1. Introduction

A great deal of work has examined the response of workers to pension incentives at the extensive margin (when to retire), and found that workers appear highly responsive to pensions when choosing their retirement date. Yet little work has examined the effects of pension incentives at the intensive margin (how much to work).

In this paper we explore this issue using an individual panel data set of payroll and pension data for workers in the city of Philadelphia. Workers in Philadelphia, especially senior ones, have discretion about whether or not to accept overtime shifts, and overtime work is pensionable. Because pensions in Philadelphia are generally based only on the worker's three highest years of pay, there are very large differences in expected compensation rates between workers for the same overtime work depending on whether that overtime pay ends up being counted in the final pension formula or not. Figure 1 shows the empirical CDF of the ratio of our estimates of the ratio of expected pension compensation to cash wages for the last hour of overtime worked in our sample. Around 25% of workers receive no extra pension in respect of this hour of work, but a small minority of workers receive expected pension compensation for that hour worth more than *four times* their hourly cash pay.[1]

Insert Figure 1 here

To test whether these incentives distort overtime work or not, we exploit differences in pension and cash compensation for each individual across time – which we argue are largely exogenous. We show that while overtime does preferentially appear to be performed by older workers, who have higher pension costs, a non-parametric approach based on Saez (2010) shows little evidence that individual workers are strategically taking account of pension rules when selecting the amount of overtime they do: our preferred estimate of the elasticity of total hours worked to expected pension

1 We calculate expected pension compensation, which is conditional on the information set available to workers (e.g. their expected pay increases, expected date of retirement or separation, likely longevity and likely future overtime). Because we do not have access to this information for individual workers, we condition on the average historical experience of workers. This implies that our estimates of expected compensation are measured with error. For the 25% of workers who receive no extra pension in respect of overtime work, however, our estimate is exact because workers – and us – know with certainty based on their observed historical compensation that this hour of work will not enter into their final pension formula.

compensation at the intensive margin is therefore 0. While consistent with prior work that finds a similar elasticity of time worked to cash wages, this is in marked distinction to the large body of work which shows that workers respond strongly to pension incentives at the extensive margin.

We therefore believe that the overtime allocation patterns we observe are driven largely by the method in which overtime is allocated in Philadelphia, which gives senior workers, who have high pension costs, preferential access to overtime shifts (and the right to refuse these shifts), rather than pension incentives themselves. We estimate the cost of this overtime allocation rule as a percentage of payroll, including pension costs, and examine its effect on pension plan funding, and show that it is significant.

Our main contribution is in estimating the response of total hours worked to pension compensation at the intensive margin. We document the value of expected pension compensation, shown in Figure 1, and use variation in the amount and composition of expected compensation across time for each individual to explore the determinants of overtime work. Our data allows us to examine two issues: first, how worker overtime responds at the intensive margin (i.e. how much to work) to changes in total compensation, in an environment where workers do have some control over their hours, and secondly, whether changes in pensions have a different effect on overtime worked to changes in cash wages, and why. Our results on the elasticity of time worked with respect to cash wages appear broadly in line with results in the literature. Surprisingly, however, standard regressions appear to show that workers are more responsive at the intensive margin to pension compensation than they are to cash wages. These findings are robust when we account for the fact that pension compensation is upward-sloping in the number of hours of overtime worked, and when we allow for the fact that different workers may be constrained in the quantity of overtime work they do.

This is puzzling because most academic work finds that workers are highly uninformed about their pension plans, and often fail to make even simple choices that would significantly increase their welfare where pension plans are concerned. We therefore investigate this issue in more detail, to test whether it could be the result of the overtime allocation rule rather than strategic choices by individual workers. We adapt the non-parametric approach of Saez (2010) to the case where kinks in budget constraints are convex rather than concave, and show that this will cause utility-maximizing workers to strategically avoid the area around the kink (which we call 'dearthing'), the opposite of the bunching effect examined by Saez (2010) around concave kinks. We select a group of workers from our data who theory suggests should have the strongest dearthing effect – those in their last full year of work before electing DROP or retiring, who work in departments where there is lots of overtime and who elect to do at least one hour of overtime in that year. We then examine the number of hours of overtime these workers work relative to the amount needed to match their own fifth-highest, fourth-highest, third-highest etc. annual wages up to that point. Our theory would predict that because pension rules cause the effective overtime wage of workers to jump discontinuously when they cross the third-highest, second-highest and highest wages, for these workers it is optimal to cross the discontinuity and keep working more overtime, avoiding the area around the kink. When workers cross the fourth- and fifth-highest wages, however, there is no discontinuity in wages and no such effect should obtain. We show, however, that in this group of workers, dearthing appears entirely absent at all wage thresholds. The implication is that worker overtime appears entirely inelastic with respect to pension compensation. We conclude from this that the response to pension incentives we document in our standard regressions is likely due to the overtime allocation rule itself, rather than strategic worker responses to pension incentives.

Starting with Fields and Mitchell (1984), a large body of work has examined the relationship between pension and social security incentives and labor supply at the extensive margin (when to retire). Excellent reviews can be found in Coile and Gruber (2007) and Coile (2015). Focusing specifically on state and local workers, Costrell and Podgursky (2009) and Koedel et al (2013) examine the incentives provided by teacher pension plans and consequences for school staffing; Kim et al (2021) examine the relationship between teacher retirement and pensions, and Fitzpatrick and Goda (2020) examine how COLA's influence retirement in public sector plans. Morrill and Westall

(2019) examine how social security and teacher retirement behavior interact in determining teacher retirement, and Ghent et al (2001) examine phased retirement among teachers. This work confirms that individuals in the public sector respond strongly to pension incentives at the extensive margin.

Likewise, there is a huge body of work examining the response in hours worked to changes in cash wages (see, for example, Bargain and Peichl, 2016, or Blundell and MaCurdy, 1999, for comprehensive reviews in the US and abroad).[2] Our preferred estimate of the cash wage elasticity of total hours worked across our whole sample is negative but close to zero, consistent with the consensus. Older workers, however, may have higher elasticities, as documented by Goda, Shoven, and Slavov (2007), and by Hudomiet, Hurd, and Rohwedder (2018), because they have the luxury of not working at all in response to lower wages, and because, many being part-time workers, they may have control over their hours. However, other than two papers dealing with social security or its equivalent,[3] we have been unable to find a paper discussing the effect of pension incentives on hours worked at the intensive margin.

Our second contribution relates to pension cost analysis and plan funding in public policy.[4] The Bureau of Labor Statistics (2020) reports that around 90% of US state and local employees are members of defined benefit (DB) pension plans, and that around 11.3% of their total compensation is in this form (although given the discount rates used, this is likely a large underestimate)[5]. State and local DB plans are, however, significantly underfunded: according to data from the Center for Retirement Research

2 Estimates vary widely from study to study, but surveys of the literature conclude that the wage elasticity of the male labor supply is around -0.1. See Chapter 2 of Borjas (2010) for an overview of the different methodologies of estimation & why estimates differ across studies.

3 Liebman et al (2009) document that the elasticity of hours worked to a measure of the change in social security benefits net of taxes is 0.42 for those earning less than the social security tax threshold in their base case (the estimated effect is smaller and less statistically significant as more controls are included), while Danzer (2013) shows that an exogenous doubling in the legal minimum pension level of Ukraine in 2004 resulted in a significant reduction in weeks worked per year, largely for women.

4 Shapiro (1985) was an early advocate for accurate cost analysis of pension plans in the public and private sectors.

5 BLS figures are based on the reported normal cost of the benefits, a figure that is calculated using a discount rate based on the expected return on plan assets. Many economists (see footnote 3) regard this as far too high. Using a lower discount rate would raise the value of DB pension benefits in absolute terms and as a proportion of overall compensation.

(CRR) at Boston College[6], the most recent estimate of state & local DB pension underfunding is $1.6trn (although again, Novy-Marx and Rauh (2009, 2010), and others argue that this is a significant under-estimate due to discount-rate choices.[7]) In recent years, required employer contributions to these plans have tripled as a percentage of payroll, to more than 19%,[8] placing municipal and state finances under strain in many places. By analyzing the available plan rules of plans in the CRR database[9], we estimate that only 6% of state and local plan liabilities are in plans where overtime is specifically excluded from pensionable compensation. In the remainder of plans, overtime is either specifically included (around 30% of liabilities are in plans where overtime is included for at least one plan section), or the precise treatment of overtime is unclear. Pensionable overtime may therefore be a significant driver of pension costs in many state and local plans in the US.

In our sample, paying pensions on overtime raises pension costs by around one-fifth, and costs around 2.9% of payroll. Of this, our work implies that around one quarter – around 0.7% of payroll, or 3.8% of total pension costs – is purely the consequence of the overtime allocation rule.[10] In Appendix A, we use our analysis of the plan rules in the CRR database to show that plans where overtime is pensionable are around 7% less funded than plans where overtime is specifically excluded, or where the treatment of overtime is unclear, even after controlling for the discount rate, past under-payment of recommended contributions and plan generosity. This would make pensionable overtime one of the more significant factors affecting plan funding identified by Munnell et al (2008) in their analysis of the funding of state and local pension plans (see their figure 7).

---

6 Data accessible at https://publicplansdata.org/public-plans-database/download-full-data-set/. Downloaded 20/11/2019.

7 Novy-Marx and Rauh (2009, 2010) argue that using a discount rate that reflects the risk of the liabilities (rather than the risk of the assets used to back them), and the same apportionment method as used in private-sector plans (the Projected Unit Credit method) indicates that deficits were around $3trn in 2008. An options-based approach that measures the extent of the risk taken on by the funding agency reaches a similar conclusion (Biggs, 2010). Other measures are summarized by Elliot (2010).

8 The figure refers to the required employer contribution rate as a percentage of payroll, averaged by fiscal year and weighted according to plan assets for all plans in the CRR state and local plan database.

9 We were able to find plan rules for 180 of the 188 plans.

10 This is measured relative to what pension costs would be if overtime were allocated equally to all employees. Employers could reduce overtime pension costs much further – even possibly to zero – by allocating overtime strategically to workers with low-pension costs.

A mechanism by which this happens is suggested by the fact that we were only able to find two examples, out of 180 plans in the CRR database, where plan actuaries makes allowance for pension spiking when projecting the costs of the plan. For the other plans, changes in worker behavior that we document here will raise pensions above the level projected by the plan actuaries, reducing long-term funding and therefore raising deficit reduction contributions. A consistent underpayment of 0.7% of payroll could increase pension underfunding by as much as 10% of liabilities over the long term. Our finding that the overtime allocation rule itself, rather than strategic choices by individual workers, drives the majority of the cost increase suggests that employers could reduce compensation costs by focusing on the total cost of overtime allocation rules when bargaining with workers, rather than needing to alter the pension rules themselves.

Several causes of underfunding have already been examined in the academic literature: insufficient contributions relative to benefits in past years (Elliot, 2010), accounting standards that allow cities and states to take advance credit for investment returns not yet earned (Novy-Marx and Rauh, 2010), governance problems in states and cities in the US related to political connections between political parties and public-sector trade unions (Anzia and Moe, 2014), the impact of scale, the complexity of pension plans, and service quality on administrative costs (Bikker *et al*, 2011), moral hazard problems associated with public finance more generally where one aspect of compensation (pensions) is opaque and not easily understood by voters (Glaeser and Ponzetto, 2014), moral hazard problems associated with the pooling of pension costs across administratively independent state entities (Fitzpatrick, 2017), and scheme-specific longevity selection risk (Fong et al, 2015). But while anecdotal evidence about pension spiking is common,[11] no-one has examined how the interaction between pension rules

11 Besides the newspaper reports in figure 1, examples known to the authors include police in a large northeastern city allocating overtime, and employees of a state university in the midwest allocating summer teaching, both explicitly on the basis of individual workers' retirement plans.

and workplace practices, such as we examine here increases pension costs in ways that may be quite difficult for policymakers (and voters) to detect.

This paper proceeds as follows. Because our paper rests on the significant non-linearities in overtime compensation caused by pension plan rules, in the first section we summarize pension plan rules, show how we derive our measure of expected pension cost, and present some indicative results to aid understanding. In the next section, we summarize how overtime is allocated in practice in Philadelphia and other cities and states. We then discuss the data we use before turning to our empirical analysis. In the empirical section, we first use standard regressions to examine the association between overtime compensation and hours worked, and perform some robustness checks. To test whether these results are the consequence of the overtime allocation rule or individual worker choices, we then use a non-parametric approach based on Saez (2010) to analyse the behavior of a select group of workers who are very close to retirement. After this, we measure the effect of changing workplace rules and compensation on employment and pension costs. The paper closes with a discussion and a conclusion. To ease readability, we relegate technical issues to appendices where appropriate.

2. Measuring the value of pension compensation

In this section, we first discuss pension plan rules, and then turn to how we use them to measure the value to workers of the additional pension payments they earn in respect of overtime work.

2.1 Description of pension plan rules

Philadelphia's plan for municipal workers has several sections, called the 1967, 1987, 2010 and 2016 plans, referring to the years in which these sections were first created. Newly-hired workers are enrolled into the currently "open" pension plan section (for hires employed after 2017, it is the 2016 section), and remain in that section for the rest of their careers. Benefits in existing plan sections have never been changed. Between 2010 and 2017, new employees were offered a choice between the 2010 and 1987 sections; most chose the 1987 section). Over time, the generosity of the pension rules in each section has fallen in response to rising pension costs and the consequent financial pressures on the city. Major features of the sections – the accrual rate, the definition of final salary used to determine pensions, and the maximum possible retirement pension receivable under the section rules – are summarized in Table 1. These were obtained from the plan summary (Philadelphia, 2017b) and the Philadelphia City Code (Philadelphia, 2017c).[12]

Insert Table 1 near here

The broad features of all section rules except the 2010 section (of which few current employees are members) are very similar, and differ only in technical details. Although we adjust the formulae we present here where necessary in our calculations to account for the rules of the different sections, to aid readability, we focus here only on the 1967 section. In that section, to receive a pension, workers must have served longer than 10 years. Once they have passed this threshold, their benefits vest, and on retirement they receive a lifetime annual pension of 2.5% of a measure of their final salary, S_i, for each year of service up to 20 years, and 2% for any longer service. For this section, S_i is defined as the average of a worker's three highest years' earnings over their entire career, including overtime pay. The pension cannot exceed 80% of S_i (so only 35 years of service are used for the calculation). The annual pension after retirement for municipal workers in this section is therefore calculated as:

$$Y_i^P = \alpha(N_i)S_i, \tag{1}$$

where Y_i^P is the annual pension benefit of the *i*'th worker after they have retired, $\alpha(N_i) = \min(0.025 \times \min(N_i, 20) + 0.02 \times \max(N_i - 20, 0), 0.8)$ if $N_i > 10$ and $\alpha(N_i) = 0$ otherwise, where N_i is the number of years of pensionable service the worker has completed on the day they retire.

12 In this paper, we focus only on plans for non-uniformed municipal employees, leaving the analysis of uniformed workers for further work. Uniformed employees are subject to different work rules, different pension arrangements and different supplementary compensation arrangements, making outright comparisons challenging.

In this and the other sections, workers can also elect what is called the Deferred Retirement Option, (DROP), once they have reached normal retirement age and their benefits have vested (so they have accrued more than 10 years' service). When an employee elects DROP, their pension benefit is calculated as of the day before they elect the option and is frozen at that level. After this date, they can continue to work for the city, and earn the same salary as before, for up to four years (more years are possible under exceptional circumstances), but their pensions are paid into an escrow account, which accrues with interest. When they actually leave the city payroll, the accrued value of the escrow is paid to them as a lump sum and they receive their monthly pension payments, as originally calculated, from that point onwards. Employees who have elected DROP can still work overtime, but because they no longer accrue new benefits in the pension plan, their pension will not change as a result of any salary they receive in respect of it (or any other changes in their salary).

2.2 Measuring workers' expected compensation for overtime work

We now turn to how these rules can be used to measure the value of the compensation that workers can expect to receive as a result of working a single extra hour. We can use payroll data to estimate the hourly cash wage worker i receives for overtime in year t, which we call $OTCashWage_{i,t}$. However, because overtime is pensionable for the group of workers we examine, their total compensation includes the present value of the extra pension benefits they expect to receive in respect of an extra hour's work. We assume that workers know the length of their past service (denoted $N_{i,t}$) and their past wage progression, including overtime work, and that they have expectations of future service, future wage progression, and future retirement behavior that are rational in the sense that they are similar to the past experience of other workers in their departments with the same tenure or time to retirement if retirement is less than 5 years away. Future mortality is assumed to be equal to the assumptions used by the actuary in valuing the plan for all workers.

If a worker knows with certainty that this year's pay will enter into her pension calculation, and has already vested, then it follows from formula (1) that an extra hour's overtime pay will increase her annual pension payments by one-third of her hourly cash overtime wage, multiplied by her total lifetime accrual, $\alpha(N_i)$ in formula (1). The pension compensation for that hour's work would then be the increase in her pension wealth in respect of it, calculated as the extra annual payment multiplied by a life annuity factor and discounted from her retirement date to the present. But most workers do not know with certainty whether any particular year's pay will be one of their three-highest years' incomes when they reach retirement. The probability that this ends up being the case is an increasing function of the number of overtime hours worked, and a decreasing function of the length of time to retirement.

It is straightforward to see why the second statement holds: the longer the time to retirement, the greater future pay is likely to be – due to inflation and promotional and other pay increases – and the lower the probability that the current pay will be one of the three highest at retirement. To see why the first statement is true, consider an individual worker whose total pensionable pay in this year is less than her third-highest total annual pay up to that point: this year's pay will never enter into her final pension calculation and an extra hour of overtime will not increase her pension at all. The pension compensation for that hour of overtime is therefore zero. However, once she has worked enough overtime to ensure that her current year's pay, including overtime pay, *is* higher than her third-highest yearly pay up to that point, there is at least a chance that this year's pay will be one of her three highest when she retires. The more overtime she works, the higher her pay in year t, the higher this probability is, and therefore the greater her expected pension compensation for this hour of overtime work. Note that the increase in this probability is not smooth: there are discontinuities when an additional hour of overtime causes her current year's pay to exceed her second highest prior yearly pay, and again when it exceeds her highest prior yearly pay, because each time this happens the conditional probability that this year's income will enter into her pension calculation rises.

To tighten these ideas, we define $P^*_{i,t}(x)$ as worker i's expected increase in year t in their pension wealth at time t as a result of the x'th hour of overtime worked in year t, conditional on worker expectations of future wages, overtime, longevity and retirement date that are equal to the historical experience of workers in their department with the same tenure as them. Formulae are given in appendix B.

We scale $P^*_{i,t}(x)$ by the hourly cash overtime wage and write:

$$P_{i,t}(x) = P^*_{i,t}(x) / OTCashWage_{i,t}\ , \tag{2}$$

where $P_{i,t}(x)$ is then the expected increase in worker *i*'s lifetime income per dollar of hourly overtime cash wage as a result of working the *x*'th overtime hour in year *t*. Total expected compensation for the *x*'th hour of overtime work for worker *i* in year *t* then equals cash compensation plus expected pension compensation, so

$$OTCashWage_{i,t} + P^*_{i,t}(x) = OTCashWage_{i,t}(1 + P_{i,t}(x))\ . \tag{3}$$

There is no analytical formula for $P_{i,t}(x)$. Instead, we estimate it using Monte-Carlo simulation of a micro-simulation model calibrated to the pension, wage and retirement data we received from the city. Details of the micro-simulation model are provided in Appendix B.

Our estimation of $P_{i,t}(x)$ is subject to various sources of error. First, expected pension compensation depends on individual worker expectations of their future wage increases, their data of retirement or separation, their likely mortality and their likely future overtime. Not having access to this information, we condition our expectation based on worker group averages, not the expectations of individuals workers themselves. Conditioning on different information could result in different estimates of expected pension compensation. Secondly, we were not provided with the age of individual employees. However, in pension fund data we received from the city, we did receive age and tenure, but only for retired workers. Therefore, for each simulation, we drew a tenure-age pair from the pension fund data, conditional on total tenure in that data being longer than the tenure observed for each worker in the payroll sample. Other measurement error is caused by the fact that we do not observe full wage histories for most of the workers in our sample (so may mis-estimate the ranking of a particular year's pay in a workers' employment history), and because pensions are based on the highest three years of pay where each year's pay is measured over any non-overlapping 12-month period, whereas we were only provided with payroll data over calendar years. Finally, we assume that the distribution of overtime remains constant in the future. As a preview, however, we note that the results of our non-parametric estimation technique are highly robust to errors in this variable.

Insert Figure 2 near here

To give the reader greater insight, the left-hand panel of Figure 2 shows a graph of $P_{i,t}(x)$ for three different workers in our data set, A, B and C, for $x \in [-150;1400]$. The graphs were calculated by running 25,000 simulated life paths for each worker and each even value of x. Worker A is very close to retirement. He needs to work around 250 hours of overtime for the current year's wage to exceed his third-highest wage up to this point, so $P_{A,t}(x)=0$ for $x<250$. Once he has worked around 500 hours, it is almost certain that each additional hour of overtime after that point will increase his pension, resulting in additional expected lifetime income with a present value of around 4.4 times his cash wage for that hour of overtime, so $P_{A,t}(x)\approx 4.4$ for $x>500$. The discontinuities when the second-highest and highest wage up to this point are crossed are significant because the worker does not have many years of work left and his future wage progression can be estimated reasonably accurately. Worker B is slightly further from retirement. She only needs to work around 60 overtime hours to exceed her third-highest wage up to this point. But because she is further from retirement, the discontinuities when she crosses her second- and first-highest wage up to this point are smaller than for worker A, she has to work more overtime than A to guarantee that this year's wage will enter into her pension calculation, and the expected value of additional pension peaks at a lower level, around 1.7 times the value of her hourly overtime cash wage. Worker C is even further from retirement. This worker earns more than his third-highest wage without doing any overtime at all, so $P_{C,t}(0)>0$. Because he is so far from retirement, the discontinuities in $P_{C,t}(x)$ are yet smaller than those of worker B, and he has to work even more overtime to guarantee that this year's wage will enter into his pension calculation. Even then, because he is currently so far from retirement, he expects that each hour of overtime work will only increase his lifetime pension income by roughly the same amount as his hourly pay, so $P_{C,t}(x)\approx 1.1\ \forall\ x>1200$.

Deriving these graphs for every worker in every year is extremely computationally demanding, and their shapes are complex. In our empirical work, we therefore slightly adjust the approach of Fields and Mitchell (1984), and approximate them with piecewise linear curves with three segments: a flat line at 0 below the individual's third-highest wage (where $P_{i,t}(x)$ starts to rise above 0), a line with positive gradient from this point to the 90th percentile[13] of the highest career annual wages, estimated off a simulation of 2,000 complete life histories for each worker, and a second flat line for overtime hours above this point. The height of this third line was estimated using equation (4) but treating $I_{i,1,r,t}(x)$ as known with certainty and equal 1/3, and $I_{i,2,r,t}$ equal to the probability that the individual vests. We focus on the middle segment, and represent its shape for each worker in each year in our sample with three variables: START, the starting point of the middle segment on the x-axis, equal to the number of overtime hours worker i has to work in year t so that their total wage that year equals their third highest wage up to that point, WIDTH, the width of the middle segment, and HEIGHT, the height of the middle segment on the vertical axis at its right-most point. START and WIDTH are measured in thousands of hours per year, while HEIGHT is a unitless quantity with maximum value in our sample of about 5. These three quantities provide a computationally feasible and reasonably accurate summary of the complete curve $P_{i,t}(x)$ for each individual in our sample in each year. Our approximations to $P_{i,t}(x)$ for workers A, B and C are shown in the figure.

There are two quantities of special interest related to $P_{i,t}(x)$. First is worker i's expected pension compensation for the final hour of overtime she worked in year t, which we write as $OTPensComp_{i,t} \equiv P_{i,t}(OT_{i,t})$, where $OT_{i,t}$ is the actual number of overtime hours worked. We estimate this quantity using the piece-wise linear approximation to the employee's budget constraint, described above. The total compensation the worker expects to receive for this last hour of work is then $OTCashWage_{i,t}(1+OTPensComp_{i,t})$.

A final point is that workers will only value pension compensation at its expected value if they are risk neutral. To allow for risk aversion across different possible pension outcomes, we also calculated the pension cost variables using the worker's certainty

13 We chose the 90th percentile because this gave the best least-squares fit to a set of full curves of $P_{i,t}(x)$ estimated off a sample of 1000 randomly-chosen individuals. In very few cases, this was lower than the value of START, in which case we simply set WIDTH to zero. For the certainty equivalent, discussed below, the best fit was obtained from the 99th percentile.

equivalent across our simulations assuming CRRA preferences and a risk-aversion coefficient of 2.5 rather than the pure expected value. Certainty equivalent curves for workers A, B and C are shown in the right-hand panel of Figure 2. The starting point is the same, but the width of the middle section of the piecewise linear curve is slightly higher (we used the 99'th percentile, as this provided a better fit), and the height is slightly lower due to variation in age and retirement age from our sampling methodology. Results (not reported for brevity) are very similar to the base case.

3. Overtime allocation rules

In this section, we discuss overtime allocation, which is usually the result of a collective bargaining process between unions and employers. The right of workers to bargain collectively with their employers was formulated in the National Labor Relations Act of 1935. For this purpose, workers divide themselves into bargaining units, each of which engages in separate negotiations with their employers. Work practices, wage rates, conditions of employment and dispute resolution mechanisms are all typically included in the negotiations. Collective Bargaining Agreements (CBA's) contain the resulting agreement between the employer and the union as it concerns that bargaining unit. CBA's have a defined term of validity and are then amended or replaced in subsequent rounds of negotiations.

State and local pension plans typically operate across a number of bargaining units (and in many cases may also operate across different employers, each of whom may have more than one bargaining unit). However, to illustrate common overtime allocation rules, we used a database maintained by the Department of Labor,[14] to match CBA's to pension plan data from the CRR. We were able to find CBA's for at least some bargaining units in 68 of the 188 plans in the CRR database. Of these, the vast majority, around 49 plans, have at least one bargaining unit where some rule for the allocation of overtime is specified in the CBA.[15] Very few (e.g. Milwaukee City

14 https://www.dol.gov/agencies/olms/regs/compliance/cba

15 Overtime pay is mandated by the Fair Labour Standards Act (1938). The act exempts highly-paid employees from the overtime provisions of the act. Bargaining units where most employees are exempt (e.g. those dealing with doctors, physicians, scientists and engineers) are unlikely to mention overtime allocation for this reason. CBA's with employees where overtime is unusual (such as educators and

Employees' Retirement System) specify that overtime allocation rests solely with the employer; one (correctional officers employed by the State of Delaware) even specifies that overtime allocation is to be determined only by the union. But most overtime allocation rules fall into two broad categories. The first specifies that overtime is to be allocated as equally or equitably as possible among employees who habitually perform the work concerned. Some CBA's even specify mechanisms by which such equality is to be measured and achieved. Around one third of plans have CBA's that only fall into this category. The second type of allocation rules uses seniority to allocate overtime in some way. Two thirds of plans (including Philadelphia) have at least one bargaining unit that use some type of seniority rule in allocating overtime.

In Philadelphia, the master CBA between the City and the main unions dates back to 1992.[16] Article 19(I), shown in full in Appendix C, prescribes how the opportunity to do overtime shall be assigned to workers. Although there have been subsequent amendments to the master agreement, this particular clause does not appear to have been renegotiated in nearly 30 years.

19(I) prescribes that overtime opportunities are to be offered to employees in each work classification and at each work site in the order of seniority, most to least, on a 'rolling basis'. Overtime is only mandatory if insufficient volunteers for a particular shift can be found, in which case it is to be allocated to employees in the order of inverse seniority. Note that in terms of 19(I) there is little or no compulsion for senior employees to do overtime: they are simply to be offered overtime shifts (which they are free to reject) before these shifts are offered to more junior staff members (which they are also free to reject, unless the overtime shift cannot be filled). Precisely how preferential this allocation rule is to senior employees depends how the rule is administered in practice (in particular, on the interpretation of the phrase 'a rolling basis'). For instance, if overtime is very frequent (so the list is cycled through many times before allocations a

teachers) also do not typically mention overtime, even though extra work (e.g. summer pay) may indeed be pensionable.

16 The agreement between the City and one of the two main unions (the American Federation of State, County and Municipal Employees, (AFSCME) district council 47), as well as subsequent amendments, can be found here: https://secureservercdn.net/104.238.71.140/630.b13.myftpupload.com/wp-content/uploads/2018/06/2187CityContracts-WEB-20180601.pdf. The overtime allocation provision is the same for DC 33, the other main city union.

reset to the top of the list, or allocations are never reset to the top of the list but simply cycle through it without end), then overtime opportunities are only slightly skewed towards senior employees, if at all. If, on the other hand, allocations reset to the top of the list over periods where not much overtime is needed, senior employees will get very preferential access to overtime opportunities indeed – although it is up to them to fill them or not. The Philadelphia CBA is silent on this important interpretive issue. Other CBA's are more forthcoming, specifying the period over which rotation is to be applied.

The Philadelphia rule seems fairly common, and is used in enterprises as diverse as the US Postal Service, cities (e.g. Chicago, San Francisco, Kansas City Missouri, Detroit) and states (e.g. Oregon, New Jersey). Interestingly, some CBA's specify that overtime is to be allocated as equitably as possible and then detail something very like the Philadelphia rule (e.g. Miami firefighters, the technical unit of Michigan State employees or non-operating employees of the Los Angeles MTA).

In Appendix A, we use our merged CBA and plan data to test for a relationship between plan funding status and CBA rules regarding overtime. Results are shown in the second column of Panel B. Unfortunately, no clear relationship emerges – possibly because of the lack of correspondence between bargaining units and plans, and the small sample size. We therefore turn to a micro-economic examination of payroll and pension data in Philadelphia to examine this issue in more depth.

The overtime allocation rule used in Philadelphia suggests that when we examine worker response to expected pension compensation, we should examine not only the worker's own expected pension compensation, but also the pension compensation of those more and less senior than themselves, since a worker's own opportunities to perform overtime may well be affected by other worker's decisions. In Appendix D, we present a theoretical model which shows that, under mild conditions, any backloaded compensation structure (such as pensions) give workers an incentive to negotiate seniority-based overtime allocation rules such as the one we observe in Philadelphia and many other places.

4. Data

In this section we briefly describe our datasets and the merging and cleaning process we followed.

4.1 Description of payroll and pension data from the City of Philadelphia

We obtained a data download of the payroll of the City of Philadelphia for years 2007 – 2018 inclusive, through a freedom of information request. The data includes for each worker *i* in each calendar year *t*: full-time or part-time status, base pay, $BP_{i,t}$, job title, $JobT_{i,t}$, department, $Dept_{i,t}$, overtime pay, $OP_{i,t}$, gross pay, $GP_{i,t}$, pension plan section, $P\mathrm{Sec}_{i,t}$, DROP details, $DROP_{i,t}=1$ if the worker is in the DROP program in year t and 0 otherwise, and length of service $N_{i,t}$. We dropped all part-time workers from the data, as we do not know what their scheduled hours of work or pension arrangements or overtime regulations are. We calculated

$$OTCashWage_{i,t} = 1.5 \times BP_{i,t} / (48 \times 40) \text{, and} \tag{4}$$

$$OT_{i,t} = OP_{i,t} / OTCashWage_{i,t}. \tag{5}$$

A separate data download was obtained from the pension system of Philadelphia, also through a freedom of information request. This data set contains data items for people who were receiving pensions from the city pension plans in 2018. Data items include pension plan section, year of birth, monthly pension, retirement date, DROP amount, among other fields. The two datasets were merged by using individual names. All records with duplicate names were dropped to ensure accurate matching. We then used social security name files to impute gender using first names, and birth years if available. We were able to obtain reliable estimates of gender for the vast majority of our dataset.

Insert Table 2 here

Together, these datasets give us a complete picture of the wage development, overtime work and retirement behavior of Philadelphia workers over the period 2007-2018.

However, much of the data was not usable for our purpose and needed to be cleaned. The data cleaning procedure we adopted is shown in Table 2. Our original dataset contains 342,368 records. We dropped uniformed workers, who are subject to different pension rules and conditions of service, dropped individuals who appeared in the data sample for less than 5 years (because we could not reasonable infer what their three highest earnings were), excluded records where base salary was listed as 0 in the data set (because, for example, they were contractual workers), individuals for whom tenure was longer than the maximum in the pension data (as we had no basis on which to simulate their retirement behavior), excluded the first three years of observations for each individual (as we could not determine whether salaries were in the highest 3 up to this point), those individuals for whom the slope of the middle segment of our approximation to the curve $P_{i,t}(x)$ was greater than the 99'th percentile (because this likely indicated some error in our estimation caused by workers who leave or join midway through the year or who take large amounts of unpaid leave), and dropped single-member workgroups and all members of work-groups in years where no member of the work-group was recorded as performing any overtime. This left 85,051 records in the database.

4.2 Group variables

Because we are interested in testing overtime allocation *within* work-groups, we divided the data into work-groups based on department and job title. Workers with the same job title in the same department were assumed to be part of a single work-group. There were around 650 work-groups (numbers differ from year to year). The largest group had about 1,400 workers, and the smallest, 2 (because we did not consider work-groups with only one worker). Each work group was allocated a number and worker i's workgroup in year t was labeled $WG_{i,t}$. We denoted the size of worker i's workgroup in year t as $NWG_{i,t}$.

We first calculated a seniority score for each worker in each year, $Sen_{i,t}$ equal to the proportion of members of that workgroup who had tenure less than or equal to worker i in year t, so the worker with the longest tenure in each work group has a seniority score of 1 and the most junior a seniority score of 0.

We then calculated, in each work-group, the total of our pension cost variables $START_{i,t}$ $HEIGHT_{i,t}$ and $WIDTH_{i,t}$ for workers other than worker i with seniority scores higher than or lower than each worker in each year, divided these by the size of the workgroup $NWG_{i,t}$ and annotated these to indicate whether workers were more senior, or less senior than worker i. Averages are taken over the whole work-group to account for both the number of workers more and less senior than a given worker and their pension variables, as well as to ensure a consistent basis for controlling for the inverse of group size in our regressions. Workers who are the most or least senior workers in their work-group (or who are tied in these positions) will not have values for these variables.

Insert Table 3 near here.

Summary statistics are shown in Table 3. We discuss only those variables that are noteworthy. First, cash compensation for overtime ranged from \$11.12 per hour to \$157.60 per hour, with a mean value of \$35.95 in our sample. The marginal hourly pension compensation had a similar large range: from 0 times cash compensation to 4.597 times cash compensation, with a mean of 0.354 times cash compensation. Note that, as shown in Figure 1, for around one quarter of workers, working an additional hour of overtime will not raise their pension at all: they are either below their third-highest wage, or have elected DROP. The range for average pension compensation for overtime was the same (the person in the sample with the highest marginal pension compensation for overtime worked no overtime in that year), but with a lower mean of 0.199 of cash compensation.[17]

In the next section we turn to our empirical work.

5. Empirical framework and results

In this section, we first perform standard regressions to estimate the relationship between compensation of different kinds and overtime hours worked. We then test whether these results are robust to the fact that pension prices are upward-sloping in the number of hours worked and the fact that workers likely face different constraints

17 In a subsequent section, we discuss why this differs from the normal cost of the plan (calculated by the plan actuaries as 9.44% of payroll) in a subsequent section.

due to the overtime allocation rules. To determine whether our results are the consequence of individual worker responses to pension incentives, or simply the result of the overtime allocation rule, we then adapt the non-parametric approach of Saez (2010) to exploit the discontinuities in pension compensation.

5.1 Uncensored regression design

Our first analytical tool is an uncensored linear regression of the log of total hours worked on the variables presented above, in logs, standard in analyses of labor supply. So:

$$\log(OT_{i,t} + 48 \times 40) = f(\text{pension variables,worker characteristics,interactions,controls}) + \varepsilon_{i,t} \quad (6)$$

In all specifications, we control for unobserved heterogeneity by including individual fixed effects. The coefficients should therefore be interpreted as the individual response in total hours worked associated with changes in each of the right-hand-side variables. In all specifications, we also include a control for the group size (the inverse of the group size) as the average group variables are all scaled by that amount, and interact all terms with this control. In all specifications, we also control for the average number of overtime hours worked in the group by individuals (other than the individual in question) to control for changes in the amount of overtime done by each group in each year. Finally, we include individual's seniority rank in their group to control for seniority. We cluster standard errors by individual.

Insert Table 4 here

Results are shown in Table 4. Our first specification, shown in column (1), regresses log of total hours worked in each year on the log of total hourly compensation for the last hour worked, including cash wage and pensions. We are cautious in interpreting these as labor supply elasticities, although the regression suggests that if we did, the elasticity of labor supply to this variable would apparently be around 0.1. As people gain in seniority, holding all else equal, this specification suggests that they perform *less* overtime, rather than more. In specification (2), we split total compensation into pension compensation and cash and examine the effect of changes in each type of compensation separately. The apparent elasticity of total hours worked to cash wages is negative (but seniority is now positive but very small), close to other estimates in this literature. Pension cost, though, has an apparent elasticity of 0.126, the opposite sign and a larger magnitude than the response to changes in cash wages: evidence that strongly suggests different mechanisms are governing individual responses to pensions and cash compensation. In the third specification, we test the individual response to the cash wages and pensions of other people in their workgroups. Individuals appear to respond positively to their own cash wages and pension compensation, but negatively to the cash wages and pension incentives of other workers in their workgroups. Once again, the coefficient on pensions is stronger than the coefficient on cash compensation. In the fourth specification, we allow the worker response to the wage and pension costs of those more and less senior to them in their workgroups to be different. There is a significant difference between the apparent individual response to the pension costs of those more and less senior to them, with much of the effect appearing to be from workers less senior (contrary to our theory).

5.2 Are these results caused by upward-sloping pension compensation?

We report the results for the specifications in the first four columns of Table 4 only because they allow us to compare standard estimates of the (apparent) elasticity of total hours worked wrt pension and cash compensation to prior literature. However, as pure measures of worker response to pension incentives they are vulnerable to the criticism that the relationships we found could simply be a consequence of the fact that pension costs rise with the number of overtime hours worked. Even if overtime hours varied across individuals for reasons totally unrelated to pension costs, these regressions would show that rising pension costs were associated with higher overtime. The more overtime workers do, the more important this effect is likely to be – meaning that a direct comparison between the worker responses to the pension incentives of those more or less senior is particularly subject to bias.

To obtain alternative estimates not vulnerable to this criticism, the next set of regressions uses the properties of the entire compensation schedule $P_{i,t}(x)$ of each individual in each year, as summarized by the variables START, WIDTH and

HEIGHT. The higher START and WIDTH, the weaker pension incentives; the higher HEIGHT, the stronger. Although they do not allow us to compare our results to prior literature, these quantities *are* a pure measure of the incentives individual face that is independent of the amount of overtime workers actually do. Specification (5) regresses the log of total hours worked on the log of each worker's cash wage and their pension incentives. The signs in specification (5) are consistent with the idea that individuals respond to these incentives: START and WIDTH have negative and significant coefficients, while HEIGHT's coefficient is positive and significant. Cash wage is slightly negative and seniority slightly positive. In the next specification, we examine the response to the cash wages and pension incentives of other individuals in each work group. Individuals appear to respond to increases in START and WIDTH of other workers (so a weakening of the pension incentives of other workers) by working more overtime, but appear to be as responsive to other HEIGHT as they are to their own. The final specification, in column (7), divides the pension incentives of other workers between those less and more senior than each worker. There are strong differences between the apparent response to other individuals, depending on whether they are more or less senior than each individual in the workgroup. In the case of START and WIDTH, individuals appear much more sensitive to the pension incentives of those more senior to them than to those less senior, although individuals do not appear to respond to changes in the HEIGHT of others more or less senior to them, suggesting that much of the effect on the coefficient of HEIGHT:AllOthers in specification (6) is driven by individuals at the same level of seniority.

The overtime allocation rule creates two challenges in interpreting the results in this section. First, any apparent response in total hours worked to pension compensation may simply be the consequence of the fact that overtime is allocated more preferentially to more senior workers, who have higher pension costs. Second, more senior workers may be less constrained in their choices due to the overtime allocation rule, biasing the results. In the next section, we test whether our results are due to quantity constraints, before testing whether they are due to individual worker choices in the subsequent section.

5.3 Are these results caused by a quantity constraint on overtime work?

A challenge in interpreting the results in the previous section is that different workers may have (unobserved) constraints on the amount of overtime worked caused by the overtime allocation rules and by a cap on the overall overtime needed by the employer that may be driving the results shown in column (7) of Table 4. In this section, we illustrate how we re-estimate these regressions to take account of these constraints to obtain 'clean' estimates of the association between total hours worked and compensation.

We first calculate a constraint score for each worker in each year, equal to the proportion of the total overtime of that work-group in that year that was worked by workers with more seniority than that worker ($j \triangleright i$ in the equation below means that worker j is more senior than worker i):

$$C_{i,t} = g\left(\frac{\sum_{\substack{j \in WG(i) \\ j \triangleright i}} OT_{j,t}}{\sum_{j \in WG(i)} OT_{j,t}}\right), \tag{7}$$

where $g(\cdot)$ is a monotonically increasing function with 0 and 1 as fixed points, and $0 \le C_{i,t} \le 1$. If a worker is the most senior worker in a work-group, the constraint score will be 0. If the worker is the most junior, and works no overtime, the constraint score will be 1. Initially, we set $g(\cdot) = \Phi[(\cdot - 0.5) \times 10]$. The first parameter (0.5) is called a shift parameter, and the second (10) a scale factor. Our results are robust to changes in these values (results not reported for brevity).

We then posit the existence of a latent variable, $OT^*_{i,t}$, which is the amount of overtime a worker would work in the absence of constraints provided by the overtime allocation practices and an overall quantity constraint provided by the employer. We write the following censored regression model, again in logs:

$$\log(OT^*_{i,t} + 48 \times 40) \begin{cases} = \log(OT_{i,t} + 48 \times 40) = f(\ldots) + \varepsilon_{i,t} \text{w.p. } 1 - C_{i,t} \\ > \log(OT_{i,t} + 48 \times 40) = f(\ldots) + \varepsilon_{i,t} \quad \text{w.p. } C_{i,t} \end{cases} \tag{8}$$

Note that this is a generalized Tobit regression. It is not pure Tobit because we do not observe whether the data is censored or not; rather, we use the seniority rank to assign a *probability* that each observation is censored[18] (we assume that all observations are censored from below; that is, that all workers except the most senior in each work-group would prefer to work more overtime than we actually observe).[19]

We estimate the model by maximum likelihood, applying the transformation of Olsen (1978), calculating the log-likelihood as follows:

$$\ell\ell(\ldots) = \sum_{i,t} \log\left[\frac{[1-C_{i,t}]}{\sigma_\varepsilon}\phi(\varepsilon_{i,t}/\sigma_\varepsilon) + C_{i,t}[1-\Phi(\varepsilon_{i,t}/\sigma_\varepsilon)]\right] \quad (9)$$

Results are shown in Table 5.

Insert Table 5 near here.

While there are some sign changes (especially on the coefficient for log of cash wages), the overall tenor of the censored results is very similar to the uncensored results. Our preferred specification is column (7) in Table 5, which can be compared with the uncensored model in column (7) of Table 4. The coefficients on log(OTCashWage) has changed sign: higher own cash wages are associated with lower total work hours, and individuals are more responsive to the cash wages of other workers (although signs haven't changed). The coefficients on START, WIDTH, and HEIGHT have the same signs as before, but slightly different magnitudes. Once the overall constraint on overtime is allowed for, workers actually appear to be more responsive to the pension incentives of those more senior to them in their work-teams, and hardly responsive to the pension incentives of those less senior than them at all – at least as far as START and WIDTH are concerned. Although there is still no significant difference between the response of individuals to the HEIGHT of those more and less senior than themselves, it

18 Note that we could randomly assign each individual to 'censored' or 'uncensored' observational categories using the constraint score. It can be readily shown that this estimator provides the same estimates as the above procedure in the case of an infinite sample size, and that therefore it inherits the consistency property of the Tobit estimator proven by Amemiya (1973).
19 Note that this ignores the possibility that junior workers may be compelled to work overtime, which we acknowledge is a possibility in terms of the overtime allocation rule. We think this is sufficiently rare in practice that we can ignore it.

is clear that constraints on the amount of available overtime work is not driving our results.

5.4 Are these results driven by individual worker choices?

In the previous section, we showed that total hours worked is positively associated with individual's own expected pension compensation and negatively associated with the expected pension compensation of more senior workers. But it is not clear that these findings can be interpreted as elasticities in the usual sense: they could be the consequence of either the overtime allocation rule or strategic responses by individual workers to pension compensation, or both. Only in the latter case would we have identified a true elasticity. In this section, we use a non-parametric discontinuity analysis to determine which of these two explanations is more likely. We first discuss some theory, and then turn to our empirical analysis. We choose a non-parametric test to ensure that our results are robust to the criticism that our estimates of expected pension compensation are subject to error.

The idea underlying our approach is very simple: a concave kink in the budget constraint for the labor-leisure choice – such as the one introduced by the pension rules in Philadelphia – will cause individuals to avoid the region around the kink point when choosing how much to work. We call this 'dearthing', as distinct from the bunching behavior generated by convex kinks in budget constraints discussed by Saez (2010) and many others. In principle, if individuals are responding to pension incentives when selecting over time rather than automatically following overtime allocation rules, then we should be able to devise an empirical strategy to observe this dearthing in our dataset. If dearthing cannot be found, we conclude that overtime allocation rules, rather than worker response to pension incentives, are likely driving the results we have documented in the previous section.

5.4.1 The theory of dearthing

To make the theory of dearthing concrete, we use the same quasi-linear and iso-elastic utility function used by Saez (2010) (and that we use in Appendix D). Here, though, we express the function in terms of wealth (w) and leisure (r) (for rest) to ensure that we retain the convention of convex indifference curves increasing to the upper right and negatively-sloped budget constraints, so:

$$u_i(w,r) = w - \frac{\beta^i}{1+\frac{1}{e}}(r^* - r)^{1+\frac{1}{e}} \tag{10}$$

We assume that $r < r^*$, the maximum amount of leisure possible. We assume that individuals are either at leisure or at work, so the amount of work l equals $r^* - r$, and (22) is a single-period version of (8) with a different argument. β^i represents individual i's disutility of work. We assume that β is distributed smoothly through the population with distribution function $F(\beta)$. For simplification, we assume that this distribution has a support that ensures that all individuals in the population will choose at least some leisure.[20] We assume that all individuals have full flexibility about the number of hours they can work, meaning that we can interpret e as a labor supply elasticity without worrying about demand constraints.

Assume first that an individual maximizes utility of wealth and leisure in period t subject to a budget constraint, assuming that all individuals face the same budget constraint, an assumption that we generalize in the empirical section. We assume that the budget constraint has constant slope $-\alpha$, where α represents the individual wage per unit of time spent at work. Let the vertical intercept be w^* and the horizontal intercept be r^*. Then the equation of the budget constraint is $w = \alpha(r^* - r)$ and $w^* = \alpha r^*$. The individual's maximization problem can then be written as:

$$\max_{w,r} u(w,r) = \max_{w,r} w - \frac{\beta_i}{1+\frac{1}{e}}(r^* - r)^{1+\frac{1}{e}} \text{ s.t. } w = \alpha(r^* - r_{i,t}) \tag{11}$$

This can be solved by substituting the budget constraint into the objective function to yield the unconstrained maximization problem, and taking first order conditions in the usual way. Note that our assumption on $F(\beta)$ ensures that we always have an internal solution. This process yields the optimal amount of leisure $\hat{r} = r^* - (\alpha/\beta_i)^e$.

[20] This simply requires a lower bound on the value of β, which depends on the slope of the budget constraint.

Now we introduce a kink into the budget constraint. We assume that the kink occurs at r_k and that the slope to the left of the kink is $\delta_{i,t} > \alpha_{i,t}$. The budget constraint now has the following equation:

$$\begin{aligned} & w = \alpha(r^* - r), r > r_k \\ & w = \delta(r_k - r) + \alpha(r^* - r_k), r \le r_k \end{aligned} \tag{12}$$

Now there will be some value of β equal to β_t, which ensure whose optimal indifference curve is tangent to the budget constraint in two places – to the right and the left of point r_k. Call the point to the left of r_k, where the slope of the budget constraint is δ, r_b and the point to the right of r_k where the slope of the budget constraint is α, r_t. We now characterize β_t, r_b and r_t as follows:

$$u(w_b, r_b) = u(w_t, r_t),$$

so

$$\delta(r_k - r_b) + \alpha(r^* - r_k) - \frac{\beta_t}{1 + \frac{1}{e}}(r^* - r_b)^{1+\frac{1}{e}} = \alpha(r^* - r_t) - \frac{\beta_t}{1 + \frac{1}{e}}(r^* - r_t)^{1+\frac{1}{e}} \tag{13}$$

$$\left.\frac{dm}{dr}\right|_{\substack{m=m_b \\ r=r_b}} = -\beta_t(r^* - r_b)^{1/e} = -\delta, \text{ so } r_b = r^* - \left(\frac{\delta}{\beta_t}\right)^e$$

$$\left.\frac{dm}{dr}\right|_{\substack{m=m_t \\ r=r_t}} = -\beta_t(r^* - r_t)^{1/e} = -\alpha, \text{ so } r_t = r^* - \left(\frac{\alpha}{\beta_t}\right)^e$$

Any individual who has a value of β less than β_t will choose less leisure than r_b and anyone who has a value of β more than β_t will choose more leisure than r_t. No individual in the population will optimally choose an amount of leisure that lies between r_t and r_b.

Figure 3 represents the model graphically. Panel A of Figure 3 shows the indifference curves of three individuals with different values of β. Individual A has $\beta > \beta_t$ and so chooses more leisure than r_t whether the budget constraint is kinked or not. Individual B has $\beta = \beta_t$ and so chooses r_t when the budget constraint is unkinked but is indifferent between r_t and r_b with the kink. Individual C has $\beta < \beta_t$ and chooses a value of r between r_b and r_k when the budget constraint is unkinked, but something less than r_b when the budget constraint is kinked.

Insert Figure 3 near here.

Panel B shows the implications for the observed distribution of labor-leisure choices in the population if the distribution of β in the population is smooth. Before the kink is introduced, the distribution of labor-leisure choices will be smooth and unimodal (shown in blue), driven by the distribution of β. But after the kink is introduced, the distribution becomes bimodal. There will be a dearth of individuals choosing leisure between r_b and r_t, and the proportion of the population choosing leisure less than r_b will rise dramatically as all individuals who would have chosen leisure between r_b and r_t will now all choose some value of leisure less than r_b. The observed distribution of leisure to the left of r_b will also not simply be a shifted version of the old distribution, because leisure is a non-linear function of β. Note that unlike the case examined by Saez (2010), where the kink in the budget constraint is concave, there will be no bunching at point r_b.

Combining results from above shows that the width of the dearth expressed as a proportion of total hours worked where there is no kink is:

$$\frac{r_t - r_b}{r^* - r_b} = \left[\left(\frac{\delta}{\beta_t} \right)^e - \left(\frac{\alpha}{\beta_t} \right)^e \right] / \left(\frac{\alpha}{\beta_t} \right)^e = (\delta / \alpha)^e - 1. \tag{14}$$

This implies that we can estimate the elasticity of total hours worked given values of alpha and delta and the observed width of the dearth.

Our basic empirical strategy is (1) to choose a set of workers for whom dearthing should be most significant (we select that group of workers who work in high-overtime groups, who perform at least one hour of overtime and who are in their last full year of work before they retire or elect DROP), and (2) to examine worker overtime behavior relative to the amount of work required to reach their fifth-, fourth-, third-, second- and highest wage up to that point. Because the discontinuities in pay rates for overtime exist at the third-, second- and first-highest wage, we would expect to see dearths around those points, but not around the fourth- or fifth-highest wages, where there is no discontinuity in wage rates. We discuss this empirical test in more detail in the next section.

5.4.2 Empirical test

In our data, we observe the number of hours of work that each person does in each year, $l_{i,t} = OT_{i,t} + 48 \times 40$. We also have an estimate of the kink point of each person's budget constraint in each year (equal to the number of hours person i would have to work year t to ensure that their entire wage that year equaled their third-highest wages up to that point), called $l_{k,i,t}^{[3]} \equiv START_{i,t} + 48 \times 40$. For each individual, we also observe their hourly wage rate for working less than the kink in year t, $\alpha_{i,t} = BP_{i,t} / (48 \times 40)$. As we have shown, the rate per hour of overtime for most individuals in our dataset increases from $\alpha_{i,t} = BP_{i,t} / (48 \times 40)$ to $\delta_{i,t} = (1 + HEIGHT_{i,t}) \times 1.5 \times BP_{i,t} / (48 \times 40)$ over $WIDTH_{i,t}$ hours once their hours of overtime exceeds $START_{i,t}$. At more than $START_{i,t} + WIDTH_{i,t} + 48 \times 40$ hours of work, the hourly rate per hour of work no longer increases and the budget constraint is linear.

This suggests an empirical strategy. We select a group of individuals for whom dearthing is likely to be most significant according to our theory. We then test for dearthing in the data in a similar way to how Saez (2010) tested for bunching. If we do not observe dearthing in this group, we can conclude that dearthing likely exists nowhere in our data.

To do this, we focus on the last full year of work for all individuals who were observed to retire or elect DROP in any year after 2013. Because these workers are very close to retirement, they are like worker A in Panel A of Figure 2: $WIDTH_{i,t}$ is small, and $HEIGHT_{i,t}$ is large, and there are significant discontinuities in the hourly wage they face for overtime at their third-highest, second highest and highest wages up to that point. There should therefore be large convex kinks in their budget constraints at their highest, second highest and third-highest wages, but not their fourth or fifth highest.

But because these individuals are in their last full year of work, the convex kink should be much larger for the third-highest wage than for the second- or first-highest.[21]

From these individuals, we select only those individuals who work in departments that do a lot of overtime, and who do at least one hour of overtime. These individuals are unlikely to be constrained in the amount of overtime that they can do, both because they are likely to be senior in their workgroups, and are therefore able to exert a greater influence over their overtime than other workers, and because there is a lot of overtime for them to choose. We also have long wage histories for this group and are therefore able to estimate $l_{k,i,t}^{[3]} \equiv START_{i,t} + 48 \times 40$ with some degree of precision. We choose only those workers who perform overtime to avoid our results being biased by corner solutions caused by people who (1) work no overtime because their disutility of work is so high, (2) whose wages do not increase over long periods of time, or (3) who had no access to overtime that year.

For individuals in this restricted sample, we calculate the ratio of the hours of work person i actually performed in year t relative to the amount of work they would need to do to earn exactly their n'th highest wage up to that point, called $r_{i,t}^{[n]}$, for $n = 1, 2, 3, 4$ and 5. So if worker i worked exactly $START_{i,t}$ hours in year t, $r_{k,i,t}^{[3]} = 1$, etc. For simplicity, we assume that individuals earn an hourly wage of $\delta_{i,t} = (1 + HEIGHT_{i,t}) \times 1.5 \times BP_{i,t} / (48 \times 40)$ for each hour of overtime once they have exceeded their third-highest wag e up to that point (in other words, we treat $SLOPE_{i,t}$ as though it is infinite or $WIDTH_{i,t}$ as though it is 0; given that $WIDTH_{i,t}$ is small for these individuals anyway this is not likely to be a significant source of error).

Panel A of Table 6 shows the number of workers in this group, by years to retirement, who did and did not work more than one hour of overtime in that year, as well as the average of $1 + HEIGHT_{i,t}$ for each group. Around 20% of workers in each group elect no overtime. $1 + HEIGHT_{i,t}$ is slightly higher on average for those who do elect overtime in each year (although the differences are not large), and, as expected, average $1 + HEIGHT_{i,t}$

21 The jumps in Figure 3 at the second-highest and first-highest wages only exist because of the possibility that some future year or years of work will cause the current second-highest or first-highest wage to become the ultimate third-highest wage; if individuals are retiring in one year with certainty, no such year can exist. Of course, unplanned retirements do happen.

falls as retirement becomes more distant. There is no jump in any variable in the third year.

Panel B shows how the distribution and average price of overtime by workers in this group varies by the lifetime ranking of their annual wage. Once again, there appears to be no jump in either the proportion of people who do overtime, nor the price of overtime in the third year. While the amount of overtime performed rises with the wage rank (unsurprising, since overtime is driving the wage rank), the jump from year 4 to year 3 is of roughly the same order as the jumps in the other years.

Insert Table 6 near here

We then plot histograms of the distribution of $r_{i,t}^{[n]}$ for $n = 1, 2, 3, 4$ and 5. Our theory would suggest that should see a dearth of this distribution around $r_{i,t}^{[n]} = 1$ for $n = 1, 2, 3,$ with the largest dearth being around $r_{i,t}^{[3]} = 1$ (because these individuals presumably mostly know that they are planning to retire and so only have one full year of working left before they retire or elect DROP), smaller ones around $r_{i,t}^{[2]} = 1$, and $r_{i,t}^{[1]} = 1$ (because the kinks there are likely to be less severe because few workers in this group anticipate working more than one further year), but no dearths around $r_{i,t}^{[i]} = 1$ for $i > 3$. If the results in column (2) of Table 5 are due entirely to individual choices, the elasticity of labor supply wrt pension compensation would be 0.1 and (24) shows that the width of the dearth should be roughly $(\delta/\alpha)^e - 1 \approx 4^{0.1} - 1 \approx 0.15$, easily noticeable.

Insert Figure 4 near here.

Results are shown in Figure 4. The figure on the left shows the histogram of $r_{i,t}^{[5]}$, and the one on the far right $r_{i,t}^{[1]}$. The first row of histograms shows the distribution of $r_{i,t}^{[1]}$ in the final full year before retirement, the second row the second full year before retirement etc. As expected, most individuals do more work than necessary to equal their fifth-highest wage (so the mean of $r_{i,t}^{[5]} > 1$), but the distribution of $r_{i,t}^{[n]}$ shifts to the left as n falls.

A simple inspection of the histogram reveals no trace of bimodality and no dearth around 1 at $r_{i,t}^{[3]}$ (or, indeed any of the other ratios), certainly on the scale that theory would suggest. To confirm this impression, we use a non-parametric test based on the number of histogram points in a small symmetric interval of width 7 points around a particular value that lie above and below an estimate of the 'true' distribution (which we estimate using a Gaussian kernel smoother with large smoothing parameter, shown as a line in the figure). Under the null hypothesis that this smoothed kernel is the true value, the number of points above the line should be binomially distributed with parameters 0.5 and 7. P-values at $r_{i,t}^{[3]}=1$ are almost never below 0.05, indicating no statistical evidence in favor of dearthing.

Given a dearth width of 0, one can infer from (24) that despite the association between expected pension compensation and overtime hours documented in Tables 5 and 6, the elasticity of total hours worked with respect to pension compensation is actually 0 or very close to it: workers do not appear to be strategically selecting overtime in their last year of work based on technicalities of pension rules at all. We conclude that the overtime allocation rule itself, rather than strategic choices by individual workers that is driving the empirical results in Tables 5 and 6. We note that because a dearth should exist for all values of for all values of δ, our simulated measure of expected pension compensation, this result is highly robust to errors in this variable.

We note that for this conclusion to be valid, we require an additional assumption. Individual workers do not choose the overtime they do in a year in a single instant, as the model above implicitly assumes. Rather, overtime shifts are presented to them gradually over the year, and they choose to accept or reject each one on a running basis. Our conclusion would hold if workers form a reasonably accurate expectation of the amount of overtime they will be offered at the beginning of each year, choose an amount of labor or leisure for the whole year in advance, and accept or reject shifts as they go to meet that target. To the extent that their expectations were incorrect, strategically rational workers may inadvertently end up inside the predicted dearth if overtime offered fell far short of what they expected in any calendar year. However, even in this case we should still see some evidence of bimodality in the distribution of $r_{i,t}^{[3]}$ provided that worker expectations were at least on average correct and there were years in which actual overtime exceeded worker expectations. This is strikingly absent.

One possible limitation of our approach is that we do not see many workers who are near retirement working so few overtime hours that they would fall below the kink. Partly, this is because salaries rise with seniority and with time (due to inflation). Thus, workers naturally have higher salaries near retirement than they do earlier in their careers, besides the fact that they work more overtime. To test dearthing, we would ideally have large numbers of workers below and above the kink. To enable the reader to assess the effect of this feature of our data on our results, in the right-most columns in Table 6 we have shown the proportion of our sample that have r inside the predicted dearth (making the assumption that the dearth is symmetric around 1) for different values of e (0.1, 0.05 and 0.01). Panel A shows these proportions for n=3, and Panel B for one year away from retirement. For the higher values of e, 0.1 and 0.05, these numbers are significant, indicating that despite the low numbers of our sample with r values below 1, our method would still detect dearthing if it were present.

We conducted a formal hypothesis test to confirm this intuition. Under H_0: $e=0$, the distribution of the number of individuals with $r_{i,t}^{[n]}, n=1...5$ inside the dearthing interval for a given H_1: $e=x$, calculated as $[1-\frac{1}{2}(4^x-1);1+\frac{1}{2}(4^x-1)]$, is binomially distributed with the number of trials equal to the total number of observed individuals, and the probability of success equal to the 'true' probability of landing in the dearthing interval, if the null hypothesis is true. The test statistic is the number of individuals actually observed with $r_{i,t}^{[n]}, n=1...5$ inside the dearthing interval, and the p-value is the probability of observing this number or fewer. Calculations of these p-values for 20 values of x between 0.1 and 0, and $n=1...5$ never fall below 0.17, and are usually much higher. Further, there is no detectable difference between the p-values for the different values of n.[22] We therefore have confidence that despite the low proportion of the sample that have r values below 1, our data is still sufficiently dispersed to detect dearthing if it were present.

We also calculated histograms for workers two and three years before retirement, shown in the lower panels of Figure 4. As expected, no dearthing is apparent in these samples

---

22 Results are not reported for brevity, but are available from the authors on request.

either, which makes sense: if individual workers do not respond to pension incentives when they are strongest, they are unlikely to be more responsive to incentives when they are weaker. We are therefore confident that our results are robust despite the small proportion of workers with $r_{i,t}^{[n]} < 1$.

6. Measuring the cost of the overtime allocation rule

We have shown that the overtime allocation rule, rather than pension incentives, is likely driving worker overtime choices in Philadelphia. In this section we calculate the cost of pensionable overtime and the effect of the overtime allocation rule on pension costs.

We calculate the average pension compensation per hour per dollar of overtime hourly wage for overtime actually worked for each worker as

$$AveP_{i,t} = \frac{1}{OT_{i,t}} \int_0^{OT_{i,t}} P_{i,t}(x)dx \, . \tag{15}$$

First, we calculate the total increase in the pension wealth of workers as a result of working overtime in each year as a proportion of payroll costs, as:

$$OCR_t^{OT} = \frac{\sum_i OT_{i,t} \times OW_{i,t} \times AveP_{i,t}}{\sum_i (BP_{i,t} + OP_{i,t})} \, . \tag{16}$$

We then calculate total worker pension wealth under the assumption that overtime is pensionable, and under the assumption that it is not, as follows:

$$PW_{i,t}^{*,OT}(x) = E_t^{r_i, S_{i,r}^{OT}} [\underset{\substack{\text{Measure of} \\ \text{final pensionable} \\ \text{earnings at} \\ \text{retirement,} \\ \text{including overtime}}}{S_{i,r}^{OT}} \underbrace{\alpha(N_{i,t} + r_i - t)}_{\substack{\text{Fraction of salary} \\ \text{replaced by pension} \\ \text{plan as a function of} \\ \text{total length of service} \\ \text{at retirement}}} \underbrace{(1+d)^{t-r} a_{i,r}}_{\substack{\text{Present value of} \\ \text{lifetime annuity} \\ \text{from retirement age}}}] \, , \tag{17}$$

and

$$PW_{i,t}^{*,NOT}(x) = E_t^{r_i, S_{i,r}^{NOT}}[\underbrace{S_{i,r}^{NOT}}_{\substack{\text{Measure of}\\ \text{final pensionable}\\ \text{earnings at}\\ \text{retirement,}\\ \text{excluding overtime}}} \underbrace{\alpha(N_{i,t} + r_i - t)}_{\substack{\text{Fraction of salary}\\ \text{replaced by pension}\\ \text{plan as a function of}\\ \text{total length of service}\\ \text{at retirement}}} \underbrace{(1+d)^{t-r} a_{i,r}}_{\substack{\text{Present value of}\\ \text{lifetime annuity}\\ \text{from retirement age}}}]. \tag{18}$$

Note that these formulae calculate an individual worker's expected pension wealth taking account of both past and likely future service. They are therefore akin to what actuaries call 'entry age method' calculations, rather than the 'projected unit credit method' calculations used in pension accounting that take account only of past service. When calculating pension costs using the 'entry age method', actuaries estimate the total liability for a 'typical' worker and then amortize it over the entire working life of the individual, usually expressed as a level percentage of the worker's pay.

These equations give the total pension liability in respect of each individual, taking account of past and future service. The ratio between the two therefore represents the proportion by which pension costs increase as a consequence of making overtime pensionable.

We call the ratio between these two amounts the overtime pension cost ratio, calculated as:

$$OTPCR_t = \frac{\sum_i PW_{i,t}^{*,OT}(x)}{\sum_i PW_{i,t}^{*,NOT}(x)}, \tag{19}$$

Note that the OTPCR makes the assumption that overtime continues to be allocated to individual workers in the future exactly as it is now. Because overtime allocation depends on a number of factors – group size, staffing levels, overtime needs, and overtime allocation methodology – this calculation implicitly assumes that all of these will remain constant in the future. We return to this important point when we estimate the cost of the overtime allocation method below.

Insert Table 7 near here.

Results are shown in Table 7. Row A shows the OCR. Pensions on overtime cost between 1.7% and 4.4% of payroll, depending on the year, with an average of 2.9% of payroll. Row B shows the OTPCR. Making overtime pensionable increases pension wealth by around 19% for this sample of workers. This indicates that making overtime pensionable is equivalent in cost terms to a plan where overtime is not pensionable, but which is around 20% more generous (e.g. because it has an accrual rate that is 20% higher), although the distributional impacts of the two schemes are likely different (because overtime work is not evenly spread across workers in different departments and different job classifications).

We use these two results to calculate the pension cost in each year, shown in row C. The average over our sample is around 18.2% of payroll. Note that this is just less than double the normal cost of the plan reported by the plan actuary, which is 9.44% of payroll (Philadelphia, 2019a). There are two reasons for this. First, we are using a discount rate of 2.5% p.a., while the plan actuaries price benefits using the expected return on plan assets, which is 7.6% p.a. This will dramatically increase reported costs. Secondly, our data sample covers only those workers in groups that do overtime work, and is limited in other ways (most importantly, it excludes uniformed workers). Note again that these cost estimates implicitly assume that group size, staffing levels, overtime needs, and overtime allocation methodology all remain in the future as they are currently.

We now turn to the conceptually more difficult task of estimating the extent to which the method by which overtime is allocated across employees changes pension costs. A key issue here is the counterfactual: what alternative allocation method are we comparing the current one with? For instance, one could imagine an employer that decides to allocate all overtime to those employees for whom the marginal pension cost is zero (those roughly 25% of employees in our sample for whom overtime does not increase pensions at all). Assuming that this were feasible (which it may not be as overtime needs may be larger than this allocation method allows), this would reduce pension costs on overtime to zero.[23] Using this counterfactual, the entire cost of

23 This allocation method would need to allow for the fact that any change in the overtime allocation method changes worker expectations of the future distribution of their wages, and therefore the likelihood that any particular year's wages will be one of their top three when each employee actually retires, and therefore our estimate of pension costs. Further, any estimate of the effect of changes in the allocation method also needs to make assumptions about staffing levels, overtime needs and group composition in the future.

pensions on overtime could therefore potentially be ascribed to the overtime allocation method.

This seems unfair. We therefore choose an alternative counterfactual, under which overtime is allocated equally to all employees, now and in the future. This is convenient for analytical purposes, because allocating overtime in this way is exactly equivalent to giving all employees a payrise equal to the proportion that overtime pay comprises of total pay. Without doing any modelling, we can see that pension costs will then rise by exactly this proportion. Admittedly driven largely by this convenient result, we therefore choose this 'equal allocation' method as our counter-factual, and ascribe the difference in pension cost between this equal allocation of overtime and the current overtime cost to the method by which overtime is currently allocated across employees.

Under this counterfactual, the ratio between OTPCR and the proportion of overtime pay in total pay represents the proportion by which actual overtime allocation increases pension costs. We therefore define the behavioral cost of overtime allocation ratio as:

$$BCOTAR_t = \frac{OTPCR_t - 1}{\sum_i OP_{i,t}} \times \sum_i (BP_{i,t} + OP_{i,t}) \tag{20}$$

Results are shown in row D of Table 7. It shows that the behavioral allocation overtime increases pension costs by 30.9% relative to what pension costs would be if overtime were allocated proportionally to all employees, although results vary by year. Row E estimates that the behavioral cost of overtime allocation alone increases pension costs by around 0.7% of payroll over our sample, or 3.8% of total pension costs. Given that, as we have discussed, almost no plans make allowance for pension spiking when projecting pension costs, these results are broadly consistent with our regression findings in Appendix A, which show that plans where overtime is pensionable are around 7% less funded on average than plans where overtime is not pensionable.

7. Conclusion

In this paper, we have analyzed the incentives provided by pension plan rules in the city of Philadelphia and examined how workers respond to them. We have shown that pension compensation for the last hour of overtime worked varies dramatically across workers, with around 25% of workers receiving no additional pension compensation for that hour, but some workers receiving expected pension compensation worth as much as 4.5 times their overtime cash wages. To the best of our knowledge, this is the first paper that has documented these extraordinary differences.

We argue that compensation differential of this magnitude gives workers significant financial incentives to perform overtime near the end of their lives, and to concentrate overtime work in as few years as possible. We first test for a relationship between expected pension compensation and overtime using across-time variation in the value of individual compensation of different types for overtime work. Strangely, these standard regressions show that total hours worked is more strongly associated with expected pension compensation than with cash wages, negatively associated with the expected pension compensation of workers more senior than them, and indifferent to the expected pension compensation of those more junior to them, once allowance for the overall constraint on overtime is allowed for. We note that because our estimates of expected pension compensation are subject to significant error, our coefficient estimates are therefore likely biased towards zero; the true association may actually be stronger than we predict.

That hours worked appears more responsive to pension incentives at the intensive margin than it is to changes in cash wages, is deeply puzzling. Standard academic work on pensions usually finds that individuals are often ill-informed about pension compensation, and that they defer making decisions with regard to pensions in ways that standard models struggle to account for, even when the consequences of inaction are highly financially significant.

We therefore investigate individual worker response to pension incentives more closely to assess whether our results are caused by the overtime allocation rule itself, or by strategic worker decisions to accept overtime shifts. We use a non-parametric approach that is highly robust to errors in variables by adapting Saez (2010) to the case where budget constraints have convex kinks, and show that this should cause workers to avoid the area around the kink. This effect should be strongest for those workers who are one year away from retirement, and so who face the most extreme kinks, and who select to do overtime. Yet we find no evidence that these workers avoid the area around the kink at all.

This suggests that, despite the strong financial incentives provided by pensions, the primary driver of the relationship between pension compensation and overtime choices that we documented in earlier sections is likely to be the overtime allocation rule itself, rather than individual worker choices about whether to accept overtime at a given price or not. In effect, the overtime allocation rule may be transmitting information about how workers can increase pension payouts across generations of workers without requiring workers to have detailed knowledge of the pension rules. We point interested readers to Appendix D, where we derive and explain a theoretical model underpinning this contention.

In the final section of the paper, we use our results to measure the cost to the employer of making overtime pensionable. We find that, in our sample, making overtime pensionable raises pension costs by almost 20%, and that pensions on overtime cost around 2.9% of payroll. Of this amount, we estimate that around 0.7% of payroll, or 3.8% of total pension costs, are the consequence of employee allocation of overtime shifts to more senior workers relative to an 'equal allocation' rule. Given that almost no valuation assumptions in a database of US plans make allowance for this behavior when projecting pension costs, this suggests around a 10% increase in the underfunding of state and local pension plans – consistent with the results of a regression we perform of plan funding on overtime status in pensionable salary using a database of state and local pension plans.

To the best of our knowledge, this is the first paper that has examined the incentives provided by pensions at the intensive margin. Our results appear consistent with prior literature that finds that workers are unresponsive to cash wages when deciding how much to work, but in marked contrast to work which shows that workers respond strongly to pension incentives at the extensive margin. From a policy perspective, our results suggest that in order to control pension costs associated with overtime, employers should focus on internalizing the true cost of overtime allocation rules, including pension costs, when bargaining with employees, but may not need to alter pension benefit formulae themselves. This approach may serve to reduce the extent to which overtime allocation rules add to pension underfunding going forward without requiring substantial changes to benefit design.

8. References

Amemiya, T., 1973, "Regression analysis when the dependent variable is truncated normal", *Econometrica,* 41, pp 997-1016.

Anzia, Sarah F., and Terry M. Moe. 2015. "Public Sector Unions and the Costs of Government." *The Journal of Politics* 77 (1): 114–27.

Babcock, Linda, John Engberg and Amihai Glazer, 1997, "Wages and Employment in Public Sector Unions", *Economic Inquiry*, XXXV, 532-543.

Bargain, Olivier and Andreas Peichl. 2016. "Own-wage labor supply elasticities: variation across time and estimation methods", *IZA Journal of Labor Economics*, Vol 5 (10), pp 1-31.

Beshears, J., Choi, J.J., Hurwitz, J., Laibson, D. and Madrian, B.C., 2015. "Liquidity in retirement savings systems: An international comparison." *American Economic Review*, *105*(5), pp.420-25.

Biggs, Andrew G. 2010. "The Market Value of Public-Sector Pension Deficits." *American Enterprise Institute for Public Policy Research*, 8.

Bikker, J.A., Steenbeek, O.W. and Torracchi, F. 2012, "The Impact of Scale, Complexity, and Service Quality on the Administrative Costs of Pension Funds: A Cross-Country Comparison." *Journal of Risk and Insurance*, 79: 477-514. https://doi.org/10.1111/j.1539-6975.2011.01439.x

Borjas, G.J., 2010. "Labor economics." *Boston: McGraw-Hill/Irwin.*

Blau, David M. 1994. "Labor Force Dynamics of Older Men." *Econometrica* 62 (1): 117–56.

Blundell, R. and Macurdy, T., 1999. "Handbook of Labor Economics, volume 3, chapter Labor Supply: a Review of Alternative Approaches."

Brown, K.M. and Laschever, R.A., 2012. “When they're sixty-four: Peer effects and the timing of retirement.” *American Economic Journal: Applied Economics*, *4*(3), pp.90-115.

Brinkman, Jeffrey, Daniele Coen-Pirani, and Holger Sieg. 2018. “The Political Economy of Municipal Pension Funding.” *American Economic Journal: Macroeconomics* 10 (3): 215–46.

Brown, Jeffrey R., Robert Clark, and Joshua Rauh. 2011. “The Economics of State and Local Pensions.” *Journal of Pension Economics & Finance* 10 (2): 161–72.

Brown, Kristine M. 2013. “The Link between Pensions and Retirement Timing: Lessons from California Teachers.” *Journal of Public Economics* 98 (February): 1–14.

Chan, Sewin, and Ann Huff Stevens. 2008. “What You Don’t Know Can’t Help You: Pension Knowledge and Retirement Decision-Making.” *The Review of Economics and Statistics* 90 (2): 253–66.

Chalmers, J., Johnson, W.T. and Reuter, J., 2014. “The effect of pension design on employer costs and employee retirement choices: Evidence from Oregon.” *Journal of Public Economics*, *116*, pp.17-34.

City of Philadelphia, 2017a *City of Philadelphia Municipal Retirement System Actuarial Valuation Report as of July 1 2017*, City of Philadelphia, https://www.phila.gov/pensions/pages/default.aspx accessed 15th February 2019.

City of Philadelphia, 2017b *City of Philadelphia Municipal Retirement System Plan Summary Documents,* City of Philadelphia, https://www.phila.gov/pensions/pages/default.aspx accessed 15th February 2019

City of Philadelphia, 2017c *City of Philadelphia City Code,* City of Philadelphia, https://www.phila.gov/pensions/pages/default.aspx accessed 15th February 2019

Coile, Courtney, and Jonathan Gruber. 2007. “Future Social Security Entitlements and the Retirement Decision.” *Review of Economics and Statistics* 89 (2): 234–46.

Crawford, Rowena, and Richard Disney. 2014. “Reform of Police Pensions in England and Wales.” *Journal of Public Economics*, NBER Pensions, 116 (August): 62–72.

Currie, Janet, 1991, “Employment Determination in a Unionized Public-Sector Labor Market: The Case of Ontario’s Teachers”, *Jounral of Labor Economics,* Vol 9, No 1, pp. 45-66.

Duflo, E. and Saez, E., 2002. “Participation and investment decisions in a retirement plan: The influence of colleagues’ choices.” *Journal of public Economics*, *85*(1), pp.121-148.

Duflo, E. and Saez, E., 2003. “The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment”. *The Quarterly journal of economics*, *118*(3), pp.815-842.

Goda, Gopi Shah, John B. Shoven, and Sita N. Slavov. 2007. “A Tax on Work for the Eldderly: Medicare as a Secondary Payer”. NBER Working Paper 13383. Cambridge, MA: National Bureau of Economic Research.

Eberts, Randall W. and Joe A. Stone, 1986, "On the Contract Curve: A Test of Alternative Models of Collective Bargaining", *Journal of Labor Economics,* 4(1): pp 66-81

Edwards, Chris. 2010. "Employee Compensation in State and Local Governments." *Cato Institue Tax and Budget Bulletin*, no. 59: 3.

Elliott, Douglas J. n.d. "State and Local Pension Funding Deficits: A Primer." *The Brookings Institution*. Available online: https://www.brookings.edu/wp-content/uploads/2016/06/1206_state_local_funding_elliott.pdf.

Farber, Henry S., 1986 "The Analysis of Union Behavior", Chapter 18 in *The Handbook of Labor Economics,* Ashenfelter, Orly and Richard Layard, eds, North Holland, Amsterdam, Netherlands.

Fong, J.H., Piggott, J. and Sherris, M., 2015, "LONGEVITY SELECTION AND LIABILITIES IN PUBLIC SECTOR PENSION FUNDS." *Journal Risk and Insurance*, 82: 33-64. https://doi.org/10.1111/j.1539-6975.2013.12005.x

Feldstein, Martin. 1974. "Social Security, Induced Retirement, and Aggregate Capital Accumulation." *Journal of Political Economy* 82 (5): 905–26.

Feldstein, Martin, and Andrew Samwick. 1992. "Social Security Rules and Marginal Tax Rates." w3962. Cambridge, MA: National Bureau of Economic Research.

Fields, Gary S., and Olivia S. Mitchell. 1984. "Economic Determinants of the Optimal Retirement Age: An Empirical Investigation." *The Journal of Human Resources* 19 (2): 245–62.

Fitzpatrick, Maria D. 2017. "Pension-Spiking, Free-Riding, and the Effects of Pension Reform on Teachers' Earnings." *Journal of Public Economics* 148 (April): 57–74.

Fitzpatrick, Maria D and Gopi Shah Goda, 2020, "The Prevalence of COLA Adjustments in Public Sector Retirement Plans", Retirement and Disability Research Center, NBER, Paper NB20-13

Fong, J.H., Piggott, J. and Sherris, M. (2015), LONGEVITY SELECTION AND LIABILITIES IN PUBLIC SECTOR PENSION FUNDS. Journal Risk and Insurance, 82: 33-64. https://doi.org/10.1111/j.1539-6975.2013.12005.x

Frank, Jeff, and James M. Malcomson, 1994, "Trade unions and seniority employment rules", *European Economic Review*, Volume 38, Issue 8, 1994,

Glaeser, Edward L., and Giacomo A. M. Ponzetto. 2014. "Shrouded Costs of Government: The Political Economy of State and Local Public Pensions." *Journal of Public Economics*, NBER Pensions, 116 (August): 89–105.

Grossman, Gene, 1983, "Union wages, Temporary layoffs and Seniority", American Economic Review Vol. 73, No. 3 (Jun., 1983), pp. 277-290 (14 pages)

Gruber, Jonathan, and David A. Wise. 1999. "Social Security and Retirement in the United States." In Social Security and Retirement around the World. A National Bureau of Economic Research Conference Report. Chicago: University of Chicago Press.

Gustman, A.L., Mitchell, O.S. and Steinmeier, T.L. (1994), 'The Role of Pensions in the Labor Market: A Survey of the Literature', Industrial and Labor Relations Review 47, pp. 417-38.

Hudomiet, Peter, Michael D. Hurd, and Susann Rohwedder. 2018. "The Causal Effects of Economic Incentives, Health and Job Characteristics on Retirement: Estimates Based on Subjective Conditional Probabilities." Paper presented at SIEPR Conference on Working Longer and Retirement, October 25–26.

Hvide, H.K. and Östberg, P., 2015. "Social interaction at work." *Journal of Financial Economics*, *117*(3), pp.628-652.

Hurd, Michael D. 1990. "Research on the Elderly: Economic Status, Retirement, and Consumption and Saving." *Journal of Economic Literature* 28 (2): 565–637.

Ippolito, Richard A. 1985. "The Labor Contract and True Economic Pension Liabilities." *The American Economic Review* 75 (5): 1031–43.

Kenney, James, Rob Dubow, and Francis X Bielli. 2016. "City Of Philadelphia Board Of Pensions And Retirement Annual Report," 155.

Kiewiet, Rod. 2010. "The Day After Tomorrow: The Politics of Public Employee Benefits." *California Journal of Politics and Policy* 2 (3): 1–31.

Kim, Dongwu, Cory Koedel, Wei Kong, Shawn Ni, Michael Podgursky, Weiwei Wu, 2021, "Pensions and Late-Career Teacher Retention". *Education Finance and Policy* 2021; 16 (1): 42–65. https://doi.org/10.1162/edfp_a_00293

Koedel, C. M Podgursky, S Shi, 2013, "Teacher pension systems, the composition of the teaching workforce, and teacher quality", *Journal of Policy Analysis and Management* 32(3): 574-596

Lazear, Edward P. 1979. "Why Is There Mandatory Retirement?" *Journal of Political Economy* 87 (6): 1261–84.

Lieber, E.M. and Skimmyhorn, W., 2018. "Peer effects in financial decision-making." *Journal of Public Economics*, *163*, pp.37-59.

Martell, Christine R., Sharon N. Kioko, and Tima Moldogaziev. 2013. "Impact of Unfunded Pension Obligations on Credit Quality of State Governments." *Public Budgeting & Finance* 33 (3): 24–54.

Maturana, Gonzalo and Nickerson, Jordan, 2018. "Teachers Teaching Teachers: The Role of Workplace Peer Effects on Financial Decisions." *Working paper,* Available at SSRN: https://ssrn.com/abstract=2686300 or http://dx.doi.org/10.2139/ssrn.2686300

Maurer, R., Mitchell, O.S., Rogalla, R. and Schimetschek, T., 2018, "Will they take the money and work? People's willingness to delay claiming social security benefits for A lump sum." *Journal Risk and Insurance*, 85: 877-909. https://doi.org/10.1111/jori.12173

Maurer, R, Mitchell, OS, Rogalla R, Schimetschek T., 2021, "Optimal social security claiming behavior under lump sum incentives: Theory and evidence. *Journal Risk and Insurance*, 88: 5- 27. https://doi.org/10.1111/jori.12302

Munnell, Alicia H., Kelly Haverstick, and Jean-Pierre Aubry. 2008. "Why Does Funding Status Vary Among State and Local Plans." State and Local Plans Issue in Brief 6. Chestnut Hill, MA: Center for Retirement Research at Boston College.

Morrill, M., & Westall, J. (2019). "Social security and retirement timing: Evidence from a national sample of teachers." *Journal of Pension Economics and Finance*, 18(4), 549-564. doi:10.1017/S1474747218000422

Novy-Marx, Robert, and Joshua D. Rauh. 2009. "The Liabilities and Risks of State-Sponsored Pension Plans." *Journal of Economic Perspectives* 23 (4): 191–210.

Novy-Marx, R. and Rauh, J., 2010. "The crisis in local government pensions in the United States. Growing Old: Paying for Retirement and Institutional Money Management after the Financial Crisis, edited by Robert Litan and Richard Herring." *Washington, DC: Brookings Institution*.

Olafsdottir, Katrin, 2013, "Efficiency of collective bargaining in the public sector", mimeo, School of Business, Reykjavik University

Oswald, Andrew J., 1993, "Efficient contracts are on the labor demand curve", *Labor Economics*, 1(1993): pp 85-113

Olsen, R., 1978, "Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model." *Econometrica, 46*(5), 1211-1215. doi:10.2307/1911445

Ouimet, P. and Tate, G., 2017. "Learning from coworkers: Peer effects on individual investment decisions (No. w24058)." *National Bureau of Economic Research*.

Peng, Jun. 2004. "Public Pension Funds and Operating Budgets: A Tale of Three States." *Public Budgeting & Finance* 24 (2): 59–73.

Ruhm, Christopher J. 1995. "Secular Changes in the Work and Retirement Patterns of Older Men." *The Journal of Human Resources* 30 (2): 362–85.

Shapiro, A., 1985, "Contributions to the Evolution of Pension Cost Analysis." *The Journal of Risk and Insurance*, 52(1), 81-99. doi:10.2307/252617

Stock, James H., and David A. Wise. 1990. "Pensions, the Option Value of Work, and Retirement." *Econometrica (1986-1998); Evanston* 58 (5): 1151.

Stock, James, and David Wise. 1990. "The Pension Inducement to Retire: An Option Value Analysis." In *Issues in the Economics of Aging*, 205–30. Cambridge, MA: National Bureau of Economic Research.

Valletta, Robert G, 1993, "Union effects on Municipal Employment and Wages", *Journal of Labor Economics,* Vol 11, No 3, 545-574

Wise, J.V., 2013. “Pension portfolio choice and peer envy.” *Journal of Risk and Insurance*, *80*(2), pp.461-489.

Wang, Qiushi, and Jun Peng. 2016. “An Empirical Analysis of State and Local Public Pension Plan Funded Ratio Change, 2001-2009.” *The American Review of Public Administration* 46 (1): 75–91.

Wise, David, and Robin Lumsdaine. 1994. “Aging and Labor Force Participation: A Review of Trends and Explanations.” In *Aging in the United States and Japan: Economic Trends*. National Bureau of Economic Research: University of Chicago Press.

9. Tables and figures

TABLE 1: Summary of important features of City of Philadelphia pension plans for non-uniformed municipal employees

Plan	Accrual rate	Definition of average final salary	Maximum benefit
1967	First 20 years of service, 2.5% p.a.; 2.0% p.a. thereafter	Average of three highest years' salary, including overtime pay	80% of average final salary
1987	First 20 years of service, 2.2% p.a.; 2.0% p.a. thereafter	Average of three highest years' salary, including overtime pay	100% of average final salary
2010	First 20 years of service, 1.25% p.a.; 0% p.a. thereafter	Average of five highest years' salary, including overtime pay	100% of average final salary
2016	First 10 years of service, 2.2% p.a.; 2% p.a. thereafter	Average of three highest years' of capped salary ($50,000 p.a.), including overtime pay	

NOTE: Plans differ in other ways as well. Most notably, the 2010 plan includes a DC section with a contribution rate equal to 0.5% of employee compensation per year, regardless of tenure. Until 2017, most new employees were offered a choice between the 2010 and the 1987 plans. Most elected the 1987 plan.

TABLE 2: Data cleaning

	Steps	PHILA PRISONS	WATER DEPARTMENT	DEPARTMENT OF HUMAN SERVICES	STREETS DEPARTMENT	OTHER	TOTAL
1	Original dataset	26,439	23,530	19,610	19,801	252,988	342,368
2	Excluding uniformed workers leaves	26,438	23,529	19,610	19,801	145,647	235,025
3	Excluding individuals observed for less than 5 years leaves	23,745	19,138	17,009	16,134	108,953	184,979
4	Excluding base salary zero leaves	23,745	19,132	17,008	16,124	100,285	176,294
5	Excluding year of service more than the max in pension data (for simulation) leaves	23,736	19,097	17,003	16,087	100,026	175,949
6	Excluding observations for first 3 years for each worker leaves	16,152	13,065	11,574	10,603	66,935	118,329
7	Excluding slope above 99th percentile leaves	16,005	12,833	11,458	10,473	65,965	116,734
8	Excluding single-member groups and groups that do no overtime leaves	15,575	11,462	10,477	9,783	37,784	85,081
9	Excluding cash wage less than the 1.5 x minimum wage leaves	15,562	11,457	10,466	9,783	37,783	85,051

	Steps	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	TOTAL
1	Original dataset	28,359	28,928	27,767	27,963	27,676	28,116	28,252	28,135	28,748	29,168	29,311	29,945	342,368
2	Excluding uniformed workers leaves	19,433	19,763	18,817	19,049	18,918	19,367	19,292	19,478	19,781	20,187	20,276	20,664	235,025
3	Excluding individuals observed for less than 5 years leaves	13,215	14,599	15,111	15,933	16,779	16,917	17,029	17,070	16,081	15,062	14,021	13,162	184,979
4	Excluding base salary zero leaves	12,619	13,894	14,439	15,109	15,932	16,029	16,175	16,217	15,316	14,395	13,478	12,691	176,294
5	Excluding year of service more than the max in pension data (for simulation) leaves	12,587	13,862	14,407	15,076	15,897	15,995	16,141	16,184	15,288	14,372	13,463	12,677	175,949
6	Excluding observations for first 3 years for each worker leaves	-	-	-	12,350	13,663	13,487	13,314	13,182	13,143	13,237	13,353	12,600	118,329
7	Excluding slope above 99th percentile leaves	-	-	-	12,282	13,447	13,260	13,066	13,017	13,001	13,129	13,186	12,346	116,734
8	Excluding single-member groups and groups that do no overtime leaves	-	-	-	8,796	9,771	9,755	9,504	9,527	9,949	9,443	9,425	8,911	85,081
9	Excluding cash wage less than the 1.5 x minimum wage leaves	-	-	-	8,791	9,768	9,751	9,498	9,526	9,947	9,440	9,420	8,910	85,051

Table 3: Summary statistics

Variable name	Description	Code	Obs	Mean	SD	Min	Max	P25	P75
	Response variable								
Overtime hours	Number of overtime hours worked	OT	85,051	225.830	294.500	0	3,135.804	3.678	345.532
	Individual non-pension variables								
Tenure	Years since first appointment for City in any job	N	85,051	14.681	8.509	3	43	8	20
Female		Female	79,598	0.446	0.497	0	1	0	1
Total compensation	Marginal total compensation for an hour's work	TotalOTComp	85,051	49.551	30.587	11.793	449.862	32.870	53.871
Overtime cash wage	Cash wage for one hour of overtime work (equals 1.5 times hourly full time wage for normal work)	OTCashWage	85,051	35.953	9.689	11.117	157.595	29.320	40.527
Work group size	Size of individual work group. Work group is all individuals with the same job title who work for the same department	NWG	85,051	266.647	451.954	2	1,432	13	181
Average overtime hours in workgroup (excluding the individual)		AveOT	85,051	225.829	199.959	0	1,914.665	38.664	390.206
Department	City department	Dept	-	-	-	-	-	-	-
Job title	Individual's job title	JobT	-	-	-	-	-	-	-
	Individual pension variables								
Pension plan section	Pension plan section	PSec	-	-	-	-	-	-	-
DROP	Individual is a member of the DROP program, meaning they still work for the city but are no longer accrue new benefits in the pension plan	DROP	85,051	0.047	0.211	0	1	0	0
Marginal hourly overtime pension compensation	Expected increase in present value of lifetime pension income for the last overtime hour worked, as a multiple of overtime cash wage	OTPensComp	85,051	0.354	0.614	0	4.597	0	0.384
Average hourly overtime pension compensation	Expected average increase in present value of lifetime pension income over all overtime hours worked, as a multiple of overtime cash wage	AveOTPensCom p	85,051	0.199	0.373	0	4.597	0	0.228
START	Overtime hours below which overtime will definitely not increase pension	START	85,051	0.102	0.388	-5.047	2.848	-0.087	0.270
HEIGHT	Maximum possible expected increase in present value of lifetime pension income for an hour of overtime worked as a fraction of hourly overtime cash wage	HEIGHT	85,051	1.913	1.086	0	4.608	1.005	2.671
WIDTH	Overtime hours need to work in addition to START to reach maximum possible expected increase in present value of lifetime pension income per hour of overtime worked	WIDTH	85,051	1.172	1.165	0.029	41.028	0.587	1.522
	Variables that depend on other individuals in a worker's work-group								
Seniority	Proportion of workers in same work-group who have lower tenure	Sen	85,051	0.535	0.289	0.002	1	0.292	0.780
Other overtime cash wage: more senior	Total cash wage for one hour of overtime work of other workers in same workgroup who are more senior, divided by group size	OTCashWage:M ore Senior	78,429	18.587	11.890	0.023	145.073	9.222	26.075
Other pension cost: more senior	Total P for others in same workgroup who are more senior divided by group size	OTPensComp:M ore Senior	78,429	0.238	0.276	0	4.560	0.084	0.300
Other START: more senior	Total START for others in same workgroup who are more senior divided by group size	START:More Senior	78,429	0.085	0.139	-1.249	2.418	-0.001	0.149
Other HEIGHT: more senior	Total HEIGHT for others in same workgroup who are more senior divided by group size	HEIGHT:More Senior	78,429	1.197	0.701	0	4.607	0.697	1.580
Other WIDTH: more senior	Total WIDTH for others in same workgroup who are more senior, divided by group size	WIDTH:More Senior	78,429	0.436	0.480	0.0001	25.597	0.130	0.623
Other overtime cash wage: less senior	Total cash wage for one hour of overtime work of other workers in same workgroup who are less senior, divided by group size	OTCashWage:Le ssSenior	74,785	18.953	12.008	0.067	144.751	9.309	26.721
Other pension cost: less senior	Total P for others in same workgroup who are less senior divided by group size	OTPensComp:Le ss Senior	74,785	0.134	0.182	0	4.588	0.037	0.161
Other START: less senior	Total START for others in same workgroup who are less senior, divided by group size	START:Less Senior	74,785	0.024	0.148	-1.424	1.926	-0.052	0.086
Other HEIGHT: less senior	Total HEIGHT for others in same workgroup who are less senior, divided by group size	HEIGHT:Less Senior	74,785	0.809	0.651	0	4.588	0.290	1.188
Other WIDTH: less senior	Total WIDTH for others in same workgroup who are less senior, divided by group size	WIDTH:Less Senior	74,785	0.771	0.710	0.006	32.677	0.395	1.013

TABLE 4: Uncensored regression results

	Dependent variable: Logged total hours = log(48x40+OT)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(TotalOTComp)	0.098***						
	(0.002)						
log(OTCashWage)		-0.062***	0.064***	-0.066***	-0.017***	0.034***	-0.049***
		(0.004)	(0.009)	(0.005)	(0.005)	(0.012)	(0.007)
log(OTCashWage:AllOthers)			-0.112***			-0.057***	
			(0.009)			(0.012)	
log(OTCashWage:MoreSenior)				0.008***			0.014***
				(0.002)			(0.003)
log(OTCashWage:LessSenior)				0.001			-0.001
				(0.001)			(0.002)
log(1 + OTPensComp)		0.126***	0.132***	0.153***			
		(0.002)	(0.002)	(0.002)			
log(1 + OTPensComp:All others)			-0.046***				
			(0.003)				
log(1 + OTPensComp:MoreSenior)				-0.025***			
				(0.003)			
log(1 + OTPensComp:LessSenior)				-0.179***			
				(0.008)			
START					-0.044***	-0.052***	-0.073***
					(0.002)	(0.003)	(0.004)
START:AllOthers						0.042***	
						(0.005)	
START:MoreSenior							0.075***
							(0.010)
STAR:LessSenior							0.029***
							(0.009)
WIDTH					-0.005***	-0.016***	-0.034***
					(0.002)	(0.004)	(0.011)
WIDTH:AllOthers						0.017***	
						(0.005)	
WIDTH:MoreSenior							0.053**
							(0.022)
WIDTH:LessSenior							0.028***
							(0.007)
HEIGHT					0.006***	0.005***	0.005***
					(0.001)	(0.001)	(0.001)
HEIGHT:AllOthers						0.006***	
						(0.002)	
HEIGHT:MoreSenior							0.004
							(0.003)
HEIGHT:LessSenior							-0.007
							(0.005)
seniorityRank	-0.054***	0.006**	-0.010***	0.051***	0.012***	0.009*	0.073***
	(0.003)	(0.003)	(0.003)	(0.009)	(0.003)	(0.005)	(0.017)
aveOTHours	0.0002***	0.0002***	0.0002***	0.0003***	0.0003***	0.0002***	0.0003***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Size control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Individual	Individual	Individual	Individual	Individual	Individual	Individual
Clustered s.e.	Individual	Individual	Individual	Individual	Individual	Individual	Individual
Observations	85,051	85,051	85,051	68,480	85,051	85,051	68,480
Adjusted R^2	0.848	0.861	0.863	0.875	0.817	0.818	0.826
NOTE:							*p**p***p<0.01

Table 5: Censored regression results

	Dependent variable: Logged total hours = log(48x40+OT)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
log(TotalOTComp)		0.079***						
		(0.001)						
log(OTCashWage)			-0.099***	-0.046***	-0.011***	-0.053***	-0.062***	0.032***
			(0.002)	(0.007)	(0.004)	(0.003)	(0.010)	(0.005)
log(OTCashWage:AllOthers)				-0.033***			0.006	
				(0.007)			(0.009)	
log(OTCashWage:MoreSenior)					0.001			0.005***
					(0.001)			(0.001)
log(OTCashWage:LessSenior)					-0.070***			-0.075***
					(0.003)			(0.004)
log(1 + OTPensComp)			0.099***	0.103***	0.117***			
			(0.001)	(0.001)	(0.001)			
log(1 + OTPensComp:All others)				-0.037***				
				(0.002)				
log(1 + OTPensComp:MoreSenior)					-0.002			
					(0.003)			
log(1 + OTPensComp:LessSenior)					-0.121***			
					(0.004)			
START						-0.052***	-0.058***	-0.065***
						(0.001)	(0.001)	(0.001)
START:AllOthers							0.015***	
							(0.004)	
START:MoreSenior								0.119***
								(0.009)
STAR:LessSenior								0.002
								(0.005)
WIDTH						-0.002**	-0.016***	-0.038***
						(0.001)	(0.002)	(0.002)
WIDTH:AllOthers							0.017***	
							(0.002)	
WIDTH:MoreSenior								0.128***
								(0.007)
WIDTH:LessSenior								0.020***
								(0.002)
HEIGHT						0.004***	0.004***	0.004***
						(0.0004)	(0.0004)	(0.0004)
HEIGHT:AllOthers							0.004**	
							(0.002)	
HEIGHT:MoreSenior								0.011***
								(0.002)
HEIGHT:LessSenior								0.015***
							(0.003)	
seniorityRank	-0.148***	-0.083***	-0.093***	0.082***	-0.109***	-0.117***	0.121***	
	(0.002)	(0.002)	(0.002)	(0.009)	(0.003)	(0.004)	(0.014)	
aveOTHr	0.0002***	0.0002***	0.0002***	0.0003***	0.0003***	0.0003***	0.0003***	
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	
Size control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Fixed effects	Individual	Individual	Individual	Individual	Individual	Individual	Individual	
Observations	85,051	85,051	85,051	68,480	85,051	85,051	68,480	
NOTE:								*p**p***p<0.01

Table 6: Overtime performed by workers in high-overtime departments by years to retirement and wage rank

Years to retirement	Number <1hr overtime	Number >1hr overtime	Average 1+*HEIGHT* <1hr overtime	Average 1+*HEIGHT* >1hr overtime	Average overtime <1hr overtime	Average overtime >1hr overtime	At third-highest wage: Proportion of workers r in (0.925,1.075) (e~0.1)	At third-highest wage: Proportion of workers r in (0.964,1.036) (e~0.05)	At third-highest wage: Proportion of workers r in (0.993,1.007) (e~0.01)
1	158	696	3.917	3.954	0.010	348.369	0.201	0.063	0.009
2	156	705	3.840	3.909	0.019	362.923	0.173	0.066	0.018
3	155	705	3.641	3.821	0.009	372.120	0.168	0.063	0.005
4	166	697	3.552	3.762	0.023	358.720	0.169	0.056	0.011
5	144	587	3.552	3.673	0.012	338.654	0.197	0.120	0.051

Wage rank	Number <1hr overtime	Number >1hr overtime	Average 1+*HEIGHT* <1hr overtime	Average 1+*HEIGHT* >1hr overtime	Average overtime <1hr overtime	Average overtime >1hr overtime	One year from retirement: Proportion of workers r in (0.925,1.075) (e~0.1)	One year from retirement: Proportion of workers r in (0.964,1.036) (e~0.1)	One year from retirement: Proportion of workers r in (0.993,1.007) (e~0.1)
1st	261	756	2.794	2.999	0.010	475.013	0.345	0.189	0.034
2nd	224	747	3.168	3.217	0.016	412.145	0.270	0.133	0.021
3rd	219	709	3.374	3.438	0.010	367.239	0.201	0.063	0.009
4th	167	713	3.504	3.499	0.017	327.571	0.132	0.046	0.002
5th	139	686	3.480	3.541	0.003	305.152	0.074	0.021	0.003

Table 7: Measures of the cost of making overtime pensionable

		2010	2011	2012	2013	2014	2015	2016	2017	2018	All years
A=	OCR	2.3%	1.7%	1.7%	1.8%	3.5%	4.4%	3.5%	3.2%	3.1%	2.9%
B=	OTPCR	120.9%	120.3%	119.8%	119.0%	118.3%	117.5%	119.0%	118.3%	118.1%	119.0%
C=	A*B/(B-1)	13.3%	10.1%	10.3%	11.3%	22.6%	29.5%	21.9%	20.7%	20.2%	18.2%
D=	BCOTAR	153.7%	146.8%	139.3%	128.5%	122.0%	120.6%	128.0%	121.6%	126.7%	130.9%
E=	(D-1)/D*A	0.8%	0.5%	0.5%	0.4%	0.6%	0.8%	0.8%	0.6%	0.7%	0.7%
F=	E/C	6.0%	5.4%	4.7%	3.5%	2.8%	2.5%	3.5%	2.7%	3.2%	3.8%

Note: The Overtime Cost Ratio (OCR) measures the cost of pensions paid on overtime in the current year as a percentage of payroll. The Overtime Pension Cost Ratio (OTPCR) measures the change in total pension wealth in respect of past and future service resulting from including overtime in the definition of pensionable salary. The Behavioral Cost of Overtime Allocation Ratio (BCOTAR) measures the increase in overtime pension costs as a result of the allocation of overtime to more senior workers within each work-group, relative to the cost if overtime were allocated proportionally to all workers. C uses the OCR and OTPCR to estimate the total cost of pensions as a percentage of payroll in each year, while E uses the OCR and BCOTAR to measure the coat of the behavioral allocation of overtime as a percentage of payroll in each year. F expresses the behavioral cost of overtime allocation as a proportion of total pension costs in each year. Formulae are in the text.

Figure 1: Empirical CDF's of the amount of pension compensation across workers for the last hour worked in each year, as a proportion of hourly overtime cash wage

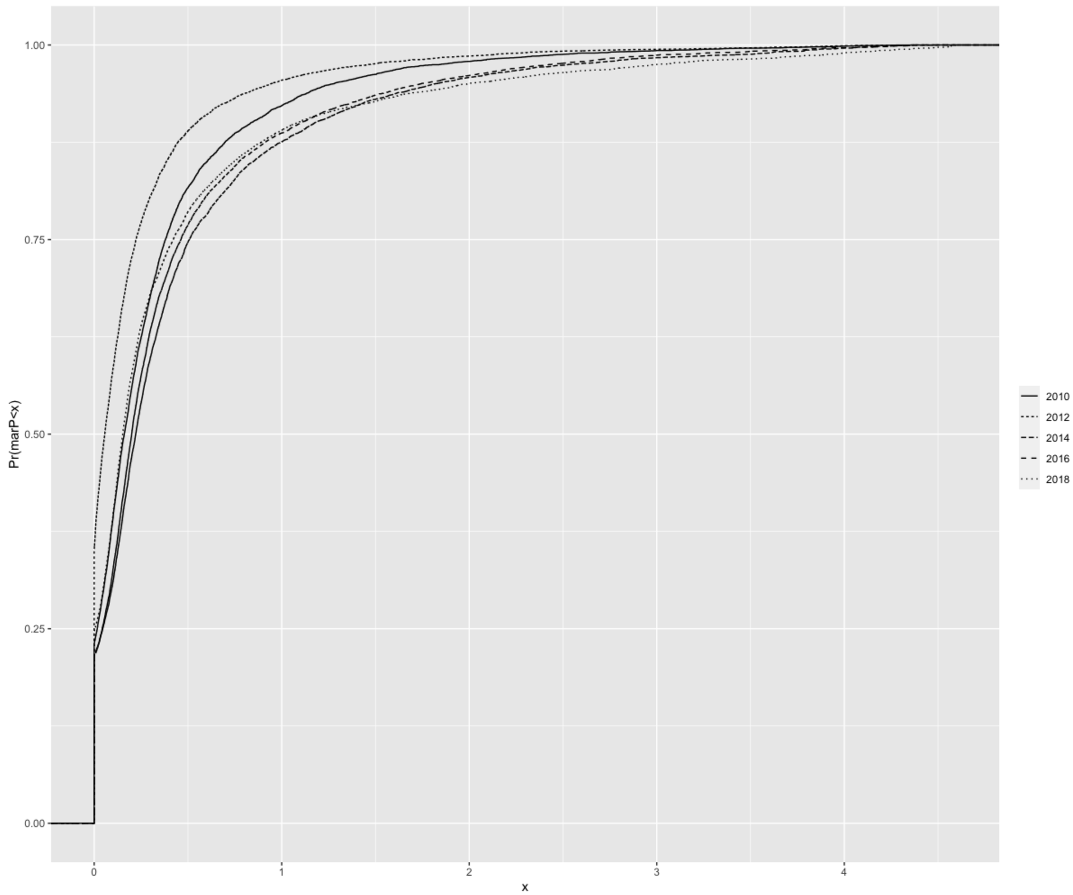
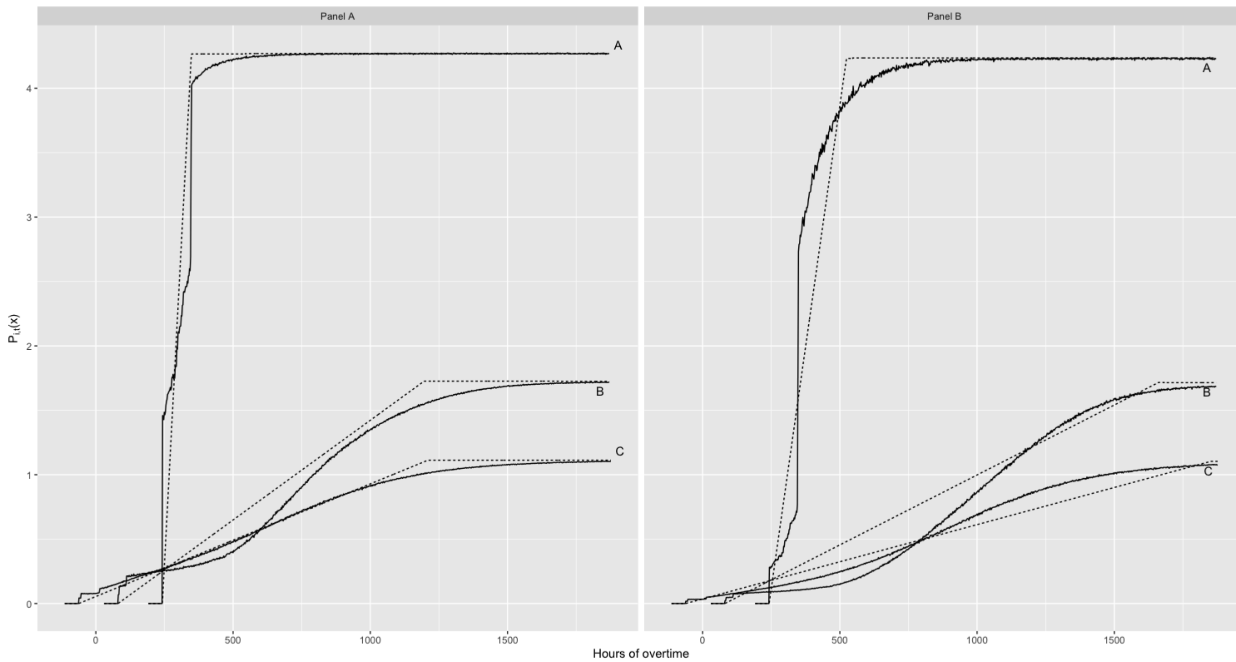


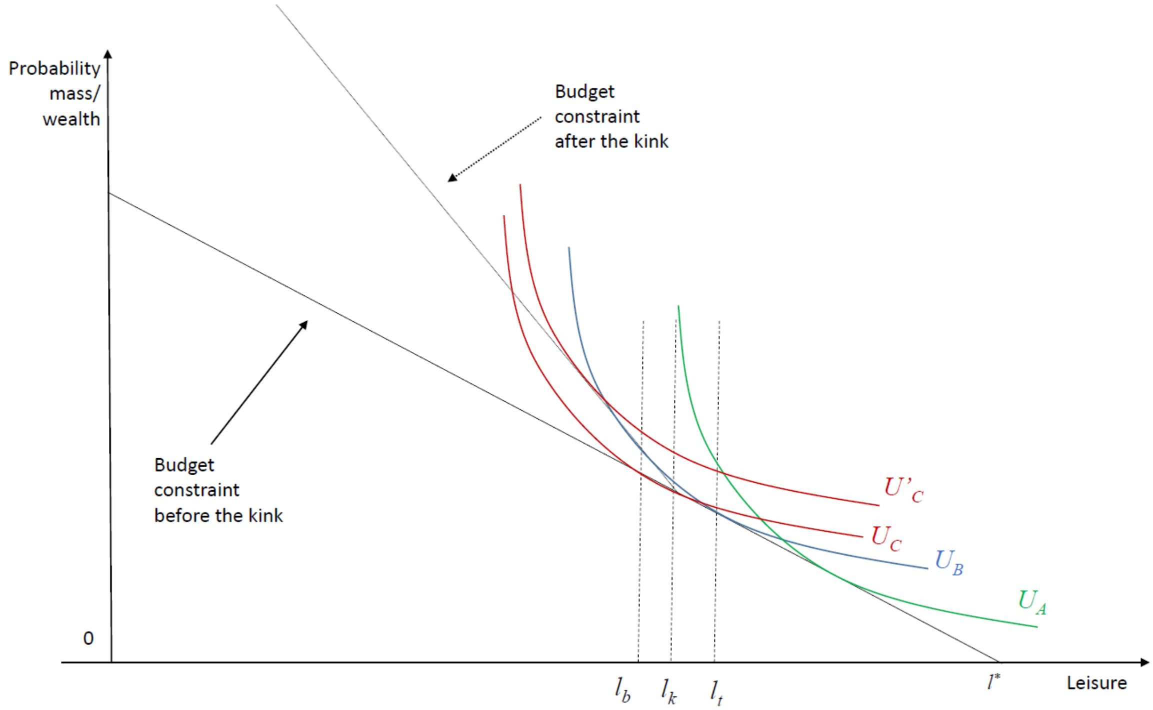
FIGURE 2: Sample curves of marginal value of pension benefits as a function of overtime hours worked, with piece-wise linear approximations



NOTE: This graph shows the sample curves (solid lines) of the marginal value of pension benefits per dollar of overtime cash wages as a function of overtime hours for three different workers, A, B, and C. The marginal value in Panel A is calculated using raw expected value and Panel B as the certainty equivalent assuming CRRA utility with gamma equals 2.5. Piecewise linear approximations are shown as dashed lines.

FIGURE 3: Indifference curve analysis for discontinuity design

Panel A: Indifference curve analysis



Panel B: Observed labor/leisure densities

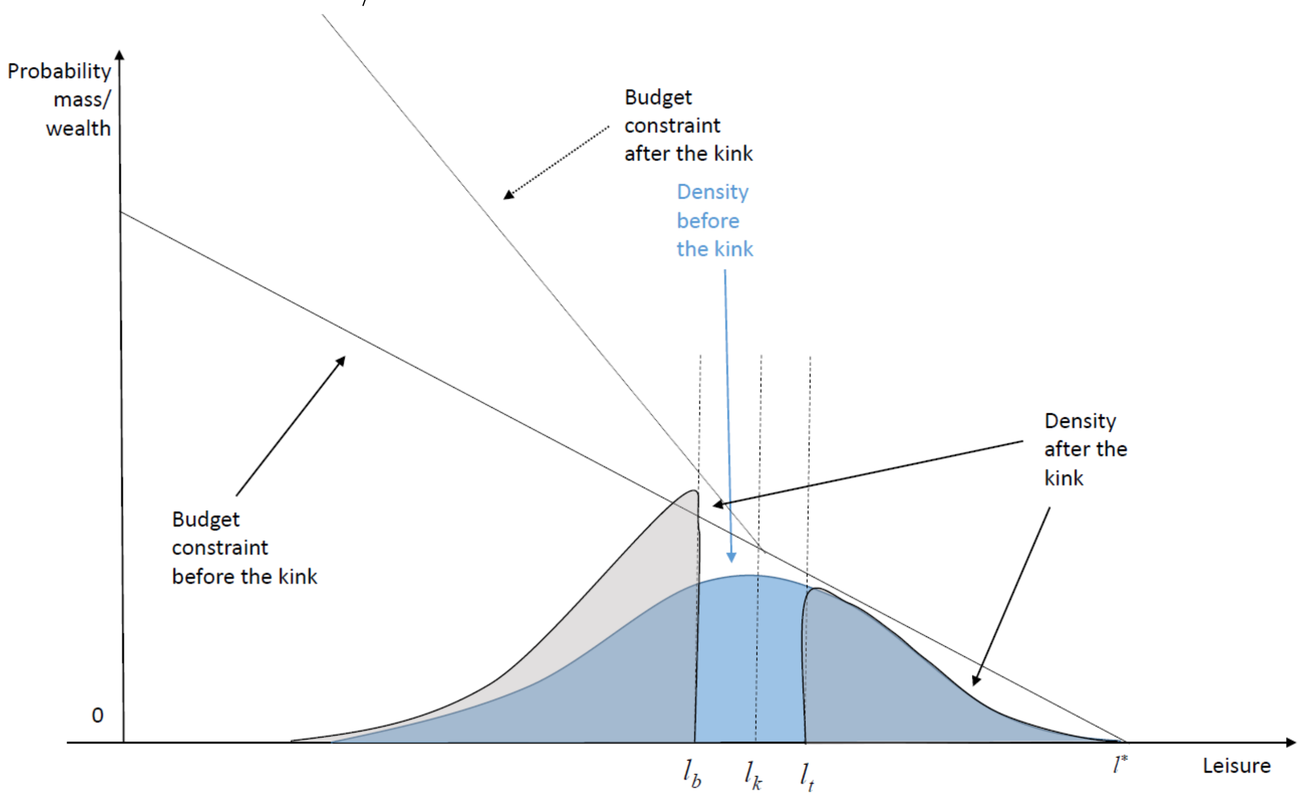
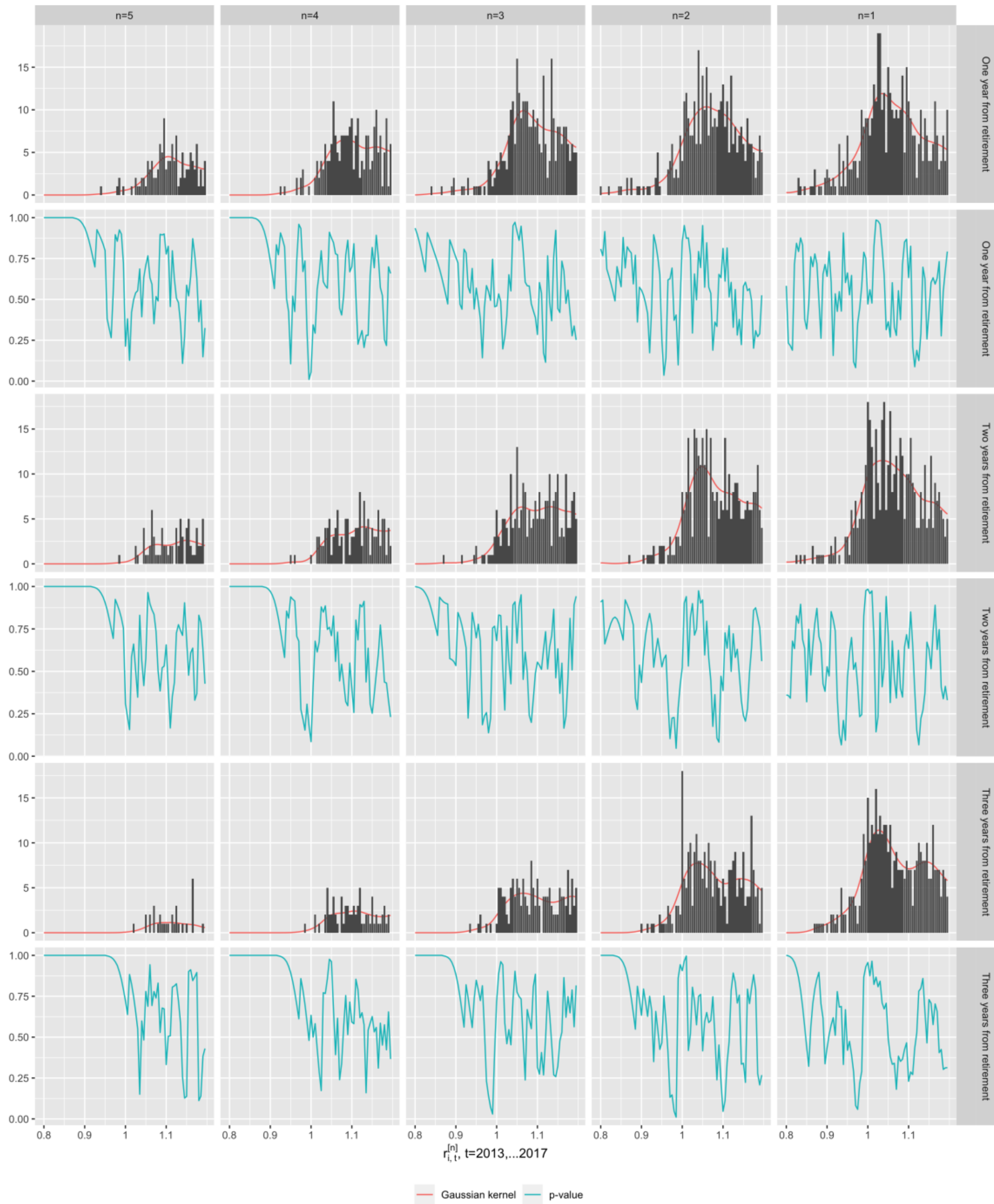


FIGURE 4: Histograms of the distribution of $r_{i,t}^{[n]}$ for $n = 1, 2, 3, 4$ and 5 for high-overtime departments



NOTE: Bottom graph represents p-values of a non-parametric test for whether the histogram in a small region around each point is drawn from the smoothed density, drawn in red.

Appendix A: Analysis of the relationship between pension overtime and plan funding

In this appendix, we use hand-collected data from the plan rules and overtime allocation practices for the 180 of 188 plans included in the CRR database where we could obtain this information. Plan rules were collected from plan summaries included in the most recent actuarial valuation reports of each plan. We show that plans where overtime is explicitly part of pensionable pay are significantly less funded than plans where the treatment of overtime is unclear. However, we are unable to find any explicit relationship between overtime allocation rules and the pension treatment of overtime pay, or between overtime allocation rules and plan funding.

Only around 6% of plan liabilities (around 30 out of 180 plans) were in plans where overtime was explicitly *excluded* from pensionable compensation, but around 29% of the total 2018 pension liabilities in the CRR database (again, roughly 30 of 180 plans) were in plans that specifically *included* overtime as part of pensionable compensation for at least one class of member. In the remainder of plans, where the precise status of overtime work was unclear, around 28% of liabilities were in plans in which "earned compensation" or "earnable compensation" (which likely includes overtime) was pensionable; and around 26% of liabilities were in plans in which "salary" was pensionable (which is less likely to include overtime, but still might). The rest used some other measure of earnings, such as "earnings" or "pay" for the purposes of calculating pensions, which again, could include overtime, or were cash balance plans.

Panel A of Table 1 compares various plan characteristics for plans in the CRR database, where the sample is split by whether overtime pay is included in the definition of pay used to determine benefits or not. The sub-samples appear to be very similar except in two areas: plans where overtime pay is explicitly used to calculate pension benefits are less well-funded on average by around 7% of plan liabilities; and sponsors whose plans explicitly excluded overtime pay in determining benefits were more likely to actually pay the pension contributions required of them by the plan's actuarial valuation. A simple regression, shown in the first column of panel B, confirms the picture: funding ratios of plans where overtime is included are lower than for other plans, even after accounting for differences in the discount rate, 10-year return on plan assets, 10-year underpayment of contributions and the generosity of benefits, at a 10% significance level. Of course, these results cannot be interpreted as causative: plan characteristics, including funding, choice of discount rate and overtime treatment, are all likely endogenous. But it should be clear that pension plan rules that include overtime are indeed associated with lower plan funding.

But even if pension rules include overtime pay, employers could reduce the impact on pension costs by allocating all overtime to workers who are far from retirement. As we will show in our theoretical model, workers have a strong incentive to formalize the allocation of overtime in order to solve a co-ordination problem among themselves. Where workers are unionized, overtime allocation is therefore often a subject of collective bargaining, and agreed-upon rules then form part of the collective bargaining agreement (CBA) between unions and the employer.

Using a database maintained by the Department of Labor,[24] we were able to find CBA's for at least some bargaining units in 68 of the 188 plans in the CRR database.[25] Of these, the vast majority, around 49 plans, have at least one bargaining unit where some rule for the allocation of overtime is specified in the CBA.[26]

A regression of plan funding that incorporates overtime allocation rules is shown in the second column of panel B of Table A1. Overtime allocation rules are not statistically significant, and their inclusion has reduced the statistical significance of 'OT: status unknown' to marginally above 10%. But any quantitative analysis of plan finances in the context of overtime allocation rules is complicated by a number of factors, besides

24 https://www.dol.gov/agencies/olms/regs/compliance/cba

25 For the purposes of collective bargaining, employees divide into groups with shared interests, called 'bargaining units'. Typically, plans operate across several bargaining units.

26 Overtime pay is mandated by the Fair Labour Standards Act (1938). The act exempts highly-paid employees from the overtime provisions of the act. Bargaining units where most employees are exempt (e.g. those dealing with doctors, physicians, scientists and engineers) are unlikely to mention overtime allocation for this reason. CBA's with employees where overtime is unusual (such as educators and teachers) also do not typically mention overtime, even though extra work (e.g. summer pay) may indeed be pensionable.

endogeneity and the small sample size. Firstly, in many cases, there are several bargaining units in the same plan, each with potentially different overtime allocation rules (as an illustration, there are 15 CBA's in the DoL database for members of the California Public Employees Retirement System (CalPers), only five of which use seniority). Plans typically operate across bargaining units. Secondly, plan rules are often extremely complex. In many large plans, as in Philadelphia, different benefit rules may apply to different classes of employees, depending on the jobs they do, when they were hired and when they retire or expect to retire. These rules do not generally coincide with bargaining units, although some may. Thirdly, the DoL database is not current. Many of the CBA's are old (some more than 20 years old), and their applicability to current employees is unknown. Fourthly, only around one-third of state and local employees are unionized. Overtime allocation practices for those employees who are not unionized are unknown. Finally, overtime allocation practices may utilize seniority even if the CBA does not require it – for instance the chief of police or the union may allocate overtime using seniority even if not required to by the CBA. Workplace practices may also differ from official rules in subtle ways that may have large consequences. For these reasons, we argue that examining micro-data, such as the data we have obtained for Philadelphia, is necessary to clarify the mechanisms connecting pension plan rules, overtime allocation and the consequent effects on pension costs.

Table A1: Comparison of mean plan characteristics by status of overtime in pensionable pay

Panel A: Summary statistics

Plan characteristic (2018 plan year)	Overtime pay included in definition of pay used to determine plan benefits: Explicitly included (n=~30)	Unknown (n=~120)	Explicitly excluded (n=~30)
Funding ratio (%)	67.0	73.9	73.2
Liability discount rate (% p.a.)	7.18	7.13	7.09
10-year average return on assets (% p.a)	7.24	7.21	7.31
Total required contributions unpaid (2008-2018) (% 2018 liabilities)	3.22	2.97	0.57
Normal cost (% of payroll)	15.6	14.5	15.2

NOTE: The funding ratio refers to the actuarial value of plan assets divided by the actuarial value of plan liabilities, calculated using Government Accounting Standards Board (GASB) rules. The liability discount rate, chosen by plan sponsors, is the expected return on plan assets used to discount plan cash flows to determine the plan liabilities under GASB. The normal cost of the plan refers to the actuary's estimate of the contributions required to meet the accruing liabilities of the plan. Total unpaid required contributions is the arithmetic sum of total required contributions between 2008 and 2018, inclusive, that were unpaid by the employer, divided by GASB plan liabilities in 2018. The status of overtime pay is taken from the plan descriptions included in the most recent actuarial valuation of the plan. The other quantities are from the CRR database of 188 pension state and local pension plans. Sample sizes differ slightly from metric to metric depending on data availability.

Panel B: Regression results

	Right-hand-side variable: Plan assets/plan liabilities * 100, 2018	
Liability discount rate (2018)	4.0052 (2.8401)	4.4121 (2.8637)
10-year average return on assets (2008-2018)	2.0734* (1.0559)	2.0887* (1.0601)
Total required contributions unpaid (2008-2018) (% 2018 liabilities)	-1.0566*** (0.1677)	-1.0361*** (0.1700)
Normal cost (% of payroll)	0.1289 (0.1779)	0.1425 (0.1789)
OT: explicitly included in plan benefits	(omitted)	(omitted)
OT: status unknown	5.752* (3.283)	5.456 (3.3314)
OT: explicitly excluded from plan benefits	2.850 (4.150)	3.1480 (4.1815)
OT allocation: by seniority in CBA		(omitted)
OT allocation: unknown		3.2946 (3.2756)
OT allocation: not by seniority in CBA		-0.5728 (5.3578)
Constant	24.73 (22.18)	27.15 (63.85)
N	168	40
R-squared	0.2342	0.2716

Appendix B: Estimation of expected pension cost

For the 1967 plan section we write:

$$P^{*}_{i,t}(x) = OW_{i,t} E_t^{r_i, I_{i,1,r}, I_{i,2,r}} [\frac{1}{3} \underbrace{I_{i,1,r,t}(x)}_{\substack{\text{Variable representing} \\ \text{whether current year's} \\ \text{pay is one of the three} \\ \text{highest at retirement}}} \underbrace{I_{i,2,r}}_{\substack{\text{Variable representing} \\ \text{whether individual works} \\ \text{long enough to vest}}} \underbrace{\alpha(N_{i,t} + r_i - t)}_{\substack{\text{Fraction of salary} \\ \text{replaced by pension} \\ \text{plan as a function of} \\ \text{total length of service} \\ \text{at retirement}}} \underbrace{(1+d)^{t-r} a_{i,r}}_{\substack{\text{Present value of} \\ \text{lifetime annuity} \\ \text{from retirement age}}} \mid OT_{i,t} = x] \quad \text{(B.1)}$$

where:

$a_{i,r}$ is an annuity factor representing the present value of annuity payments to individual i, starting at their retirement or DROP date r_i (a random variable), (we calculate this annuity factor using mortality assumptions used by the plan and an annual discount rate of 2.5% p.a.);

d is an annual discount rate (we use 2.5% p.a.);

$N_{i,t} + r_i - t$ is the individual's total length of service at retirement;

$I_{i,1,r,t}(x)$ is a random variable measuring whether year t pay is used to calculate the pension benefit at date r_i (if, at retirement, year t is one of the highest three year's earnings and $t < r_i$, then $I_{i,1,r,t}(x) = 1$, else it is 0);

$I_{i,2,r}$ is a random variable equal to 1 if the individual vests (that is, has served more than 10 years when they retire) and zero otherwise; and

$OT_{i,t}$ is the number of overtime hours the worker is assumed to have done in year t.

Note that for employees who have elected DROP, $t > r_i$, $I_{i,1,r,t}(x) \equiv 0$ and hence $P^{*}_{i,t}(x) = 0 \; \forall \, x$. The expectation is joint over random variables r_i, $I_{i,1,r,t}(x)$ and $I_{i,2,r,t}$, and conditional on all information known to the employee up to time t.

We first use the wage data to obtain the following empirical distributions for each department (when simulating, we draw from the appropriate empirical distribution):

1. Distributions of increases in base salary, by tenure. For employees of each tenure, we obtain a sample of the 1-year increase in base salary.
2. Distribution of wage increases, by time to retirement (or DROP). For employees with each number of years before (known) retirement or DROP, we obtain a sample of the 1-year increase in base salary.
3. Distribution of overtime pay (as a % of base salary), by tenure. For employees of each tenure, we obtain a sample of the overtime pay divided by the base salary in that year.
4. Distribution of overtime pay (as a % of base salary), by time to retirement (or DROP). For employees with each number of years before (known) retirement or DROP, we obtain a sample of the overtime pay divided by the base salary in that year.
5. A general random variable equal to the ratio between base salary and gross pay less overtime. This is intended to model issues such as unpaid leave.

Next, we turn to the procedure of estimating the variables measuring pension incentives, START, WIDTH and HEIGHT, and pension wealth, $PW_{i,t}^{*,OT}(x)$ and $PW_{i,t}^{*,NOT}(x)$. For each individual in the payroll data:

1. Take a draw of the joint variable (r,N) from the pension fund data. Note that r is the employee's age at retirement (or DROP) and N is the employee's tenure at retirement or DROP (whichever is earlier). Check that this variable is consistent with the (known) maximum tenure up to that point of this individual (using the original appointment date, which is in the payroll data). If not, redraw until the draw is consistent with the known tenure of the individual. Use N and the (known) pension plan the individual is a member of to calculate $\alpha(N_{i,r})$.
2. To estimate the highest three (or five) possible career annual wages, we simulate the employee's wage increases, overtime and unpaid leave from the current year,

in each year up to the date of retirement. We use the wage increases and overtime samples by year of tenure up to 5 years before retirement, and the wage increases and overtime samples by years to retirement thereafter, for workers in the current worker's department. This gives the full path of wage/overtime history for the individual until retirement.

3. If the employee has not vested (vesting occurs after 10 years' service in all Philadelphia plans), calculate the probability that the employee will vest, using the termination assumptions made by the actuary in the plan valuation (City of Philadelphia, 2018). This is $E[I_{i,2,r,t}]$. Note that we assume that $I_{i,2,r,t}$ is independent of all the other random variables.

4. With (r,N) and the employees original appointment date from the payroll data, we calculate their remaining working life and their (implied) current age. r gives us $a_{i,r}$ (which we calculate using the mortality assumptions used by the actuary in the plan valuation, and a discount rate of 2.5% p.a.) and $(1+d)^{t-r}$. We use a discount rate of 2.5% p.a.

5. With the path of wage history, we can also record the highest 3 (or 5) annual earnings before retirement, with and without pensionable overtime, denoted as $highest1_{i,t}^{*,OT}(x)$, …, $highest3_{i,t}^{*,OT}(x)$, and $highest1_{i,t}^{*,NOT}(x)$, …, $highest3_{i,t}^{*,NOT}(x)$[17].

All of these items together allow one instance of $(1+d)^{t-r}$, $a_{i,r}$, $I_{i,1,r,t}(x)$, $I_{i,2,r,t}$, $\alpha(N_{i,r})$, and $highest1_{i,t}^{*,OT}(x)$, …, $highest3_{i,t}^{*,OT}(x)$, $highest1_{i,t}^{*,NOT}(x)$, …, $highest3_{i,t}^{*,NOT}(x)$. For each individual in the sample, we start from the first year of observation and repeat step 1-6 for 1,000 runs. Then we use the (r,N) generated in step 1 to repeat steps 2-6 for 1,000 runs for each of the following years. We estimate HEIGHT as the simple average of $\frac{1}{3}(1+d)^{t-r}a_{i,r}I_{i,2,r}(t)\alpha(N_{i,r})$ across the 1,000 runs, $PW_{i,t}^{*,OT}$ the average of $\frac{1}{3}(highest1_{i,t}^{*,OT}+highest2_{i,t}^{*,OT}+highest3_{i,t}^{*,OT})(1+d)^{t-r}a_{i,r}\alpha(N_{i,r})$, and $PW_{i,t}^{*,NOT}$ the average of $\frac{1}{3}(highest1_{i,t}^{*,NOT}+highest2_{i,t}^{*,NOT}+highest3_{i,t}^{*,NOT})(1+d)^{t-r}a_{i,r}\alpha(N_{i,r})$. We use different formulae for each section of the pension plan, as appropriate.

Appendix C: Philadelphia overtime allocation rule

19(I) states:

> *ASSIGNMENT OF OVERTIME – Overtime work for regular full-time employees shall be assigned in accordance with the following:*
>
> 1. *Each department shall establish departmental work site and shift volunteer overtime lists in the work location where the employees regularly work.*
> 2. *Employees on the overtime desired lists shall be selected in order of their seniority within each classification on a rolling basis.*
> 3. *If the voluntary overtime list does not provide sufficient volunteers, the department may require other departmental employees to work overtime. Said overtime shall be assigned on the basis of inverse seniority within each classification*
> 4. *A departmental labor management committee comprised of equal numbers of management and union representatives shall, by majority vote, grant to individuals (upon application) temporary or permanent exemption from the assignment of mandatory overtime based on objective standards developed by the committee*
> 5. *No full-time regular employee will be required to work overtime on more than four (4) of the employee's five (5) scheduled days, or work over ten (10) hours on a regularly scheduled day or over eight (8) hours on a non-scheduled day.*
> 6. *Notwithstanding this provision, individual employees shall retain the right to volunteer to work overtime to complete work assignments in progress, and the City may require overtime in cases affecting public health or safety.*

Appendix D: The theory and practice of overtime allocation within work-groups

In the previous section, we showed how the interaction between pension rules and worker histories gives rise to substantial heterogeneity in the amount of pension compensation for an hour's overtime work. In particular, workers with greater tenure are likely paid much more for overtime than younger workers, even if they earn the same cash wages and have the same job title, and, for these more senior workers, the more overtime they do in a year, the greater the value of their pension compensation for each additional hour. It should be obvious that this creates strong financial incentives for workers: firstly, the longer the tenure of a worker, the greater their incentive to work overtime, and, secondly, when they do work overtime, they have an incentive to concentrate it in as few years as possible. These two together implies that it is optimal for workers to try and concentrate overtime work into as few years as possible near the end of their working careers.[27]

In this section, we use a simple OLG dynamic game to show how these incentives give rise to a co-ordination problem between different generations of workers. We argue that this dilemma can only be resolved by collective action by workers that is binding on future generations of workers as well as current ones. We use our model to show why in Philadelphia, as in many other unionized states and cities, overtime allocation rules are a subject of collective bargaining, and how backloaded compensation in general, and pension incentives in particular, influence the overtime allocation rules that result from that process. We then discuss how overtime is actually allocated in Philadelphia and other cities and states and show how this is consistent with the theory.

The fundamental cause of the dilemma is that the total amount of overtime that workers can do in any year is limited by the needs of the employer. We show that under certain mild conditions, young workers are better off if they allow current older

27 This incentive holds even though the Philadelphia plan, unlike many DB plans, bases pensions on the highest three years, rather than the final three (or five) year's salary. For those plans, these incentives would be even stronger.

workers to choose how much overtime to perform, and do whatever is left, but only provided future young workers do the same to them when they are old. If young workers choose their overtime first and leave the slack to the old, in the absence of any compulsion on future young workers, the first-best equilibrium is unstable and reverts to a myopic equilibrium under which all workers are worse off than if the old are allowed to choose first.

D.1 An inter-generational model of the allocation of overtime work within work-groups

Our basic theoretical framework is an open group of workers tasked with performing particular duties for an employer. An example of such a functional work-group might be a group of call center workers with the responsibility of answering a particular telephone number, or a group of workers responsible for cleaning or maintaining a particular building (e.g. washing the windows of Philadelphia Airport). Within this group, we make two basic assumptions. First, we assume that there is some overtime work which allows workers to earn income over and above their regular full-time salaries. Second, we assume that workers receive differential compensation in respect of this overtime work that is, in part, related to their seniority. In our analysis, these differentials are provided largely by pensions, as shown in the previous section, although any back-loaded compensation system (such as seniority pay systems) could provide another example and lead to the same conclusion.

To abstract from unnecessary details, we propose an overlapping generations model of a functional work-group with $2N$ members. Each member works for two periods: when they are young and when they are older. They spend one further period retired. There is no attrition and no mortality, so N of the workers are young, N are older and N retired in each period. There is a constant flow of MN period-equivalents of overtime work per period, which must be completed. We assume in this stylized example that all workers receive the same rate of pay per period worked, including overtime, W, but that income in the retired period is provided by a DB-type plan that provides a payment equal to a fraction α of cash compensation during the second working period. For simplicity, we assume that W is constant. The flow of MN overtime period-equivalents each period must be divided between the $2N$ working members of the group. Generations are assumed to be indifferent to one another, so there is no inter-generational altruism.

We first examine the overtime allocation that a benevolent social planner who ignored the interests of the employer would choose. Treating each worker identically, a stable equilibrium would require each worker to perform M hours of overtime over their life. Because our focus is on labor supply, we follow Saez (2010), abstract from risk aversion over wealth and assume that the central planner would allocate overtime to workers over their lives to maximize a lifetime utility function equal to the sum of felicity functions in each period that are quasi-linear and iso-elastic, with elasticity e:

$$\max_{\{l_Y,l_O\}} u(l_Y,l_O) = a_L - \frac{\beta}{1+\frac{1}{e}}(T+l_Y)^{1+\frac{1}{e}} - \frac{\beta}{1+\frac{1}{e}}(T+l_O)^{1+\frac{1}{e}} \quad , \qquad \text{(D.1)}$$

subject to their lifetime budget constraint, $a_L = WT + Wl_Y + (W+\alpha)T + (W+\alpha)l_O$, and the lifetime constraint on total overtime $l_Y + l_O = M$, where subscripts refer to young (Y) or old (O), a represents assets, and l the amount of overtime they work, β is the worker's disutility of work and T is the total hours of full-time work. We denote the solution to this lifetime problem l_O^{FB} and l_Y^{FB}, the first-best outcome.

Note that for simplicity we assume that workers are indifferent to the allocation of income between periods (perhaps because they face no credit constraints and are free to borrow and lend at the same rate). The optimal allocation across periods depends on the worker's elasticity of labor supply, e. The first order condition of the model can be derived in the usual way as:

$$(T + M - l_Y^{FB})^{1/e} - (T + l_Y^{FB})^{1/e} = \alpha / \beta . \qquad \text{(D.2)}$$

We do not need to go further to note that if labor supply is infinitely elastic, workers are indifferent to the allocation of labor across periods and the optimal solution is to do all overtime in the second period. As labor supply becomes more inelastic (e falls), labor supply is shifted away from the old toward the young: l_Y^{FB} is a decreasing function of e. However, unless labor supply is perfectly inelastic, it is always optimal for workers to do more overtime when they are old than when they are young (and if labor supply is perfectly inelastic, workers would voluntarily work no overtime at any price in this model in any case).

Note that the employer's constraint that total overtime equals M in each period does not imply a lifetime constraint on overtime for any particular generation (except in the sense that it must be less than M in each period). One particularly powerful generation might be able to extort overtime from the old when they are young, and from the next generation of young when they are old.

Such a generation, caring only about the current period, would solve the following one-period problem when they are young:

$$\max_{\{l_Y\}} u_Y^M(l_Y) = a_Y - \frac{\beta}{1+\frac{1}{e}}(T+l_Y)^{1+\frac{1}{e}} \text{ s.t. } a_Y = WT + Wl_Y \text{ and } l_Y \le M, \tag{D.3}$$

and this one

$$\max_{\{l_O\}} u_Y^O(l_O) = a_O - \frac{\beta}{1+\frac{1}{e}}(T+l_O)^{1+\frac{1}{e}}, \text{ s.t. } a_O = (W+\alpha)T + (W+\alpha)l_O \text{ and } l_Y \le M, \tag{D.4}$$

when they are old. Solving these models in the usual way yields the myopic optimal amount of overtime work for young workers of $l_Y^M = \min((W/\beta)^e - T, M)$ and for old workers of $l_O^M = \min(((W+\alpha)/\beta)^e - T, M)$. We call these the myopic optimum overtime allocations because they are the overtime allocations that each generation would try to achieve, if they were powerful enough and cared only about the current period.

From now on we assume that the values of M, W, α, β, and e are such that $0 < l_Y^M < M$, $0 < l_O^M < M$ and $l_Y^M + l_O^M > M$, where M is the number of overtime hours per worker in each age group. This assumption simply requires that staffing levels and current overtime wages (and pensions) are set high enough that the entire overtime needs of the group could be met at this wage (and pension) level without resorting to compulsion, and that neither group would voluntarily perform all the overtime on its own, given the option.[28] However, this assumption implies that the generations compete to perform overtime, and provided that $l_Y^M + l_O^M > M$, the myopic solutions above can never give rise to the first-best outcome. Some non-market rule is needed to allocate overtime between the young and the old.

We examine two such rules: 'young preference' and 'old preference'. Under 'old preference', older workers are given the opportunity to perform overtime first. We denote the overtime choices of old and young under this rule as l_O^{OP} and l_Y^{OP} respectively. In our model above, under the 'old preference' rule, old workers, indifferent to the preferences of current young workers, face a one-period problem, and they will choose the myopic optimum $l_O^{OP} = l_O^M$, leaving the remaining overtime to be done by the young, so $l_Y^{OP} = M - l_O^M$. When old, the current young will choose $l_O^{OP} = l_O^M$ again, for the same reason, giving them a lifetime allocation of overtime of M hours. No party has an incentive to deviate, and the myopic 'old preference' allocation is therefore stable.

Under 'young preference', the young are given the option to do overtime first. A rational young worker internalizing the lifetime overtime constraint might think about choosing the same lifetime optimum as the central planner, so denoting the overtime choices of young and old under this rule as l_Y^{YP} and l_O^{YP} respectively, $l_Y^{YP} = l_Y^{FB}$, leaving the remaining overtime to be done by the old, so $l_O^{YP} = M - l_Y^{YP} = l_O^{FB}$. But this allocation is not stable. In fact, regardless of what future young generations choose to do, it is always optimal for current young generations to deviate from the first-best outcome and choose the myopic optimum. To see why, note that the current young have to choose:

---

28 Note that this assumption implies that the employer is paying wages that are too high given the preferences of employees for labor and leisure and their elasticity of labor. A profit-maximizing employer would choose values of W and α that allowed its demand for labor, including overtime, to be satisfied at the lowest possible price. In this paper, we abstract from this requirement, based on two grounds. First, around 30% of state and local workers, including those in Philadelphia, are unionized. As is well known (see, for example, Oswald, 1993), unions drive up the price of labor above market-clearing rates. Secondly, hourly overtime rates are not determined by employers, but are set at 1.5 times base hourly pay by the Fair Labor Standards Act (1938), meaning that the employer may have no ability to reduce overtime wages to ensure that labor supply equals demand (although it could reduce regular wages). Thirdly, overtime needs fluctuate and conscientious employers might pay higher wages to avoid being unable to cover overtime needs in extraordinary situations.

$$\max_{\{l_Y\}} u(l_Y) = a_L - \frac{\beta}{1+\frac{1}{e}}(T+l_Y)^{1+\frac{1}{e}} - \frac{\beta}{1+\frac{1}{e}}(T+M-x)^{1+\frac{1}{e}}, \tag{D.5}$$

subject to $a_L = WT + Wl_Y + (W+\alpha)T + (W+\alpha)(M-x)$ where x is what they think future young generations will choose. But the first order condition of this problem is exactly the same as the myopic problem, regardless of the value of x. So the current young will always choose the myopic solution. Because future young will make exactly the same calculations, this myopic equilibrium is the only stable outcome under the 'young preference' rule. Hence, $l_Y^{YP} = l_Y^M$ and $l_O^{YP} = M - l_Y^M$. So both the 'old preference' and the 'young preference' rules lead to the myopic equilibria, but driven by different generations and so with different lifetime distributions of overtime.

The overtime allocations across life under the 'young preference' and 'old preference' rules are both stable, and are shown in Table D.1.

Insert Table D.1 near here.

We now show that under mild assumptions, the 'old preference' rule always gives rise to higher lifetime utility than the 'young preference' rule.

Lemma: If $l_Y^M + l_O^M > M$, and $\alpha > 0$ then the 'old preference' rule always gives rise to higher lifetime utility than the 'young preference' rule.

Proof:

The lifetime utility is given by:

$$u(l_Y, l_O) = a_L - \frac{\beta}{1+\frac{1}{e}}(T+l_Y)^{1+\frac{1}{e}} - \frac{\beta}{1+\frac{1}{e}}(T+l_O)^{1+\frac{1}{e}} \quad , \tag{D.6}$$

where $a_L = WT + Wl_Y + (W+\alpha)T + (W+\alpha)l_O$.

Substituting the values off Table 2 in the text into the lifetime utility gives:

$$u(l_Y^{OP}, l_O^{OP}) = WT + Wl_Y^{OP} + (W+\alpha)T + (W+\alpha)l_O - \frac{\beta}{1+\frac{1}{e}}(T+l_Y^{OP})^{1+\frac{1}{e}} - \frac{\beta}{1+\frac{1}{e}}(T+l_O^{OP})^{1+\frac{1}{e}} \tag{D.7}$$

$$u(l_Y^{YP}, l_O^{YP}) = WT + Wl_Y^{YP} + (W+\alpha)T + (W+\alpha)l_O^{YP} - \frac{\beta}{1+\frac{1}{e}}(T + l_Y^{YP})^{1+\frac{1}{e}} - \frac{\beta}{1+\frac{1}{e}}(T + l_O^{YP})^{1+\frac{1}{e}} \qquad \text{(D.8)}$$

The difference is:

$$\begin{aligned} u(l_Y^{OP}, l_O^{OP}) - u(l_Y^{YP}, l_O^{YP}) = Wl_Y^{OP} + (W+\alpha)l_O^{OP} - \frac{\beta}{1+\frac{1}{e}}(T + l_Y^{OP})^{1+\frac{1}{e}} - \frac{\beta}{1+\frac{1}{e}}(T + l_O^{OP})^{1+\frac{1}{e}} \\ - Wl_Y^{YP} - (W+\alpha)l_O^{YP} + \frac{\beta}{1+\frac{1}{e}}(T + l_Y^{YP})^{1+\frac{1}{e}} + \frac{\beta}{1+\frac{1}{e}}(T + l_O^{YP})^{1+\frac{1}{e}} \end{aligned}. \qquad \text{(D.9)}$$

Substituting values off the table and collecting terms gives:

$$\begin{aligned} u(l_Y^{OP}, l_O^{OP}) - u(l_Y^{YP}, l_O^{YP}) &= \alpha((W+\alpha)/\beta)^e + \alpha(W/\beta)^e - \alpha(2T+M) \\ &- \frac{\beta}{1+\frac{1}{e}}\left[(W/\beta)^{1+e} - ((W+\alpha)/\beta)^{1+e}\right] \\ &- \frac{\beta}{1+\frac{1}{e}}\left[(2T+M-((W+\alpha)/\beta)^e)^{1+\frac{1}{e}} - (2T+M-(W/\beta)^e)^{1+\frac{1}{e}}\right] \end{aligned} \qquad \text{(D.10)}$$

Because $\alpha > 0$, the second two terms are both positive. So if we can show that the first term is positive then the lemma is proved.

But the first term equals:

$$\alpha(l_O^{OP} + T) + \alpha(l_Y^{YP} + T) - \alpha(2T+M) = \alpha(l_O^M + l_Y^M - M) > \alpha(M - M) > 0, \qquad \text{(D.11)}$$

with the last inequality from the assumption of the lemma. QED

The implication here is that if workers were to sit down together and choose an allocation rule that was stable, both young and old workers would prefer the 'old preference' rule over the 'young preference' rule. Under the 'old preference' rule, however, young workers do less overtime than they would do either under the 'young preference' rule or under the first-best allocation. In this sense, they sacrifice overtime when they are young, in exchange for doing more overtime when they are old. This sacrifice can be viewed as posting a bond, similar to backloaded compensation models studied by Lazear (1979). However, the bond only has value if the current young can be sure that future young will sacrifice overtime to them when they are old. For the young to accept the 'old preference' rule, they therefore need to have confidence that the choice of allocation rule itself is stable. In the next section, we argue that it is for this reason that overtime allocation rules are included in most collective bargaining agreements: doing so solves the co-ordination problem between young and old that we have highlighted here, where compensation is backloaded. If the inter-temporal elasticity of substitution of labor is sufficiently high (in our model it is constrained to equal the intra-temporal elasticity for ease of explication) then the 'old preference' rule actually approaches the first-best outcome. We suggest that the theory we have presented above explains how backloaded compensation patterns create an incentive for overtime to be allocated by seniority, as happens in many cities and states in our data.

It is important to note that our model does not imply that pension compensation is the *only* factor underlying the 'old preference' rule. In fact, any back-loaded compensation system (such as seniority wages, retiree healthcare, and others) will give the same incentive. Other explanations for the overtime allocation rule are also possible, such as inter-union political dynamics. We simply argue that the model shows that pension rules create additional incentives for workers to allocate overtime on the basis of seniority at the margin.[29]

Note also that the 'old preference' rule will lead to $dl_O^{OP} / d\alpha > 0$, but $dl_Y^{OP} / d\alpha < 0$, while the 'young preference' rule will have $dl_Y^{YP} / d\alpha = 0 = dl_O^{YP} / d\alpha$. Under the 'old preference' rule, the employer's overall constraint on overtime induces an overtime response in young workers to changes in the pension costs of old workers, even though these changes in wages do not directly affect them (in the current period). We use this condition as a basis for our empirical work in a subsequent section by examining worker's response to their own pension compensation (our model predicts that workers will respond positively) and to the pension compensation of workers more and less senior than them (we predict that workers will respond negatively to the pension compensation of other workers more senior to them, but not to the pension compensation of workers more junior to them).

29 Of course the greater the fraction of workers that leave before their pension benefits vest (see Munnell et al, 2012), the weaker the pension incentive structure analyzed by the model will be.

TABLE D.1: Allocation of overtime hours across life under 'young preference' and 'old preference' rules

	Young	Old
'Young preference'	$l_Y^{YP} = (W / \beta)^e - T$	$l_O^{YP} = M + T - (W / \beta)^e$
'Old preference'	$l_Y^{OP} = M + T - ((W + \alpha) / \beta)^e$	$l_O^{OP} = ((W + \alpha) / \beta)^e - T$

NOTE: $l_Y^{YP} = l_Y^M$ and $l_O^{OP} = l_O^M$, the myopic allocations discussed in the text.