticity. Importantly, increasing g has a larger effect on the elasticity at higher values of Δ . Since the estimated point of indifference is relatively close to the threshold, the implied elasticity is relatively unaffected by increasing g ; with $g = 2.0$, the implied elasticity is still less than 0.5.

To further estimate the examine the plausibility of higher intertemporal labor supply elasticities, we simulate the theoretical model using different elasticities and then compare the simulated outcomes against the data. We simulate 76,400 individuals to match the sample size of the restricted sample used in Figure 10. We simulate the model using quasi-linear utility as in the estimation, so the parameters T and x become irrelevant and do not need to be specified. Annual earnings and the severance payments are set as above ($y = 20,000$ and $dx = 0.333 * y$). Lastly, the distributional assumptions for θ , \bar{R} and z are as follows. We assume that θ is drawn from an truncated exponential distribution with parameter $\mu = 0.25$; the truncation points are set so that counterfactual retirement choices are between 0 and 10. To set the distribution of threshold retirement ages, we first draw a threshold distance from uniform distribution on $[-3, 3]$ for each individual. Next, we draw a task completion length $z \in \{0, 1, \dots, 9\}$; the probabilities for each possible z value are chosen to approximate the observed fraction of excess retirements at each number of months beyond the tenure thresholds. We then add the threshold distance and task completion length to the individual's counterfactual retirement age to determine the threshold retirement age.

Figure 17 presents a plot of the data and simulated outcomes using elasticities $e = 0.30$ and $e = 1.0$. The figure highlights that a larger elasticity implies that the declines in the frequencies of retirements by distance to the threshold start much earlier. In particular, with an elasticity of $e = 1.0$, we would expect to start seeing declines at roughly 2 years before the thresholds. Because the point of indifference is much further before the threshold, the higher elasticity also predicts a much larger spike at the tenure threshold since more individuals would prefer to delay their retirements. The lower elasticity implies a point of indifference closer to the threshold and hence a smaller excess mass just after the threshold as well. Thus, higher elasticities appear to be at odds with the observed patterns in the data.

6 Conclusions

This paper presents evidence on individuals' willingness to delay exiting the labor force in response to anticipated increases in retirement benefits. This evidence is based on discontinuous increases in retirement benefits upon completion of 10, 15, 20, and 25 years of tenure by retirement. The graphical evidence illustrates a relatively modest willingness to delay retirement in response to the discontinuous increases in benefits at the tenure thresholds. Based on this evidence, we estimate a low intertemporal labor supply elasticity of 0.30. Because of the large sample size and minimal measurement error from the administrative data, there is high statistical precision in this estimate. Additionally, there is high economic precision in the estimate since the estimation is based on relatively large changes in financial incentives. Intuitively, large changes in financial incentives may be more likely to permit identification of underlying structural parameters since individuals are less likely to be overwhelmed by optimization frictions and adjustment costs.²²

Based on the empirical evidence from the current analysis, we conclude that it is difficult to reconcile the difference between intertemporal labor supply elasticities in macroeconomic and microeconomic models based only on the difference between the intensive and extensive margins of labor supply. In particular, elasticities in macroeconomic models may be larger than elasticities from microeconomic models due to other factors such as adjustment costs or search frictions (see Chetty et al (2010) and Hall (2006) respectively).

²²See Chetty (2009).

References

- Altonji, J. G. (1986). Intertemporal substitution in labor supply: Evidence from micro data. *Journal of Political Economy* 94(3), S176–S215.
- Blundell, R. and T. MaCurdy (1999). Labor supply: A review of alternative approaches. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 27, pp. 1559–1695. Elsevier.
- Browning, M., A. Deaton, and M. Irish (1985). A profitable approach to labor supply and commodity demands over the life-cycle. *Econometrica* 53(3), 503–543.
- Browning, M., L. Hansen, and J. Heckman (1999). Micro data and general equilibrium models. In J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1, Chapter 8, pp. 543–633. Elsevier.
- Card, D. (1994). Intertemporal labor supply: An assessment. In C. Sims (Ed.), *Advances in Econometrics, Sixth World Congress*. Cambridge University Press.
- Card, D., R. Chetty, and A. Weber (2007). Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market. *Quarterly Journal of Economics* 122(4), 1511–1560.
- Chetty, R. (2009). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. Working Paper 15616, National Bureau of Economic Research.
- Chetty, R., J. Friedman, T. Olsen, and L. Pistaferri (2010). Adjustment costs, firm responses, and labor supply elasticities: Evidence from danish tax records. *Quarterly Journal of Economics* (forthcoming).
- Hall, R. (2006). The labor market and macro volatility: A nonstationary general-equilibrium analysis. Working paper.
- Heckman, J. J. (1993). What has been learned about labor supply in the past twenty years? *American Economic Review* 83(2), 116–121.
- Heckman, J. J. and T. E. MaCurdy (1980). A life cycle model of female labour supply. *Review of Economic Studies* 47(1), 47–74.
- Imai, S. and M. P. Keane (2004). Intertemporal Labor Supply and Human Capital Accumulation. *International Economic Review* 45(2), 601–641.

- Keane, M. and R. Rogerson (2010). Reconciling micro and macro labor supply elasticities. Working paper.
- Ljungqvist, L. and T. Sargent (2010). Career length: Effects of curvature of earnings profiles, earnings shocks, and social security. Working paper.
- Ljungqvist, L., T. J. Sargent, O. Blanchard, and E. C. Prescott (2006). Do Taxes Explain European Employment? Indivisible Labor, Human Capital, Lotteries, and Savings [with Comments and Discussion]. *NBER Macroeconomics Annual 21*, 181–246.
- MaCurdy, T. E. (1981). An empirical model of labor supply in a life-cycle setting. *Journal of Political Economy* 89(6), 1059–85.
- Mulligan, C. B. (1999). Substitution over time: Another look at life-cycle labor supply. In B. Bernanke and J. Rotemberg (Eds.), *NBER Macroeconomics Annual 1998*, Volume 13, pp. 75–152. MIT Press.
- Ohanian, L., A. Raffo, and R. Rogerson (2008). Long-term Changes in Labor Supply and Taxes: Evidence from OECD Countries, 1956-2004. *Journal of Monetary Economics* 55(8), 1353–1362.
- Prescott, E. C. (2006). Nobel Lecture: The Transformation of Macroeconomic Policy and Research. *Journal of Political Economy* 114(2), 203–235.
- Saez, E. (1999). Do taxpayers bunch at kink points? Working Paper 7366, National Bureau of Economic Research.
- Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy* (forthcoming).
- Zweimüller, J., R. Winter-Ebmer, R. Lalive, A. Kuhn, J.-P. Wuellrich, O. Ruf, and S. Büchi (2009). Austrian social security database. *NRN: The Austrian Center for Labor Economics and the Analysis of the Welfare State* (Working Paper 0903).

Table 1
Sample Selection

	Number of Observations	Percentage change
Individuals in cohorts born 1930 - 1940	1,578,549	
Still employed at age 55	651,336	-59%
More than one year employment experience before age 55	625,251	-4%
Excluding workers ever employed as civil servant	546,308	-13%
Excluding workers with last job not in construction	487,019	-11%
Excluding left censored tenure in last job	380,737	-22%
Workers retiring within 6 months of their last job	269,411	-29%
<u>Restricted sample of highly responsive individuals</u>		
Excluding individuals with high number of sick leave days	233,976	-13%
Excluding individuals with high or low average earnings	197,726	-15%
Excluding workers from firms with less than 10 employees	159,186	-19%
Excluding workers with in jobs with highest rigidity index	154,484	-3%

Notes: Numbers based on the ASSD

Table 2
Summary Statistics

	Full Sample	Tenure at Retirement ≥ 10
# of Individuals	269,411	142,332
Retirement Age	58.64 59.00 2.58	58.71 59.25 2.61
Tenure	11.11 10.50 7.76	17.27 16.50 5.06
Annual Earnings	18459.69 17983.05 13458.51	19572.48 19649.52 13571.58
Fraction Top-Coded	0.14 - 0.34	0.15 - 0.36
Years of Employment	32.71 34.54 9.59	34.08 35.28 8.36
Years of Sick Leave	0.48 0.12 0.88	0.43 0.09 0.81
Fractions:		
Claiming Disability Pensions	0.282	0.233
Claiming Early Retirement Pensions	0.375	0.401
Claiming Old Age Pensions	0.343	0.366
Agriculture & Mining	0.045	0.041
Manufacturing	0.255	0.240
Sales	0.193	0.167
Tourism	0.044	0.022
Transportation	0.051	0.045
Services	0.412	0.485

Notes: Except for the Fractions, the mean, median and standard deviations are reported for each variable. All earnings variables are expressed in 2008 euros. Summary statistics for lifetime earnings are based on birth cohorts beyond 1935. Employment Time and Sick Leave are measured in years.

Table 3
Elasticity Estimation Results

Excess Workers	Reduced Workers	Δ Workers	Excess Fraction
5198.7270 (106.7238)	5200.1261 (112.8092)	-1.3991 (35.4562)	0.3859 (0.0111)
Δ	R_H		
1.2500 (0.1015)	0.7500 (0.0199)	e	0.2995 (0.0491)

Notes: Bootstrapped standard errors based on 100 replications are in parentheses.

Fig. 1. Payment Amounts based on Tenure at Retirement

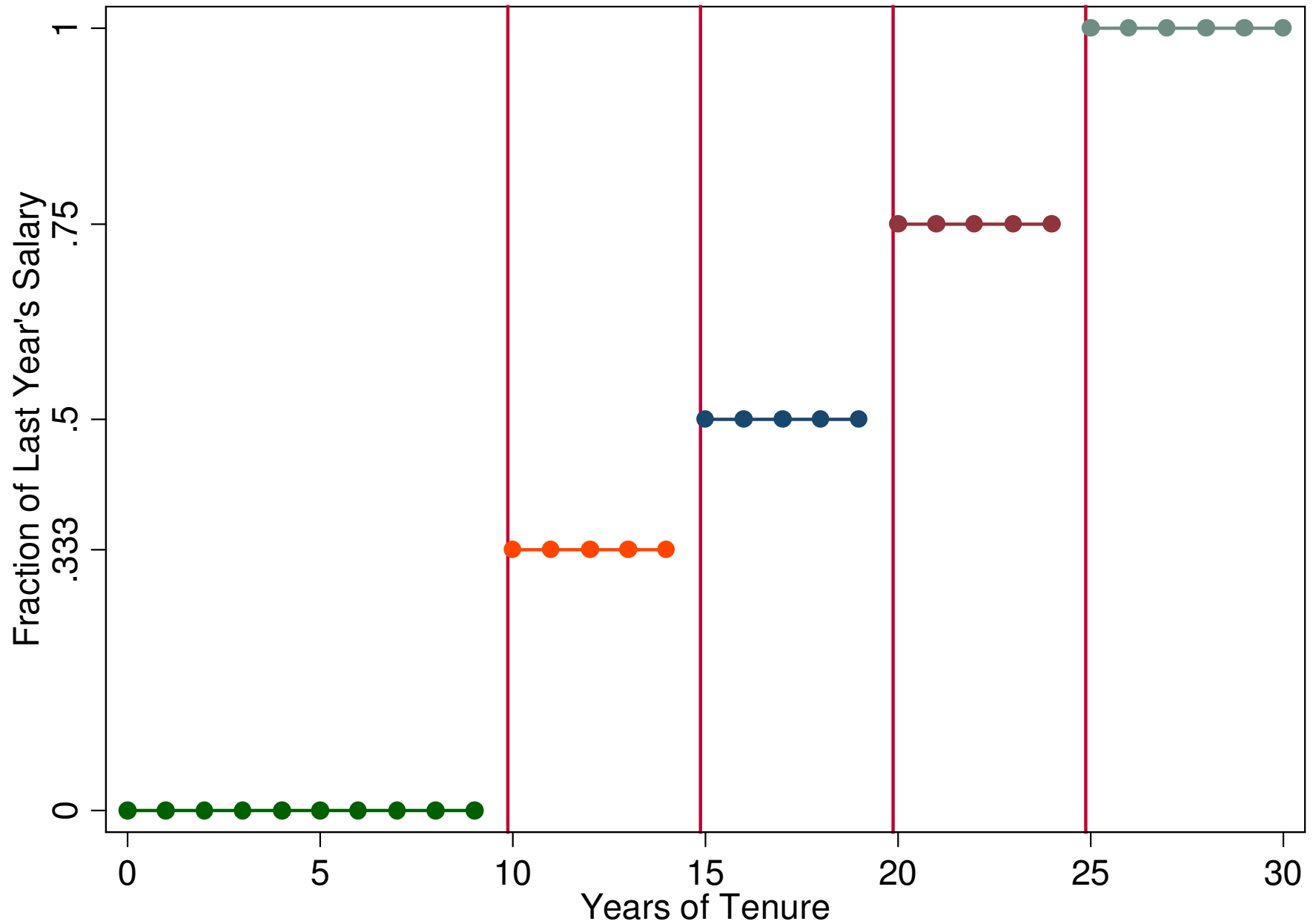


Fig. 2. Exits from Labor Force into Retirement

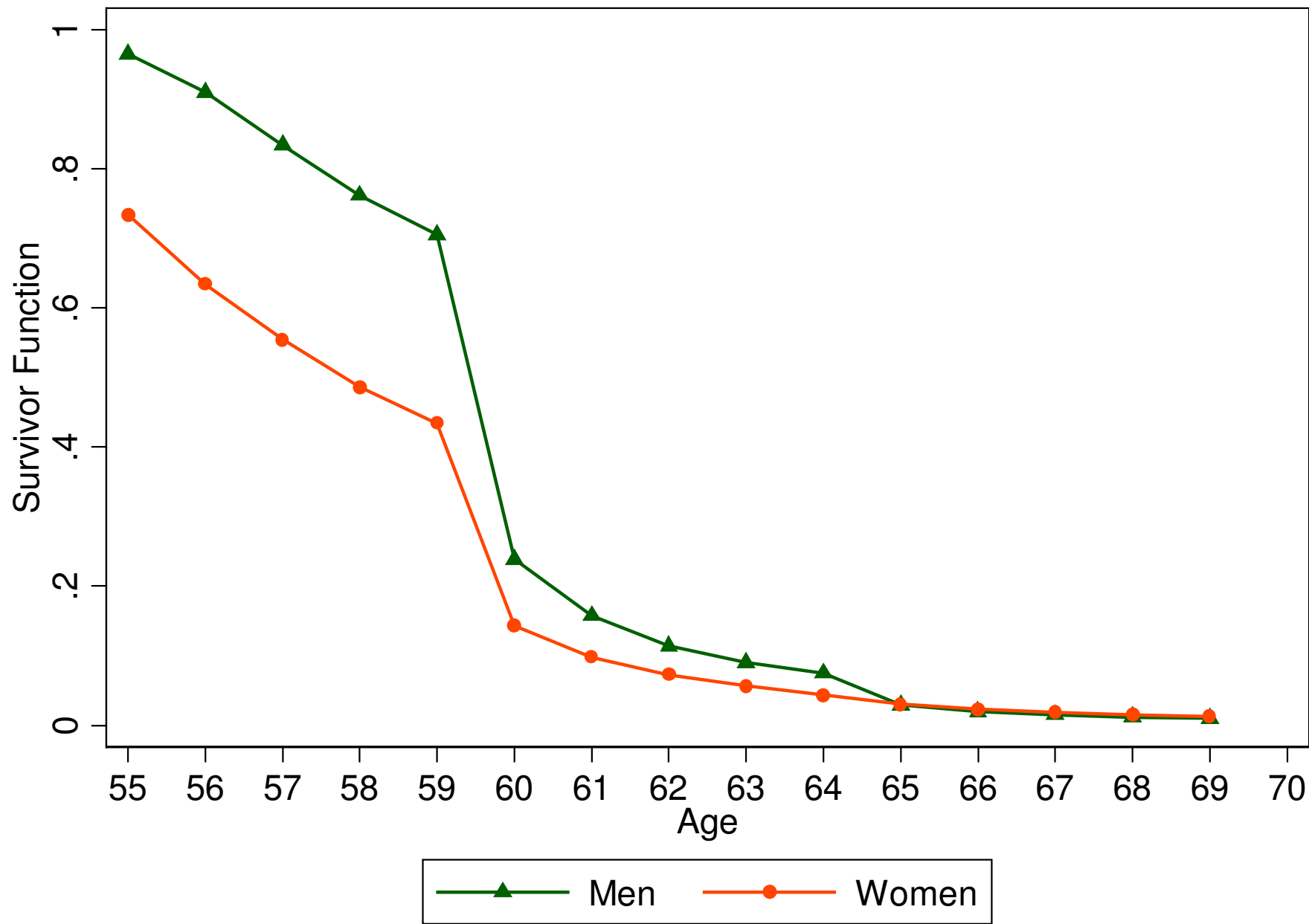


Fig. 3. Distribution of Tenure at Retirement, Full Sample

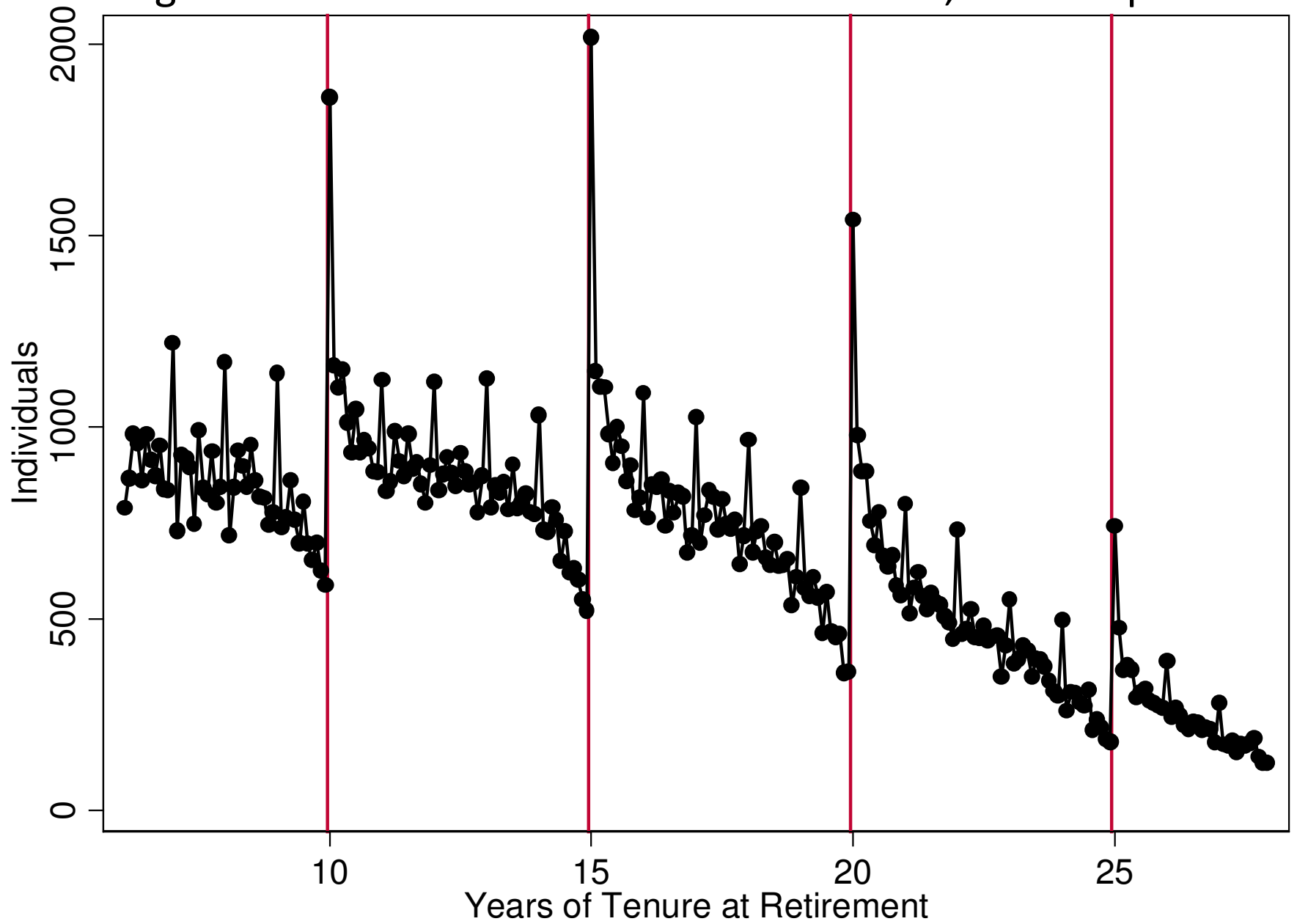


Fig. 4. Controlling for Covariates

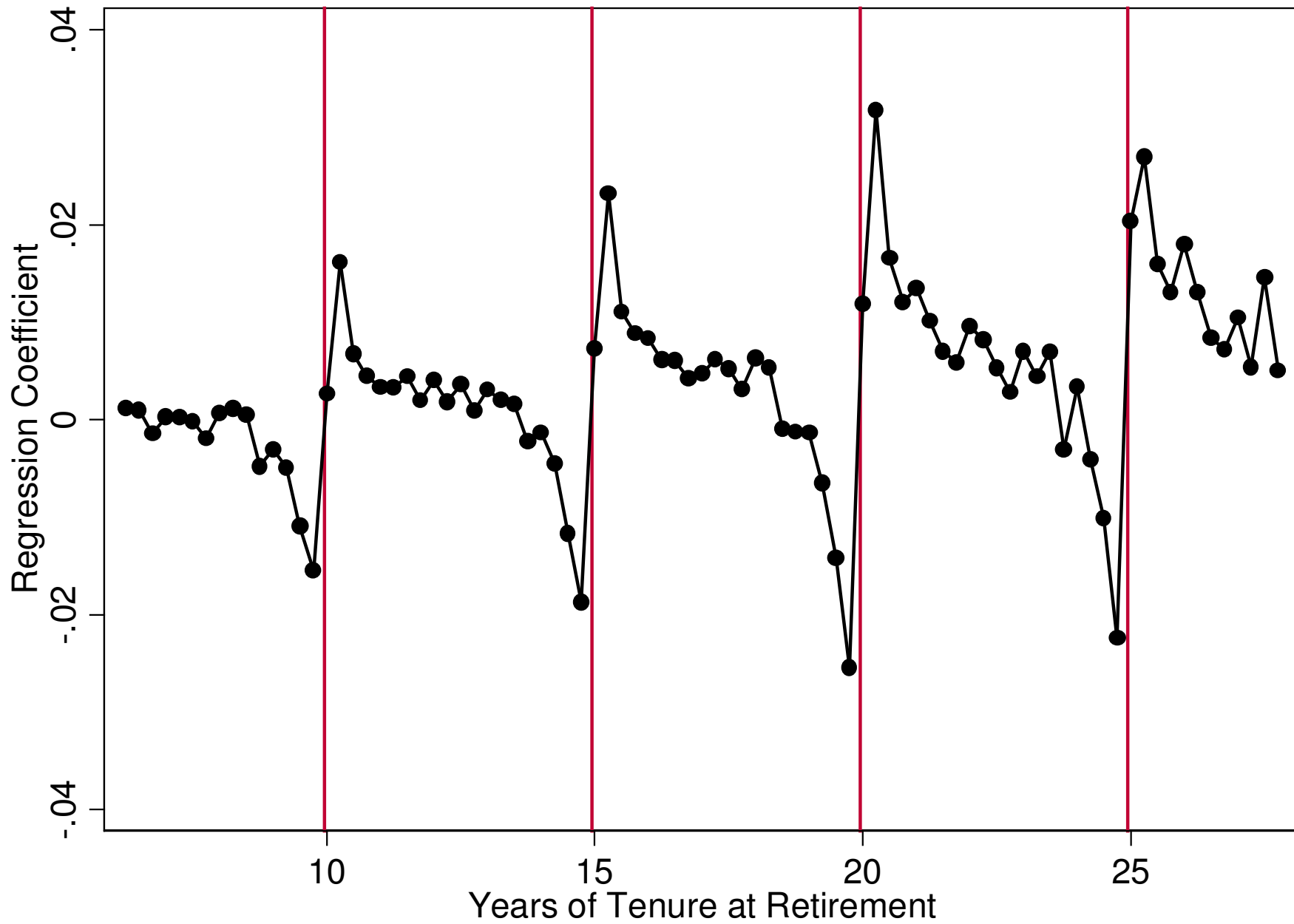


Fig. 5. Job Starts by Age

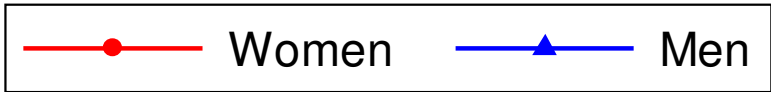
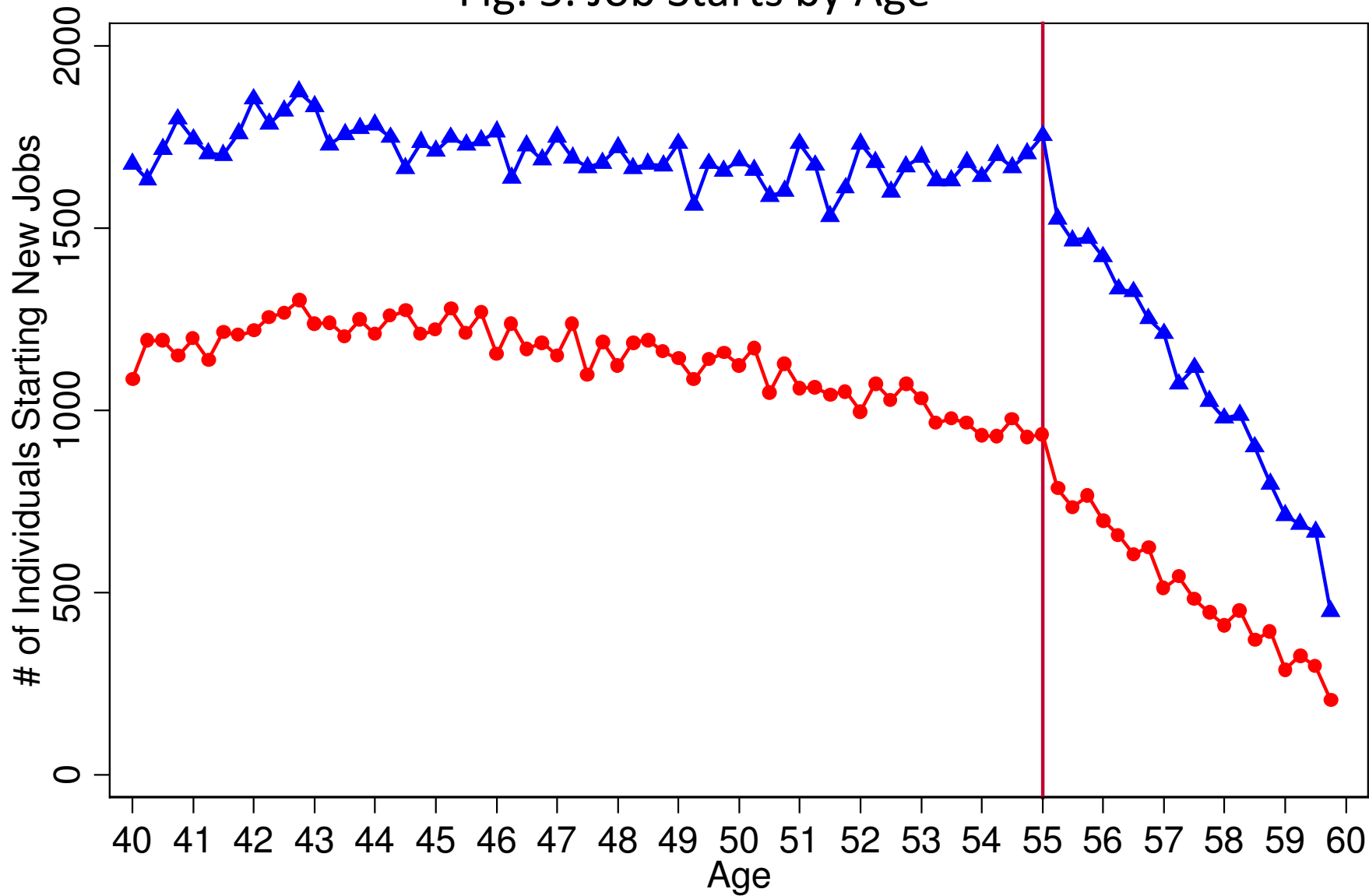


Fig. 6. Distribution of Tenure at Retirement by Health Status

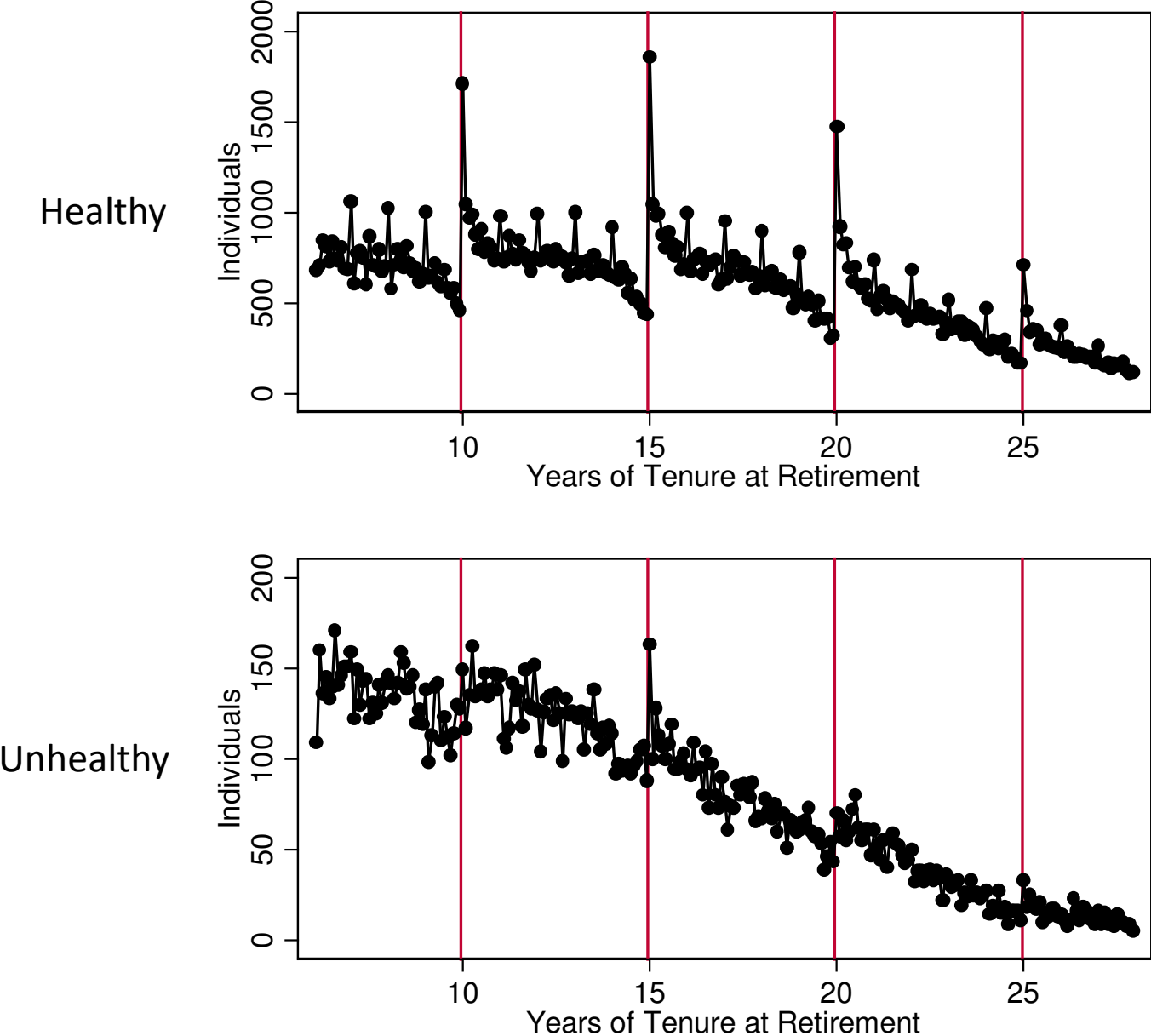


Fig. 7. Retirement at Thresholds by Earnings Groups

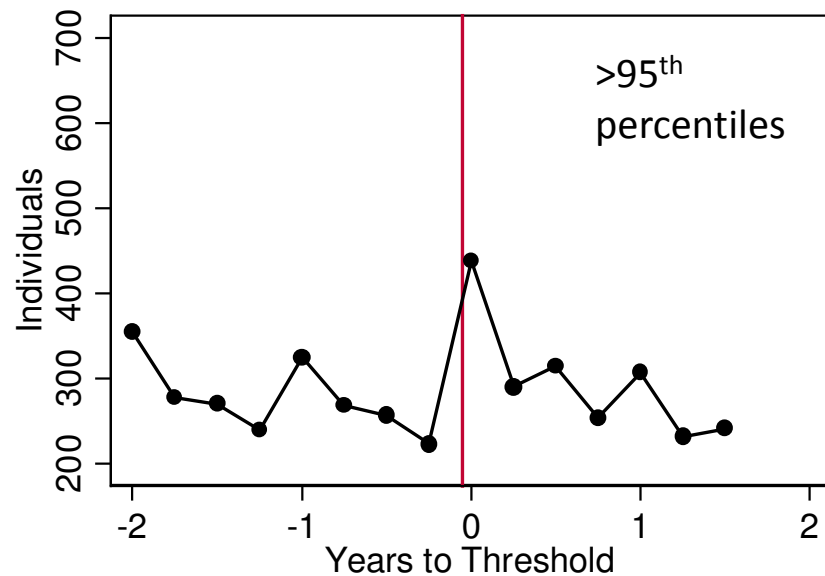
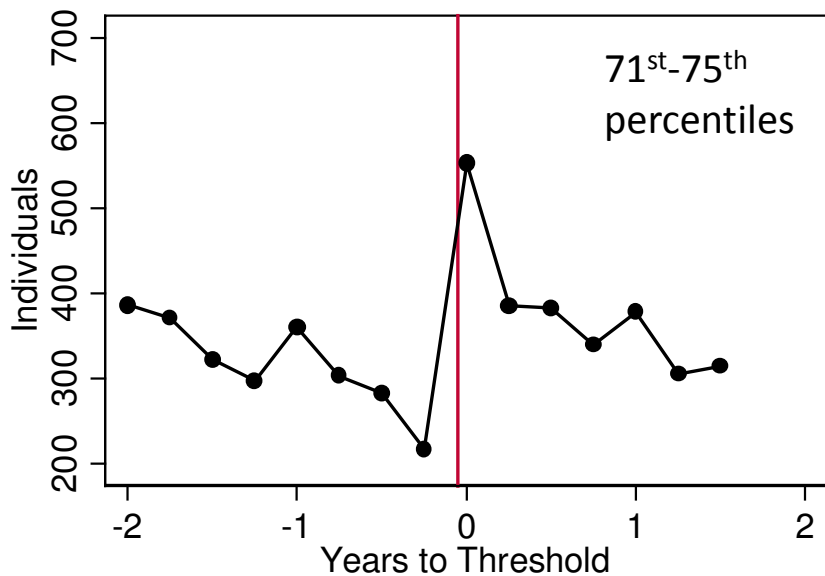
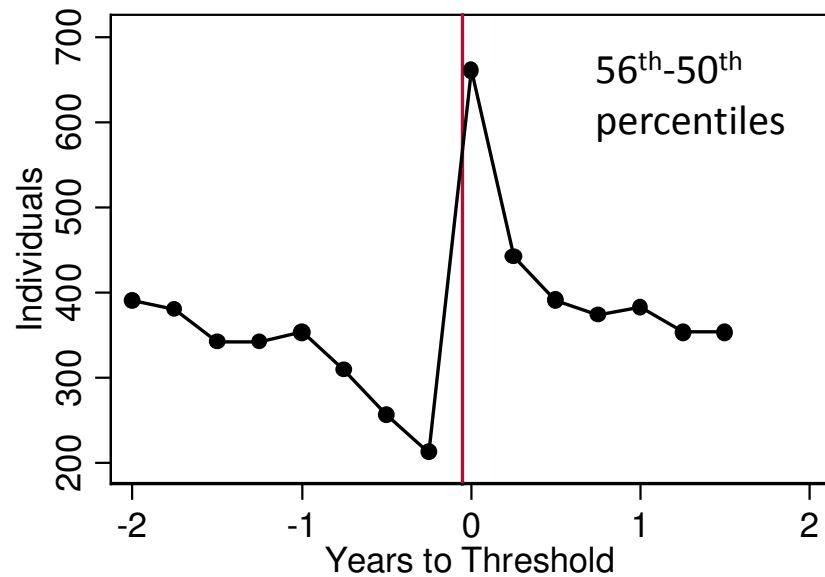
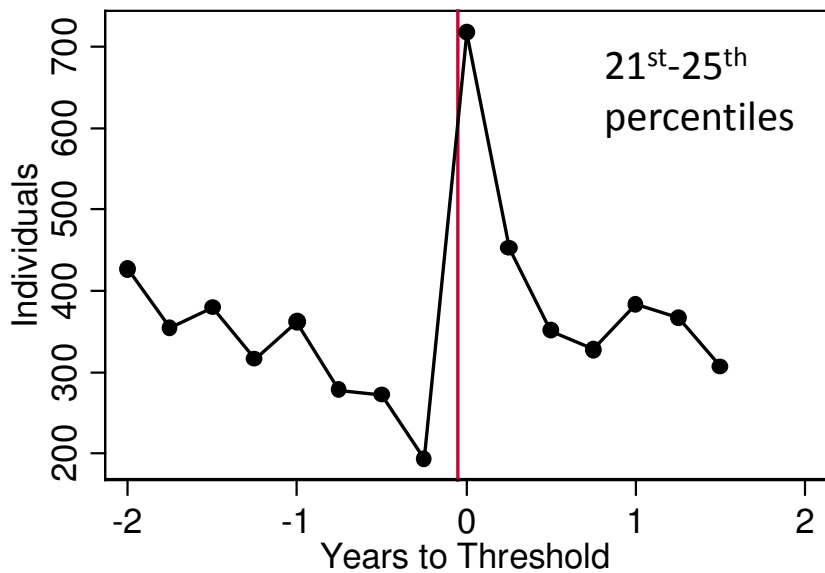


Fig. 8. Distribution of Tenure at Retirement by Firm Size

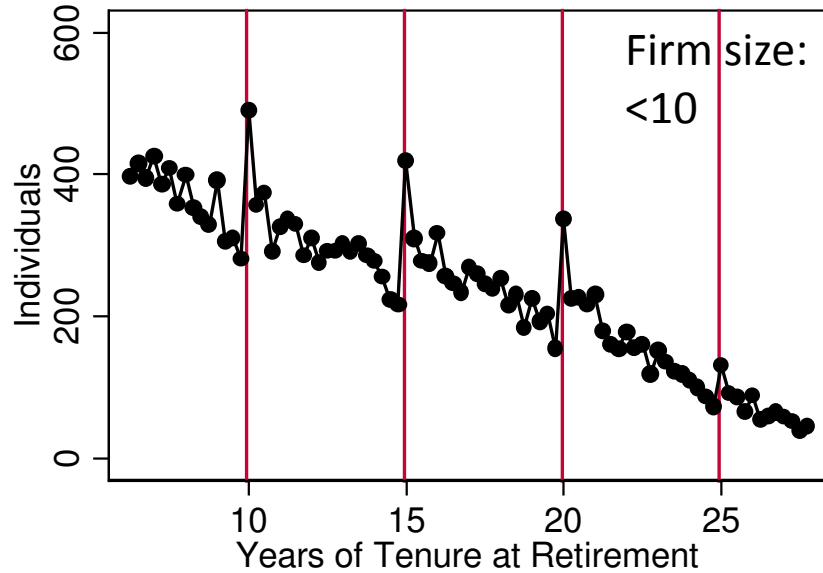
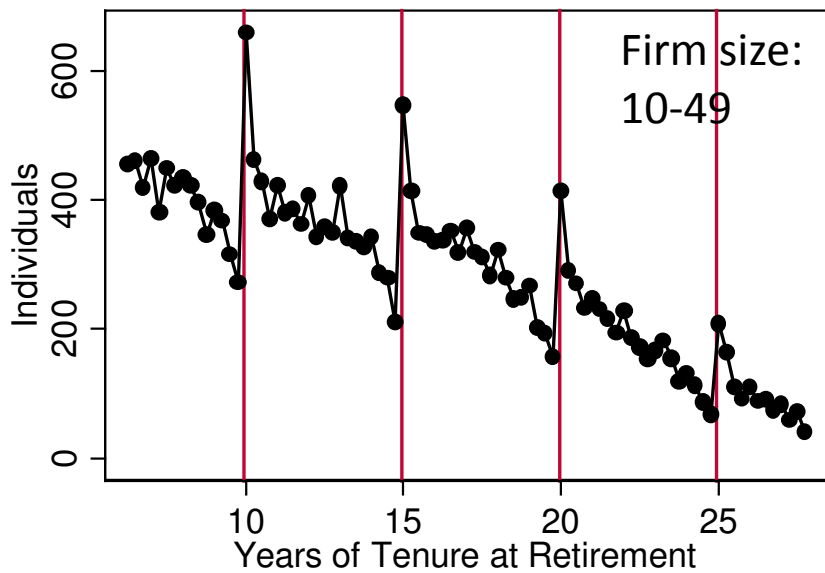
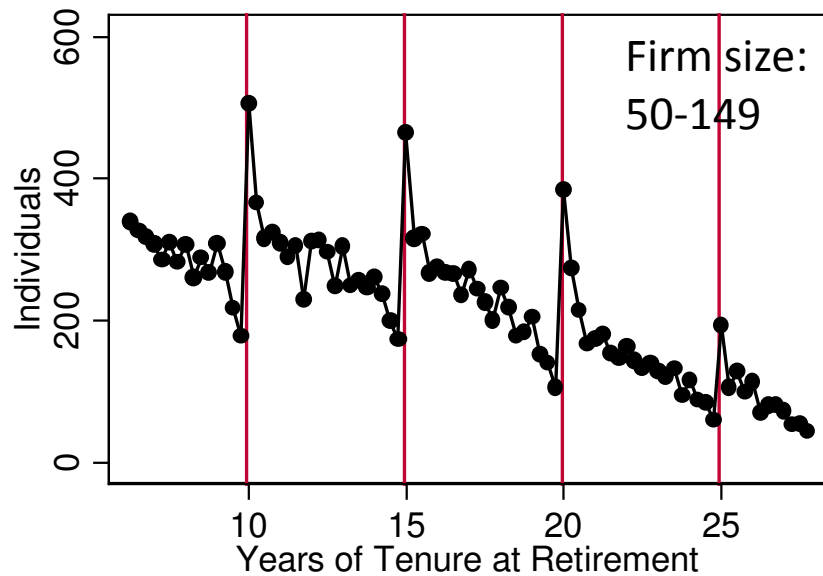
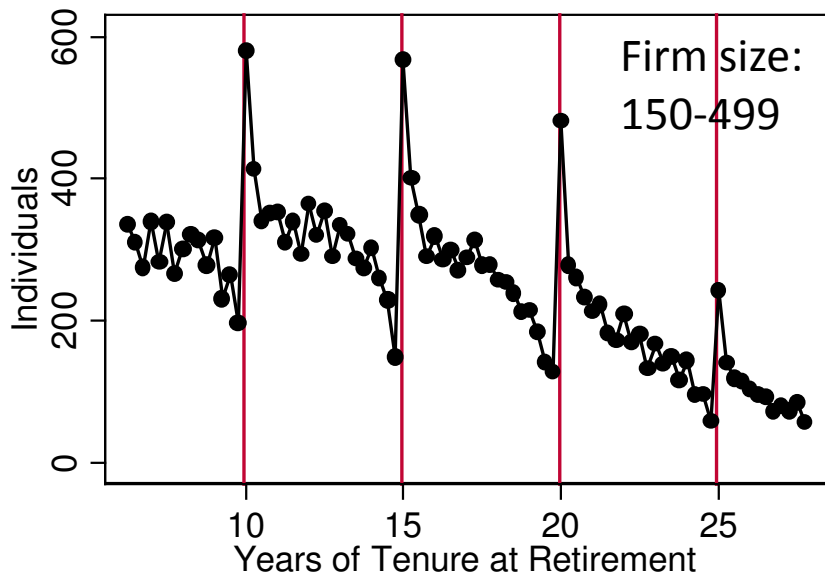


Fig. 9. Retirement at Thresholds by Job Rigidity

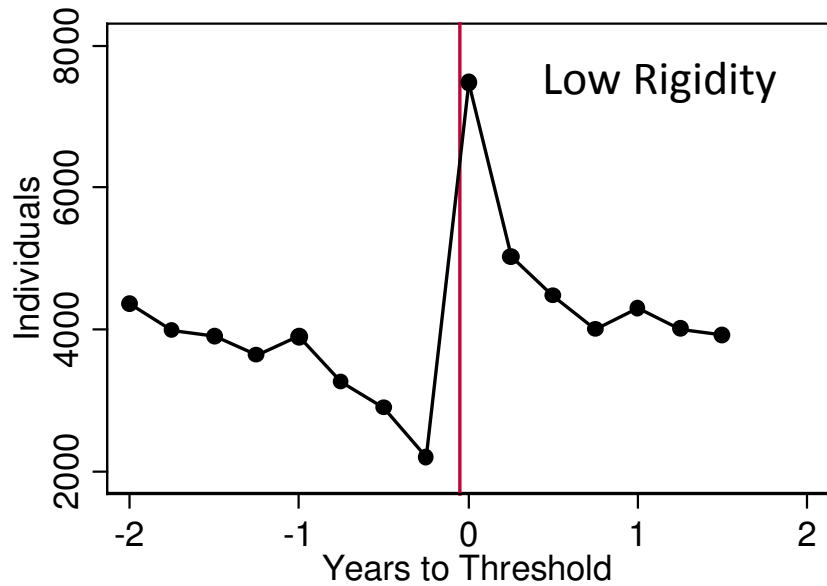
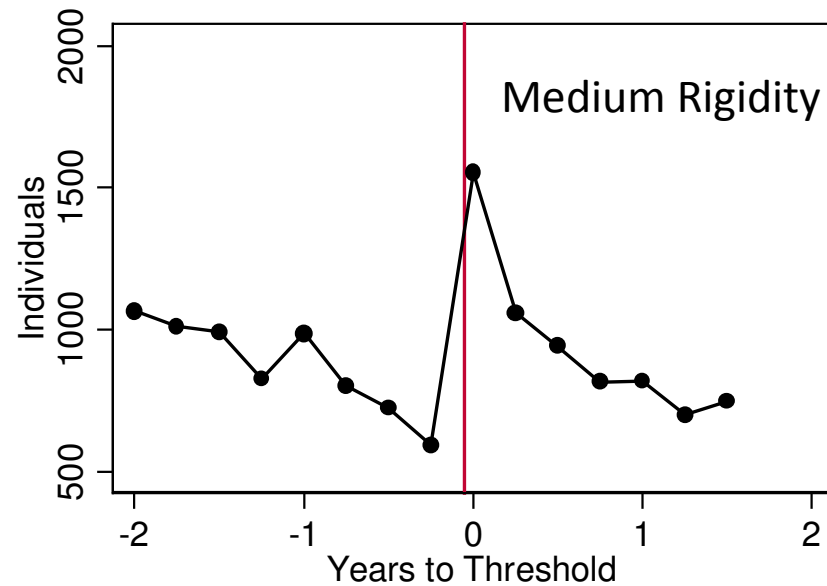
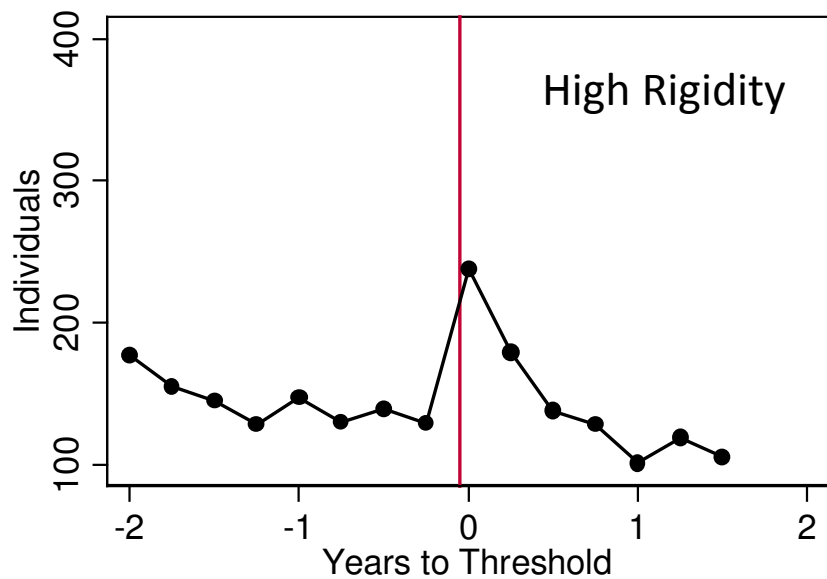


Fig. 10. Retirement at Thresholds, Restricted Sample

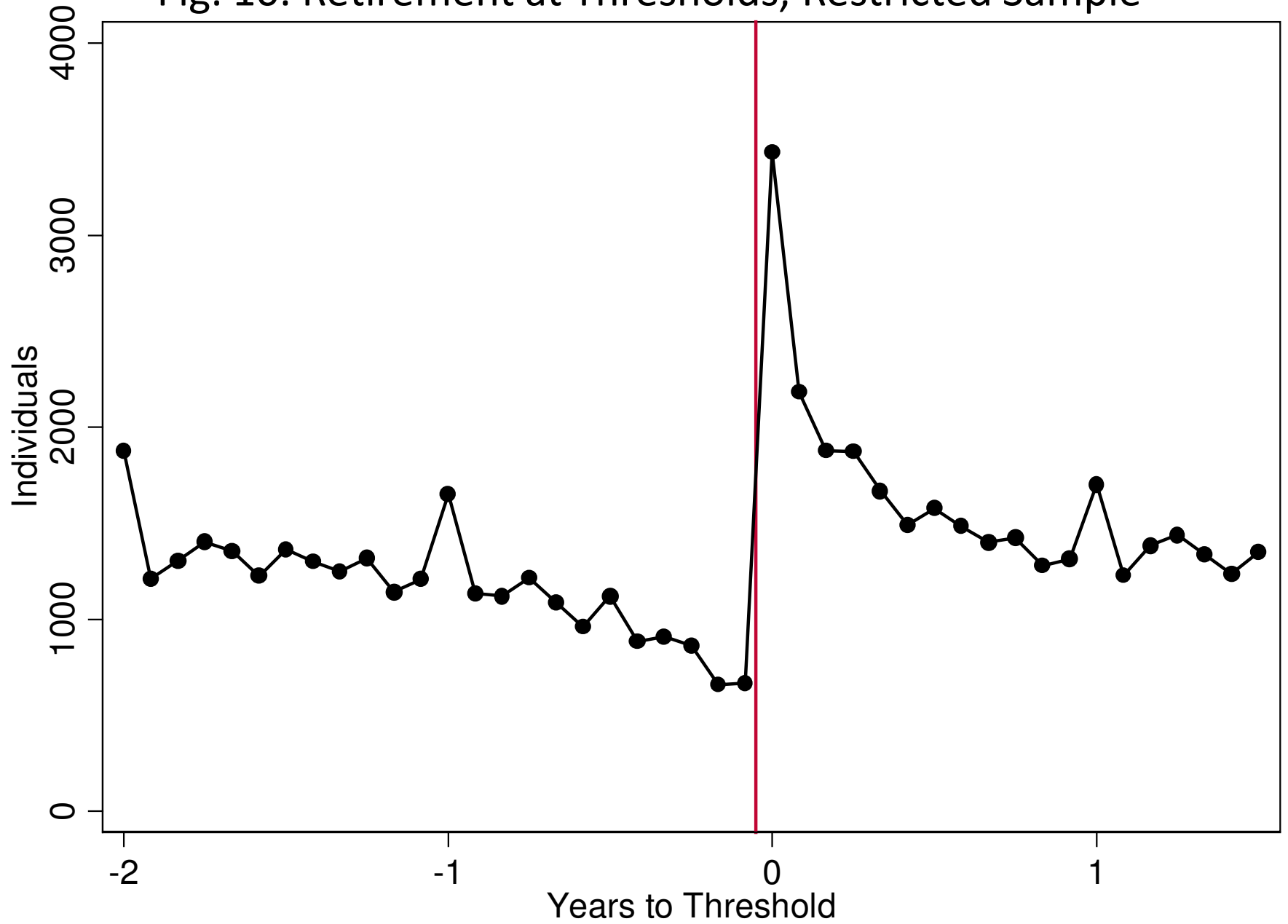
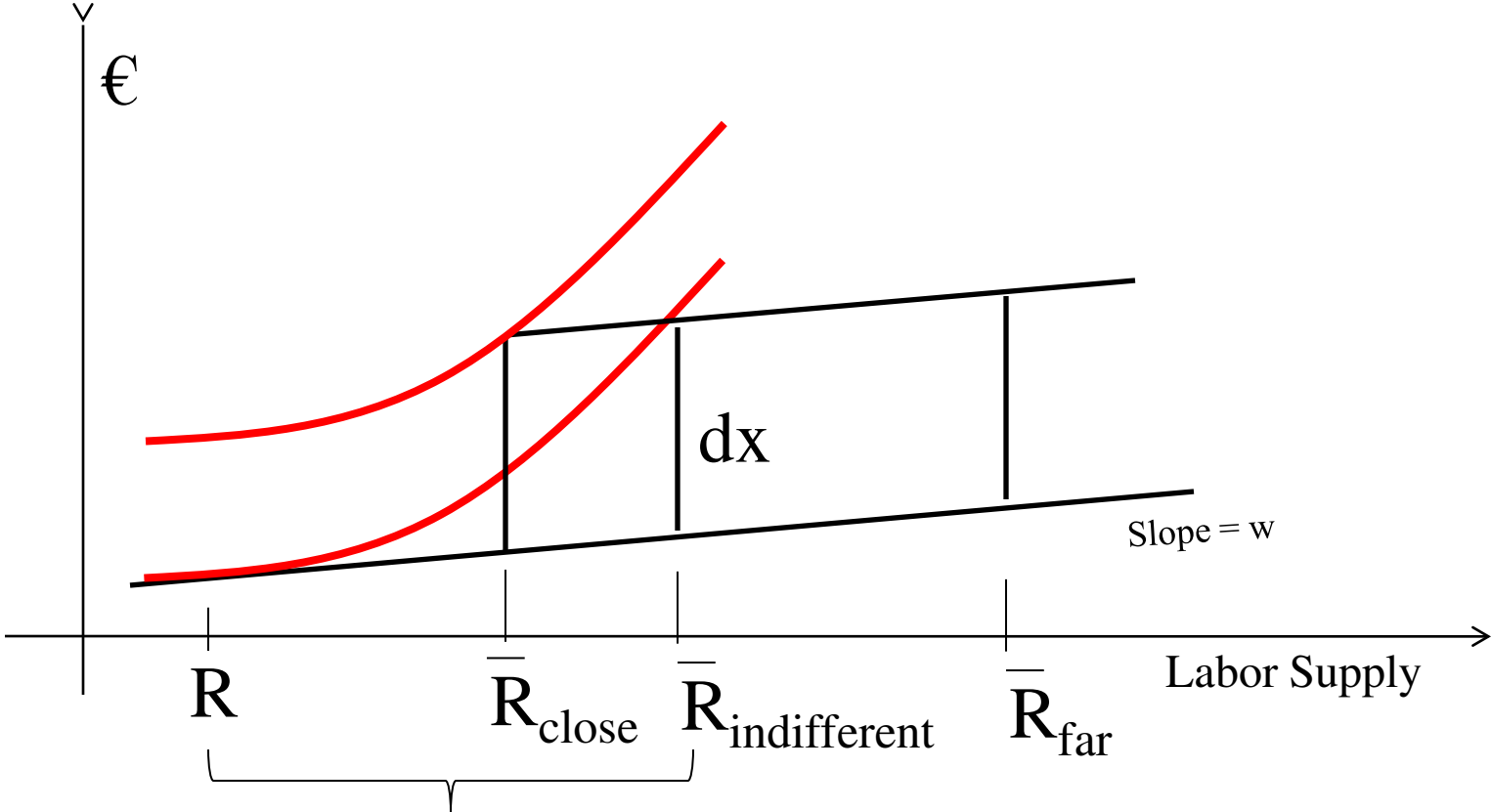


Fig. 11. Optimal Retirement Choices with Severance Pay



Point of Indifference (Max Length of Delay) :
 $\Delta = \bar{R}_{indifferent} - R$

Fig. 12.
Model Simulation

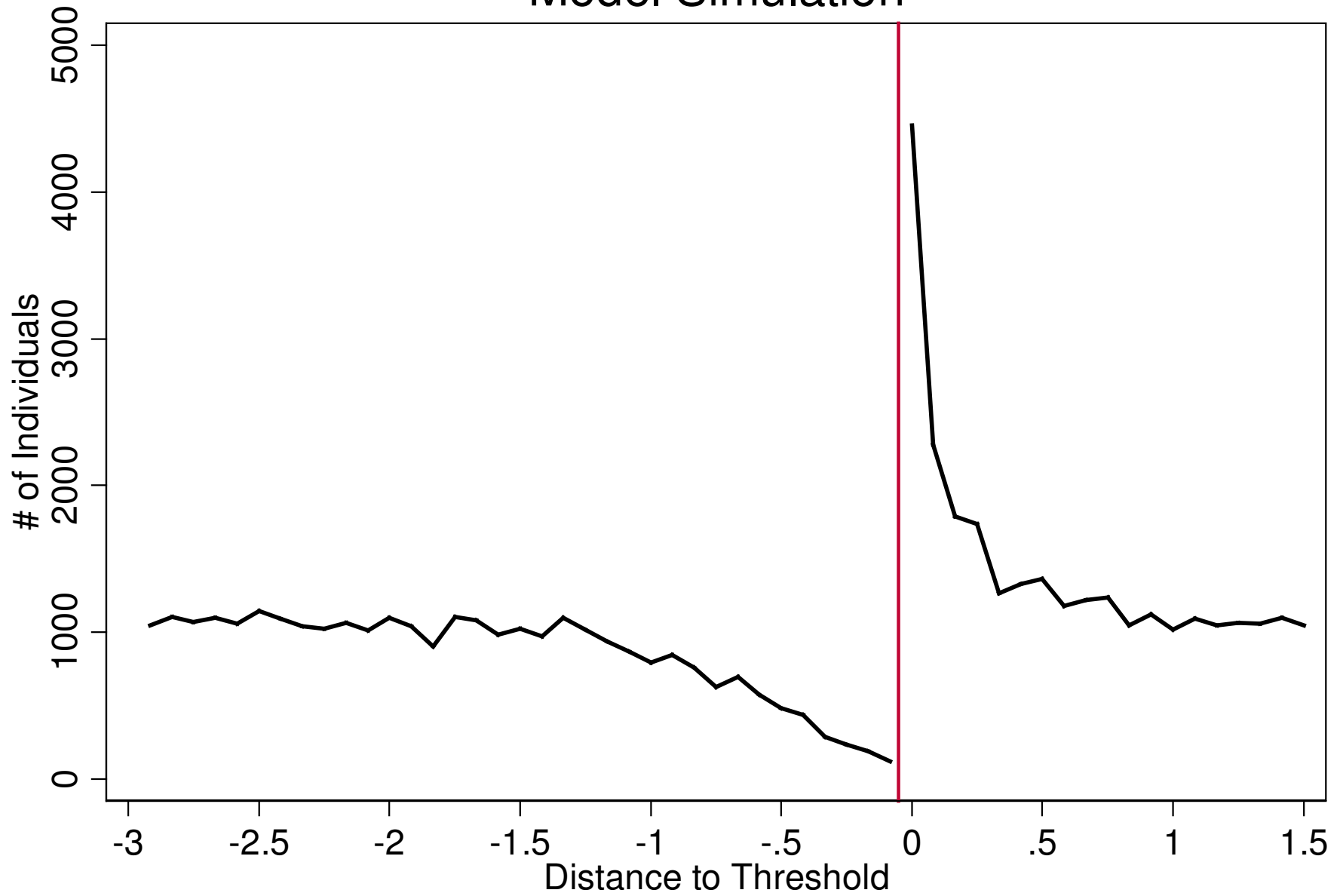


Fig. 13.

Spike & Polynomial Approximation of Frequencies

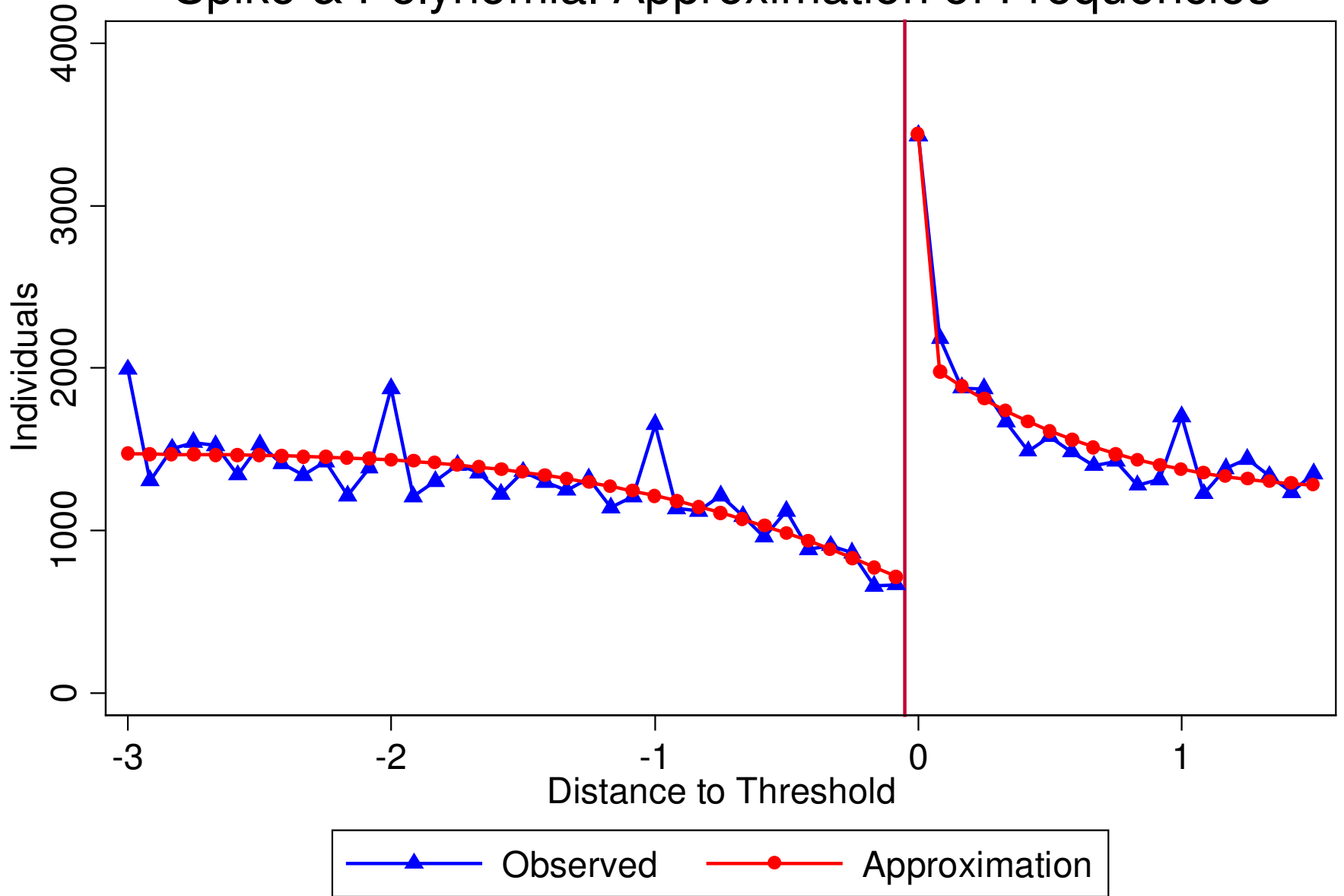


Fig. 14.

Actual vs. Counterfactual Frequencies

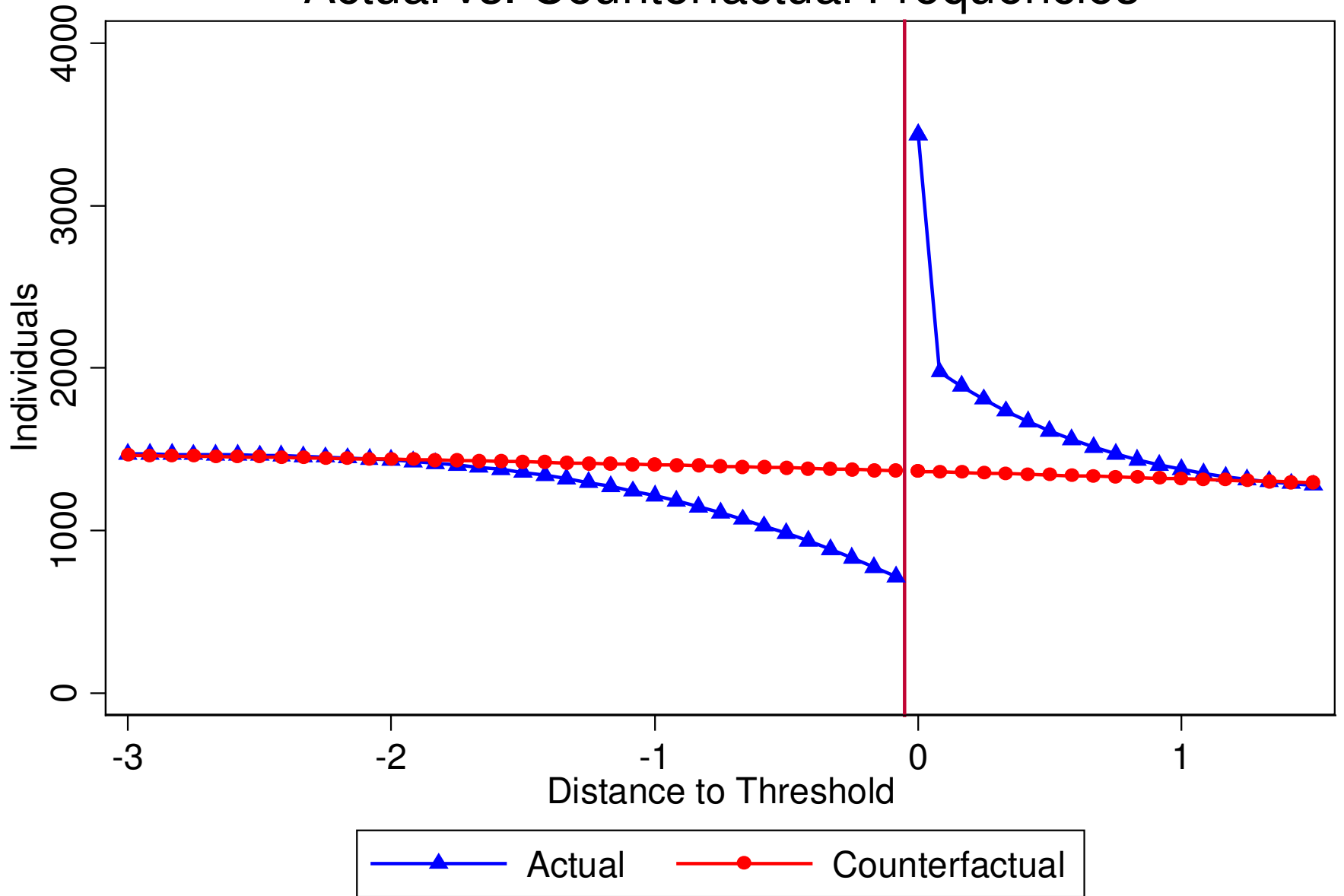


Fig. 15.
Estimation Procedures

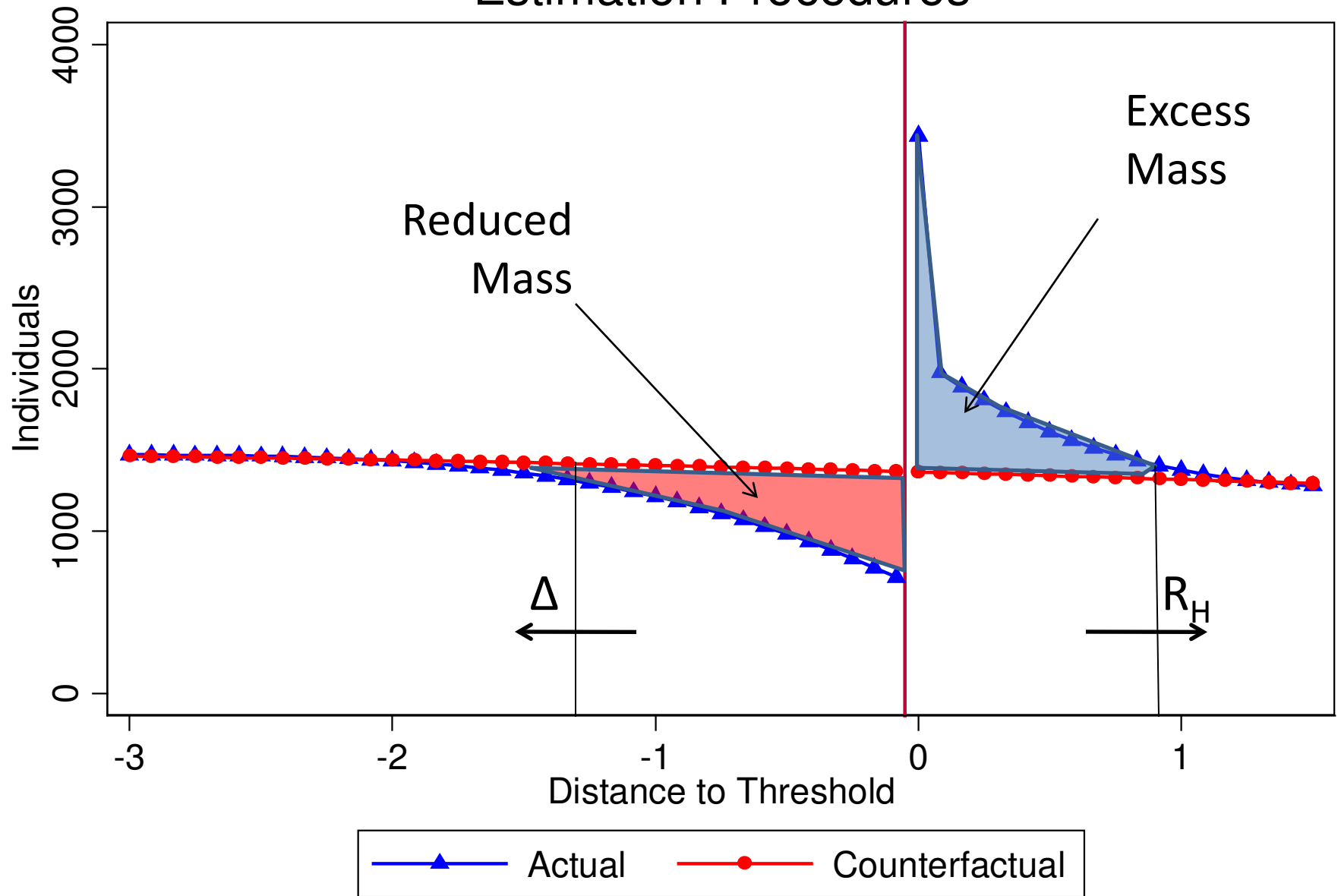


Fig. 16.

Elasticities with Different Utility Specifications

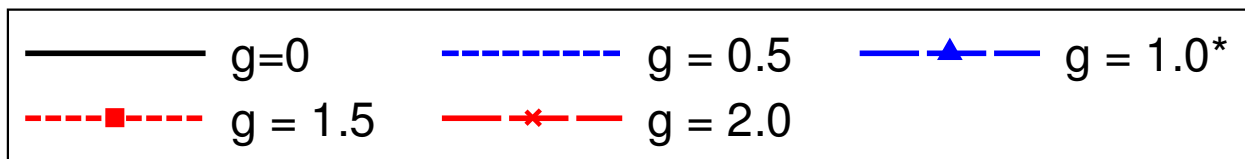
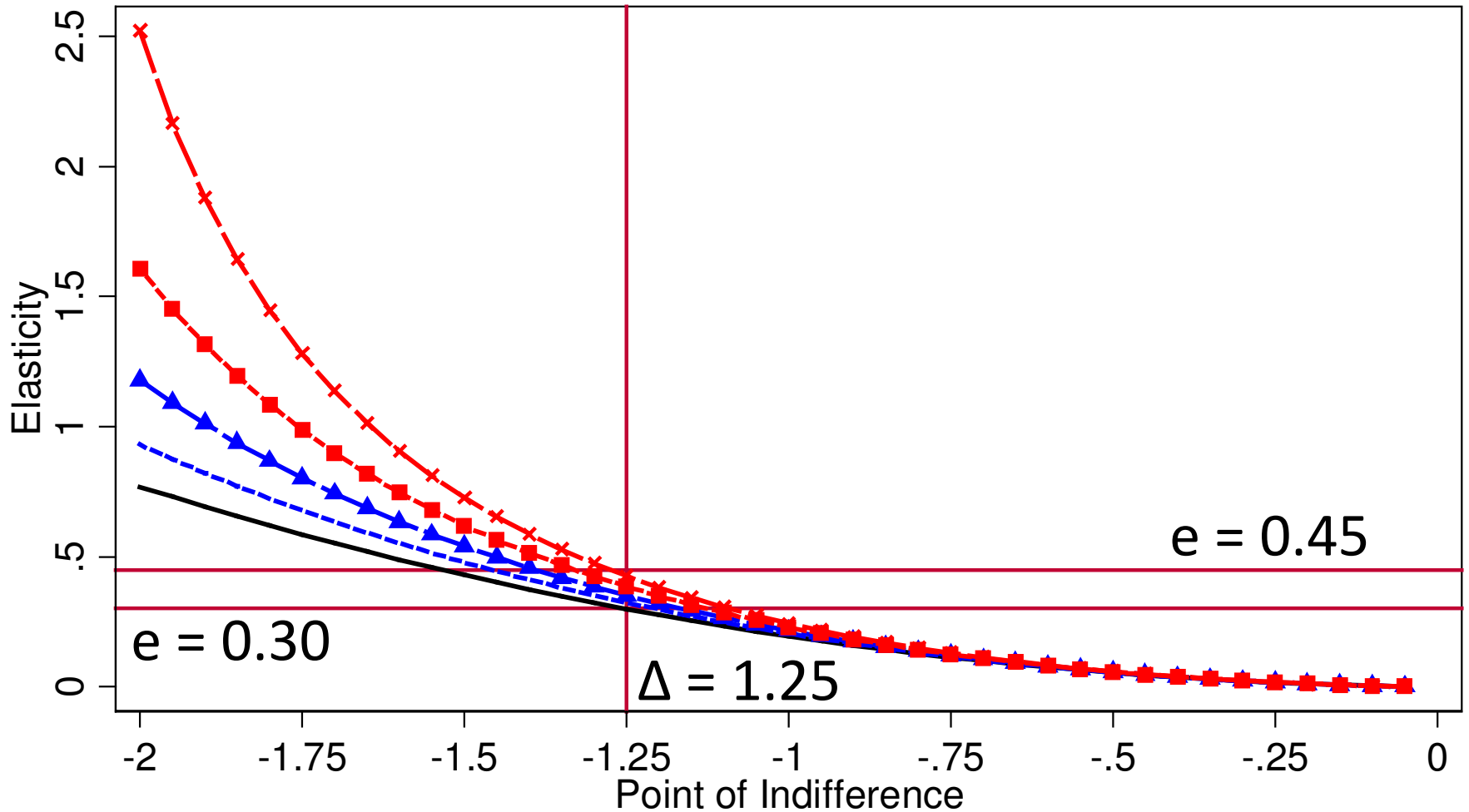


Fig. 17.
Model Simulations

