



Economics Working Papers

2015-06

Documenting and Improving the Hourly Wage Measure in the
Danish IDA Database

Christian Gjødesen Lund and Rune Vejlin



AARHUS
UNIVERSITY

BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

Documenting and Improving the Hourly Wage Measure in the Danish IDA Database

Christian Giødesen Lund[†], Rune Vejlin^{*}

Abstract: This paper overhauls the hourly wage measure that is most often used in Danish research, the TIMELON variable in the IDA database. Based on a replication that we have constructed, we provide a documentation of the wage variable, the first of its kind, then continue with a performance analysis. We find four puzzles. 1) The wages of part-timers fall steeply from 1992 to 1993, 2) the wages of full-timers fall from 2003 to 2004, 3) the level of the part-timer wages is around 12.5% higher than it should be, and 4) the wages of new hires fall steeply from the first year of employment to the second year. We analyse these puzzles in depth and solve almost all of them. Finally, we propose a new hourly wage measure that incorporates all the solutions and we show that it performs much better.

JEL classification: J00, J31

Keywords: Danish hourly wages, IDA data

Acknowledgements: We would like to thank Søren Leth and Jørn Schmidt from Statistics Denmark for help with old pieces of code and variable definitions. We would also like to thank Annette Mortensen for proof-reading. Finally, this paper could not have been written if it was not for the help of Henning Bunzel, who was instrumental in establishing the contact with Statistics Denmark and obtaining the data. All errors are our own.

[†] Copenhagen Economics A/S. email: cgl@copenhageneconomics.com

^{*} Corresponding author: Department of Economics and Business, Business and Social Sciences, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V., Denmark. Email: rvejlin@econ.au.dk

Chapter 1: Documentation and Puzzles of the IDA Wage Measure

1 Introduction

This chapter provides a thorough documentation of how the hourly wage measure in the Danish IDA database is computed. So far, only sparse documentation exists in Danish, such as Statistics Denmark (1991) and the readily available but quite cursory online documentation (Online (1) in the References section). Unfortunately, even after a careful study of these sources one is left with the feeling that the estimation method is an impenetrable black box. In order to fully understand how the hourly wage is actually computed, what its advantages and disadvantages are, and if there is room for improvements, it is necessary to dig much deeper than what the two sources do. This chapter aims at giving the reader a good understanding of what Statistics Denmark actually does and a sense of some of the short-comings inherent in the estimation method. First, the chapter seeks to present the essentials and rationale of the estimation method in a simple and coherent way, drawing on the relevant institutional details when necessary. To convey the idea of the method, the account will deviate slightly from what Statistics Denmark actually does in their program code, and then go on to consider the differences in Section 5. This approach is harmless as the differences only boil down to rounding of some of the variables. Second, the chapter provides four simple examples in which the hourly wage measure behaves too suspiciously to not reflect fundamental problems with the estimation method. The first puzzle concerns the time series of part-timer wages. In most years, the wages of part-timers track the wages of full-timers perfectly, but not in 1992-1993 where the wages of the full-timers stagnate and the wages of the part-timers take a big hit. The second puzzle concerns the time series of the full-timer wages. From 2003 to 2004, their wages fall on average whereas the wages of the part-timers rise, giving

rise to a falling aggregate wage. Third, in a comparison with another source of wage information we find that the level of the full-timer wages is roughly right in most years, but the level of the part-timer wages seems to be completely off the mark, with a difference of roughly 20 DKK or around 12.5%. Finally, tenure profiles are unrealistic. In the second year at a firm, a worker's wage falls. We argue that the problem is one of composition since the wage cuts are not seen for part-timers or full-timers on their own. All these puzzles are solved or almost solved in later chapters.

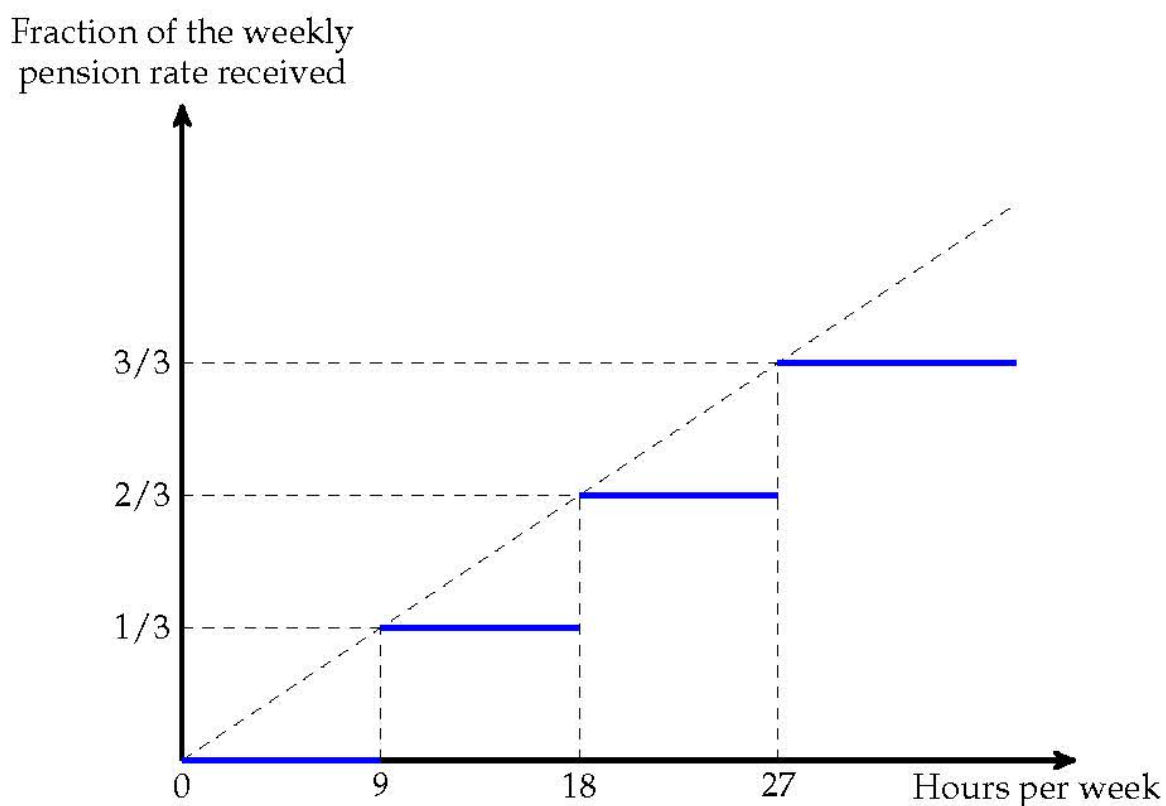
The source of wage information that we use for the comparison is a survey called "Lønstatistik" ("Wage and Salary Statistics") produced from 1997 by Statistics Denmark. We are among the first to complete a formal comparison, possibly due to the fact that the survey consists of several wage components that must be selected and combined in a specific way in order to match the IDA wage concept, which hitherto was almost undocumented. We are aware of one other attempt in DØRS (2003). However, our results are very different even though we use the same components. In particular, one implication of the finding in DØRS (2003) is that the IDA wages of part-timers should be closer to the surveyed wages than the IDA wages of full-timers. We find that quite unlikely to be true in practice. Unfortunately DØRS (2003) is only a very brief draft that does not specify the sample that was used, which precludes a definitive analysis of what drives the differences. Yet we argue that some elements in our approach are more appropriate, and that overall ours is a useful benchmark that is difficult to improve upon within the constraints of the Lønstatistik.

This chapter is structured as follows. Section 2 explains how Statistics Denmark estimates the hourly wage in the IDA database, Section 3 identifies the historical values of the parameters that enter the estimation, and Section 4 presents the specific examples of the shortcomings of the wage measure. Finally, some details about the practical implementation of the estimation method are described in Section 5, and the Appendix contains extra details on some historical values of parameters that are related to the ones used in the estimation.

2 Estimation of the Hourly Wage

The hourly wages provided in the IDA data are estimated by Statistics Denmark based on accurate yearly earnings records from the tax authorities and an estimate of the number of hours worked during the year. The earnings include overtime and compensation while on vacation. In many cases, they also include compensation during illness, but pensions are not included. Hours are estimated based on (accurate) yearly pension contribution records for workers who are employed in either main or second jobs in the last week of November. The estimation is possible because the accumulated pensions depend on hours in a known way, and on hours only. See Figure 1.1 for an example with weekly compensated workers.

Figure 1.1: Weekly Hours and the Received Fraction of the Weekly Pension Rate



As is seen, the relationship is not invertible for these workers, but Statistics Denmark backs out a lower bound for hours that is the minimal number of hours consistent with the accumulated pensions in the year, and an upper bound reflecting an assumption on how many hours employees can maximally work without earning extra pensions. For example, from the figure it can be seen that an employee working 18 hours could have worked 9 more hours and still earned the same pension amount, and an employee working 27 hours could have worked more than 27 hours without accumulating more pensions. Other restrictions shrink the interval of possible hours. In the following, we carefully explain the method currently in use. The first section is based on Statistics Denmark (1991), the pension laws contained in the References section, online resources for firms on types of rates and historical values (Online 2, 3) and a sample of an original program.

Institutional setting

In Denmark, workers save for their old age in a number of ways. One is through the “Additional Pension from the Labour Market”, called ATP. Employers make contributions for each employee to a pension fund and they increase with hours worked. The specific relationship between amounts and hours depends on whether the worker is a casual labourer compensated by the hour or is in more stable employment and compensated on a weekly, bi-weekly, or monthly basis. The relationship also depends on whether the worker receives the A-, B-, C-, D-, E-, or F-pension rate. Within each mode of compensation and type of rate, the relationships are the same for all workers. In particular, they are the same for all levels of wages and salaries. How useful these relationships are for estimating a universal measure of hours would seem to depend on what information were available to distinguish workers with different pay cycles and rates. Fortunately, Statistics Denmark does have access to information on the types of rates and converts all reported amounts to the B-rate. As a result, the only source of heterogeneity in the relationship between

pension amounts and hours is the compensation mode. As it turns out, for most workers there is a way to deal with that, too, even without resorting to information on the compensation mode.

For workers compensated weekly, bi-weekly, or monthly, at the end of the pay cycle the employers determine what hours-bracket each employee is in. There are four brackets, the top-bracket, middle-bracket, low-bracket, and zero-bracket. The time thresholds defining the brackets for bi-weekly and monthly compensated workers are the same as the ones for weekly compensated workers, just scaled by 2 and $4 \frac{1}{3}$ respectively. This guarantees that a worker ends up in the same bracket regardless of how frequently she is paid. Based on the bracket, the employer pays a fraction of the weekly, bi-weekly, or monthly pension rate. Workers in the top-bracket receive the full rate, workers in the middle-bracket receive two-thirds of the full rate, and workers in the low-bracket receive $\frac{1}{3}$ of the full rate. Workers in the zero-bracket receive nothing in that cycle. Now, while the bi-weekly B-pension rate has always been twice the weekly rate, the monthly rate was 4 times the weekly rate until 1992 and lower than that from 1993. Thus, a worker who has worked for a couple of months will receive a higher pension if paid weekly than if paid monthly, simply because the conversion factor used to obtain the monthly pension rate is lower than the number of weeks in a month. However, from 1993 these differences in the ways hours and pensions are related virtually disappear once annual pension rates are considered. This is because monthly compensated workers earn pensions while on vacation but weekly or bi-weekly compensated workers do not. In fact, in order to treat workers equally, the pension rates for weekly and bi-weekly compensated workers are correspondingly higher. This means that if vacation is held as intended, the annually earned pension amounts are not too different for the three compensation modes. As a result, from a yearly viewpoint, the relationships between pension amounts and hours are roughly the same across compensation modes. Indeed, employees subject to the different frequencies of pay who work identical hours end up in the same hours-

brackets over the course of a year, their share of pensions are the same, and the annual pension rates are almost the same, too.

The final group consists of casual labourers (“løsarbejdere” in Danish). These are individuals who are loosely attached to the workplace and not employed there on a consistent basis, and often compensated by the hour or day. The casual labourers accumulate pensions in a different way, namely according to a perfectly linear scheme. Unfortunately Statistics Denmark does not have information on whether an individual is a casual labourer or not, but it is well known that only a very tiny portion of the workforce are employed in this way. We have had conversations with current and former employees at Statistics Denmark who confirmed this impression in no ambiguous terms. As a result, Statistics Denmark estimates the hourly wages as if all workers were paid weekly, bi-weekly, or monthly. This approach should not cause any problems at all for the performance of the estimation method. In sum, by considering annual quantities, the heterogeneity in compensation modes does not matter for the determination of hours from 1993. Whether the heterogeneity matters in 1980-1992 depends on how many workers are compensated on a weekly and bi-weekly basis and will be investigated in Chapter 2.

While the earnings that the hourly wage was estimated from did include vacation pay, overtime, and some illness allowances, the estimated hours do not. Legal holidays and lunch breaks are also not included in the hours estimates. In the next section we will describe how Statistics Denmark obtains the estimates of hours from the reported pensions. We begin with the lower and upper bounds, t_{min} and t_{max} and proceed to how the bounds are used in the estimation. The sources are Statistics Denmark (1991), a sample of an original program, and the pension laws contained in the References section.

The Lower Bound on Hours

It is convenient to begin with some notation and define a number of variables. The specific values taken on by the variables defined in the following will depend on whether the pay cycle is a week, fortnight, or month. But for notational simplicity, that dependence is suppressed. Whenever two variables enter the same expression they are meant to refer to the same pay cycle even if the notation does not explicitly indicate it. Note also that it is very easy to convert the values of one pay cycle to another. To convert from weeks to fortnights, simply multiply by two. To convert from weeks to months, multiply by $4 \frac{1}{3}$.

Full pay cycle: A week, fortnight, or month that is not shortened due to legal holidays. The alternative is a calendar pay cycle. Some but not all calendar pay cycles are full.

h: Effective number of full-time hours in a full pay cycle, as stipulated by agreements between unions and employers. Lunch breaks not included since no work is done during lunch.

$\tau_1, \tau_2, \tau_3:$ The time thresholds defining the hours-brackets in a pay cycle. If a worker works less than τ_1 hours in a cycle, she earns no pensions in that cycle. If she works τ_1 hours or more but less than τ_2 hours, she earns $1/3$ of the pension rate pertaining to the pay cycle. If she works τ_2 hours or more but less than τ_3 hours, she earns $2/3$ of the pension rate. If a worker works τ_3 hours or more, she earns the full pension rate in that cycle. The following relations hold: $0 < \tau_1$, $\tau_2 = 2\tau_1$, $\tau_3 = 3\tau_1$, and $\tau_3 < h$.

A: The yearly pension rate. From 1993, employers are not obligated to pay more than this amount in total for the year, even if the sum of earned pensions in each cycle calls for it; and workers who are in the top-bracket for the entire year have a right to receive this amount, even if the sum of earned pensions in each cycle is lower, as long as the workers are always in the top-bracket¹. From 1964-1992 no yearly benchmark existed and *A* was just twelve times the monthly rate.

¹ Promulgation BKG 1992-09-29 nr 822.

ATP: Total pension contributions from an employer to an employee in a particular job, accumulated over the year and reported to the tax authorities. Statistics Denmark caps accumulated pensions at the yearly rate if they exceed that rate. Thus, $ATP \leq A$. In 1981, *ATP* was capped for around 5% of the employees in main jobs.

T: “Normaltimetallet”, that is, the effective number of hours of work in a year for full-time employees. Vacation, weekends, legal holidays, or lunch breaks are not included in *T* because no work is done during these periods. *T* is effectively agreed on by the social partners and enshrined in collective agreements when they decide on the weekly hours. Note that the collective agreements offer considerable flexibility for the employer to allocate time, especially in industries where this is necessary. *T* is just the reference point of full-time work, and not the exact hours that all full-timers work in a year.

k: An estimate of the number of hours a worker can work in a full pay cycle without earning extra pensions. In the zero-, lower-, and middle-brackets, a worker can maximally work τ_1 hours without passing the threshold to the next bracket and earning more pensions. Since the top-bracket has no upper bound, in that bracket an assumption on how many hours a worker can maximally work in a pay cycle is necessary to obtain a number for *k*. How *k* is chosen in practice by Statistics Denmark will be described in Section 3.

α_k : The maximum possible number of hours for which no ATP is earned if the worker works for a year. An official year is T/h full pay cycles so $\alpha_k = kT/h$. Note that α_k does not depend on the pay cycle since *k* and *h* refer to the same pay cycle.

With these definitions, the lower bound on hours, t_{min} , is computed as

$$t_{min} = \frac{ATP}{A} \cdot (T - \alpha_k).$$

The intuition is that the lower bound is the earned share of the yearly pension rate multiplied by the minimum number of hours one needs to work to obtain the yearly rate. Notice how the bound is computed from an annual perspective. From 1993 this approach effectively neutralizes the impact of the different ways that monthly paid and weekly or bi-weekly paid workers accumulate pensions.

The Upper Bound on Hours

We now proceed to explain how Statistics Denmark computes the upper bound of hours, t_{max} . The upper bound adds to the lower bound an estimate of the number of pay cycles the worker could have been employed in the year multiplied by how many hours in each of those pay cycles the worker could maximally have worked without earning extra pensions. It is convenient to define the following variables.

u_{min} : The minimum number of full pay cycles consistent with ATP and the annual effective pension rate. If one works τ_3 hours or more and receives the full rate, one reaches ATP by only working in u_{min} full pay cycles. The formula is

$$u_{min} = \frac{ATP}{\frac{A}{\frac{T}{h}}},$$

that is, the earned pension amount divided by the full, effective pension rate.

V : The official number of weeks of vacation as guaranteed by law. There is no information in the IDA database about held vacation so the next best alternative is to use the official number for everyone.

u_{max} : An upper bound on the number of calendar weeks of employment in which there is actual work. The bound takes into account that individuals do not work while they are on vacation or full-time unemployed or receive social security allowances for

illness. There is some information in the IDA database about unemployment and illness. Specifically, u_{max} is computed as the number of weeks in a year over and above the number of weeks with vacation, illness allowance, or full-time unemployment insurance.

p : The period of employment reported in weeks by employers. The employers indicate if the employee was employed throughout the entire year, in a continuous period including begin and end dates, or in more than one period without indicating any dates. Note that this number is not so reliable since the employers have no incentive to report it correctly.

u : A mix between a lowest available upper bound and an estimate of the number of full pay cycles in an employment spell. With a slight abuse of notation, it is obtained as $\max\{u_{min}, \min\{p, u_{max}\}\}$ where $\min\{p, u_{max}\}$ is now measured in full pay cycles. (To make the conversion, first convert $\min\{p, u_{max}\}$ into the same pay cycle as u_{min} , then multiply by $T/(h(52 - V))$ to make it a full pay cycle.) Days with vacation or legal holidays are not included in u . Note that in principle $u_{min} > \min\{p, u_{max}\}$ cannot happen, but when it does in the data, then u_{min} is considered the most reliable number, and the contradiction is resolved in favour of u_{min} . Similarly, if $p > u_{max}$ happens then u_{max} wins.

$\alpha_{k,u}$: The maximum possible number of hours for which no ATP is earned if the worker works in u pay cycles, ku . The variable k is the same as the one used to compute α_k .

With these definitions, the estimated upper bound on hours in a job in a year is equal to

$$t_{max} = t_{min} + \alpha_{k,u}.$$

There are two things to note. First, for a given job spell length, it does not matter what the pay cycle is. Second, strictly speaking, t_{max} is not always an upper bound. If $u_{min} < p < u_{max}$ (again abusing notation slightly) and p is lower than the true period length, t_{max} may be smaller than the

true hours worked. Nevertheless, we will stick to referring to t_{max} as the upper bound on hours, and, similarly, to u as the upper bound on pay cycles.

Having established the bounds, we proceed to show how Statistics Denmark computes the actual estimate of hours and a measure of the relative uncertainty of the estimates of hours. The sources are Statistics Denmark (1991), a sample of an original program, and the pension laws contained in the References section.

Estimation of the actual number of hours

Given the bounds on hours, Statistics Denmark uses additional information to determine where in the interval the estimate of hours should be. The main consideration here is if it can be determined whether the worker is a full-timer or part-timer (part-timer does not mean half-timer; it means non-full-timer). Surveys have indicated that individuals who self-report being full-time employed also work full-time. This observation suggests that if it can be identified who are full-timers, their hours should lie close to h in an average pay cycle. On the other hand, individuals who self-report having part-time work are more equally distributed within and across the hours-brackets. Consequentially, the group of workers who are in the top-bracket in all pay cycles consists mostly of full-timers, but not exclusively of full-timers. Similarly, not all full-timers can be found among workers always in the top-bracket. It follows that a good but not perfect tool to determine who are part-timers and full-timers is the top-bracket. As a departure point, Statistics Denmark classifies workers always in the top-bracket as full-timers and the remaining workers as part-timers. On top of this criterion, four other restrictions are in place to better isolate the full-timers among all the workers who are always in the top-bracket. First, if information about part-time insurance is found to apply to a worker always in the top-bracket, then this worker will also not be considered a full-timer but a part-timer instead. The same happens if information about part-time unemployment or high minimum hours in a second job is found to apply. Finally, all

second jobs are part-time jobs. Unfortunately, Statistics Denmark has not found a way to locate the full-timers who are not in the top-bracket in all pay cycles. In practice, to identify workers who are always in the top-bracket, Statistics Denmark uses the condition $u = u_{min}$. If it is satisfied, then the worker has earned her pension amount in the number of pay cycles it takes to earn that amount when always working in the top-bracket, so she must have been in the top-bracket in all pay cycles. Now, for the workers who are classified as full-timers, Statistics Denmark chooses an estimate of hours in the interval by using $u = u_{min}$ to obtain

$$t_{min} = \frac{ATPT}{A} \frac{1}{h} (h - k) = u_{min}(h - k) = u(h - k),$$

$$t_{max} = t_{min} + ku = hu.$$

Thus, hours should equal t_{max} for this group of selected full-time workers. Table 1 illustrates the division between full-timers and part-timers and how frequently each criterion is applied.

It is more complicated to determine the hours of the part-timers since more types of workers comprise this group, and because indications are lacking of where in the interval of hours the true number of hours really lies. A distinction is made between workers in their main and second jobs. The hours in a main job spell are quantified as the average of the bounds, unless the lower bound on the number of hours in a second job is so high that, in all likelihood, the hours in the main job are closer to their lower bound. To operationalize this idea, Statistics Denmark computes the following for main part-time jobs:

T' : The full-time hours of a year that for an employee may be shortened due to illness or full-time unemployment. T' is proportional to T , with a factor of proportionality defined as u_{max} divided by $52 - V$.

t'_{min} : The lower bound on hours in the second job. Equal to zero if there is no second job.

T_{red} : Reduced number of working hours. $T_{red} = T' - t'_{min}$.

If the reduced number of working hours due to employment elsewhere is greater than the lower bound, $T_{red} > t_{min}$, then the second job is not sufficiently constraining, and the midpoint $0.5(t_{min} + t_{max})$ of the interval (t_{min}, t_{max}) is used as the number of hours in the main part-time job. On the other hand, if the reduced number of working hours is smaller than the lower bound on hours, i.e. if $T_{red} < t_{min}$, then it is very likely that the second job is a constraining factor, and the lower bound t_{min} on hours in the main part-time job is used as the estimate of hours. Note that u_{max} used in the computation of T' is an upper bound of the spell length, which makes T_{red} a conservative estimate of the reduced hours. In the other direction counts that workers with two or more jobs may work more than full-time yet full-time hours are assumed in the computation of T_{red} .

Hours of second job spells are again determined differently. Since the information on the main jobs is deemed more reliable², and since the hours of the main jobs are likely to be constraining, more weight is attached to the lower bound on hours in the other job spells. Statistics Denmark again computes T' , t'_{min} , and T_{red} , this time for a worker's other jobs than her second job. Then hours are computed as

$$\min\left(\max(T_{red}, t_{min}), \frac{t_{min} + t_{max}}{2}\right).$$

In other words, if $T_{red} > t_{min}$, then T_{red} may be used as the estimate of hours if T_{red} is also lower than the average of the bounds. Otherwise the hours of the second part-time job are computed as for the main part-time job.

Table 1. Criteria for Full-Time and Part-Time Employment, and Their Frequencies in Percent

| | Criterion | Men | Women |
|------------------|--|------------|--------------|
| Full-time | Always top-bracket, main job, no part-time insurance, no part-time unemployment, no time reduction | 51.9 | 42.9 |
| Part-time | One or more pay cycles below top-bracket, main job | 33.7 | 41.1 |

² Statistics Denmark (1991), p. 27 footnote 17.

| | | | |
|----------------|---|------|-----|
| | Always top-bracket, part-time insured, main job | 0.0 | 1.5 |
| | Always top-bracket, full-time insured, part-time unemployed, main job | 3.7 | 4.5 |
| | Always top-bracket, full-time insured, no part-time unemployment, time reduction (i.e. $T_{red} \leq t_{min}$), main job | 0.1 | 0.2 |
| | Second job | 10.2 | 9.4 |
| Unknown | No ATP earned and period of employment unreported or zero | 0.4 | 0.4 |
| Total | - | 100 | 100 |

Source: Authors' calculations using the sample described in Section 4.

Note: Reduced hours are not only used in part-timer category 4 but also in the other part-timer categories if the appropriate conditions are satisfied.

The first thing to note from Table 1 is that the assumptions regarding reduced hours really do not matter for main jobs. It can be discussed if they are appropriate, but in the end, it turns out that they are not important, since the additional constraint only rarely binds in the sample. On the other hand, the criterion for being a full-timer really is restrictive. Only half of the men are employed full-time, and less than half of the women. In contrast, in the "Registerbaseret Arbejdsstyrkestatistik" ("Register-based Workforce Statistics") a much larger fraction is full-time employed. Although the full-timer concept in these statistics is designed to capture the employment relation rather than hours, the discrepancy should still be indicative of a criterion that is quite conservative.

Relative Uncertainty of the Estimated Wages

The final step in the process of estimating hours and wages is to gauge the uncertainties associated with the estimates. Given that the hours are just one possible choice out of an interval of options, the estimated wage falls within a similar range. How useful the wage estimate is depends on the magnitude of the wage relative to the width of the interval it is taken from. Statistics Denmark calls this measure the "relative uncertainty" and makes it available along with

the estimated hourly wage. At present, it is not known with certainty how Statistics Denmark calculates the relative uncertainty of the wages of workers in second jobs. But for main jobs, the formula is

$$r_w = \frac{w_{max} - w_{min}}{w} = t \left(\frac{1}{t_{min}} - \frac{1}{t_{max}} \right),$$

where w is the actual imputed wage defined as earnings divided by the imputed hours t .

For full-timers, $t = t_{max} = t_{min} + \alpha_{k,u}$ and $u_{min} = u$ so the formula reduces to

$$r_w = \frac{\alpha_{k,u}}{t_{min}} = \frac{ku}{\frac{t_{min}}{u_{min}} \cdot u_{min}} = \frac{k}{\frac{t_{min}}{u_{min}}} = \frac{k}{h - k}.$$

If the worker is a part-timer in a main job and the number of hours is reduced because of other jobs, so that $t = t_{min}$, then

$$r_w = \frac{ku}{\frac{t_{min}}{u_{min}} \cdot u_{min} + ku} = \frac{k \frac{u}{u_{min}}}{h - k + k \frac{u}{u_{min}}}$$

If the worker is also always in the top-bracket, $r_w = k/(h - k + k) = k/h$.

If the worker is a part-timer in a main job and there is no reduction in the number of hours due to second jobs, then the hours estimate is the average between the bounds, and

$$r_w = \frac{1}{2} \left(\frac{k}{\frac{t_{min}}{u_{min}}} \cdot \frac{u}{u_{min}} + \frac{k \frac{u}{u_{min}}}{\frac{t_{min}}{u_{min}} + k \frac{u}{u_{min}}} \right) = \frac{1}{2} \left(\frac{k}{h - k} \cdot \frac{u}{u_{min}} + \frac{k \frac{u}{u_{min}}}{h - k + k \frac{u}{u_{min}}} \right).$$

Thus, the relative uncertainty is the average between the relative uncertainties for the two preceding groups of workers.

It can be seen from the formulas for the three groups of workers that the only thing that matters for the relative uncertainty is the relation between the upper bound and lower bound on the number of pay cycles. The higher this ratio, the higher is the uncertainty. Intuitively,

the width of the interval of possible hours is determined by the maximum number of pay cycles, u . On the other hand, the chosen number of hours partly reflects the lower bound on the number of pay cycles, u_{min} . Thus, if u_{min} is large relative to u , the possible range of hours is narrow in comparison to the chosen number of hours, and vice versa. Consider a janitor who works few hours every week. Her hours could instead have been reached by working full-time in few weeks, and thus the relative distance between the minimum and maximum number of weeks is big. Since she could have worked k extra hours for many pay cycles, the uncertainty is high. Or consider a fisherman who works many hours but very intermittently. Again the relative distance between the minimum and maximum number of pay cycles is big, and as a result the number of extra hours the fisherman could have worked is relatively high. On the other hand, for an employee working close to full-time, the number of extra hours is small compared to the chosen number of hours. It follows from this reasoning that what determines the relative uncertainty is the average hours worked in a pay cycle. Few hours on average yield uncertain wages and full-time hours yield wages that are determined quite well.

3 Historical Values of the Parameters

Any documentation of the IDA hourly wage measure would be incomplete without the recent and historical values of the parameters that were used to estimate the wage. The parameter values are essential for understanding the wage measure and its properties better, and for replicating or correcting it if that need should arise. Unfortunately, no complete records of the historical parameters seem to exist, so we try to infer the actually used parameters from many different sources. We use Statistics Denmark (1991), a sample of an original program, Online (3-13), the most recent laws on vacation, a law on working hours and pay of 1985, historical laws on the labour market pensions, and old calendars marked with legal holidays. In addition, we have access to an Excel-file of Statistics Denmark containing the historic number of work days in each month,

the official number of weeks of vacation in each year, the weekly full-time hours, the yearly full-time hours, and values for the yearly pension rate. Unfortunately, the sources provide conflicting information. Below we discuss the evidence and if there are obvious choices of the parameters, and we compare to our own identification exercises that we run when in doubt. Then we decide on the values of the parameters that we deem most likely. Note that we pick the values that apply with weeks as the pay cycle. The other values are easily obtained by multiplication by 2 or 4 1/3. In terms of notation, a “wk” or “w” subscript indicates the weekly value when necessary.

k , the number of hours a worker can maximally work without earning extra pensions. There is no obvious choice for this parameter. In the zero-, low-, and middle-bracket, an employee can maximally work $\tau_3 - \tau_2 = \tau_2 - \tau_1 = \tau_1$ hours without crossing the threshold to the next hours-bracket. That calls for choosing $k = \tau_1$. But workers in the top-bracket have no upper limit on how much they can work without earning extra pensions, so an upper bound must be assumed and the ideal upper bound need not be $\tau_1 + \tau_3$, the value for which the width of the bracket is τ_1 . Thus, potentially there is a tension between choosing k to match the right hours of workers in the lower brackets and workers in the top-bracket. The years 1980-1985 turn out to have an obvious choice of $k_{wk} = 10$. In these years, $h_{wk} - \tau_{3,w} = \tau_{1,w} = 10$, so with an assumed upper bound of h_{wk} the hours-brackets are equally wide. However, in all other years $h_{wk} - \tau_{3,w} \neq \tau_{1,w}$ and mistakes will invariably be made. In practice, the original program code reveals that $k_{wk} = 10$ in all years. A similar result was found in a number of correlation exercises using the IDA hourly wage variable and our own reconstruction of it using the sample described in Section 4. As a result, we are quite confident about this value. Improvements to how k is chosen are suggested in Chapter 2.

Weeks of vacation, V. Both the Excel-file and Statistics Denmark (1991) set the number of weeks of vacation to 4.5 in 1980 and 5 weeks thereafter. So it is very likely that these were also the actual, used numbers. However, Online (4) suggests that the number of weeks of vacation as secured by

law was five weeks already from 1980. The difference could be due to the fact that 1980 was a transition year between 4 weeks in 1979 and 5 weeks from 1980. The number 4.5 could be a judgement deemed the most fitting to the actual numbers of weeks of vacation held by the workforce, but the exact circumstances are not clear. As a result, even if Online (4) suggests 5 weeks, the most likely choice of Statistics Denmark is 4.5 weeks in 1980. Then, by the end of the 1990s and the beginning of the 2000s the vacation concept in the Danish labour market became more flexible and this development could have an impact on how the number of weeks of vacation was chosen. In the initial period from 1980, as a rule of thumb all workers were required to use all their weeks of vacation so the estimation of it was quite reliable. But from the 2000s, the rules for converting the fifth vacation week into a higher wage were made more flexible. As a result, if any adjustment should be made to the number of weeks of vacation used in the estimation, it should go down. However, at the same time, from around 1999 two extra days off work were introduced for many but not all workers. These extra days applied to workers covered by collective agreements and who had earned them, in contrast to the vacation weeks which were secured by law and the same for everybody. Moreover, the extra days off were governed by more flexible rules regarding their placement in the year and their conversion to a higher wage. The arrangement was gradually expanded, and many workers now effectively have the choice of an additional week off. The net result of the two parallel developments is a more flexible vacation concept, and if there is an obvious choice of the number of weeks of vacation, it is probably not too far from 5. Moreover, given that the Excel file must naturally be a recent construction, it is likely to actually reflect the recent practice. For these reasons, we are quite confident that Statistics Denmark uses 5 weeks from 1981 and 4.5 in 1980.

Full-time hours h. The obvious choice of the weekly full-time hours is 37 hours per week from 1991, and 40 hours per week from 1980-1985. All sources agree in these years and there is no

ambiguity. However, from 1986 through 1990, hours were gradually reduced, and unfortunately not on January 1st each year. Paragraphs 1 and 8 of the law on working hours and pay of 1985 makes it clear that the weekly hours dropped from 40 to 39 on December 1st 1986 for most workers and on January 1st 1987 for a small remaining group. In 1987, Online (5, 6, 7, 8) show that the weekly hours fell from 39 to 38.5 on 1st of September for most workers and on 1st of October for some workers. Then, from 1988 the weekly hours fell by a half every 1st of September until they reached 37 in the second part of 1990. As a result, there is no obvious choice of the weekly hours in any of the years 1986-1990. The following possibilities exist.

1. The weighted average of weekly hours throughout the year. To take 1986 as an example, there were 230 work days before the change and 21 after, in total 251 days. The weighted average is $230/251*40+21/251*39=39.92$ hours per week. The weighted averages of the other years can be found in the Appendix.
2. The weighted average of weekly hours throughout the year, rounded to one decimal.
3. The weighted average of weekly hours throughout the year, rounded to an integer.
4. Weekly hours as they were in the beginning and longest part of the year, "January hours".
5. Weekly hours as they were from September 1st, "September hours".
6. Weekly hours as they were at the end of the year, from December 1st, "December hours."

Arguably, in principle the best way of choosing the weekly hours should be the un-rounded weighted average. The worst methods could be Methods 3 and 6. In years in which the January hours are an integer, Method 3 will yield the same hours as the January hours, and in the other years it will yield the September hours. In principle, though, one should not alternate but stick to one method. Similarly, the September and December hours also perform worse than the January hours since they are only correct in the shortest part of the year. Thus, we should expect Statistics Denmark to have used one of Methods 1, 2, or 4, and if not, then perhaps 5.

It turns out that the Excel file uses January hours. Statistics Denmark (1991) agrees in 1986 and 1987, but does not tabulate hours in the years 1988-1990. Thus, it seems that the most likely choice is January hours. But, since the Excel file has been constructed roughly a decade after the actual choices were made, it is not certain that the information in this file is correct. Getting the weekly full-time hours right is crucial for any correction of the estimation method, so to be perfectly sure we run our own identification exercises, using the fact that the different methods will agree in some years and disagree in others.

Consider a full-time employee in the period 1986-1990, which were the years in which weekly hours changed. We know that the relative uncertainty of the wage of such a worker is

$$r_w = \frac{k}{h - k} = \frac{10}{h_{wk} - 10}.$$

In other words, the relative uncertainty is only a function of the weekly hours. A variable that comes very close to the relative uncertainty exists and is readily available in the data. It is called *tlonkval* and the only difference between that and the relative uncertainty is rounding error and a multiplication by a factor of 100. Section 5 describes exactly how *tlonkval* is computed, but the key point for now is that all the other variables entering the calculation will cancel out in the absence of rounding error and simplify to the expression for the relative uncertainty. There are two strategies one can follow to identify the weekly hours using *tlonkval*. The first is to feed the expression for the relative uncertainty with some chosen values of the weekly hours and compare to *tlonkval*. The second is to replicate the exact calculation of Statistics Denmark for the same values of the weekly hours and compare to *tlonkval*. The second approach is much safer because it is not directly vulnerable to rounding error. The only weakness is if somehow the rounding causes the other variables entering the calculation to no longer completely cancel out. But even in that case, the values of the other variables are almost all precisely known and do not pose much of a threat to identification. In effect, errors made with the second approach are second order effects

whereas they are first order with the first approach. In the following, we test the six ways of computing the weekly hours throughout 1986-1990 and reconstruct *tlonkval* as Statistics Denmark has done it, drawing on their raw registers. Deviations from *tlonkval* will indicate that the wrong weekly hours have been chosen. To limit our vulnerability to the (second order) effects of the presence of other variables, we only replicate *tlonkval* for the part of the population for which our replication is the safest, that is, for full-timers with hours higher than 1000, $0 < tlonkval < 100$, $0 < timelon < 1000$, and no period with illness allowances. We also only use an observation if the total hours estimated by Statistics Denmark are not significantly lower than what we would predict using our replication. The reason is that in a few cases we cannot replicate the classification between part-timers and full-timers, so a full-timer from our perspective might be a part-timer from the perspective of Statistics Denmark and thus have *tlonkval* computed differently. The estimate of total hours of Statistics Denmark reveals when we mistakenly have gauged a worker to work full-time. We compare for each of the years 1986-1990 separately using the sums of squared deviations criterion. The results are reported in Table 2.

Table 2. Evaluation of the Six Ways of Choosing Weekly Hours

| Year | Method 1 | Method 2 | Method 3 | Method 4 | Method 5 | Method 6 |
|------|----------|----------|----------|----------|----------|----------|
| 1986 | 0 | 0 | 0 | 0 | 0 | 972363 |
| 1987 | 8149 | 8149 | 0 | 0 | 996822 | 996822 |
| 1988 | 961717 | 961717 | 0 | 0 | 0 | 0 |
| 1989 | 4454 | 4454 | 4454 | 4454 | 985757 | 985757 |
| 1990 | 974552 | 974552 | 0 | 0 | 0 | 0 |

Note: The table shows the Sums of Squared Deviations between predicted and actual *tlonkval*.

The results are clear. Except in 1989, January hours (Method 4) and the rounded weighted average (Method 3) have zero deviations in 100% of the close to one million cases. In 1989, half of a percent of the observations are one integer less than *tlonkval* for unknown reasons. All the other methods get it right or almost right in some years but terribly wrong in others. Note that it is a little surprising that Methods 3 and 4 perform identically since their weekly hours differ in some

years. It must be an artefact of the rounding. Given that Method 3 is not a good method as argued above, the conclusion from Table 2 is that Statistics Denmark uses January hours. Overall, the results are in accordance with the Excel file and Statistics Denmark (1991).

Yearly hours of full-timers, T. There are two questions of interest. The first is what yearly full-time hours were used by Statistics Denmark, and the second is how the agency arrived at the numbers. The latter question is important if one were to think about improvements. We only have two sources, the Excel-file and Statistics Denmark (1991), and the latter source has information on the years 1980-1987 only. The starting point is Table 3 comparing the two. The idea is to use the years in which they agree to figure out the method, and then try to find reasons for divergence in any years with deviations.

Table 3. Yearly Full-Time Hours According to the Excel File and Statistics Denmark (1991)

| Year | Excel file | Statistics Denmark (1991) |
|------|------------|---------------------------|
| 1980 | 1836 | 1840 |
| 1981 | 1816 | 1816 |
| 1982 | 1832 | 1832 |
| 1983 | 1824 | 1824 |
| 1984 | 1816 | 1816 |
| 1985 | 1808 | 1808 |
| 1986 | 1808 | 1804 |
| 1987 | 1771 | 1764 |

As is seen, the two sources present conflicting information. They agree in the years 1981-1985, but disagree in 1980, 1986, and 1987. To uncover the method, we first look at the years 1981-1985. The yearly hours of full-timers can be reproduced by experimentation with some help from the number of work days in each month found in the Excel-file. The outcome is,

$$T = \tilde{h}_{wk} \left(\frac{(D - 6 - H)}{5} - V \right),$$

where D is the number of week days in each year, 6 is the number of legal holidays that always fall on a week day (Skærtorsdag, Langfredag, 2. Påskedag, Store Bededag, Kristelig Himmelfart, 2. Pinsedag), H is the number of legal holidays that sometimes but not always fall on weekdays (Nytårsdag, 1. maj (half a day until and including 2001), Grundlovsdag (half a day until and including 2005), 1. juledag, 2. juledag), and \tilde{h}_{wk} is the weekly hours used in the calculation of T . They might potentially be different from h_{wk} , hence the \sim . For 1981-1985, $\tilde{h}_{wk} = 40$ replicates T as Table 4 shows.

Table 4. Yearly Full-Time Hours According to the Excel file, Statistics Denmark (1991), And Our Replication

| Year | Excel-file | Statistics Denmark (1991) | Replication |
|------|------------|---------------------------|-------------|
| 1981 | 1816 | 1816 | 1816 |
| 1982 | 1832 | 1832 | 1832 |
| 1983 | 1824 | 1824 | 1824 |
| 1984 | 1816 | 1816 | 1816 |
| 1985 | 1808 | 1808 | 1808 |

Knowing the method, we can then look into where the deviations in the years 1980, 1986, and 1987 come from. In 1986-1987, an obvious candidate is the weekly hours since there are six different ways to pick the weekly hours in these years, cf. the discussion above. To get an idea, we look into which of the sources is the most reliable. The advantage of Statistics Denmark (1991) is that it was written in 1991, only one year after the period 1986-1990 in which the weekly hours fell gradually. In contrast, the Excel-file is a reconstruction. Hence, Statistics Denmark (1991) should be the more reliable source for what actually happened in the period 1986-1990. There is one key claim in it. Footnote six mentions that Statistics Denmark obtained the number of yearly full-time hours from the Wage and Salary Statistics of Dansk Arbejdsgiverforening, the association of employers in Denmark. This means that Statistics Denmark themselves never had to calculate the yearly number of hours of full-timers. As a result, in the years 1986-1990 with ambiguous weekly hours, we should expect no connection between the weekly hours used by Statistics

Denmark in their hourly wage calculations, and the weekly hours that were used to compute the yearly full-time hours. There should also not be any correspondence with the weekly hours used to reconstruct the yearly full-time hours as they are found in the Excel-file. Consequentially, it should not be a surprise that the two estimates of the yearly full-time hours differ in 1986 and 1987 and are equal in 1981-1985. For these reasons, we check if any of the six methods to choose the weekly hours in 1986 and 1987 can replicate the yearly full-time hours of the Excel file and Statistics Denmark (1991). It turns out that the un-rounded weighted average and January hours can.

Table 5. Differences in Yearly and Weekly Full-Time Hours

| Year | Method 1 | Method 4 | Statistics Denmark (1991) | Excel file |
|------|----------|----------|---------------------------|------------|
| 1986 | 1804 | 1808 | 1804 | 1808 |
| 1987 | 1763 | 1771 | 1764 | 1771 |

Table 5 shows that the differences in the yearly full-time hours between Statistics Denmark (1991) and the Excel file in 1986 and 1987 indeed are due to differing weekly hours. Statistics Denmark (1991) reports yearly full-time hours that are obtained from an un-rounded weighted average, whereas the numbers in the Excel file are based on January hours. The difference of one between using the weighted average and Statistics Denmark (1991) is explainable by the fact that for some worker groups, the fall from 39 to 38.5 hours took place one month later (Online, 6). Indeed, if 15th of September is used as the average threshold, one arrives at 1764 hours in 1987.

We then turn to examining 1980 where there is no doubt about the weekly hours equalling 40. The results are contained in Table 6.

Table 6. Yearly Full-Time Hours According to the Excel File, Statistics Denmark (1991), And Our Replication

| Year | Excel file | Statistics Denmark (1991) | Replication |
|------|------------|---------------------------|-------------|
| 1980 | 1836 | 1840 | 1836 |

We are not sure what causes the numbers of Statistics Denmark (1991) to be off in 1980 by four hours. It could have to do with mistakenly not counting 1. maj or Grundlovsdag as half a day off in 1980. Given that the weekly hours are unambiguously equal to 40 in that year, the discrepancy can only arise from counting the work days incorrectly. Whether Statistics Denmark actually used 1836 or 1840 as reported we cannot say because we do not have access to the raw registers prior to 1984 so we cannot test it. Determining the exact number is left for a future project or for the interested reader.

Now, since our main source for the years 1988-1990 is the incorrect Excel file, we instead test in the data the same six methods for choosing the weekly hours that were tried when identifying h in 1986-1990. We also use the same sample but this time we compare predicted hours to actual hours, and use the Mean Squared Error criterion. The result appears in Table 7.

Table 7. Weekly Hours Consistent With the Yearly Full-Timer Hours Actually Used

| Year | Method 1 | Method 2 | Method 3 | Method 4 | Method 5 | Method 6 |
|------|----------|----------|----------|----------|----------|----------|
| 1986 | 49.3 | 49.3 | 49.3 | 49.3 | 49.3 | 1977.7 |
| 1987 | 37.0 | 37.0 | 323.2 | 323.2 | 292.6 | 292.6 |
| 1988 | 31.9 | 31.9 | 57.0 | 151.9 | 57.0 | 57.0 |
| 1989 | 51.4 | 51.4 | 74.7 | 74.7 | 426.1 | 426.1 |
| 1990 | 24.7 | 24.7 | 311.8 | 24.7 | 311.8 | 311.8 |

Two things stand out in the table. First, the criterion values are actually quite low given that the total hours are all above 1000 and any deviations are squared. Second, the result of the comparison of the methods is clear. Method 1 and Method 2 are identical and yield the lowest total mean squared error. In other words, Statistics Denmark used the yearly full-time hours that were computed from the weighted average of the weekly hours. A separate exercise not reported here shows that there is no difference between using 1764 and 1763 in 1987 since a comparison yields the exact same mean squared error. Thus it is impossible to determine if Statistics Denmark

actually used 1763 or 1764 but it also does not matter. Given that the Excel file is wrong and Statistics Denmark (1991) reports 1764, that is then the most likely number.

From 1991 and on, there is only the Excel-file and our replication of it, and they perfectly agree. The yearly full-time hours of all years are listed in Table 8.

The yearly labour market pension rate, A. The Excel file from Statistics Denmark directly tells us this amount, and a verification of historical yearly pension rates confirms the numbers in the Excel-file (except that Statistics Denmark rounds down). Notice that the B-rate is used to calculate the hourly wages. If a worker earns another rate, it is converted to the B-rate.

Table 8. Overview of Historical Values of Parameters used by Statistics Denmark, 1980-2007

| Year | k_{wk} | V | h_{wk} | T | A | Year | k_{wk} | V | h_{wk} | T | A |
|------|----------|-----|----------|-----------|------|------|----------|-----|----------|------|------|
| 1980 | 10 | 4.5 | 40 | 1836/1840 | 432 | 1994 | 10 | 5 | 37 | 1687 | 1166 |
| 1981 | 10 | 5 | 40 | 1816 | 432 | 1995 | 10 | 5 | 37 | 1676 | 1166 |
| 1982 | 10 | 5 | 40 | 1832 | 1166 | 1996 | 10 | 5 | 37 | 1680 | 1166 |
| 1983 | 10 | 5 | 40 | 1824 | 1166 | 1997 | 10 | 5 | 37 | 1672 | 1166 |
| 1984 | 10 | 5 | 40 | 1816 | 1166 | 1998 | 10 | 5 | 37 | 1680 | 1166 |
| 1985 | 10 | 5 | 40 | 1808 | 1166 | 1999 | 10 | 5 | 37 | 1695 | 1166 |
| 1986 | 10 | 5 | 40 | 1804 | 1166 | 2000 | 10 | 5 | 37 | 1672 | 1166 |
| 1987 | 10 | 5 | 39 | 1764 | 1166 | 2001 | 10 | 5 | 37 | 1665 | 1166 |
| 1988 | 10 | 5 | 38.5 | 1748 | 1166 | 2002 | 10 | 5 | 37 | 1676 | 1166 |
| 1989 | 10 | 5 | 38 | 1710 | 1166 | 2003 | 10 | 5 | 37 | 1676 | 1166 |
| 1990 | 10 | 5 | 37.5 | 1687 | 1166 | 2004 | 10 | 5 | 37 | 1702 | 1166 |
| 1991 | 10 | 5 | 37 | 1672 | 1166 | 2005 | 10 | 5 | 37 | 1687 | 1166 |
| 1992 | 10 | 5 | 37 | 1687 | 1166 | 2006 | 10 | 5 | 37 | 1680 | 1166 |
| 1993 | 10 | 5 | 37 | 1695 | 1166 | 2007 | 10 | 5 | 37 | 1672 | 1166 |

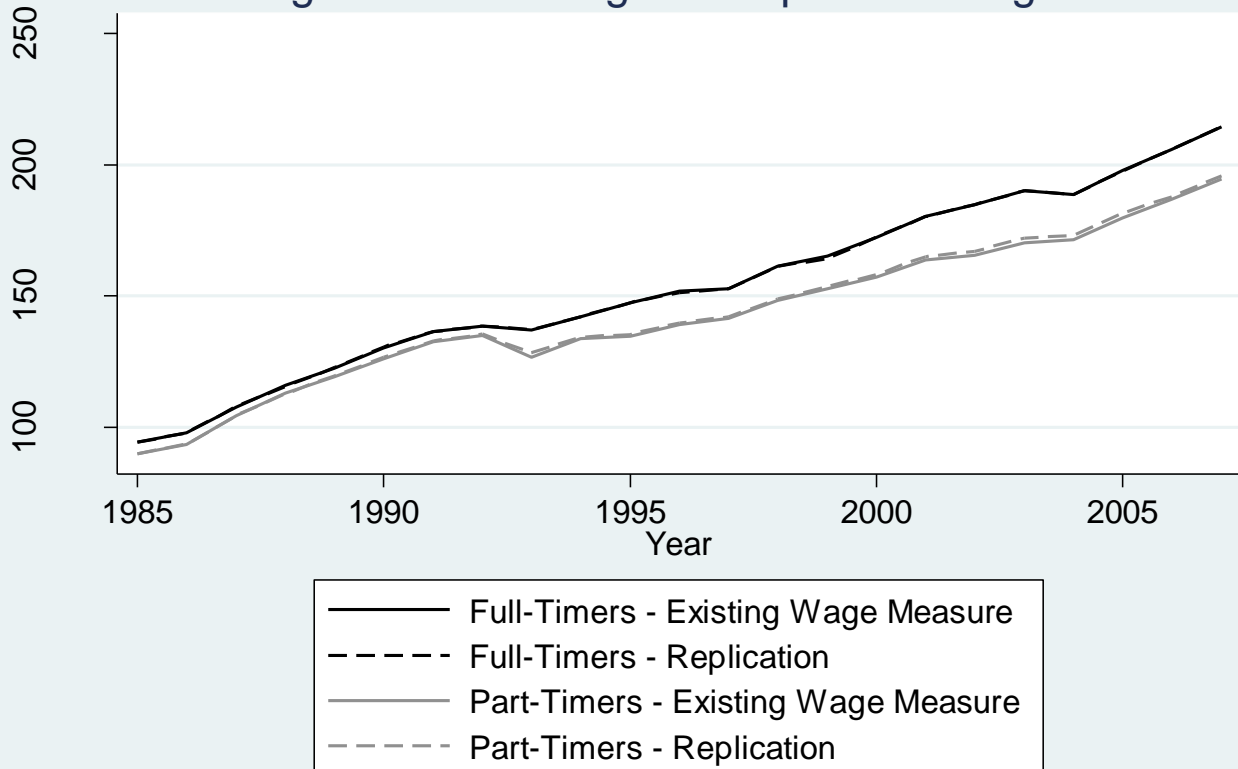
4 Four Cases of Mediocre Performance of the Hourly Wage

Having established the parametric form of the estimation method of Statistics Denmark and the historical parameter values, we now turn to documenting four disconcerting behaviours of the hourly wage measure. They will illustrate some deeper flaws in the way the hourly wage is estimated. We pick the most comprehensive population possible from the universe of all Danish

establishments and workers with a job in the last week of November. In principle, we are interested in how the estimation method of the hourly wage performs for all the workers it is applied to. Since it is applied to everyone, we do not exclude any groups of workers from our analysis. For example, we include all workers in the public sector and all other sectors, full-timers and part-timers of all hours, and workers of all ages between 15 and 74 that are the age limits in the IDA database. The exception is that we do not include second jobs (type B-jobs) in our analysis. In principle we should, but there are some data access issues and technical complications which make an expanded analysis infeasible at this point. By only focusing on the main jobs, we lose 6.1 million observations or 9.8%. The time frame is the largest possible, 1985-2007. We do not have access to the registers outside of this period. Finally, we exclude workers with zero pensions, periods of employment that are reported as zero or have been zero due to sickness or unemployment, and observations that are similar to those once rounding has been accounted for. We lose 6.8 million observations with these exclusions, or 12.0%. In total, we have 4.1 million unique workers, 0.51 million unique establishments, and 49.7 million pooled observations.

With the sample, we replicate the IDA wage measure. Throughout this and subsequent chapters we use our replication as the benchmark and not the original version since we will need to modify the wage measure along several dimensions. The modifications can best be done to the replication so for consistency reasons we prefer the replication as the benchmark. This approach should not compromise our results as the replication performs very well, cf. Figure 1.2.

Figure 1.2: Existing and Replicated Wages



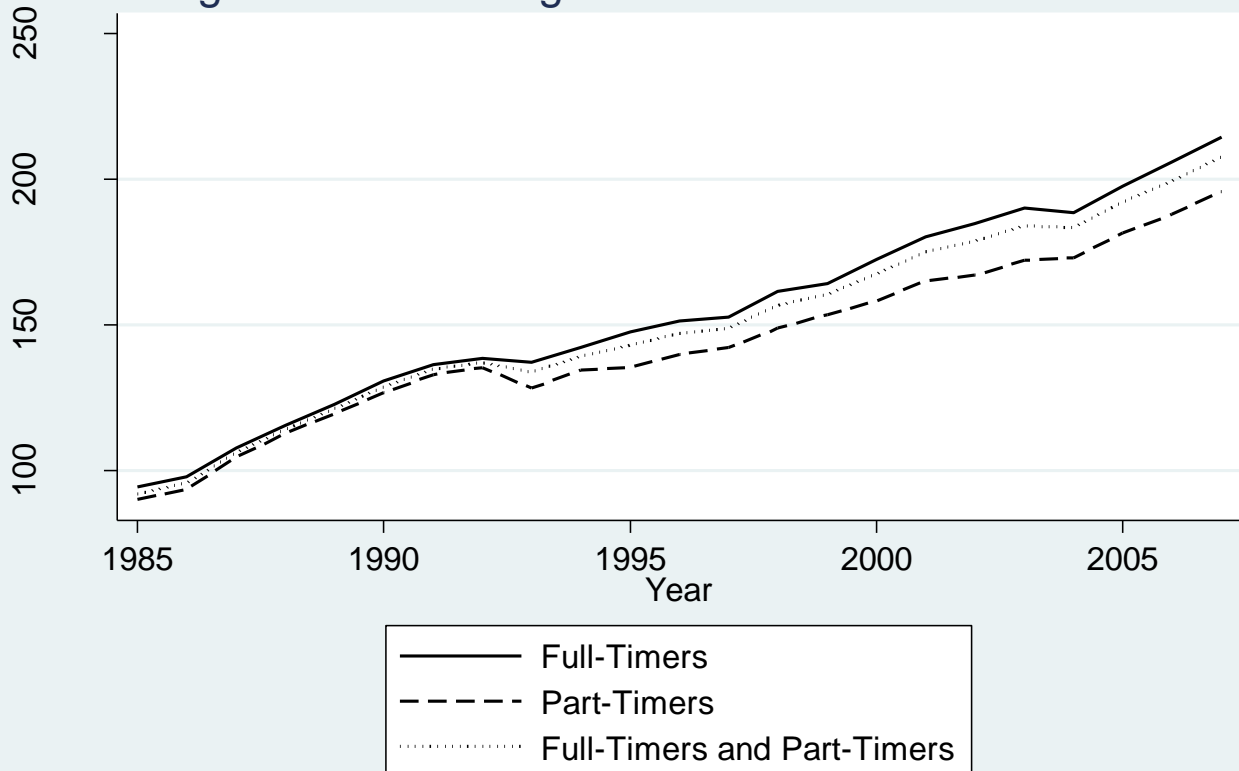
Source: Authors' calculations of wages in the Danish IDA database.

Indeed, Figure 1.2 shows that the wages of both full-timers and part-timers are close to perfectly replicated. There are very minor differences among the part-timers in the 2000s, but otherwise the graphs are completely on top of each other. For completeness, in the Appendix we investigate our replication in more detail and find that the (small and negligible) differences are to a large extent caused by our use of an alternative measurement of illness.

Puzzles 1 through 3 – Time Series and Level Issues

The first case in which the IDA hourly wage performs poorly is from 1987 to 1993. Figure 1.3 shows the time series of the average (replicated) wages of part-timers and full-timers and both groups together.

Figure 1.3: IDA Wages of Full-Timers & Part-Timers



Source: Authors' calculations of wages in the Danish IDA database.

The wages of the part-timers and full-timers track each other in an absolute way from 1985 to 1992. But then the wages of the full-timers stagnate whereas the wages of the part-timers take a big hit. From 1993, the wages of the two groups track each other nicely again, with roughly the same growth rates in every year. But the wages of the part-timers never make up the lost ground. This is unusual in the Danish labour market where the different groups of workers observe each others' wages and differences often are smoothed out when the collective bargaining takes place. It thus seems that there is a fundamental problem with how the wages are estimated for the part-timers around 1993. To get a sense of the magnitudes we show the average wages from 1985-1993 in Table 9

Table 9. Average Hourly Wages of Part-Timers and Full-Timers, 1985-1993

| Year | Full-Timers, in DKK | Part-Timers, in DKK | Relative Difference, in % |
|------|---------------------|---------------------|---------------------------|
| 1985 | 94.3 | 90.1 | -4.5 |
| 1986 | 98.0 | 93.7 | -4.4 |
| 1987 | 107.9 | 104.5 | -3.1 |
| 1988 | 115.5 | 112.8 | -2.4 |
| 1989 | 122.8 | 119.7 | -2.5 |
| 1990 | 130.7 | 126.7 | -3.1 |
| 1991 | 136.5 | 133.0 | -2.5 |
| 1992 | 138.6 | 135.4 | -2.3 |
| 1993 | 137.2 | 128.3 | -6.5 |

Note: The average Relative Difference from 1994-2007 was -8.0%.

Table 9 shows that the part-timers earned very high wages in the years prior to 1993. For example, in 1992 on average they earned almost the same wage as the full-timers, only 2.3% less. In general, there was a declining relative difference from -4.5% in 1985 to -2.3% in 1992. That changed in 1993 where the difference suddenly jumped to -6.5%, the result of a big fall in the wages of the part-timers from 1992 to 1993 of 5.2%. The relative difference in 1993 seems closer to normal if judged by the average relative difference from 1994-2007 of -8.0%. Thus, we have the following Puzzle.

Puzzle 1, the Time Series of Part-timers. From 1985-1992, the wages of the part-timers began high relative to the wages of the full-timers and rose further until they ended up suspiciously close with only a 2.3% difference. Then, from 1992 to 1993 the wages of the part-timers fell by 5.2% whereas the wages of the full-timers stagnated. The relative difference corrected itself and was closer to the historic average thereafter.

Not surprisingly, Figure 1.3 shows that also the average wage of everybody falls from 1992 to 1993. This anomaly in the growth rate of the average wage in 1993 has been pondered before by Statistics Denmark (un-dated memo). However, no explicit distinction between part-timers and full-timers was made, and the phenomenon was mainly cast as lower ranking workers

experiencing falling wages. A brief diagnosis was made based on some descriptive statistics. They attributed the falls to an overhaul of the pension rules that took effect in 1993. However, a full formal analysis of the estimation method and how it responds to changes in parameters was never undertaken, and, as a result, it seems that Statistics Denmark never got to the root of the problem. In Chapter 2, we solve the puzzle with such an analysis and propose some adaptations of the estimation method. Without additional assumptions, the changes make the method more flexible and accommodative to the changes in the labour market that took place in 1986-1993. As a result, the anomalies in the time series completely disappear.

The second time-span that looks suspicious is from 2003 to 2004 when both the aggregate wage and the wage of the full-timers fall while the wage of the part-timers rises. If that happened in reality, it would be quite unusual. Statistics Denmark has noticed the fall in the aggregate wage in Online (1) but no explanation is given there. To assess whether the wages fell in reality, we use a survey of hourly wages conducted from 1997 by Statistics Denmark called “Lønstatistik”. The survey contains a number of pay and time components that can be assembled to different kinds of hourly wage concepts. Given that this study is the first thorough documentation of the IDA wage, we should be in a very good position to pick the right components and assemble them so that the concept is as close as possible to that of the IDA wage. We define what we call the “Survey wage” as

$$wage_{survey} = \frac{sftj - pens}{timprae + timfra - timover}$$

where *sftj* is “samlet fortjeneste” i.e. a gross earnings measure, “*pens*” is pensions, “*timprae*” is hours worked, “*timfra*” is hours of certain kinds of absence, and “*timover*” is over-time. While there is no guarantee that the components perfectly correspond to those that enter the estimation of the IDA wage, we still deem the correspondence quite good (see below).

Now, there are a number of sample issues that we must deal with since the samples in the IDA database and the Lønstatistik are far from identical. Firstly, the Lønstatistik is not a representative sample of the labour market, at least not in the form that we have access to. Indeed, firms run by the government and large firms are over-represented and only 50% of all jobs are included. We use the skewed sample anyway. Secondly, since we only work with November jobs in the IDA database, we select the jobs in the survey that seem to be filled in November. However, it often happens that there are multiple such jobs for each worker in the survey despite the fact that the comprehensive IDA database only contains one job. In that case we simply choose the job from the survey with the highest wage since we deem that job the most likely to exist in reality. Thirdly, there are large outliers in the survey that we trim away by discarding 0.5% of both tails. Fourthly, the firm identifier *cvnrnr* is sometimes missing or incorrect, compromising the sample match with the IDA database. The lack of correct firm identifiers is especially a problem in the initial years and accordingly our matched sample will improve over the years. We discard the year 1997 for this reason. Finally we match the two samples. Since the idea is to strictly compare the IDA and Survey wages for each worker and not compare the different samples, we discard all observations for which there is no counterpart in the other dataset.

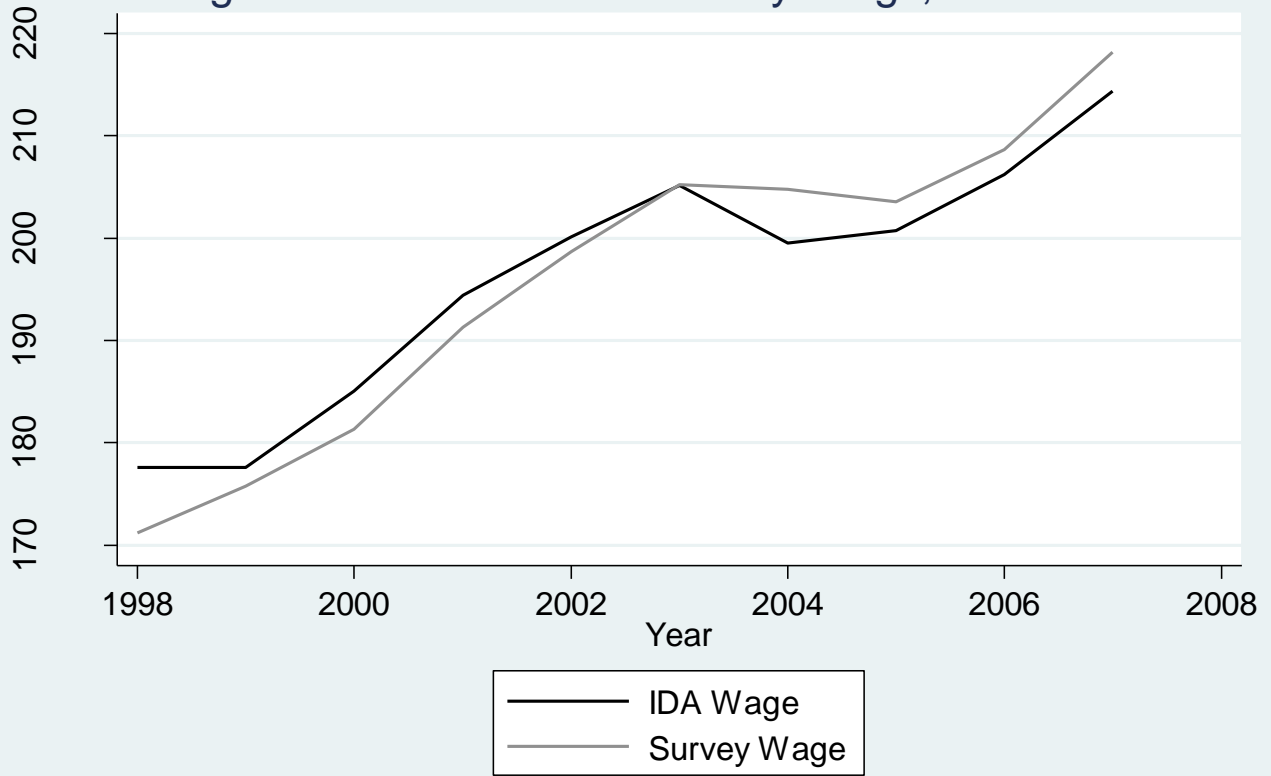
Our constructed benchmark appears in Figures 1.4a-c for the years 1998-2007. Notice how the wage levels have risen compared to Figure 1.3 due to the use of the matched sample, and how the numerical difference shrinks with time in Figure 1.4a due to the improving sample. The average difference between the aggregate wages is 6.7 DKK or 3.6% when weighting each year equally, whereas it is 6.0 DKK or 3.2% when weighting by the number of observations in each year which is larger in recent years.

Figure 1.4a: IDA Versus Survey Wage, All Workers

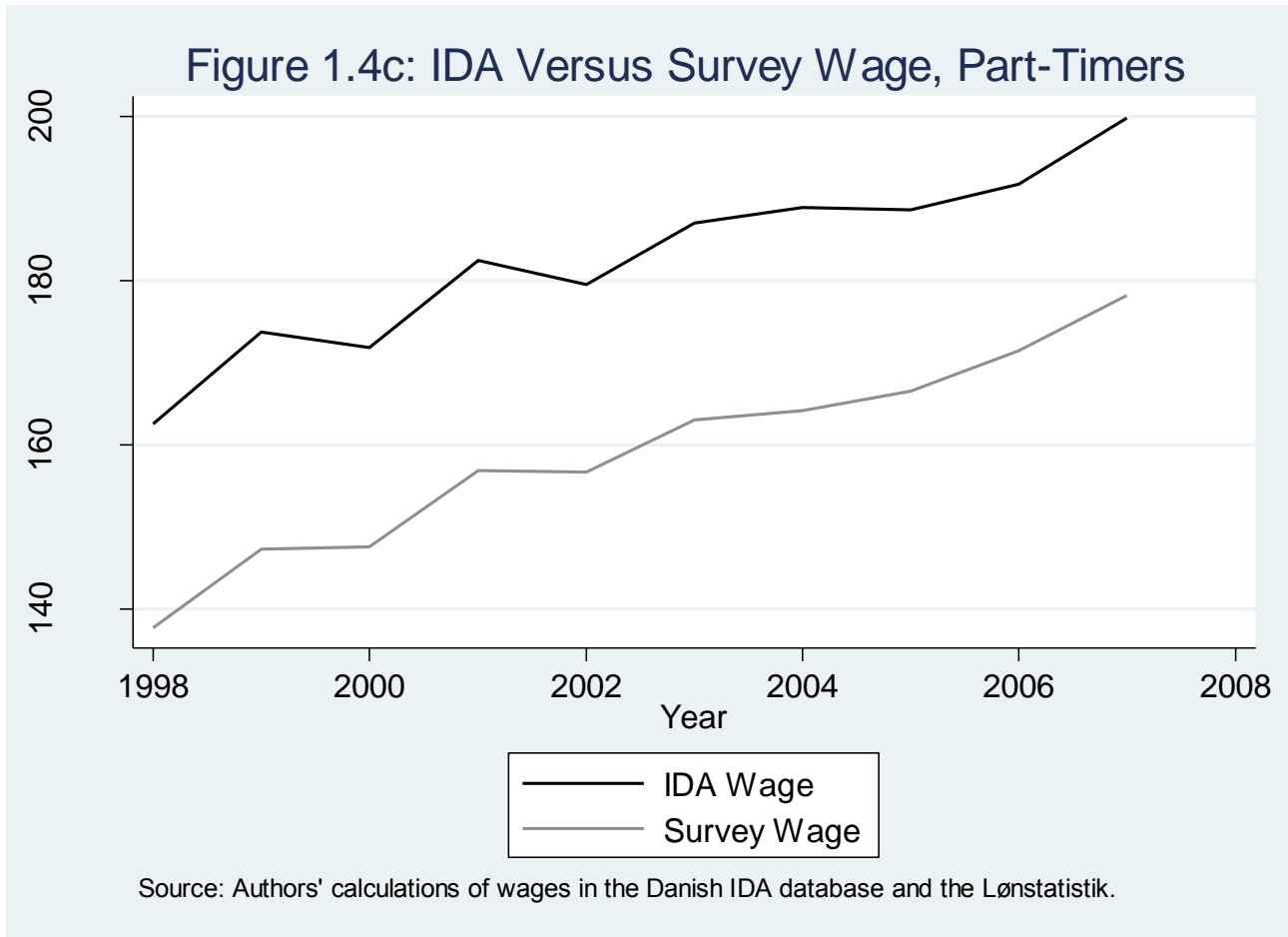


Source: Authors' calculations of wages in the Danish IDA database and the Lønstatistik.

Figure 1.4b: IDA Versus Survey Wage, Full-Timers



Source: Authors' calculations of wages in the Danish IDA database and the Lønstatistik.



There are three main takeaways from Figures 1.4a-c. Firstly, overall the full-timer wages are very close to each other. The fact that the difference is so small for the group of workers with the most precisely estimated wages in IDA (the full-timers) gives credence to the Survey wage as a useful and reliable benchmark despite its problems. Secondly, from 2003 to 2004 the Survey wages stay roughly constant, especially the full-timer and aggregate Survey wages, while the estimated full-timer wage falls significantly and the estimated part-timer wage tracks its survey counterpart. We thus have

Puzzle 2, the Time Series of Full-timers. The fall in the aggregate IDA wage from 2003 to 2004 is very likely an artefact of the estimation method which does not reflect what happens in practice in

the labour market. It can be traced to an unlikely fall in the wages of the full-timers. We advance on this puzzle in Chapter 4 but do not solve it completely.

Thirdly, there is a very big difference between the surveyed and estimated wages of the part-timers, of magnitude 20 DKK. Thus it appears that the part-timer wages are very poorly estimated in the IDA database. It is indeed unlikely that all of the difference is due to a skewed sample that systematically distorts the comparison in one direction for the part-timers but not at all for the full-timers. It is also unlikely that a different wage concept can explain the difference since the same comparison for the full-timers yielded very small differences. Thus,

Puzzle 3, the Level of Estimated IDA Wages of Part-timers. The estimated IDA part-timer wages seem much too high. In Chapter 3, we will investigate further the sources of the numerical difference and, as it will turn out, our solutions to Puzzle 4 below partly solves this Puzzle 3 as well.

In order to know more about the limitations of our Survey wage as a benchmark, whether other authors have performed similar comparisons, and whether other more reliable benchmarks exist, we have searched for other studies that have successfully performed a comparison. We found one such study, DØRS (2003). Unfortunately the two are not directly comparable. While DØRS (2003) uses the same components from the survey and manipulates them in the same way as we do to arrive at the same wage concept, that study deals with the sample issues differently. Firstly, for some unknown reason if a worker has had more than one job during the year, DØRS (2003) averages the worker's wages across the jobs within the year. This seems inappropriate for the comparison that is undertaken as the IDA wages are estimated for each job. Worse, it seems that the averaging is performed even if the jobs are held across different firms. Secondly, the IDA sample used for the matching is bound to be different from ours but exactly on

what dimensions is hard to say since the sample of DØRS (2003) is never described. Thirdly, the output statistics differ in the sense that DØRS (2003) only reports for 1999 the difference between the Survey wage and IDA wage when the part-timers are included and excluded. Judging from these statistics, it seems that the comparison in DØRS (2003) cannot be very accurate. Indeed, they obtain a wage difference that increases when the part-timers are excluded, implying that the IDA wage is better estimated for the part-timers, in stark contrast to our findings. Given that Statistics Denmark puts priority on getting the full-timer wages right as argued earlier in this chapter, and given the higher uncertainty inherent in the estimation of the hours of the part-timers, it seems a quite implausible result. Thus, we are led to believe that ours is currently the most accurate wage benchmark.

Puzzle 4 – Composition Issues

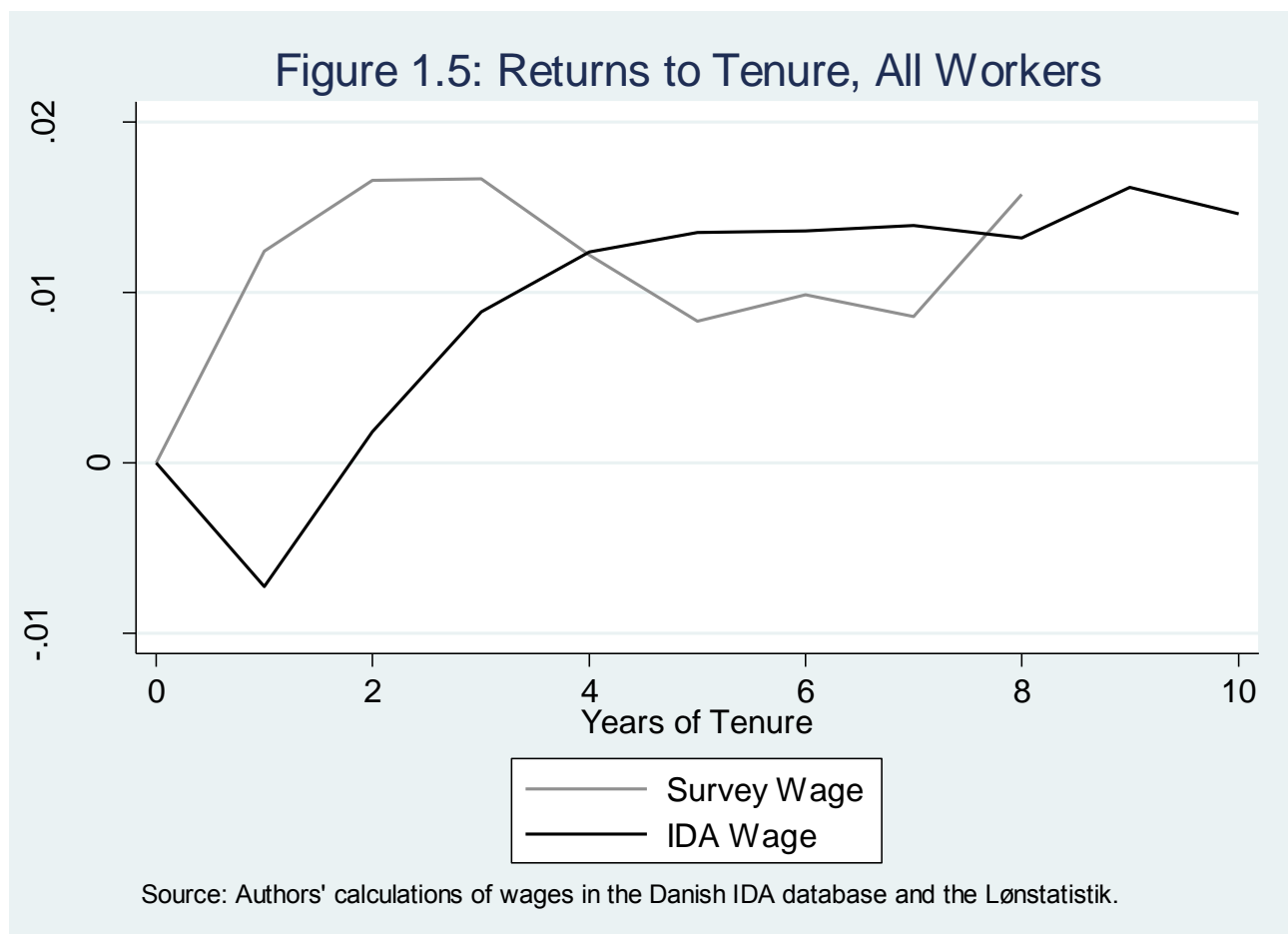
We now turn to the fourth example of discomforting behaviour of the hourly wage measure in the IDA database. Usually wages increase with the duration of a job spell, and we check if that is the case with the IDA wage. In order to not let Puzzle 1 interfere, we restrict ourselves to the time period 1994-2007. (We still include 2003-2004 since that problem appears to be minor.) We first look at all workers together and run the following regression.

$$\log w_{it} = b_0 \cdot e_{it} + b_1 \cdot \tau_{it} + y_t + \alpha_i + \epsilon_{it}.$$

Here, w_{it} is the hourly wage of individual i , e_{it} is a vector of dummies capturing the overall labor market experience of worker i in year t , τ_{it} is a vector of dummies capturing the tenure of individual i in her job in year t , y_t is a year fixed effect, α_i is a worker fixed effect, and ϵ_{it} is an error term. Note that since we focus only on the main jobs, each worker has only one job in each year. The overall labour market experience is essentially measured as the total pensions accumulated since 1964 converted into full-time years and rounded down. Tenure is measured as

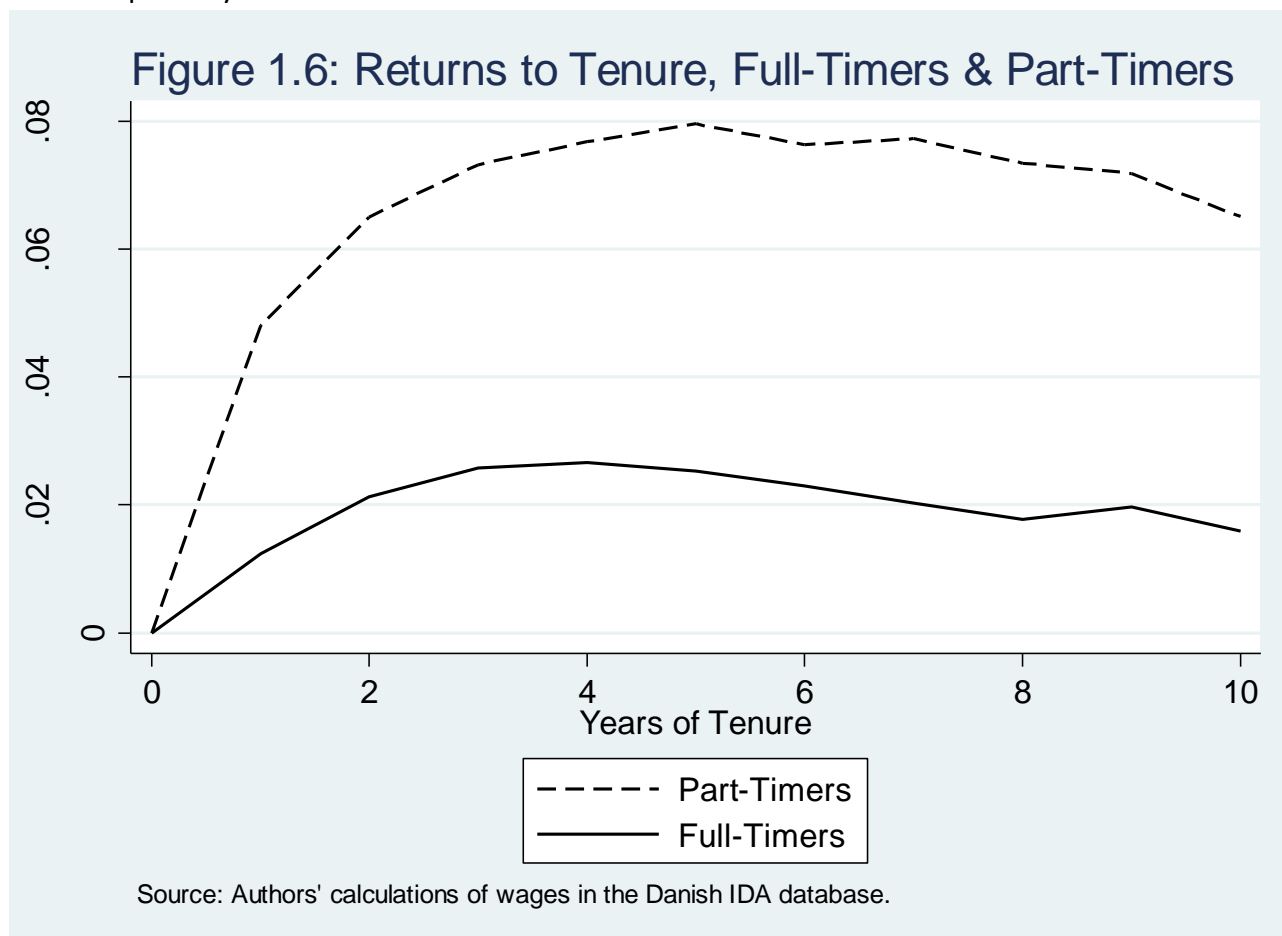
the number of calendar years since hiring, so 0 in the first year and rising by one every year after that.

The object of interest from the regression is the vector of coefficients b_1 that quantifies the relative increases in the wages since the year of hiring. The resulting tenure profile is depicted in Figure 1.5 together with a similar tenure profile for the Survey wage. Note that the samples used for the two tenure profiles in the figure are different. We prefer using our standard sample for the IDA tenure profile (except that we only use the years 1994-2007) and not the matched sample used in Figures 1.4a-c. For the Survey wage we also use the un-matched sample³. Thus, the Survey wage tenure profile should not be seen as the exact true profile, but nonetheless it does indicate the size of the error.



³ We still pick the highest income job, but since there is no matching we now keep also the jobs without a cvnr.

Evidently, there is something wrong with the IDA wage measure. It is not plausible that wages fall by 0.75 percent in the year after hiring. Given the findings in Puzzles 1-3, one could speculate that perhaps the fall had to do with inherent differences among the part-timers and full-timers. We check this hypothesis in the data by estimating similar tenure profiles for the full-timers and part-timers separately.



As Figure 1.6 shows, the wages of both groups behave perfectly naturally in all years. There is no puzzle for the part-timers on their own, and no puzzle for the full-timers on their own, only when they are together. The most likely explanation is then that the fall in the wages observed when all workers are bunched together must come from a composition effect whereby the shares of the

part-timers and the full-timers change systematically from the first year to the second year on the job. We thus have Puzzle 4.

Puzzle 4 – An Abnormal Tenure Profile Due to a Changing Composition of Part-Timers and Full-Timers. The wage measure in the Danish IDA database gives rise to a tenure profile according to which the wages fall in the second year. The effect should stem from a systematic change in the composition of part-timers and full-timers from the first year to the second since the wages of each group separately rise with tenure. Puzzle 4 is solved in Chapter 3.

Puzzle 4 is robust to different definitions of tenure. For instance, we also tried defining tenure as the increase in effective experience (measured by mandatory pension contributions, ATP) over the job spell. That is, instead of defining experience by calendar years, we defined it in terms of years of full-time work. Defining tenure this way accentuates the fall after the first year with a fall of 2.5 percent. Nevertheless, we prefer measuring tenure in calendar years since we show in Chapter 3 that Puzzle 4 stems from how spell lengths are measured in the first and second (calendar) years at a firm.

5 The Estimation Method in Practice

In order to explain the estimation method of hourly wages and the reasoning behind it in as understandable and clear a way as possible, it was necessary to stylize the presentation of the method a bit. This section attempts to bridge the gap between the description of the estimation method contained in the previous sections and how the estimation is carried out in practice. The differences are minor and only pertain to rounding issues but nonetheless the documentation would not be complete without.

Differences Between the Description in this Chapter and What Statistics Denmark Does

First of all, Statistics Denmark measures $h, k, u_{min}, u, u_{max}, p$ in calendar weeks. The presentation in the previous sections made clear that the estimation method did not hinge on using one particular pay cycle but gave identical results for all. The reason for choosing one pay cycle throughout the calculations is simplicity, and the reason for choosing weeks rather than months is that e.g. 5.4 weeks is less vulnerable to rounding than its equivalent 1.2 months. As we will see, Statistics Denmark rounds the numbers and bigger mistakes would be made if longer pay cycles were used. There is no particular motivation for using calendar weeks instead of full weeks. The sole consequence is that the correction to full weeks just happens at a later stage and that the reading and interpretation of the code becomes more difficult.

The first reason for why Statistics Denmark rounds all period lengths is that the condition $u_{min} = u$ in this chapter is not very practical because both quantities are measured in full pay cycles. This means that they are not integers and exact equality is only likely to happen if $u_{min} > \min(p, u_{max})$ where the right-hand side abuses notation slightly as in Section 2. To avoid this problem, Statistics Denmark rounds all period lengths to the closest integer. However, these are not the only instances of rounding that occur in the program code. Other variables are rounded as well due to a historical need for minimizing the data storage. The precise calculations performed by Statistics Denmark are the following. First u_{min} is computed as $aumin$, defined by

$$aumin = \text{round}\left(\frac{ATP}{\frac{A}{52 - V}}\right).$$

V is the number of weeks of vacation, as before. The upper bound on weeks is

$$aumax = \max(aumin, \min(p, u_{max})).$$

The lower bound on hours is computed as

$$\tilde{t}_{min} = \text{round}(hald \cdot ATP),$$

where

$$hald = round\left(100 \frac{T}{h_{wk}A} (h_{wk} - k_{wk})\right) \frac{1}{100}.$$

If *hald* had not been rounded, the only difference between the lower bound on hours computed by Statistics Denmark and t_{min} would have been counting the lower bound in integers. Instead *hald* is rounded to the second decimal, and if *ATP* is large, then the total difference can be several integers. $\alpha_{k,u}$ is computed as

$$\tilde{\alpha}_{k,u} = round(korr \cdot aumax \cdot k_{wk}),$$

where

$$korr = round\left(100 \frac{T}{h_{wk}(52 - V)}\right) \frac{1}{100}.$$

Here there are three differences compared to $\alpha_{k,u}$. The first is that *aumax* is based on *aumin* which due to rounding is computed differently than u_{min} , even if the latter were counted in calendar weeks as *p* and u_{max} and not in full weeks. The second is that the conversion to full weeks is imprecise because *korr* is rounded to the second decimal, and the third is that the final result itself is rounded. In terms of sequencing, it is seen that the conversion of calendar weeks into full weeks is postponed to a later stage than what is done in this chapter. Finally, the relative uncertainty is computed as

$$tlonkval = round\left(\frac{\tilde{t}}{\tilde{t}_{min}}\right) - round\left(\frac{\tilde{t}}{\tilde{t}_{max}}\right).$$

Here, \tilde{t} is Statistics Denmark's estimate of the hours worked.

Appendix C1

For completeness, this Appendix contains more information about the weighted average of weekly hours in 1986-1990 that are used to compute the yearly full-time hours, T . The official thresholds for the hours-brackets and official pension rates have been included as well.

Details on the Calculation of T and Official Values of Pension Rates and Hours-Thresholds

Weekly hours \tilde{h}_{wk} used to calculate the yearly full-time hours, T .

- 1986: 230 days before the change from 40 to 39 hours, 21 after, 251 in total. On average, $\tilde{h}_{wk} = 230/251 \cdot 40 + 21/251 \cdot 39 = 39.92$.
- 1987: 165 days with 39 hours pr. week, 87 with 38.5, 252 days in total. $\tilde{h}_{wk} = 38.83$. Alternatively, if the cut-off date is September 15th and not September 1st (Section 3 of this chapter), the weighted average is $\tilde{h}_{wk} = 38.86$.
- 1988: 167 days with 38.5, 86 with 38, 253 days total, $\tilde{h}_{wk} = 38.33$.
- 1989: 167 days with 38, 84 with 37.5, 251 days total, $\tilde{h}_{wk} = 37.83$.
- 1990: 167 days with 37.5, 84 with 37, 251 days total, $\tilde{h}_{wk} = 37.33$.

In Table A.1 I present an overview of the historical values of parameters related to the determination of the yearly full-time hours and the hourly wage. Although the method to estimate hours does not use information on the thresholds of the hours-brackets, nor the official historical pension rates that apply to different pay cycles, they are nonetheless important to keep in mind. Later chapters will show how the estimation method ought to be augmented to take into consideration that the hours-thresholds change over time, and Section 2 of this chapter mentioned how changing ratios between weekly, monthly, and yearly pension rates could bias the wages of some groups of workers.

Table A.1 Parameters Related to the Calculation of T and the Hourly Wage

| Year | \tilde{h}_{wk} | $\tau_{1,w}, \tau_{2,w}, \tau_{3,w}$ | atp_w | atp_{2w} | atp_m | atp_{yr} |
|------|------------------|--------------------------------------|---------|------------|---------|------------|
| 1980 | 40 | 10,20,30 | 9.00 | 18.00 | 36.00 | 432.00 |
| 1981 | 40 | 10,20,30 | 9.00 | 18.00 | 36.00 | 432.00 |
| 1982 | 40 | 10,20,30 | 24.30 | 48.60 | 97.20 | 1166.40 |
| 1983 | 40 | 10,20,30 | 24.30 | 48.60 | 97.20 | 1166.40 |
| 1984 | 40 | 10,20,30 | 24.30 | 48.60 | 97.20 | 1166.40 |
| 1985 | 40 | 10,20,30 | 24.30 | 48.60 | 97.20 | 1166.40 |
| 1986 | 39.92 | 10,20,30 | 24.30 | 48.60 | 97.20 | 1166.40 |
| 1987 | 38.83* | 10,20,30 | 24.30 | 48.60 | 97.20 | 1166.40 |
| 1988 | 38.33* | 10,20,30 | 24.30 | 48.60 | 97.20 | 1166.40 |
| 1989 | 37.83* | 10,20,30 | 24.30 | 48.60 | 97.20 | 1166.40 |
| 1990 | 37.33* | 10,20,30 | 24.30 | 48.60 | 97.20 | 1166.40 |
| 1991 | 37 | 10,20,30 | 24.30 | 48.60 | 97.20 | 1166.40 |
| 1992 | 37 | 10,20,30 | 24.30 | 48.60 | 97.20 | 1166.40 |
| 1993 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 1994 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 1995 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 1996 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 1997 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 1998 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 1999 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 2000 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 2001 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 2002 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 2003 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 2004 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 2005 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 2006 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |
| 2007 | 37 | 9,18,27 | 25.65 | 51.30 | 97.20 | 1166.40 |

Note: * Indicates our best estimate as argued in Section 3. atp_{wk} , atp_{2w} , atp_m denote the actual periodical rates. atp_{yr} denotes the official yearly pension rate. The yearly rate used by Statistics Denmark is $A = round(atp_{yr})$. No official yearly pension rate existed in 1964-1992. For these years, it has been officially calculated ex-post by multiplying the monthly rate by 12.

One important observation from Table A.1 is that until 1992, the ratio between the weekly and monthly pension rates was always 4, whereas after 1993 it was adjusted to 3.79. This will be important in Chapter 2.

The Replication of the Hourly Wage in the IDA Database

In Figure 1.2 of this chapter we found that we could replicate the hourly wage in the IDA database very well. However, there were some small differences especially in the early 2000s. Here we look into the causes of the differences and to what extent they can be remedied. To begin, we note that according to the old SAS programs there can only be one candidate explanation for the differences. Indeed, following the results of this chapter we should be using the right parameter values and our estimation code should exactly copy that of Statistics Denmark, except for one difference: We measure illness periods using the variable “varmms” whereas Statistics Denmark uses the variable “adagp”. In principle they should be identical but they are not in practice. It has been necessary for us to use varmms even if we do have access to adagp, simply because there is a parameter associated with adagp whose historical values are unknown. Without the parameter values the adagp variable is useless. Luckily, from our conversations with former employees of Statistics Denmark we have recovered the parameter values for the years 1985-1991. We use those values here and replicate the hourly wage using both varmms and adagp and compare the results. The latter replication is close to perfect and as good as it can possibly get.

Table A.2 Replications With and Without Adagp

| Wage Measure | Mean | Standard deviation | Minimum | Maximum |
|--------------------------|-------------|---------------------------|----------------|----------------|
| Existing | 113.025 | 49.391 | 1 | 7657 |
| Replication using varmms | 113.153 | 49.527 | 1 | 7657 |
| Existing | 113.027 | 49.398 | 1 | 7657 |
| Replication using adagp | 113.048 | 49.441 | 1 | 7657 |

Note: The first and third rows differ slightly in columns one and two because the samples differ slightly. Some exclusions are made in order to make the samples similar to the one used in the main parts of this chapter, and the exclusions depend slightly on whether varmms or adagp is used. The samples in the first and second rows and in the third and fourth rows are the same.

Table A.2 shows that both replications work fine, but the one with adagp performs better, being accurate to the first decimal.

We then dig a bit deeper and describe the deviations in Table A.3.

Table A.3 The Two Replications and Their Deviations

| | |
|---|--------|
| Perfectly replicated using varmms | 82.67% |
| Non-zero deviations using varmms | 17.33% |
| - Of which the difference is -1 or 1 | 90.37% |
| Perfectly replicated using adagp | 84.67% |
| Non-zero deviations using adagp | 15.33% |
| - Of which the difference is -1 or 1 | 98.50% |

Note: The table shows the fraction of the observations that fall into each category.

The share of perfectly replicated observations as measured by a zero difference to the existing wage measure rises by two percentage points when adagp is used. In other words, there is a slight improvement in the number of observations that are perfectly replicated. However, more tellingly the deviations themselves shrink so that with adagp, 98.50% of the deviations are only one or minus one. It looks as if some rounding issue is responsible for most of the deviations that already are somewhat fewer in number. The same cannot be said for the replication that uses varmms where almost 10% of the deviations are outside the one or minus one range.

Chapter 2: Solving the Time Series Puzzle of the Part-Timers

1 Introduction

In the Danish labour market, it is abnormal to observe that some groups of workers take drastic wage cuts while others do not. It may happen if a firm is in distress, but not to major segments of the workforce. Chapter 1 documented how the wages of part-timers were suspiciously close to the wages of full-timers from 1985-1992, then drastically dropped by 5.2% from 1992 to 1993 while the wages of full-timers stagnated. From then on, the wages of part-timers tracked the wages of full-timers almost perfectly and the lost ground was never made up.

Statistics Denmark (un-dated memo) gives some clues to what could have happened. A number of changes to the pension rules took place from 1992-1993. The first change concerned the pension rates of casual labourers and workers who were compensated weekly and bi-weekly. In 1992 and in earlier years their annual rates were effectively lower because they did not earn pensions when on vacation. But from 1993 on, their rates were raised in order to compensate. The second change concerned the thresholds between the hours-brackets. From 1964-1992 the weekly ones were 10, 20, and 30 hours per week. Then they were lowered to 9, 18, and 27 hours per week in 1993, reflecting the gradual decline in working hours from 1986 to 1990. While the changes were found to cause a drop in the aggregate wage from 1992 to 1993, the conclusions of Statistics Denmark were based on descriptive statistics without any formal analysis to underpin the hypotheses.

In this chapter, I go to the root of the problem and investigate what factors could potentially cause biases in the wages. Since the hours of part-timers are computed in the same

way in all years, any strange patterns over time must pertain to the bounds on hours and not to how the bounds are used in the estimation. I first derive what all the expressions of the bounds from Chapter 1 should have been in order to truly capture what they were supposed to. I then use the true expressions as a benchmark to which I compare the actual expressions. In doing so, I take into account all the changes to the environment that happened in the period from 1985 to 1993, not only the overhaul of the pension rules that took effect in 1993 and was mentioned by Statistics Denmark. As it turns out, the changes in the pension rates of the casual labourers and the weekly and bi-weekly compensated workers affect hours in a way that is completely consistent with Puzzle 1⁴. Unfortunately, it is not possible to distinguish between pay cycles in the data and for this reason there is no way for me to quantitatively assess the impact of the change in rates. But judging from conversations with former employees at Statistics Denmark, most workers at the time were compensated on a monthly basis so it is hard to see how the rate changes could significantly affect the wages.

After considering the change in pension rates I turn to the change in the thresholds of the hours-brackets and the decline in full-time hours. I judge that they are much more important for wages since they affect everyone as opposed to a somewhat narrow group of workers. Interestingly, the effect of the change in the thresholds is to increase the estimated hours from 1992 to 1993, but only because the lower bound is incorrectly computed. Instead of being based on the thresholds between the hours-brackets, it is based on the prevailing full-time hours minus a constant. When the full-time hours fell gradually from 1986 to 1990, the lower bound on hours incorrectly decreased with it. As a result, hours were *too low* prior to 1992, and the change in the thresholds from 1992 to 1993 simply restored the lower bound to what it should have been. If the lower bound had been based on the thresholds instead, it would have been correct from 1986 to 1992 and not much would have happened on average from 1992 to 1993.

⁴ See Chapter 1 for a description of all the puzzles. Puzzle 1 is the issue with the time-series of the part-timers that we deal with in this chapter.

The fact that the lower bound on hours was too low from 1986-1992 partly explains why the wages of part-timers were so close to those of full-timers, but not after 1993. The reason is that the full-timers were not affected by this error in their lower bound because another source of bias in their upper bound exactly cancelled out with it. In effect, the hours that a full-timer can work without earning extra pensions equal the width of the top-bracket but were estimated as the widths of the lower brackets. So the decline in full-time hours from 1986 to 1990 that mistakenly decreased the lower bound should have decreased the upper bound instead, but did not. The widening of the top-bracket and the narrowing of the lower brackets that took place in 1993 erased both errors and both bounds were correctly computed as in 1985. The net result was estimates of hours that equalled the true values throughout. In contrast, the two biases were not equally important for part-timers and the net effect was too low hours in the period 1986-1992 with the biggest shortfalls in 1991-1992.

After deriving the biases in the bounds I proceed to propose improvements that eliminate them. Since I do not have access to data that distinguish between the compensation modes of workers I ignore the less important rate changes in 1993 and focus on the biases stemming from the fall in the full-time hours and the thresholds of the hours-brackets. While the correction of the lower bound is straightforward, the correction of the upper bound relies on a key innovation. Specifically, I show how to compute a lower and upper bound on the number of pay cycles that a worker could have been in the top-bracket during her employment spell. With these bounds, I construct a new upper bound on hours that takes into account the different widths of the top-bracket and lower brackets after 1986. The wage measure that results is a conservative improvement over the IDA wage in the sense that it makes no additional assumptions and offers a guaranteed improvement that is 100% safe to use.

The outline of this chapter is the following. Section 2 diagnoses the biases of the current method of computing the bounds. Section 3 offers a way to eliminate the key biases by

means of a correction to each of the bounds. Section 4 then evaluates the performance of the corrections and whether Puzzle 1 has been resolved. Finally, Section 5 is for readers who wish to use a wage variable that does not suffer from the issues described here but who do not have access to my comprehensive dataset or the wage variables it produces. That section provides an easily implementable procedure that relies on very standard variables to modify the IDA wage according to the recommendations of this chapter. Most users should have access to these variables and the procedure should effectively give the same results. The Appendix provides the SAS code of the procedure.

2 Diagnosis of the Puzzle of Part-Timer Wages

The aim in this section is to compare how the lower and upper bounds on hours are computed with how they should have been computed in order to work as intended. I will emphasize the three most important biases but there are several others that one could naturally analyse within the framework. For example, I do not consider casual labourers at all, nor the derivative effects of potential biases in u_{min} on u or the classification of workers as full-timers and part-timers. The starting point is Section 2 of Chapter 1 which describes how the bounds on hours are estimated.

The Lower Bound

Basically, there are two possible sources of error in the estimation of the lower bound. To see this, write

$$t_{min} = ATP \cdot \frac{T}{A} \cdot (h - k) = \frac{ATP}{\frac{T}{h}} (h - k).$$

Since ATP is correctly measured, the only potential sources of error are the annualized effective pension rate in the denominator on the right-hand side, and the factor $(h - k)$. Regarding the first source, Section 2 of Chapter 1 mentioned that the ratio between the actual weekly and monthly

pension rates changed in 1993. From this year, the former group was intended to earn the same annual pension amount as the latter group and their weekly rates went up to compensate for the fact that they did not earn pensions while on vacation. In contrast, the ratio between the effective weekly and monthly pension rates is always the same. Intuitively, the effective rates cannot be correct for all compensation modes both before and after 1993. Regarding the second source of error, it does not seem intuitive that the minimum number of hours it takes to earn a given pension amount is computed using the maximum possible hours in a pay cycle, albeit reduced by a constant. It would make more sense if they were computed using the minimum number of hours it takes to earn a given share of the pension rate.

To formally diagnose the problems with the lower bound, I derive what the lower bound ideally should be (the “true lower bound”) and compare with the one used by Statistics Denmark. To do that, after 1993 it is necessary to distinguish between workers who are in the top-bracket in all pay cycles of the year and everybody else. The reason is that after 1993, the yearly pension rate became a benchmark that all workers always in the top-bracket were intended to receive irrespective of how frequently they were paid. Thus, supposedly the accumulated pensions of such workers are adjusted if necessary in the last pay cycle of the year to equal the yearly rate.⁵ This adjustment should not take place for other workers, and as a result the lower bounds on hours are derived differently. First consider the group that is not affected by the introduction of the yearly rate. Let atp denote the pension rate received by the worker each pay cycle, and let n_0, n_l, n_m, n_t be the number of pay cycles in the brackets zero, low, middle, and top. The accumulated ATP is

$$ATP = n_l \cdot \frac{1}{3} atp + n_m \cdot \frac{2}{3} atp + n_t \cdot \frac{3}{3} atp = (n_l + 2n_m + 3n_t) \cdot \frac{atp}{3}.$$

⁵ BKG 1992-09-29 nr 822 paragraph 1. Later promulgations specified that employers were not required to pay more than the yearly amount.

Using this expression, one can calculate the lower bound on hours. To do this, one has to further distinguish between monthly compensated workers on the one hand, and weekly and bi-weekly compensated workers on the other. The reason is that the former group receives pensions even for the hours they are on vacation, but the latter group does not⁶. First consider the group of weekly and bi-weekly compensated workers.

$$t_{min}^{(1)} = n_l \cdot \tau_1 + n_m \cdot \tau_2 + n_t \cdot \tau_3 = (n_l + 2n_m + 3n_t)\tau_1 = \frac{ATP}{atp} 3\tau_1 = \frac{ATP}{atp} \tau_3.$$

The true lower bound on hours equals the effective number of weeks or fortnights it takes to accumulate the earned pensions with the full rate, multiplied by the lowest possible hours consistent with the full rate. Next, perform the same calculation for monthly paid workers. Define Y as the hours of a full-timer in a full year when not subtracting vacation and legal holidays, and let an “m”-subscript denote months. Then,

$$\begin{aligned} t_{min}^{(2)} &= n_l \cdot \tau_{1,m} \frac{T}{Y} + n_m \cdot \tau_{2,m} \frac{T}{Y} + n_t \cdot \tau_{3,m} \frac{T}{Y} = (n_l + 2n_m + 3n_t)\tau_{1,m} \frac{T}{Y} = \frac{ATP}{atp_m} 3\tau_{1,m} \frac{T}{Y} \\ &= \frac{T}{Y} \frac{ATP}{atp_m} \tau_{3,m}. \end{aligned}$$

Here, the first equality follows from the fact that to end up in a given bracket, one needs only to work a share T/Y of the lower bound of the bracket, since the rest is accounted for by vacation or legal holidays during which monthly paid workers also accumulate hours. Note the implicit key assumption that workers accumulate hours at the same pace during vacation as during work. A part-timer should only earn hours as if she is on part-time vacation. This might not hold if vacation is held in slack periods, but overall it is probably not a bad assumption. In any case it is the law and therefore the best assumption one can make. The third equality follows from the pension

⁶ For 1993-present, BKG 1992-09-29 nr 822 or later promulgations. For 1980-1992, Statistics Denmark (1991) and Statistics Denmark (un-dated memo). The former implicitly makes this assumption for the period 1980-1991. The latter investigates the impacts of the overhaul to the pension rules but never mentions any changes to how vacation counts. Together, these two sources imply that there were no significant changes to how vacation counts.

accumulation equation applied to a monthly pay cycle. The right-hand side is the effective number of months it would take with the full rate to accumulate the earned pensions, multiplied with the minimum number of hours of work it takes to earn the full rate in each month.

I now proceed to express the lower bound on hours of workers who are in the top-bracket in all pay cycles of the year, as it applies after 1993. It is quite simple. The number of full pay cycles worked is the official number of full pay cycles of the year, T/h . In each full pay cycle, the minimum number of hours to stay in the top-bracket is τ_3 . So

$$t_{min}^{(3)} = \tau_3 \frac{T}{h}.$$

Having derived the lower bounds, I turn to comparing the bounds with the one used by Statistics Denmark. A direct comparison of $t_{min}^{(1)}$ and t_{min} confirms our intuition. The first bias is if the annualized effective rates,

$$\frac{A}{T} \text{ and } \frac{A}{T} = 2 \frac{A}{T},$$

$$\frac{h_{wk}}{h_{wk}} \text{ and } \frac{h_{2w}}{h_{2w}} = 2 \frac{h_{wk}}{h_{wk}},$$

do not equal the actual rates that weekly or bi-weekly compensated workers receive, atp_{wk} and atp_{2w} . The annualized effective rates can be computed from Table 8 in Chapter 1, and the actual rates can be found in the Appendix to Chapter 1. The actual bi-weekly rate is always twice the weekly rate, so it is only necessary to consider the latter. It was 25.65 from 1993 and on, comparable in magnitude to the annualized effective weekly rate that equalled 25.45 in 1993 and varied between 25.34 and 25.91 in future years. The similarity in magnitudes means that any bias from this source should be quite small after 1993. However, this was not the case before 1993 as the following table shows.

Table 1. Actual and Annual Effective Weekly Rates Before 1993

| Year | Actual Weekly Rate | Annual Effective Weekly Rate |
|------|--------------------|------------------------------|
| 1985 | 24.30 | 25.80 |
| 1986 | 24.30 | 25.85 |
| 1987 | 24.30 | 25.79 |
| 1988 | 24.30 | 25.68 |
| 1989 | 24.30 | 25.91 |
| 1990 | 24.30 | 25.92 |
| 1991 | 24.30 | 25.80 |
| 1992 | 24.30 | 25.57 |

The discrepancies illustrated in Table 1 imply biases in all years prior to 1993. In 1992, the lower bound on hours is 5.0% too small for weekly and bi-weekly compensated workers who did not work enough hours to be in the top-bracket in all pay cycles of the year. In 1990 it was 6% too small.

The second source of bias when comparing $t_{min}^{(1)}$ with t_{min} is that $\tau_3 \neq h - k$ in the years 1987-1992 where the weekly numbers were $\tau_{3,w} = 30$ and $h_{wk} - k_{wk} < 30$. In 1991-1992, the lower bound on hours was 10% too small as a result of this bias since $(h_{wk} - k_{wk})/\tau_{3,w}$ equaled $(37 - 10)/30 = 0.9$.

Next I compare $t_{min}^{(2)}$ with t_{min} . Except for minor timing issues, the bias due to the use of the annualized effective rate is not present for monthly compensated workers. Imagine that all months are exactly equally long and equal to $4 \frac{1}{3}$ weeks and to $1/12$ of a year. In that case,

$$t_{min} = ATP \cdot \frac{T}{A} \cdot \frac{h_m}{A} \cdot (h_m - k_m) = \frac{T}{Y} ATP \cdot \frac{h_m}{A} \cdot (h_m - k_m) = \frac{T}{\frac{Y}{12}} \frac{ATP}{A} \cdot (h_m - k_m).$$

Here, $A/12 = atp_m$ is the monthly pension rate. Monthly paid workers in the top-bracket do earn this amount in a month. It follows that the bias from different pension rates does not apply to monthly compensated workers, but the second bias involving the factors $h - k$ and τ_3 does.

Finally, I compare $t_{min}^{(3)}$ to t_{min} . The key observation is that $ATP = A$ so that $t_{min} = T/h \cdot (h - k)$. It follows that the lower bound of the group of workers always in the top-

bracket after 1993 is not sensitive in any way to the relative magnitudes of the actual and annualized effective rates, but a difference between $h - k$ and τ_3 does matter as before. To summarize, I have identified the following two biases to the lower bound on hours.

Bias 1. Weekly and bi-weekly compensated workers have a lower bound on hours that is too low in 1980-1992. The reason is that the annualized effective pension rates used to compute the lower bounds are higher than the actual rates. The difference in 1992 is about 5%. From 1993 there are only minor differences so the resulting biases in the lower bound are negligible from then on.

Bias 2. Due to the fact that $h - k$ declined in the years 1987-1990 as the full-time hours gradually fell (see Table 8 in Chapter 1), t_{min} was biased downwards by factors 0.97, 0.95, 0.93, 0.92, 0.90, and 0.90 in years 1987, 1988, 1989, 1990, 1991, and 1992, respectively. These factors applied to all workers. The thresholds between the hours-brackets were adjusted down in 1993 so that from that year, $h - k = \tau_3$ again and the bias disappeared.

The Upper Bound

Regarding the errors in the upper bound, first notice that any biases in the lower bound on hours will carry over to the upper bound as well, since

$$t_{max} = t_{min} + ku.$$

The additional errors in t_{max} concern the choice of the constant k and the computation of u that uses the potentially contaminated ingredients u_{min} , p , and u_{max} . There is not much to do about how the period of employment is reported and the available data that determine u_{max} . Furthermore, since u_{min} is lower than u for most part-timers, u_{min} should not have a big impact on the upper bound on hours even if it is not 100% correctly computed due to the use of the

annual effective pension rates. For these reasons, when investigating the upper bound on hours I will only focus on k .

The problem with k is that it serves as the number of hours in a pay cycle that an employee can work without earning extra pensions, for all workers in all of their pay cycles. In practice, some workers work more than others, and hours may vary over time for the same worker. At the same time, the number of hours that a worker can work without earning extra pensions depends on whether she is in the top-bracket or not. To cross the threshold to the next hours-bracket in general takes a different amount of hours than reaching the assumed limit of h full-time hours in the top-bracket. This means that usually there is no ideal value for k . From 1980-1986 (see Table 8 in Chapter 1), however, $k_{wk} = h_{wk} - \tau_{3,w} = \tau_{3,w} - \tau_{2,w} = \tau_{2,w} - \tau_{1,w} = \tau_{1,w} - 0 = 10$, so in these years it did not matter what bracket the worker was in in each pay cycle. $k_{wk} = 10$ was an ideal choice in these years. However, this was no longer true from 1987. From this year to 1992, it was still the case that $\tau_{3,w} - \tau_{2,w} = \tau_{2,w} - \tau_{1,w} = \tau_{1,w} - 0 = 10$, but $h_{wk} - \tau_{3,w} < 10$. Then from 1993, $\tau_{3,w} - \tau_{2,w} = \tau_{2,w} - \tau_{1,w} = \tau_{1,w} - 0 = 9$, but $h_{wk} - \tau_{3,w} = 10$ again. With $k_{wk} = 10$ in all years, essentially from 1987 to 1992 $\alpha_{k,u}$ was correct only for workers who were never in the top-bracket. Then from 1993, $\alpha_{k,u}$ was correct only for workers who were always in the top-bracket.

Bias 3. Due to the fact that from 1987 the number of hours that can be worked without earning extra pensions depends on what hours-bracket the employee is in, and due to the fact that there is no information in the data on how often an employee is in each bracket, with Statistics Denmark's choice of $k_{wk} = 10$ the difference $t_{max} - t_{min}$ is necessarily too high on average. Workers in the top-bracket have too high $\alpha_{k,u}$ in 1987-1992, and workers outside of the top-bracket have too high $\alpha_{k,u}$ after 1993.

The Expected Impacts on Wages

All the three identified biases have an impact on part-timers and full-timers alike. Bias 1 cuts across the full-timer and part-timer categories, but many of the workers subject to Bias 1 are probably part-timers with shorter or longer job spells because weekly and bi-weekly compensated workers should be more loosely attached to the workplace than monthly paid workers. Overall, the expectation is that workers subject to Bias 1 will comprise a much larger fraction of the part-timers than of the full-timers. As a result, in relation to Puzzle 1 we should expect too high wages of part-timers in 1980-1992, but not after. The effect of the bias is thus a sharp drop in wages from 1992-1993, but only because they were too high prior to 1992. There could potentially be a small negative impact on the wage growth of full-timers but any such effect should be quite small. Both predictions are perfectly in accordance with Puzzle 1, but Bias 1 is not the whole story. In terms of magnitudes, the effect of Bias 1 on wages will be smaller than the factor with which the annual effective rate is off, simply because the total hours are larger than the lower bounds. In addition, Bias 1 only applies to weekly and bi-weekly compensated workers. The steep fall in the wages of part-timers of 5.2% found in Chapter 1 is much higher than what a narrow group of workers with wages that are strictly less than 5% too high can accomplish.

Biases 2 and 3 are also relevant for both part-timers and full-timers. But contrary to Bias 1, everybody is affected. The explanatory power of Bias 2 and 3 could therefore be much larger than that of Bias 1. In terms of the qualitative effects of the biases, Bias 2 increases the wages of everyone from 1987-1992 with the biggest effects found in 1991 and 1992. The lower bounds on hours were all 10% too low in these years, so there is a big positive effect on wages albeit less than 10% since in percentage terms the total hours are affected less than the lower bounds. There is no effect of Bias 2 from 1993. Once we turn to Bias 3, things get more complicated. The effect on the wages of full-timers is negative in the years 1987-1992 and zero from 1993 since full-timers are always in the top-bracket. In other words, Bias 3 should induce an

unambiguous jump in wages from 1992-1993. On the other hand, the effect on the wages of part-timers is slightly negative both up to 1992 and after 1993, since part-timers are both in the top-bracket and in the lower brackets. The exact effect of Bias 3 on part-timers will depend on the distribution of part-timer pay cycles across hours-brackets. Any jump or drop in the wages of part-timers from 1992-1993 arising from Bias 3 will be a second-order effect, but throughout the period their wages will be slightly depressed.

Considering the effects of Bias 2 and 3 together it is clear that the wages of part-timers are too high overall in 1987-1992, with the biggest exaggeration in 1991 and 1992 and a steep fall after that. Bias 2 and 3 are thus perfectly consistent with Puzzle 1. For full-timers, the question is how much Bias 2 and 3 cancel out with each other in 1987-1992. The exact derivation will come in the next section, but for now it suffices to note that it is the declining full-time hours that cause both biases but in different directions. Since the difference between the upper bound on hours and the lower bound is computed based on the same number of pay cycles as the lower bound on hours ($u_{min} = u$), we should not be surprised if there was no effect at all.

3 Two Corrections to Eliminate the Key Biases

The analysis in Section 2 identified three biases that were consistent with Puzzle 1. They emerged from a comparison of the lower and upper bounds on hours with their so-called true values. In this section, I correct the bounds in such a way that Bias 2 completely disappears and a good portion of Bias 3 is eliminated. I will not consider Bias 1 further as it is impossible to deal with it without making the problem worse. Indeed, without data on the compensation mode, any changes must use the same pension rate for everybody. The question is then who should have the right rate. Since monthly compensated workers are supposedly more numerous, it is better that they get it right than the weekly and bi-weekly compensated workers. In the following, I outline the corrections that eliminate Bias 2 and (part of) 3.

Correction of the Lower Bound

To begin, I replace α_k defined in Section 2 of Chapter 1 with α_y , defined as follows.

α_y : The highest possible number of hours that a worker can work less and still obtain A if she works T hours in a year and at least τ_3 hours in every full pay cycle (and proportionally less in shorter cycles).

Notice the difference to α_k that did not incorporate a requirement of always being in the top-bracket. Written out, the definition says

$$\alpha_y = \frac{T}{h}(h - \tau_3).$$

The second factor is how many hours the worker can work less in an average pay cycle, given that she reaches T hours in total and cannot go below τ_3 hours in a full pay cycle. The first factor is the number of full pay cycles. With this definition, I define the new lower bound on hours as

$$t_{min}^* = \frac{ATP}{A} \cdot (T - \alpha_y).$$

To demonstrate that Bias 2 has disappeared, simply observe that

$$t_{min}^* = \frac{ATP}{\frac{T}{h}} \tau_3 \neq \frac{ATP}{\frac{T}{h}} (h - k) = t_{min}.$$

In practice, I correct t_{min} by multiplying it by the factor $\tau_3/(h - k)$.

Correction of the Upper Bound

Bias 3 is a bit more involved. Complete elimination of it would require a modification of k reflecting the frequencies in all hours-brackets of each individual worker. Unfortunately, data on how many pay cycles each worker spends in each hours-bracket do not exist. Instead, I compute upper and lower bounds on the number of pay cycles in and outside of the top-bracket using

existing information. I then couple that information with the true widths of the brackets. The result is a simple and elegant solution that reduces the bias to a large extent while not relying on additional assumptions. When it cannot be determined what hours-bracket the worker is in, I just follow Statistics Denmark and assume that the width of the bracket is 10 in a week (cf. Section 3 of Chapter 1).

The first step is to realize that what really matters is the frequency of pay cycles spent in the top-bracket, and the frequency spent in all other hours-brackets together. Since $\tau_3 - \tau_2 = \tau_2 - \tau_1 = \tau_1$, the widths of all of the lower hours-brackets are the same. So the number of additional hours one can work without crossing the threshold to the next hours-bracket is the same in all of the lower brackets. Hence, a distinction between them is not necessary. The second observation is that the total pensions earned in the period of the employment spell necessarily sets a lower bound on the number of pay cycles in the top-bracket. Indeed, the minimum number of such pay cycles that is consistent with the earned *ATP* must satisfy the condition that enough pensions are accumulated in them to fill the minimum pension gap that may arise if the worker is in the middle-bracket in all other pay cycles. To see this more clearly, consider the following definition.

u_{top} : The smallest possible number of pay cycles in which hours worked surpass τ_3 if the worker works in u pay cycles.

Then, abbreviating the annual effective pension rate by x , u_{top} satisfies the following equation.

$$ATP = u_{top} \cdot x + (u - u_{top}) \cdot \frac{2}{3}x \text{ if } u \frac{2}{3}x \leq ATP, \quad u_{top} \geq 0.$$

The condition $u \frac{2}{3}x \leq ATP$ determines if there is a pension gap at all to be met by working in the top-bracket in some of the pay cycles. Using that $u_{min} = ATP/x$, the solution to the equation is

$$u_{top} = (3u_{min} - 2u) \cdot 1 \left[u_{min} \geq \frac{2}{3}u \right].$$

If u is only a little larger than u_{min} , then the pensions coming from the $u - u_{min}$ pay cycles are minor and the remaining u_{min} pay cycles must almost all be spent in the top-bracket in order to reach the total pension amount. If u is a lot larger, then it is possible to accumulate ATP without working in the top-bracket at all.

The upper bound on the number of pay cycles in the top-bracket is u_{min} . More pay cycles than u_{min} in the top-bracket would lead to a higher ATP . Hence, given that the worker works in u pay cycles, $u - u_{min}$ is the minimum number of pay cycles outside of the top-bracket. Equipped with these bounds and nothing else, the best correction that can be made to k is

$$k^* = \frac{u_{top}}{u} \cdot (h - \tau_3) + \left(1 - \frac{u_{min}}{u}\right) \cdot (\tau_3 - \tau_2) + \frac{u_{min} - u_{top}}{u} \cdot k.$$

The first term retains the assumption of h as an upper bound on hours in an average pay cycle, and replaces k with the true number $h - \tau_3$ for the guaranteed pay cycles in the top-bracket. The second term replaces k with the true number $\tau_3 - \tau_2$ for the guaranteed pay cycles in other brackets than the top-bracket. And the third term does not make any correction for the maximum number of pay cycles in ambiguous brackets. Notice that no extra assumptions are made meaning that k^* is a certain improvement over k . Less cautious modifications of k can be made also, for example by correcting the value in the ambiguous pay cycles to an average of the widths of each bracket with an assumption that each bracket is equally likely, as in $\frac{3}{4} \cdot (\tau_3 - \tau_2) + \frac{1}{4} \cdot (h - \tau_3)$. A bolder correction would estimate the number of pay cycles in each bracket as the average between the upper and lower bound, as in

$$\frac{1}{2} \left(\frac{u_{top}}{u} + \frac{u_{min}}{u} \right) \cdot (h - \tau_3) + \left(1 - \frac{1}{2} \left(\frac{u_{top}}{u} + \frac{u_{min}}{u} \right) \right) \cdot (\tau_3 - \tau_2).$$

Which one of the three corrections is the most desirable depends on how they perform empirically and on how strong their assumptions are. Post-1993 they will almost perform the same since the difference between the widths of the hours-brackets in weekly terms is just one hour, or

only ten percent. However, in 1987-1992 the difference between $(h - \tau_3)$, $(\tau_3 - \tau_2)$, and k is up to three times bigger and a credibility trade-off emerges. k^* on the one hand is the most conservative correction and offers a guaranteed improvement over the current approach but also leaves some pay cycles un-corrected. In contrast, the two other possible methods correct all pay cycles but also make additional assumptions that could perform worse than the current method. Especially the third method that uses the estimated frequencies in each bracket makes a strong assumption and will yield the biggest changes to hours. An empirical comparison and recommendation will be made in Section 4.

Define the corrected difference between the bounds on hours as follows.

α_u : The corrected $\alpha_{k,u}$. Equal to k^*u , $k^{**}u$, or $k^{***}u$ where $k^{(\cdot)}$ is the corrected number of hours in an average pay cycle for which no pension is earned. The three options for $k^{(\cdot)}$ are called the Conservative, Intermediate, and Bold option, respectively. The first uses the derived bounds on the fraction of pay cycles in and out of the top-bracket. The second uses the same bounds but also uses the average bracket width for the share of ambiguous pay cycles. The third uses the estimated frequencies in and out of the top-bracket.

The Impact of the Corrections on the Hours of Full-Timers

As the next step, I calculate what the impact of the changes to the bounds is for the hours of full-timers. The concern is that Puzzle 1 is really about part-timers since no strange patterns were found in the wages of full-timers in Chapter 1. If there were big changes to the hours of full-timers, then that would indicate that Puzzle 1 was not completely solved and other biases were waiting to be discovered. Fortunately, the hours of full-timers are unchanged. To see this, simply note that for full-timers there is no difference between the different ways of computing k^* , and calculate

$$\begin{aligned}
t_{max}^* &= t_{min}^* + \alpha_u = t_{min} \frac{\tau_3}{h-k} + (h - \tau_3)u = u_{min}\tau_3 + (h - \tau_3)u_{min} = hu_{min} \\
&= u_{min}(h - k) + ku_{min} = t_{min} + ku = t_{max}.
\end{aligned}$$

So there is no change to the hours of full-timers, as desired.

The Impact of the Corrections on the Hours of Part-Timers

The final step is to calculate how the hours of part-timers change. I only consider part-timers in main jobs who do not have reduced hours, since this is the main group of part-time workers. I show only the calculation for the Conservative option, but the others are similar.

$$\begin{aligned}
t_{min}^* + \frac{1}{2}k^*u &= t_{min} \frac{\tau_3}{h-k} + \frac{1}{2}(u_{top}(h - \tau_3) + (u - u_{min})(\tau_3 - \tau_2) + (u_{min} - u_{top})k) \\
&= t_{min} + \frac{1}{2}ku + u_{min}(\tau_3 - (h - k)) \\
&\quad + \frac{1}{2}(u_{top}(h - \tau_3 - k) + (u - u_{min})(\tau_3 - \tau_2 - k)).
\end{aligned}$$

The first two terms comprise the uncorrected hours of part-timers. The third term is the correction to the lower bound. The last term is the correction to the upper bound and it equals $\frac{1}{2}u_{top}(h - \tau_3 - k)$ in 1987-1992 and $\frac{1}{2}(u - u_{min})(\tau_3 - \tau_2 - k)$ after 1993. To see the impact of the corrections in relation to Puzzle 1, define $t^* = t_{min}^* + \frac{1}{2}k^*u$ and write for 1987-1992,

$$t^* - t = u_{min}(\tau_3 - (h - k)) - \frac{1}{2}u_{top}(\tau_3 - (h - k)) = (u_{min} - \frac{1}{2}u_{top})(\tau_3 - (h - k)) > 0$$

since $u_{top} \leq u_{min}$ and $\tau_3 - (h - k) > 0$ in 1987-1992. In other words, the lower bound strictly dominates as expected. On the other hand, in 1993 and after, the lower bound is unchanged, and

$$t^* - t = 0 - \frac{1}{2}(u - u_{min}) \cdot 1 < 0.$$

So wages will be adjusted down before 1993 and up after, improving on Puzzle 1.

4 Performance of the Solution to the Puzzle of Part-Timer Wages

The formal analysis of the biases and how to eliminate them yielded one correction to the lower bound on hours and three options for correcting the difference between the upper and lower bounds. To see what their effects are, I first plot the changes to the bounds and then the effects on wages, using the same sample as in Chapter 1. Since the Intermediate option lies in between the Conservative and Bold ones for the period 1987-1992, for clarity I omit it and show only the extremes. It will become clear that no essential information is omitted in this way. The first figure shows what happens to the average lower bound on hours.

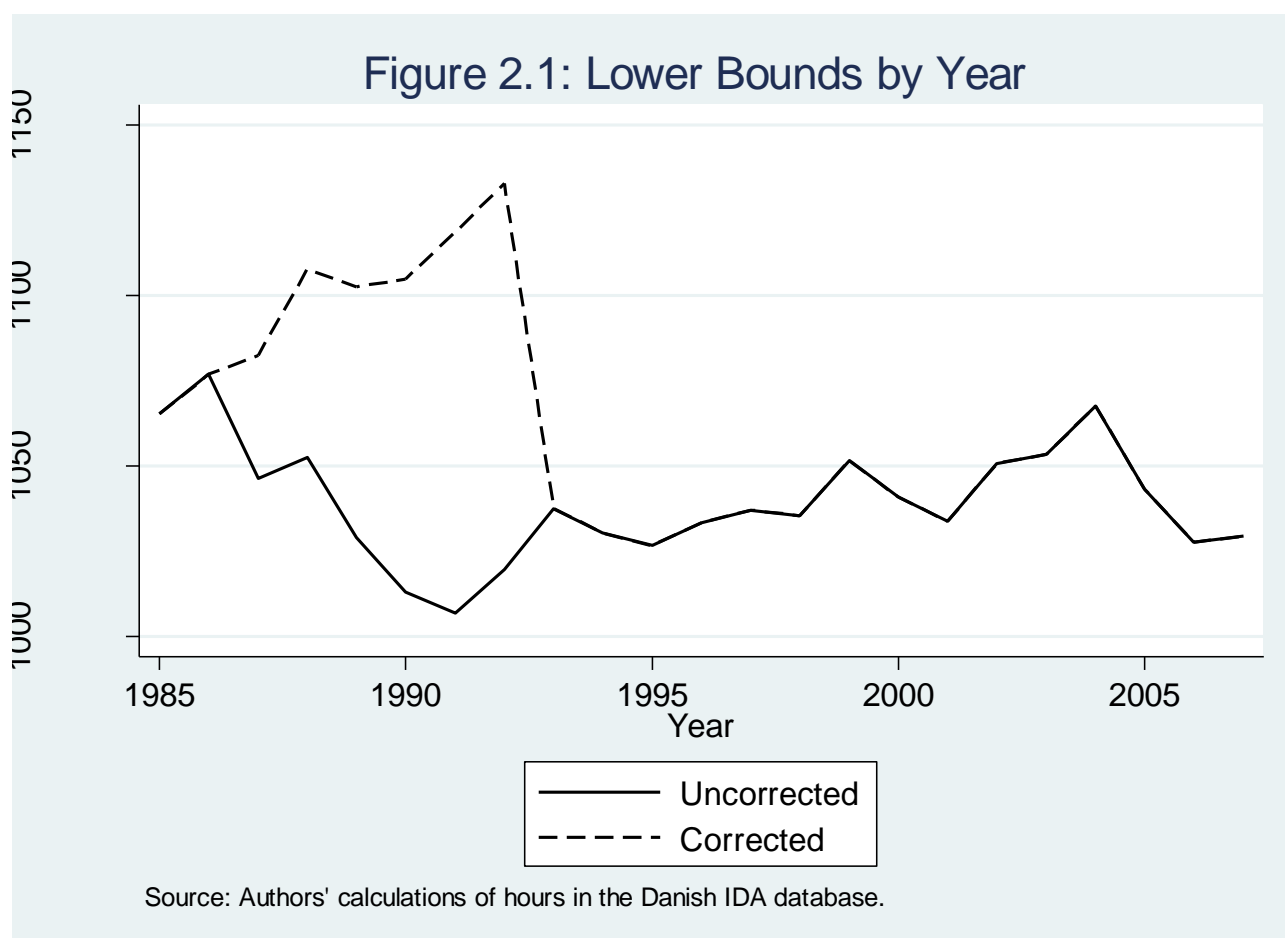
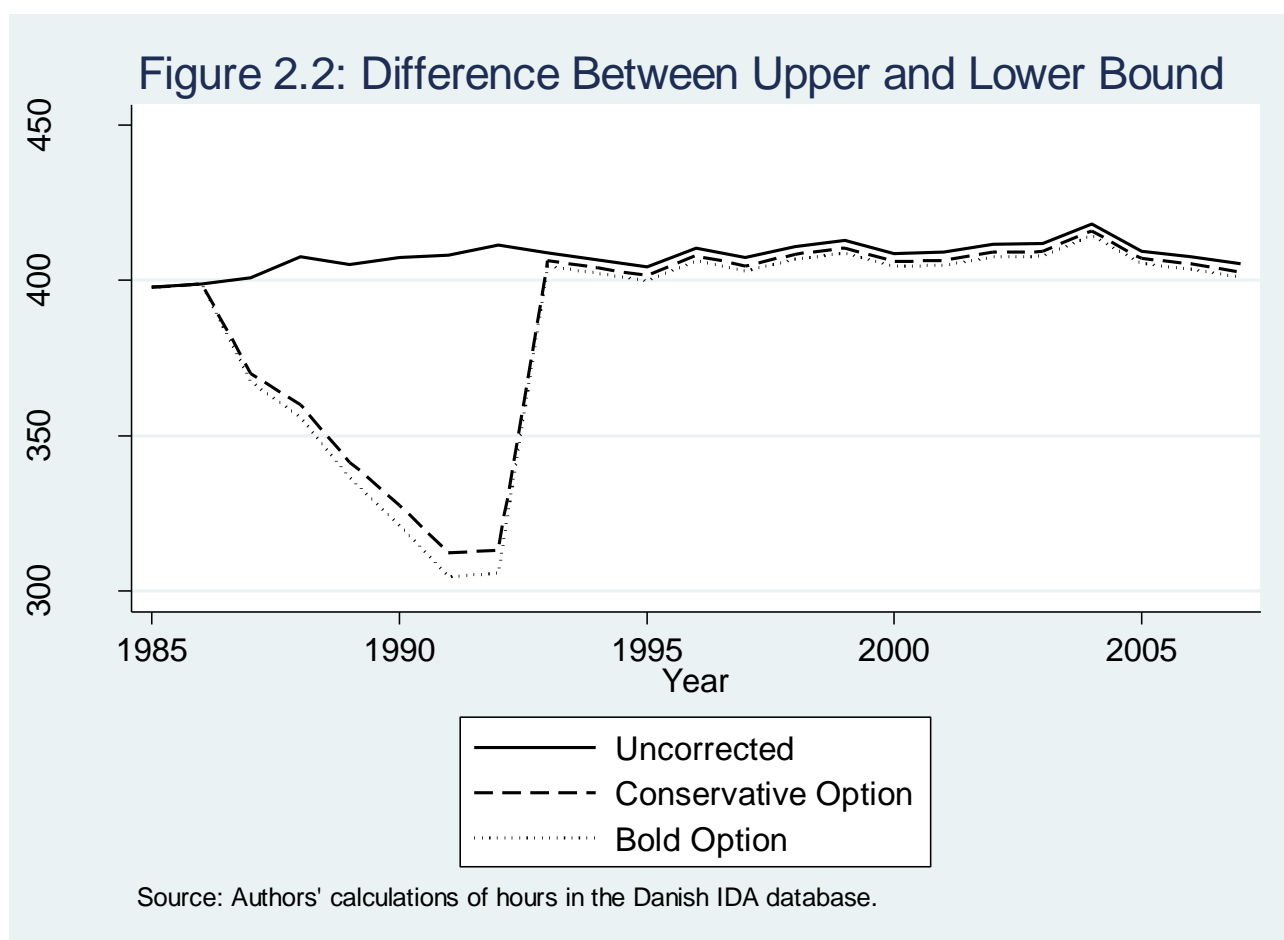


Figure 2.1 shows the dramatic effect of the correction. As expected, the impact rises from 1987 through 1991 as a result of the decline in the full-time hours. The largest effect can be found in 1991-1992 in which it equals 10%. As explained in Section 2, in these years the weekly full-time

hours were three hours lower than in 1985, yet the minimum weekly hours required for the top-bracket had remained at thirty, necessitating a correction of 10%. From 1993 the current lower bound is the right one and it is kept as it is.

In Figure 2.2 below, I show the average effect on the difference between the lower and upper bound when the Conservative and Bold options are used. Contrary to what was the case with the lower bound, the effect is not identical for everyone since it depends on the frequency in each hours-bracket. The average is computed across all workers, part-timers and full-timers alike.



First of all, Figure 2.2 shows that it does not make much of a difference whether the Conservative or Bold option is used to correct k , especially in the period after 1993 for which it was predicted that their difference would be minimal. Second, Figure 2.2 shows that the corrected average

difference between the upper and lower bound on hours is slightly lower than the current difference after 1993, and much lower prior to that. The reason is that the correction factor for full-timers reaches $(h - \tau_3)/k = 0.7$ in 1991 and 1992, and full-timers is the most important group since they are many, always in the corrected top-bracket, and in all likelihood have the longest job spells. We should not be surprised to see that the changes closely mirror the changes to the lower bound since both were the results of the fall in full-time hours and are affected in opposite ways by that fall.

I then turn to displaying the effects on wages. First, in Figure 2.3 I compare the modified series with each other and with the original series. I do not show the effects on the wages of full-timers since they are known to be zero.

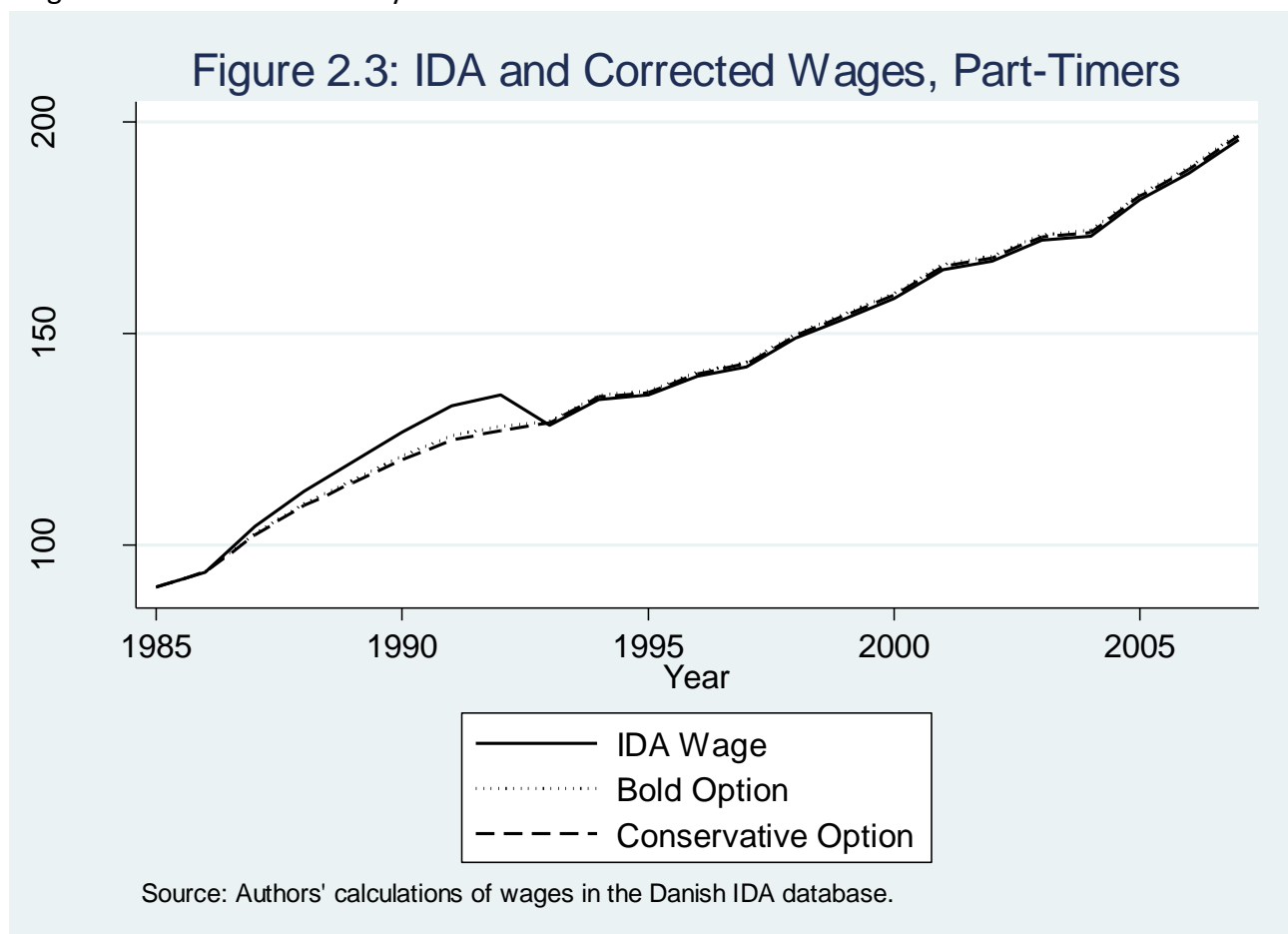
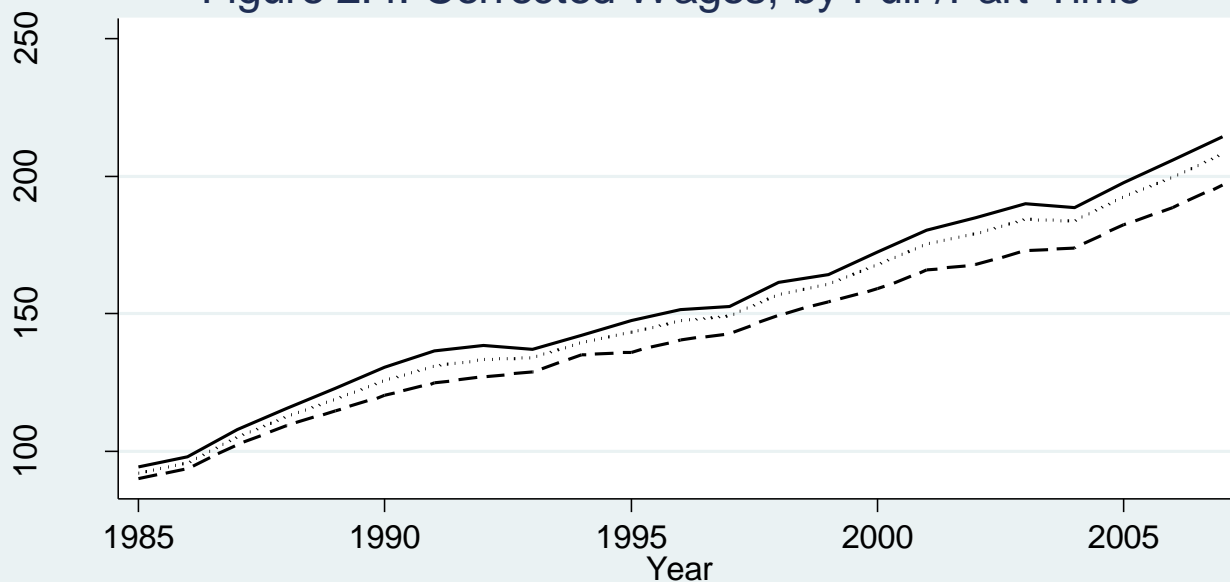


Figure 2.3 shows that the results are the same whether one uses the Conservative or Bold option. Given that the Intermediate option (not shown) would produce a series that lay in between the two others for 1987-1992, in practice all three methods yield identical outcomes. Since the Conservative option is known to offer a guaranteed improvement without additional assumptions and does not perform worse, it is the recommended correction. For this reason, in the following figures I display only those series. Second, Figure 2.3 shows the effects of the improvements on the wages of part-timers. As can be seen, the elimination of Bias 2 and 3 completely solves the identified problems with the wages of part-timers in Puzzle 1. The dip in 1992-1993 disappears and the wages of part-timers are no longer as high in the period 1987-1992, with the biggest revisions happening exactly in the years they should happen. Moreover, as predicted, the wages were a bit too low after 1993. Even though Bias 1 was also completely consistent with Puzzle 1, the somewhat narrow group of workers that were subject to it was apparently not big enough to really have an impact. If Bias 1 had had an impact, the truth would have resembled an *upward* jump in wages in 1992-1993 which is very unlikely for an economy emerging from a recession, as it happened.

To see the modified wages of part-timers in comparison with the wages of full-timers, in Figure 2.4 I replicate Figure 1.3 in Chapter 1.

Figure 2.4: Corrected Wages, by Full-/Part-Time



— Full-Timers
 - - - Part-Timers
 Full-Timers and Part-Timers

Note: The wage series uses the Conservative option.
 Source: Authors' calculations of wages in the Danish IDA database.

Indeed, the relative differences between the wages of part-timers and full-timers and the growth rates over time look much better than in Figure 1.3 of Chapter 1. To get a better sense of the numbers, in Table 2 I replicate Table 9 in Chapter 1 and add the corrected wages.

Table 2. Average Wages of Part-Timers and Full-Timers, 1985-1993

| Year | Full-Timers | Part-Timers | Relative Difference | Conservative | Relative Difference |
|------|-------------|-------------|---------------------|--------------|---------------------|
| 1985 | 94.3 | 90.1 | -4.5 | 90.1 | -4.5 |
| 1986 | 98.0 | 93.7 | -4.4 | 93.7 | -4.4 |
| 1987 | 107.9 | 104.5 | -3.1 | 102.4 | -5.1 |
| 1988 | 115.5 | 112.8 | -2.4 | 109.3 | -5.4 |
| 1989 | 122.8 | 119.7 | -2.5 | 114.8 | -6.5 |
| 1990 | 130.7 | 126.7 | -3.1 | 120.2 | -8.0 |
| 1991 | 136.5 | 133.0 | -2.5 | 124.8 | -8.6 |
| 1992 | 138.6 | 135.4 | -2.3 | 127.1 | -8.3 |
| 1993 | 137.2 | 128.3 | -6.5 | 128.9 | -6.0 |

Note: The third column shows the difference in percent between columns one and two. The fourth column shows the wage series using the Conservative option. The fifth column shows the difference in percent between the first and fourth columns. The average of column five is -7.6% from 1994-2007. The average of column three in the same time span is -8.0%.

Judging from Table 2, it looks like Puzzle 1 has been completely solved except that the relative differences seem a bit too close to zero in 1985 and 1986 and nothing is done about that. In the other years, the wages of part-timers are no longer as close to the wages of full-timers nor approaching them. Moreover, the big 5.2% drop from 1992-1993 has been replaced with a small increase. It is not clear what could cause the high wages of part-timers in 1985 and 1986 and if it is a problem at all. The high wages of part-timers in these years are not the result of Bias 1 since its effect was the same in all years prior to 1993.

5 Identification and Correction of the Bounds With Standard Variables

The performance analysis of Section 4 of this chapter used a reconstructed dataset that was based on the raw registers of Statistics Denmark. Only very few users have access to the registers but nevertheless it is possible for everyone to apply the corrections proposed in this chapter. In fact, from readily available standard variables one can easily identify the lower and upper bounds that were used in the estimation of hours. All one needs is *timelon*, *joblon*, *tilknyt*, *tlonkval*, *type*, *aar*. Here, *timelon* is the hourly wage, *joblon* is the total earnings of the year in the job, *tilknyt* is a variable that separates part-timers from full-timers, *type* distinguishes between main and second jobs, and *aar* is the year. The actual estimate of hours is easily recovered by dividing *joblon* with *timelon*. Then, since *tlonkval* is the relation between the hours estimate and the width of the interval of possible hours, by combining *tlonkval* with *joblon* and *timelon* one obtains information about the width of the interval. The second step is then to use whether a worker is a full-timer or part-timer to infer where in the interval the estimate of hours is. Combining the width of the interval, the actual estimate of hours, and where in the interval the estimate is, it is easy to recover the bounds of the interval.

There are some groups of workers whose hours need not or cannot be corrected. I can keep the hours of full-time workers since the corrections that would apply would cancel out anyway, but in principle I should correct the bounds of all part-timers. At present, however, it is not possible to correct the bounds of part-time workers in second jobs. There are two reasons for this. First, it has not been possible to verify in the old program codes given to us by Statistics Denmark that part-timers in second jobs actually have *tlonkval* computed as the relative uncertainty. Second, even if they have, from standard variables I cannot identify who among the part-timers in second jobs have t_{min} , T_{red} , or $\frac{1}{2}(t_{min} + t_{max})$ as their hours (see Section 2 of Chapter 1). Unless the fraction of one of them is very high (which I also do not know), it will not make any sense to try to identify and correct the bounds since it is not known where in the interval of hours the actual hours estimate is. Finally, the procedure mistakenly modifies the hours of part-timers in main jobs with reduced hours. The reason is the same as for second jobs, it is impossible to distinguish workers with reduced hours. However, at least we know that the fraction of workers with reduced hours approximately equals 5.4% of part-timers. With this small number, I choose to correct the bounds of all part-timers in main jobs, even if the result is not ideal for some observations.

Consider then a part-time worker in a main job defined as $tilknyt \notin \{01, 11, 21, 31\}$ and $type = H$. The estimate of hours is conveniently located exactly in the middle of the interval (provided that the worker does not have reduced hours). From Section 2 of Chapter 1, it is known that

$$r_w = \frac{1}{2} \left(\frac{\alpha_{k,u}}{t_{min}} + \frac{\alpha_{k,u}}{t_{min} + \alpha_{k,u}} \right).$$

The variable *tlonkval* expresses the relative uncertainty in percentage terms, that is,

$$tlonkval = 100r_w.$$

Solving for $\alpha_{k,u}/t_{min}$ and discarding the negative root,

$$\frac{\alpha_{k,u}}{t_{min}} = \frac{t_{lonkval}}{100} - 1 + \sqrt{1 + \left(\frac{t_{lonkval}}{100}\right)^2} = r.$$

At the same time, it is known that $t = t_{min} + \frac{1}{2}\alpha_{k,u}$, and t is observed as $joblon/timelon$. Hence,

I have two equations in two unknowns that are easy to solve. The result is

$$t_{min} = \frac{\frac{joblon}{timelon}}{1 + \frac{1}{2} \cdot r},$$

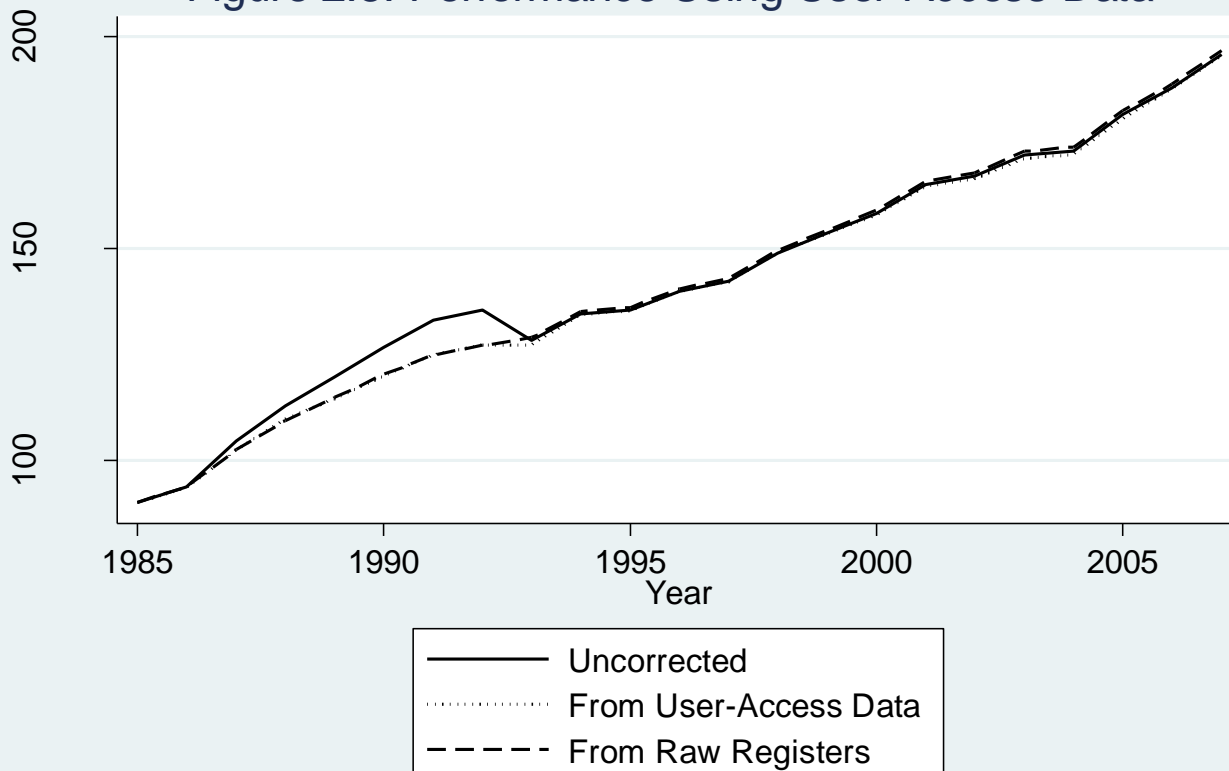
$$\alpha_{k,u} = r \frac{\frac{joblon}{timelon}}{1 + \frac{1}{2} \cdot r}.$$

The lower bound t_{min} is then corrected by multiplying by $\tau_3/(h - k)$. In order to correct $\alpha_{k,u}$ it is necessary to identify u_{min}/u . This is done by computing

$$\left(\frac{u}{u_{min}}\right)^{-1} = \left(\frac{\alpha_{k,u}}{t_{min}} \frac{h - k}{k}\right)^{-1} = \left(r \frac{h_{wk} - 10}{10}\right)^{-1}.$$

From u_{min}/u one computes u_{top}/u and k^*/k and multiplies $\alpha_{k,u}$ by k^*/k to obtain α_u . Then the corrected estimates of hours are computed and the corrected wages are obtained by dividing $joblon$ by t^* , the new measure of hours. The SAS code for this program can be found in the Appendix. The performance of it is very good, as can be seen in Figure 2.5. Indeed, the series based on user-access data almost replicates the more elaborate solution based on the raw registers.

Figure 2.5: Performance Using User-Access Data



Source: Authors' calculations of wages in the Danish IDA database.

There are only minor differences from 1993 on where the series based on the raw registers outperforms the accessible solution. In these years, the latter is a few DKK too low.

Appendix C2

Section 5 of this chapter mentioned how to correct hours and wages only with access to standard variables. One key relation was

$$t\text{lonkval} = 100r_w.$$

However, this was a simplification. The real *tlonkval* in the data has some rounding issues that are described in detail in Section 5 of Chapter 1. The reader is referred to that section for more

information. In practice, the recipe in this Chapter is the best one can do and the corrections still perform very well.

Code Using the Conservative Option, Using only Standard Variables

```

data w_corrections;
set sample;
  if type='H' & (tilknyt not in('01','11','21','31')) then do; /*Main jobs and
part-timers*/
    r=tlonkval/100-1+sqrt(1+(tlonkval/100)^2);
    hours=jobblon/timelon;
    t_min=hours / (1+r/2);
    alpha_u=hours*r/(1+r/2);
    if aar<=1986 then do;
      u_min_share=1/(3*r); /*k/(h-k)=1/3, and r=alpha_u/t_min*/
      indi=(u_min_share>=2/3);
      u_full_share=(3*u_min_share-2)*indi;
      u_small_share=1-u_min_share;
      h_cor_factor=1*u_full_share+1*u_small_share+1*(1-u_full_share-
        u_small_share);
      new_alpha_u=h_cor_factor*alpha_u;
      new_hours=t_min+1/2*new_alpha_u;
    end;
    if aar=1987 then do;
      new_t_min=(30/29)*t_min;
      u_min_share=1/r*10/29;
      indi=(u_min_share>=2/3);
      u_full_share=(3*u_min_share-2)*indi;
      u_small_share=1-u_min_share;
      h_cor_factor= 9/10*u_full_share+1*u_small_share+1*(1-
        u_full_share-u_small_share);
      new_alpha_u=h_cor_factor*alpha_u;
      new_hours=new_t_min+1/2*new_alpha_u;
    end;
    if aar=1988 then do;
      new_t_min=(30/28.5)*t_min;
      u_min_share=1/r*10/28.5;
      indi=(u_min_share>=2/3);
      u_full_share=(3*u_min_share-2)*indi;
      u_small_share=1-u_min_share;
      h_cor_factor= 8.5/10*u_full_share+1*u_small_share+1*(1-
        u_full_share-u_small_share);
      new_alpha_u=h_cor_factor*alpha_u;
      new_hours=new_t_min+1/2*new_alpha_u;
    end;
    if aar=1989 then do;
      new_t_min=(30/28)*t_min;
      u_min_share=1/r*10/28;
      indi=(u_min_share>=2/3);
      u_full_share=(3*u_min_share-2)*indi;
      u_small_share=1-u_min_share;
      h_cor_factor= 8/10*u_full_share+1*u_small_share+1*(1-
        u_full_share-u_small_share);
      new_alpha_u=h_cor_factor*alpha_u;

```

```

    new_hours=new_t_min+1/2*new_alpha_u;
end;
if aar=1990 then do;
    new_t_min=(30/27.5)*t_min;
    u_min_share=1/r*10/27.5;
    indi=(u_min_share>=2/3);
    u_full_share=(3*u_min_share-2)*indi;
    u_small_share=1-u_min_share;
    h_cor_factor= 7.5/10*u_full_share+1*u_small_share+1*(1-
        u_full_share-u_small_share);
    new_alpha_u=h_cor_factor*alpha_u;
    new_hours=new_t_min+1/2*new_alpha_u;
end;
if aar in(1991,1992) then do;
    new_t_min=(30/27)*t_min;
    u_min_share=1/r*10/27;
    indi=(u_min_share>=2/3);
    u_full_share=(3*u_min_share-2)*indi;
    u_small_share=1-u_min_share;
    h_cor_factor= 7/10*u_full_share+1*u_small_share+1*(1-
        u_full_share-u_small_share);
    new_alpha_u=h_cor_factor*alpha_u;
    new_hours=new_t_min+1/2*new_alpha_u;
end;
if aar>=1993 then do;
    u_min_share=1/r*10/27;
    indi=(u_min_share>=2/3);
    u_full_share=(3*u_min_share-2)*indi;
    u_small_share=1-u_min_share;
    h_cor_factor= u_full_share+9/10*u_small_share+1*(1-
        u_full_share-u_small_share);
    new_alpha_u=h_cor_factor*alpha_u;
    new_hours=t_min+1/2*new_alpha_u;
end;
rename timelon=old_timelon;
new_timelon=joblon/new_hours;
drop hours new_hours;
end;
/*Full-timers: No correction necessary as the corrections cancel out.*/
/*Type='B': I'm not able to correct type Bs;*/
if type^='H' | (tilknyt in('01','11','21','31')) then do;
    new_timelon=timelon;
    rename timelon=old_timelon;
end;
run;

```

Chapter 3: The Tenure Profile Puzzle and Other Compositional Issues

1 Introduction

Usually wages rise over the course of a job spell. Countless studies have measured the effects of tenure, seniority, and human capital accumulation and found positive relationships. The Danish labour market should be no exception. However, Puzzle 4 in Chapter 1 documented how the wages in the Danish IDA database fell significantly from the first year of employment to the second year. This pattern is worrying because many studies use the IDA wage to study the returns to tenure as well as other related questions (Bagger et al, 2014; Buhai et al, 2014). Chapter 1 offered some clues as to what could be wrong with the wage measure. It examined if a fall occurred among both full-timers and part-timers but found that the individual tenure profiles for the two groups looked perfectly normal. As a result, the only likely explanation was that the proportions of the part-timers and full-timers changed systematically from the year of hiring to the second year of employment. One can think of two explanations for such a compositional change of the workforce. One is that part-timers and full-timers are unequally likely to stay in their jobs from the first year to the second, and another is that there is a transitioning between the two groups. In this chapter, we show that the important explanation is the latter one. Many workers classified as full-timers in their second year of employment are in fact classified as part-timers in their first year, whereas only very few part-timers in their second year were full-timers in their first year. Moreover, we argue that the transition is mostly an unfortunate artefact of a classification method that does not reflect what happens in the labour market in practice.

The mis-classification happens for two reasons. Most importantly, Statistics Denmark makes a mistake when computing the duration of employment within each year. This can be seen by the fact that lower and upper bounds on the number of pay cycles within an employment always take vacation into account, but an estimate of the number of pay cycles does not. Whenever the estimate is inside the bounds, the classification procedure compares it to the lower bound, and judges a worker to be a full-timer if there is equality. Since the estimate includes vacation and the lower bound does not, equality will not obtain and full-timers will appear as part-timers. This mistake happens systematically for new hires in their first year, since the upper bound on the number of pay cycles for most of them is close to an entire year of work, whereas the estimate in most cases picks up the fact that the new hire has only worked part of the year. As a result, for these workers the estimate is inside the bounds, and they are excessively classified as part-timers.

Secondly, there is a big potential for erroneous reporting of the period of employment of new hires, but not for more senior workers. The commencement date of new hires can be any day of the year, and there is ample room for the reported date to be far from the correct one. In contrast, by definition the workers in their second year and beyond work from the beginning of the year, so for them we have a very precise indication of the employment period. Moreover, the estimation method detects and corrects under-reporting of the employment period of the full-timers. So for full-timers, there is a lot of over-reporting in the first year, and that is the only measurement error there is. For part-timers, however, there is both under- and over-reporting in the first year. Now, full-timers and part-timers are affected asymmetrically by erroneous reporting. An epsilon of over-reporting pushes the full-timers into the part-timer category, but under-reporting only seldom changes the status of the part-timers. As a result, 27% of continuing workers seem to transition from part-timer to full-timer, and only 5% in the other direction.

The effect of the mis-classification is to raise wages artificially in the first year. The hours of a full-timer turned part-timer will be discontinuously too low since they are computed by averaging the bounds on hours instead of using the upper bound. While it is true that an over-reported employment period tends to increase the total hours if one sticks to either the average or the upper bound, in practice the switch to the average dwarfs the increase. In effect, over-reporting seems to be frequent but often not huge in magnitude. The bottom-line is that hours are too low and wages too high in the first year.

As it turns out, similar composition effects may appear in some industries. Full-time workers in industries with varying work loads risk being classified as part-timers simply because they in some pay cycles do not work enough hours to be in the top-bracket, even if they compensate in other pay cycles by working over-time. Our analysis shows that Agriculture, Construction, and Manufacturing have the highest proportions of workers just outside the top-bracket and the lowest proportions of workers in the top-bracket, and they are all known to be industries with varying work loads. More stable industries such as Utilities and some Services are less prone to excluding true full-timers from the full-timer category. It follows that the volatile industries should record too high wages and the stable industries should not. We use the wage survey benchmark that we constructed in Chapter 1 and indeed find that Agriculture and Construction pay higher wages relative to the benchmark than other industries. However, there is no such effect in Manufacturing.

The same logic extends to occupations and time series. For example, in some occupations the hours of each employee are strictly monitored but not in others. Although we do not have direct evidence, if hours are not precisely measured, there is a certain variance in the reporting which means that workers in some pay cycles could end up outside of the top-bracket. Such occupations would appear as high-wage. Similarly, during the last two decades some rules on working hours have softened and it has become common to place working hours more flexibly. As

a result, it is a possibility that more full-time workers in recent years have been pushed into the part-timer category compared to in previous years, with artificially higher wages as a result.

In all the mentioned examples, the wages of the misclassified full-timers were too high while the wages of the remaining workers seemed not to be. An obvious implication is that the measured aggregate and part-timer wages should be too high as well. This hypothesis was incidentally tried in Figures 1.4a-c of Chapter 1, using the survey wage measure constructed there. We found that the estimated wages were indeed too high and phrased the result as Puzzle 3. In this chapter, we show that our solution to the compositional issues can explain about half of that puzzle as well.

We argue that the fundamental problem with the estimation method is the discrete classification of workers into part-timers and full-timers. Two or any finite number of categories will not work since some workers will always be mis-classified, creating biases in the wages. It turns out that there is a straightforward way of working with a continuum of worker types. The key step is to estimate the *degree* to which a worker is a part-timer or full-timer, and here we take advantage of an elegant method that was developed in Chapter 2. There, it was estimated how often a worker was in the top-bracket, and we use that criterion to distinguish workers on the continuum. We then estimate hours as a convex combination of the current estimation methods for part-timers and full-timers, using the frequency in the top-bracket as the weight. As it turns out, our approach is micro-founded if one accepts the idea that most workers should work full-time hours when and only when they are in the top-bracket. We argue that that is indeed the case for the workers whose wages we correct.

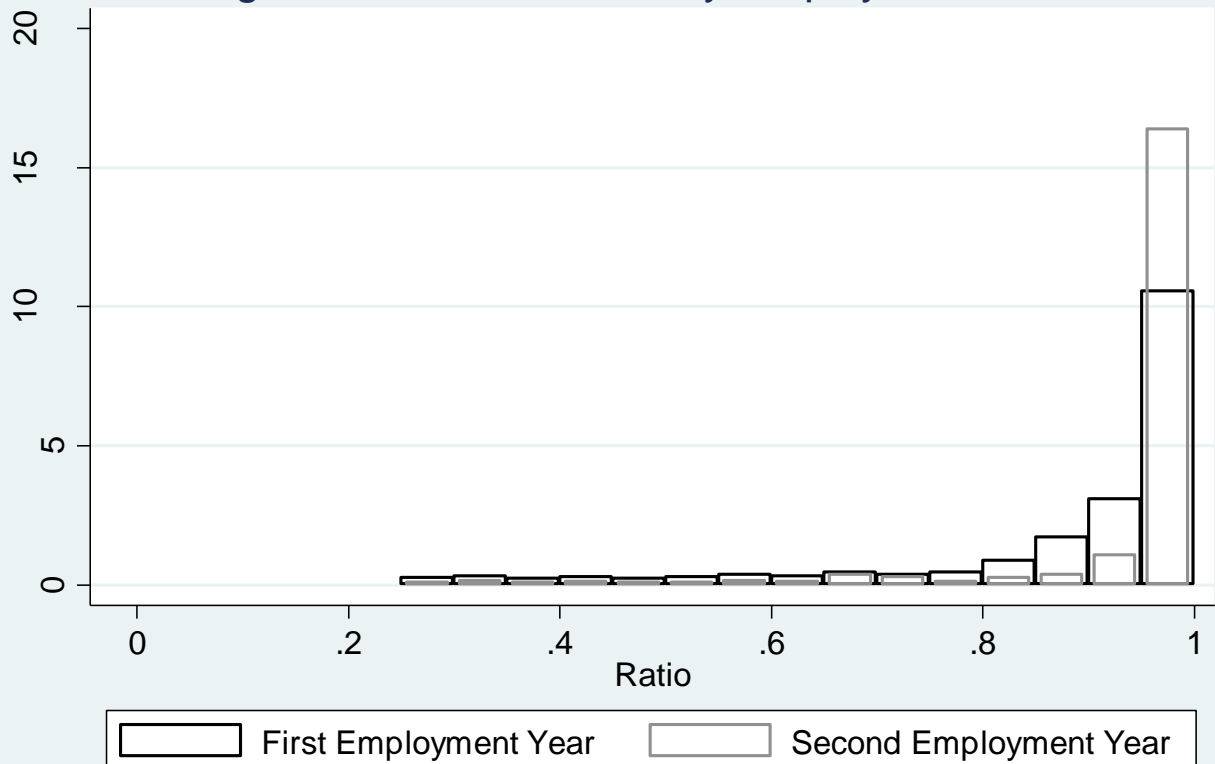
This chapter consists of four main sections. First Section 2 offers a diagnosis of the changing composition of part-timers and full-timers from the first year of employment to the second. Then the solution is presented in Section 3 and evaluated in Section 4. Finally, Section 5 combines the changes to the IDA wage of this chapter with those of Chapter 2.

2 Diagnosis of the Composition Puzzle and Its Effects

It was established in Chapter 1 that the abnormal tenure profile presented there must stem from a compositional change in the shares of part-timers and full-timers from the year of hiring to the second year. However, the driving force behind the changing composition was not identified. Here we dig deeper and isolate the driving force as well as ponder how it can create the abnormal tenure profile and perhaps other biases as well. To begin, first recall that part-timers and full-timers are primarily distinguished by comparing the ratio between the variables u_{min} and u to one. Any changes in the composition of part-timers and full-timers from the year of hiring to the second year will be reflected in the distributions of this ratio. Figure 3.1 compares the two distributions.⁷

⁷ Throughout this chapter, we only look at the years 1994-2007 to avoid possible contamination from issues with the weekly hours in 1986-1992 (see Chapters 1 and 2).

Figure 3.1: u_{min}/u Ratio by Employment Year

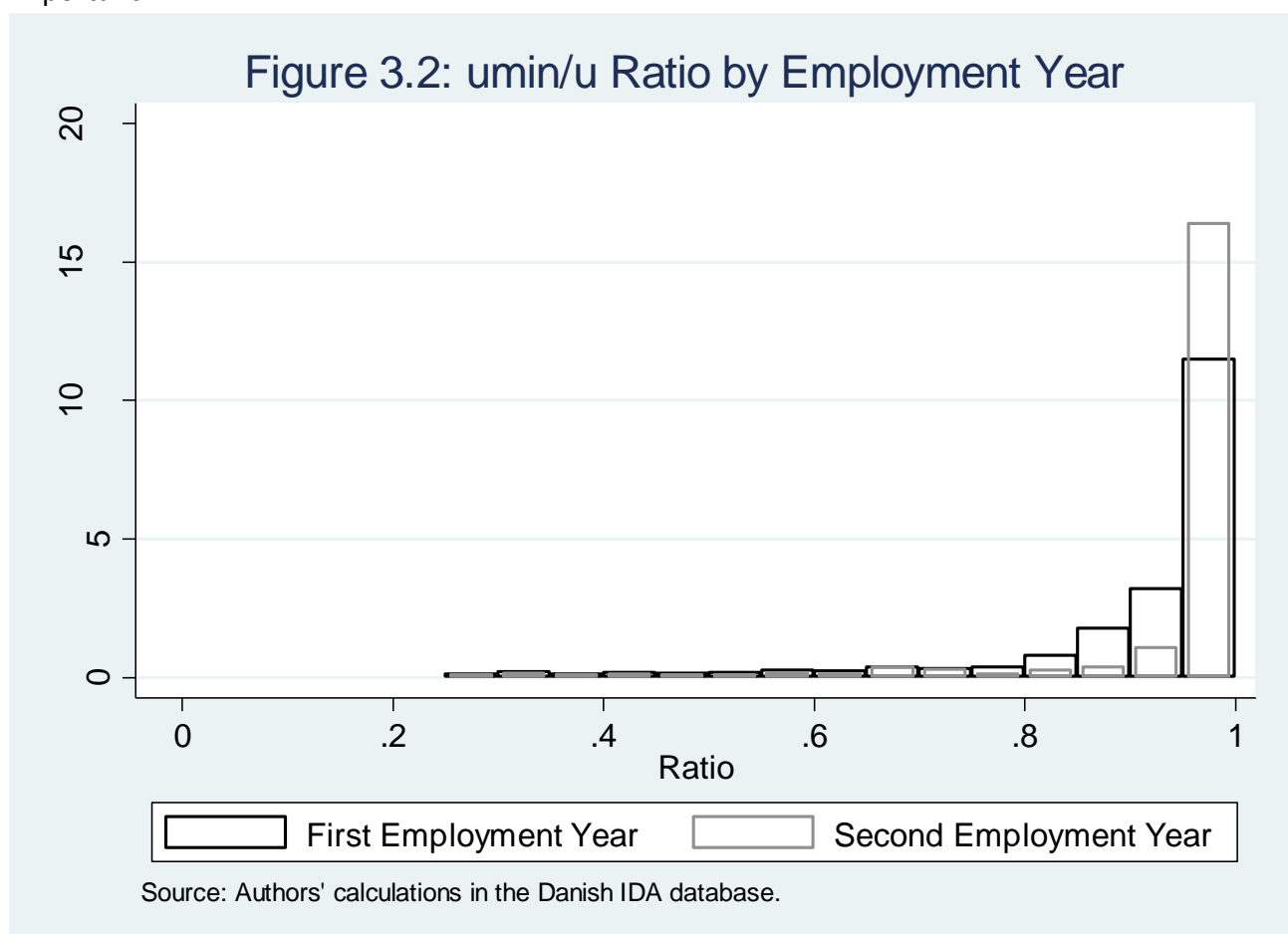


Source: Authors' calculations in the Danish IDA database.

Clearly, the second year is characterized by many full-time workers and few part-time workers relative to the year of hiring. Indeed, there is a much higher mass on the values of the ratio below one in the first year, and a correspondingly lower mass in the right-most column that includes the full-timers.

There are three competing explanations for the compositional change that we see, namely attrition and real and perceived transitions between part-time and full-time work. If part-timers leave their jobs to a larger extent than full-timers, then full-timers will appear as more frequent in the second year. Alternatively, some workers may work more or less than in their first year and change the composition as a result. Finally, it could be that many workers were systematically classified in the wrong category in one year, with too many part-timers in the first year or too many full-timers in the second year. Unfortunately it is impossible to completely

separate attrition from the two other explanations since some of the workers who separate could also have changed classification from part-time to full-time if they had continued. Such interaction effects could potentially be non-negligible since only 74% of workers in their first year continue into their second. Nonetheless, in Figure 3.2 below we attempt to visually assess the effect of attrition. We show a version of Figure 3.1 that is conditional on workers who continue into their second year. If Figures 3.1 and 3.2 turn out to be roughly similar, attrition should not be important.



Comparing Figure 3.2 with Figure 3.1 suggests that attrition is only a small contributor to the compositional change. Indeed, in the right-most bar a clear majority of the difference in probability masses between the first and second year remains after controlling for attrition. Accordingly, many of the full-timers in year two were classified as part-timers in year one.

In order to further distinguish between the two remaining explanations (real transitioning and misclassified workers) it is informative to compute how many workers change from part-time to full-time and from full-time to part-time. Doing so, we find that 27% of workers who continue into their second year are recorded as changing from part-time to full-time, whereas only 5% move in the other direction, from full-time to part-time. These numbers indicate that some mis-classification is the most likely explanation for the changes in work intensity. Indeed, it seems quite suspicious that more than a quarter of continuing workers work part-time in their first year and full-time in their second. Employers usually hire full-time workers if they have a full-time position to fill and part-timers if only part-time work is available. In the data, 48% of new, continuing positions are part-time initially, and it does not seem realistic that the workload expands in 56% of those cases while in full-time positions it almost never shrinks (10%). A similar argument can be made from the perspective of the supply side. If the work intensity reflects the worker's preference, then in 56% of the cases the part-timers change their minds. This is quite a high number, and it seems even higher when compared to the corresponding statistic for the full-timers, which again is 10%. Even if the work intensity is not a question of preference but is imposed on the worker by outside circumstances, then changes in such circumstances should also affect the full-timers. Finally, one theory that at first glance seems consistent with the facts is that of the springboard, meaning that some new hires learn so much while being part-time employed that they are able to fulfil the duties of other (full-time) positions within the firm after the first year. However, that outcome is a non-automatic, random development that cannot be predicted by employers and employees in advance, so it seems unrealistic that 56% of the new part-timers end up making the jump. In any event, if an employer wants to try out an employee, the employer can do just that – hire the new employee in the position she is supposed to fill, and make a decision when it becomes clear whether the employee performs well enough.

We then investigate where the mis-classification might come from. In doing so, we identify two explanations that fit the facts perfectly. To begin, recall from earlier chapters that u_{min} should be correctly reported in the years 1994-2007 that concern us here, hence we examine the reported employment period u more closely. Our first probe reveals a systematic measurement error in u . Consider an employer who is about to report the period of employment of an employee during the past year. The employer has three options. The first is to tick off an “entire year”-option, the second is to report exact commencement and termination dates of the employment period, and the third is to tick off a “multiple periods throughout the year”-option. New hires have a huge margin of possible error in u because there is no way of verifying the commencement dates which could be any time within the year. In contrast, workers in their second year and beyond always work at the beginning of the year. Indeed, in practice they are identified as those who worked in that position also at the end of the previous year. In most cases they do not interrupt their employment and thus cannot possibly have an over-reported period of employment within that year. If they do interrupt their employment, the “multiple-periods throughout the year”-option should be ticked, also preventing over-reporting from happening. Under-reporting should be quite rare as well, given how easy it is to just tick off “the entire year”-box or “multiple periods throughout the year”⁸ when they apply. As a result, for the most part the second year employment periods should be correctly reported. Now, u_{min} is a theoretical lower bound on the period of employment (see Chapter 1) and the ratio u_{min}/u should never exceed one. If it does, a cap is put in place during the estimation stage. The cap affects full-timers and part-timers differently. Indeed, full-timers are completely invulnerable to under-reporting since their u_{min}/u -ratios are already one, while the ratios of the part-timers adjust to the under-reporting. As a result, in the year of hiring many of the full-timers will have over-reported

⁸ The “multiple periods throughout the year”-option is rarely used. In a sample from 1981, Statistics Denmark found that among main jobs, only 5% were reported to have worked multiple periods. This is fortunate for us as it complicates the wage estimation.

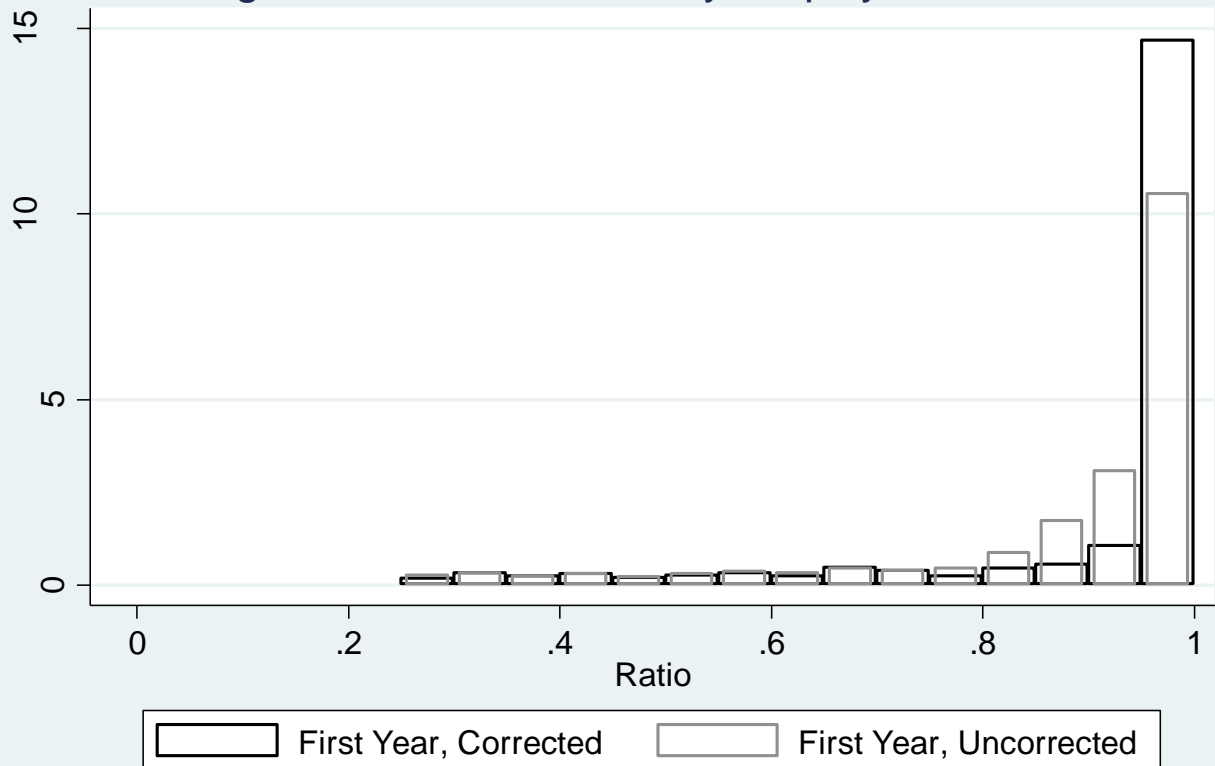
employment periods but none of them will have under-reported employment periods. In contrast, part-timers will experience both under-reporting and over-reporting (with under-reporting happening to a larger extent among part-timers who work less). Finally, even the slightest over-reporting is enough to make full-timers appear as part-timers, while it takes a sufficiently big under-estimate of the employment period plus the satisfaction of other conditions (Chapter 1) for part-timers to appear as full-timers. It follows that there will be way too many part-timers in the first year, and only in the first year. The exact effects of the over-reporting will be quantified later.

In our second probe, we examine the variables that u is computed from and find a coding error. Recall from Chapter 1 that

$$u = \max \left\{ u_{min}, \min \{ p, u_{max} \} \frac{T}{h(52 - V)} \right\},$$

where u_{min} as mentioned should be accurate, p is the period of employment as reported by employers, and u_{max} is an upper bound on the period with actual work, computed as one year minus vacation and detected periods with illness allowances or full-time unemployment insurance benefits. The problem with this formula is that the units of the period lengths are not the same. u_{min} is measured in full pay cycles, and so is $u_{max} \cdot T / (h(52 - V))$ since vacation has been netted out from u_{max} and the conversion factor $T / (h(52 - V))$ simply accounts for legal holidays (see Chapter 1). However, vacation is included in p . The effect is that for employment spells with $u_{min} < p < u_{max}$, u is off by a factor of roughly $52/47 = 1.1064$, implying that u_{min}/u is off by a factor of 0.904. The most common cases in which this happens is exactly the new hires who by definition work less than a year. The effect of removing this coding error is huge and can be seen in Figure 3.3. After the removal, the distribution resembles much more that of the second year displayed in Figures 3.1 and 3.2.

Figure 3.3: u_{min}/u Ratio by Employment Year



Source: Authors' calculations in the Danish IDA database.

From now on, we work only with the correct units of p . No illustration or statistic below will be contaminated with an employment duration measure that includes vacation.

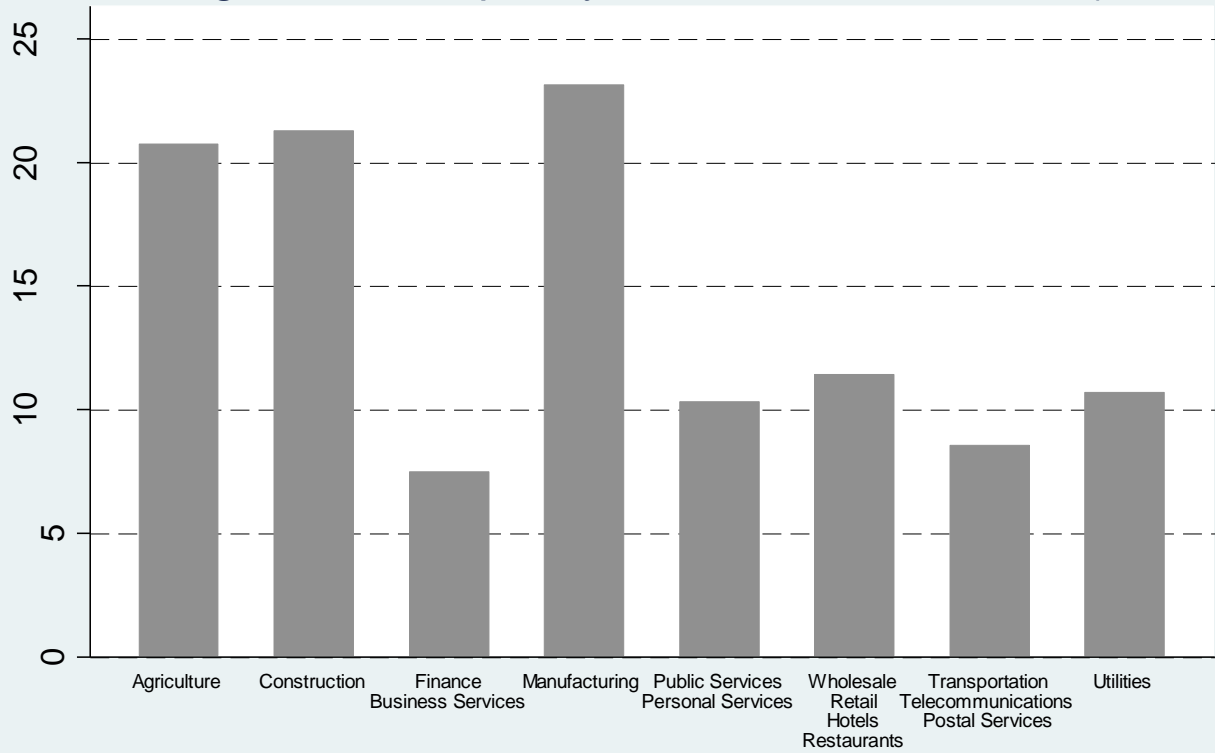
The final step in evaluating the explanations for the changing composition that we have put forth is to consider their effects on the tenure profile. The key observation that catches the eye is that the average wage is higher in the first year when the share of part-timers is high. How can that be if the wages of full-timers are higher, as found in Chapter 1? The answer is misclassification, as opposed to attrition and transitioning workers. Everything else equal, separating or transitioning, correctly classified part-timers would make the average wage rise from the first to the second year. On the other hand, the measurement and coding error explanations are completely consistent with the abnormal tenure profile. Since the hours of part-timers are calculated as the average of the bounds on hours and the hours of full-timers are set to the upper

bound, a full-timer classified as a part-timer will experience a discrete increase in the wage. Of course, this effect is only present if the over-reporting or coding error of the employment period of new hires is small in magnitude. If it is big, then the longer period will more than compensate for the change in estimation method. However, Figure 3.3 indicates that the change in the u_{min}/u ratio as a result of the coding error is indeed most often small, typically less than 20%, compared to the 50% needed for no effect on the wages.

Other areas where mis-classification might have effects

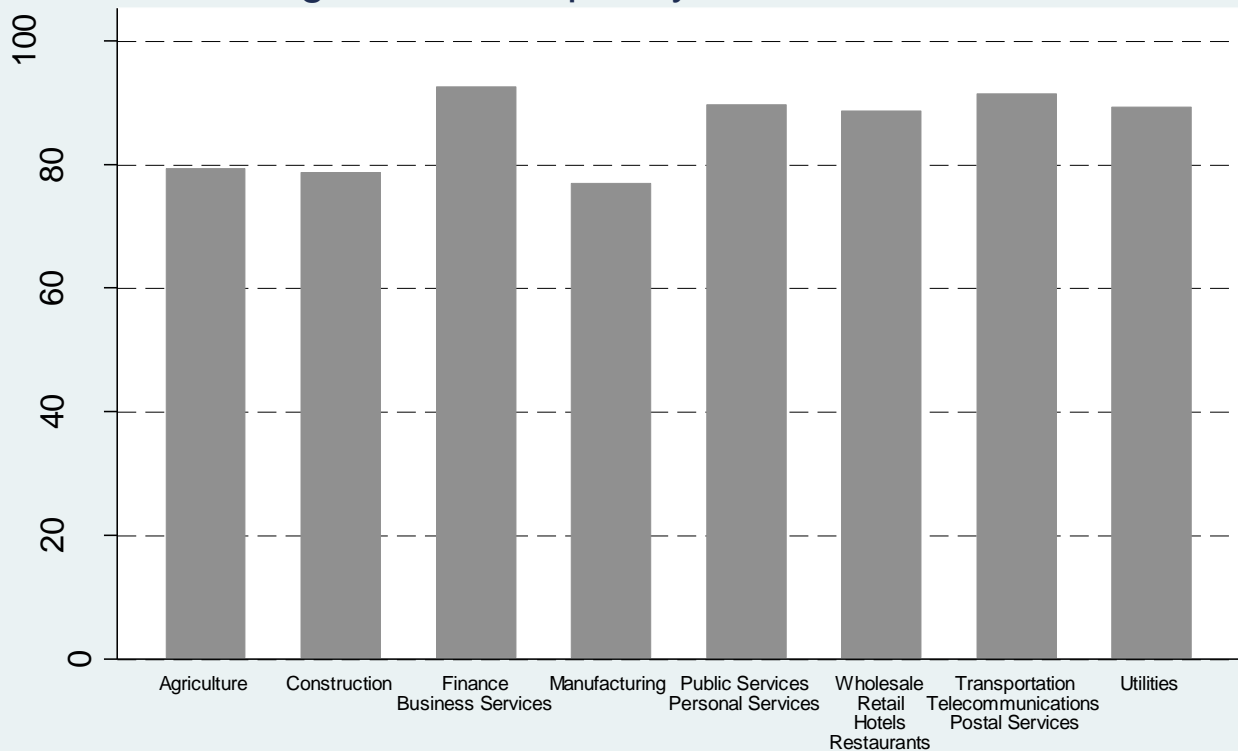
The concern with composition effects is not limited to effects that vary with tenure. For instance, there could be effects across occupations and industries as well. In occupations and industries with flexible hours or in which demand fluctuates a lot, it is likely that a higher share of the workers will sometimes be outside of the top-bracket even if they are full-timers. They just compensate by working over-time in other pay cycles. As a result, they will incorrectly be classified as part-timers even if they work full-time on average, and have too high wages in the data. Figures 3.4 and 3.5 check this hypothesis. In Figure 3.4 we show the share of the workforce with u_{min}/u -ratio in the interval $[0.8,1)$ across industries, while in Figure 3.5 we show the fraction in each industry with u_{min}/u -ratio equal to one. In both figures, we condition on job spells that are into their second year or above, to avoid contamination from over-reporting. The coding error has been fixed as well.

Figure 3.4: Frequency of u_{min}/u -Ratio in $[0.8,1)$



Note: Personal Services cover: Entertainment, Culture, Sports, and Other Services.
Source: Authors' calculations in the Danish IDA database and Dansk Branchekode 2003.

Figure 3.5: Frequency of u_{min}/u -Ratio = 1

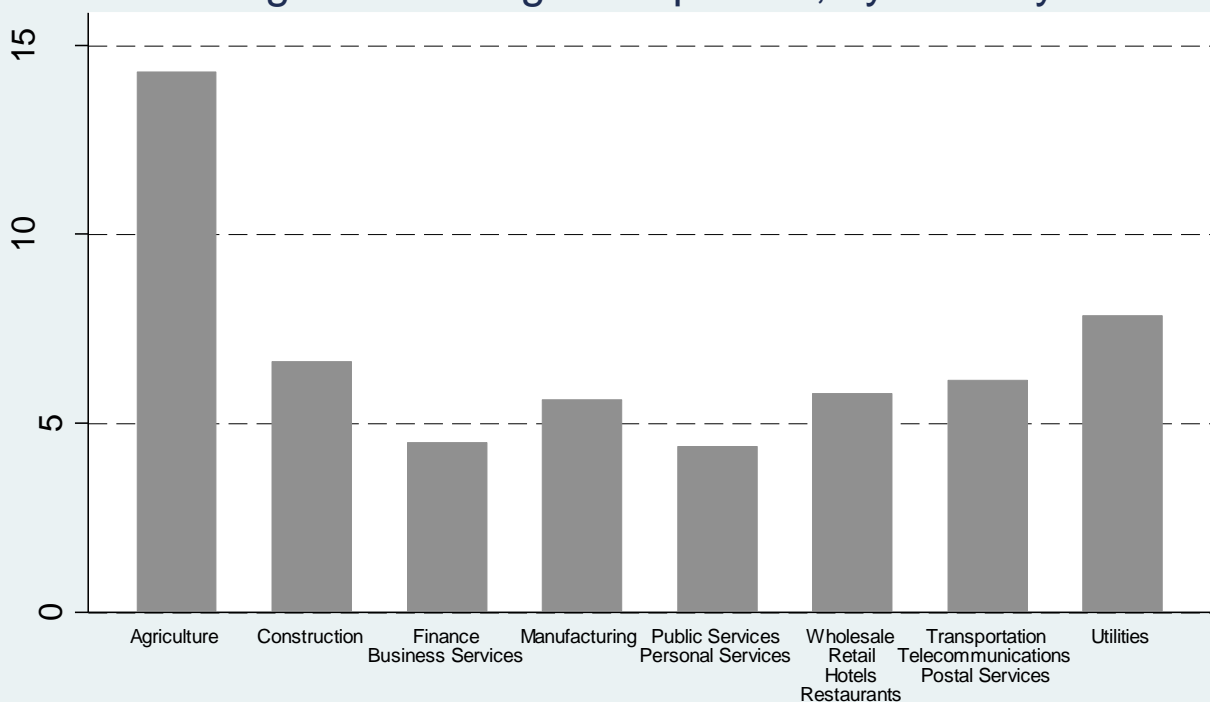


Note: Personal Services cover: Entertainment, Culture, Sports, and Other Services.
 Source: Authors' calculations in the Danish IDA database and Dansk Branchekode 2003.

The evidence here seems squarely in accordance with the hypothesis. Agriculture, Construction, and Manufacturing are all known to be very dependent on the season or on market conditions. Both Construction and Manufacturing are highly demand-driven and agricultural activity is concentrated in parts of the year. Figure 3.4 shows that exactly in these industries the shares of workers close to always being in the top-bracket are very high. Second, Utilities, Telecommunications and Postal Services, and Finance and other services have low shares of part-timers who are close to always being in the top-bracket. These are industries with supposedly more stable hours. Figure 3.5 confirms the impression by essentially showing the opposite pattern for the share of workers always in the top-bracket. The figure shows that the differences between the sectors in Figure 3.4 are due to different shares in the top-bracket, and not to different shares in the interval $(0,0.8)$. In terms of the impacts on wages, recall that a full-timer with u_{min}/u -ratio

just below one will have too high wages. Thus, the IDA wage measure should be biased across industries with too high wages in industries prone to demand-shocks or seasonality. We test the hypothesis in a careful comparison with a wage benchmark that we constructed in Chapter 1 from Statistics Denmark’s wage survey “Lønstatistik”. As described there, we matched the observations in the survey with the ones in our population from IDA, picked the components of the survey that we believed would match the hourly wage in our data, constructed a new wage measure (named “Survey wage”) from the components, and trimmed away 0.5% of the tails. Here we compute the difference by industry between our replication of the IDA wage and the new Survey wage, conditioning on job spells that have lasted at least one year. The results are as follows.

Figure 3.6: Wage Comparison, by Industry



Note: Each bar shows the difference between the IDA wage and the Survey wage.
 Personal Services cover: Entertainment, Culture, Sports, and Other Services.
 Source: Authors' calculations of wages in the Danish IDA database and the Lønstatistik, Dansk Branchekode 2003.

As appears clearly in the figure, the estimated wage in Agriculture is much higher than the Survey wage, and the difference is higher than in other sectors. The difference in Construction is the third-highest (after Utilities), but disregarding Agriculture and Construction the Manufacturing sector is only in the middle of the pack. The pattern is roughly but not completely in accordance with the hypothesis. Of course, the components of the Survey or IDA wage could be systematically biased across industries so the evidence in Figure 3.6 should be used with caution.

One could ask if similar composition biases appear over time. For instance, working hours without a fixed schedule and other flexible work arrangements have become more and more popular over time. Nowadays a larger share of the workforce could sometimes fall out of the top-bracket in the data without really being part-timers, simply because their hours vary more from pay cycle to pay cycle. However, in preliminary examinations of this question we did not find any supportive evidence. The reason could be that most workers are paid monthly, and any variations are likely to happen more from week to week. It is not so likely that a worker works more than 10 hours less per week in all 4 1/3 weeks in a month, and then compensates by working more than 10 hours more per week in the next, just because of a flexible schedule.

Another possible effect is that mistakes are made real-time at the workplace. In some occupations hours are not very carefully monitored, and some firms may not know how to properly count the hours of an employee when she is off work for various reasons. There are several types of absence that count as working hours, ranging from vacation for monthly paid employees, days off when the child is sick, paid maternity leave, to days off for further education. The somewhat complex system for counting hours increases the risk of mistakes. A higher variance and incorrectly reported hours would channel too many workers into the part-timer group, perhaps systematically across occupations. Unfortunately we do not have evidence in favour of or against this conjecture.

Notice that in all of these examples, the problem is always that the true full-timers are pushed out of the top-bracket, and that we observe too high part-timer and aggregate wages as a result. The opposite mistake is unlikely to happen since the criteria for being a full-timer are so strict. Thus, the effects of a distorted composition of part-timers and full-timers just go in that one direction. To find evidence for the hypothesis, it suffices to go back to Puzzle 3 and Figure 1.4a-c in Chapter 1 that unambiguously show how the IDA hourly wage is everywhere higher than the Survey wage. We will see below in Figures 3.9 and 3.10 how the solution to Puzzle 4 (the tenure profile puzzle) also significantly reduces the differences in the aggregate and part-timer wage levels.

3 Improving the Way the Bounds are Used to Estimate Hours

The significant impact on the tenure profile of the mis-classification of full-time workers as well as the potential for harm to other wage statistics calls for a major revision to the notions of part-timers and full-timers. The goal here is to find an estimation method that not only works well for the subgroups of workers who fit perfectly into the categories of part-timers and full-timers, but is also appealing for everybody else in between. At the same time, the new method should be reliable and easy to implement, and produce a realistic tenure profile. Finally, in dealing with the over-reporting we would like it to not rule out the issues of attrition and true transitions between part-time and full-time work.

The starting point is that a part-timer/full-timer designation is only a tool, not an end in itself. If we can estimate the hours in a better way without such a classification, there is no need to use one. Second, any classification of workers into a discrete number of groups seems to suffer from the fact that many workers will never fit perfectly in any category and some workers that naturally belong to one group will mistakenly be placed in another. As a result, unavoidable trade-offs will emerge whereby a stricter allocation of workers to one group will make it more

representable but also increase the mistakes made in other groups. For instance, Statistics Denmark has chosen to prioritize the full-timers by using a very strict criterion for that group (Chapter 1, Table 1). As a result, they are very homogenous and should give an accurate picture of the true full-timers. However, the group of part-timers becomes a melting pot of wrongly allocated full-timers, true part-timers, and workers in between, with implications for tenure profiles, aggregate wages etc.

We propose a continuum of groups instead of just two. Our refinement describes workers according to *how much* of a full-timer or part-timer they are. For example, the workers that are full-time according to the current designation are for us full-timers to *the maximum extent*. The idea is then to estimate hours in a way that changes continuously with the degree to which a worker is a full-timer. Since the current method of estimating the hours of the full-timers seems to perform quite well, we calibrate the new method so that it yields the same results for workers with the highest possible degree of full-time work. Similarly, to the left on the continuum we also do not want to deviate too much from the current method when calculating the wages of the part-timers. In between though, we let workers accumulate hours according to both of the current methods, with weights that reflect how often the worker is in the top-bracket. As a result, the workers who are most often in the top-bracket and have the most biased wages will also receive the biggest wage corrections.

Of course, this program can only be carried out if it is possible to estimate how frequent a worker is in the top-bracket. Luckily, that problem was solved in Chapter 2 and here we take advantage of the method developed there. The basic insight was that the ratio between u_{min} and u determined both the lower and upper bounds on the fraction of pay cycles in which the worker was in and outside of the top-bracket. u_{top}/u was the lower bound on the share in the top-bracket, given by

$$\frac{u_{top}}{u} = \left(3 \frac{u_{min}}{u} - 2\right) 1 \left[\frac{u_{min}}{u} \geq \frac{2}{3} \right].$$

u_{min}/u itself was the upper bound. Thus, the key innovation is to compute hours as a linear combination of $1/2(t_{min} + t_{max})$ and t_{max} where the weights are functions of the ratio between u_{min} and u . Specifically, we compute hours as

$$\left(1 - f\left(\frac{u_{min}}{u}\right)\right)\frac{1}{2}(t_{min} + t_{max}) + f\left(\frac{u_{min}}{u}\right)t_{max} = t_{min} + \frac{1}{2}\left(f\left(\frac{u_{min}}{u}\right) + 1\right)ku$$

where $f' \geq 0$, $f \in [0,1]$ are minimum requirements on the function f . Thus, f is chosen so that hours equal $1/2(t_{min} + t_{max})$ or t_{max} or convex combinations of them.

This method may seem more ad hoc than it actually is. In fact, if overtime is allowed for in the top-bracket, it is micro-founded. To see this, recall that k is the maximum number of hours a worker can work without earning extra pensions. Then rewrite our new expression of hours as

$$\frac{1}{2}(t_{min} + \tilde{t}_{max}),$$

$$\tilde{t}_{max} = t_{min} + f2ku + (1 - f)ku.$$

In other words, we compute hours as the midpoint of the interval spanned by the lower bound on hours and an upper bound that reflects that the worker cannot exceed k extra hours without earning extra pensions when she is outside of the top-bracket, but can work up to $2k$ extra hours when in the top-bracket. Now, in the top-bracket, there really is no way of increasing one's pensions, and the extra hours one can work reflects an artificially imposed ceiling on the number of hours. Recall that the limit on hours in the top-bracket imposed by Statistics Denmark corresponded to full-time work and was reached when a worker worked the $1 \cdot k$ extra hours. Hence, compared to Statistics Denmark the additional $(2 - 1) \cdot k$ hours that we allow for can simply be interpreted as allowing for k hours of overtime.

In turn, the k hours of over-time are not just a convenient change in the upper bound on the number of hours in the top-bracket. Our improvement does not hinge on an

arbitrary choice of the maximum hours in a pay cycle. Instead, our solution preserves the choice of Statistics Denmark and extends it to other workers when it is sensible to do so. To see this point more clearly, first consider a full-time worker. The conceptual difference between the estimation methods is that we do not impose the full-time hours as the upper limit on the number of hours and assign probability one to that limit. Instead, our way of computing hours is simply the expected hours when hours are uniformly distributed over an interval, the midpoint of which is the full-time hours. Thus our method explicitly recognizes that full-time workers do not hit full-time hours right on target in every pay cycle, yet the permissible range of hours in the top-bracket by construction yields the exact same estimate of the number of hours. Then consider the part-timers. What our method does is simply to treat them as full-timers when they are in the top-bracket, and only then.

One might be concerned that some “true” part-timers randomly end up in the top-bracket in some pay cycles, and are incorrectly assigned the same hours in those pay cycles as if they had been full-timers. Perhaps the range of likely hours in the top-bracket for such workers should only extend up to the full-time hours and not to the full-time *plus* over-time hours. Put differently, one might be concerned that our changes go too far. There are two answers to this concern. The first is that, as it happens, good criteria to distinguish “true” part-timers in the top-bracket from other workers exist and are already used by Statistics Denmark. Part-time insurance, part-time unemployment at some point in the year, and reduced hours are deemed sufficiently strong indicators of part-time work that workers always in the top-bracket satisfying these conditions are considered part-timers. We use the same criteria, for workers always in the top-bracket as well as for everybody else, to distinguish “true” part-timers. We set $f = 0$ for the identified workers, that is, we copy Statistics Denmark and do not correct at all when a worker is

deemed a true part-timer.⁹ The second answer is that true part-timers should not be in the top-bracket very often. If some of them are not caught by the three criteria, at least the changes that we make are small since they only apply in the rare event that the workers actually are in the top-bracket. The resulting estimates of hours should not deviate significantly from the ones of Statistics Denmark.

There are two main groups of workers whose hours are corrected, the “true” full-timers who for one reason or another are outside of the top-bracket only in few periods, and the workers who are both a bit of full-timer and part-timer. Our method does what seems to be the most sensible thing to do for these workers, which is to assign them the hours of full-timers in the pay cycles in which they actually are in the top-bracket, and only in those pay cycles. In this way, we remove the imposed constraint that there must be two discrete groups with hours calculated in their own ways. There are no one-size-fits-all and subsequent problems of finding good criteria for separating workers into artificial groups. It would seem that the only reason for why ours is not already the standard method is the fact that Statistics Denmark has not known of the method to distinguish pay cycles in and out of the top-bracket that was developed in Chapter 2. Indeed, without that method one is forced to make the same choices as Statistics Denmark and select k as the number of hours one can work without earning extra pensions, for everyone.

The challenge of course is to gauge the share of pay cycles in the top-bracket. In principle the share could be any convex combination of u_{min}/u and u_{top}/u , depending on how conservative one wishes the estimation of it to be. The most cautious estimate is given by

$$f\left(\frac{u_{min}}{u}\right) = \frac{u_{top}}{u} 1[G],$$

⁹ Workers with reduced hours are assigned t_{min} by Statistics Denmark but we assign them $\frac{1}{2}(t_{min} + t_{max})$. What is meant is that we do not assign them $\frac{1}{2}(t_{min} + \tilde{t}_{max})$. Arguably $\frac{1}{2}(t_{min} + t_{max})$ is as realistic as t_{min} , but it does not make any difference anyway since only 5.4% of part-timers have reduced hours.

where $1[G]$ is an indicator function that equals one if a worker is not part-time insured, is not part-time unemployed in the year, and does not have reduced hours due to second jobs. The factor u_{top}/u guarantees that changes are only made when the worker for sure was in the top-bracket, since it is the absolute lower bound on the share of pay cycles in the top-bracket. In reality, workers will be in the top-bracket more often than captured by f . Compared to the estimation method of Statistics Denmark, this choice of f in effect makes no additional assumptions except that some part-timers can sometimes work full-time hours. To us, the lack of extra assumptions is a significant plus and we consider this way of choosing f a credible improvement over the current estimation method. We refer to it as the “Conservative” estimation method, one of two methods that will be presented in this chapter. The final formula is,

$$t_0^* = t_{min} + \frac{1}{2} \left(\frac{u_{top}}{u} 1[G] + 1 \right) ku.$$

Even if the Conservative method offers an improvement over the current estimation method, a lingering concern is that it is too cautious. To see why, consider the following alternative interpretation of the bounds on the fraction of pay cycles in the top-bracket. Consider first the lower bound. It was obtained (Chapter 2) by counting the maximum amount of pensions earned outside of the top-bracket and comparing this maximum with the total accumulated pensions. When $u_{top}/u > 0$, that maximum was attained if a worker worked in the middle-bracket in all other pay cycles than those in the top-bracket. It follows that, if $u_{top}/u > 0$, the actual fraction of pay cycles in the top-bracket equals u_{top}/u if and only if there is maximal smoothing of working hours across pay cycles. Then consider the upper bound. A higher fraction of pay cycles in the top-bracket than u_{min}/u was impossible because then the earned pensions would exceed the amount recorded in the data. A fraction equal to u_{min}/u meant that all work had been lumped together in a few intense pay cycles. It follows that the upper bound corresponds to minimal smoothing of working hours across pay cycles. Now, we know that over-

reporting of the employment periods of full-timers in the first year of employment is a problem. We also know that no work is done between the incorrect and correct commencement dates. In other words, smoothing is minimal when there is over-reporting, and the upper bound should be used in these cases. In effect, the upper bound and not the conservative lower bound is the appropriate one.

The bolder estimation method that we propose has the Conservative method as the baseline, but in the first year of all employment spells it uses a weighted average of u_{min}/u and u_{top}/u for the fraction of pay cycles in the top-bracket, instead of just u_{top}/u . The weight on the upper bound u_{min}/u , the bound corresponding to minimal smoothing, is an estimate of the probability of being an over-reported full-timer. We compute these probabilities conditionally on the observed ratios u_{min}/u , in this context interpreted as the degree to which a worker is a full-timer. Notice that this bolder approach to estimating the wages is, in fact, still somewhat cautious, since the residual weight is put on the lower bound u_{top}/u in the first year, and in all other years full weight is given to u_{top}/u .

The idea behind the estimation of the probabilities is to find a set of workers who are thought to work full-time in their first year of employment and observe their reported u_{min}/u -distribution. The conditional probabilities of reporting u_{min}/u given full-time work are obtained in this way. Then Bayes' Rule can be used to infer the probabilities of full-time work given a reported value of u_{min}/u . The trick to identify the set of true full-time workers is to observe what happens in the second year of employment in which the spell lengths are very likely to be correctly reported, and then assess the fraction of this group who also work full-time in the first year. The specific notation and assumptions are as follows.

Notation.

- If no subscript the year is the year of hiring. Otherwise the subscript indicates the year.

- x is the observed u_{min}/u -ratio. x^* is the true u_{min}/u -ratio. x_2^* is the true ratio in the second year of the job spell, etc.
- f is the (inverse of the) factor of over-reporting, defined as $f = x/x^*$.
- $T = \{SL > 1, x_2^* = 1\}$ is the set of observations whose job spells will last longer than one year and whose true u_{min}/u -ratio in the second year equals 1. The complement of T is $T^c = \{SL = 1 \text{ or } (x_2^* < 1 \text{ and } SL > 1)\}$.
- $T' = \{SL > 2, x_3^* = 1\}$, i.e., the set T shifted forward one period.

Assumptions.

- (i) The spell length is correctly reported in year 2 and 3, $x_2^* = x_2$ and $x_3^* = x_3$.
- (ii) The share of full-timers among the workers continuing full-time in the next period does not vary with time, $P(x^* = 1|T) = P(x_2^* = 1|T')$.
- (iii) In the T^c -group, the share of true full-timers is larger than observed, i.e., $P(x^* = 1|T^c) \geq P(x = 1|T^c)$.
- (iv) The future of a newly hired worker as captured by T and T^c does not matter for the distribution of factor errors once the effect of the future on current work intensity is accounted for, i.e., $P(f = a|x^* = b, T) = P(f = a|x^* = b, T^c)$.
- (v) We assume that $P(x = a|x^* = 1, T) = P(x = a|(obs. char. = 1), T)$ where $(obs. char. = 1)$ means that all of the following are satisfied:
 - a. Age between (and including) 35 and 55 in the year of estimation.
 - b. Male.
 - c. Full-time insured in the year of estimation.
 - d. No period of part-time unemployment in the year of estimation.
 - e. Employed in all of the past 4 years.

- f. Within the past 4 years, whenever a year is not the year of hiring, the worker must have $u_{min} = u$, i.e. $x_j^* = 1$ where j is the year of employment in the spell and $j > 1$.
- g. At least 2 years of the past 4 must be continuing years, i.e. $j > 1$ happens in at least 2 instances within the past 4 years.

As argued elsewhere, Assumption (i) is likely to hold in practice. So is Assumption (iii), simply because over-reporting is more likely than under-reporting since “The Entire Year” is an option for employers and “None of the Year” is not, and because any over-reporting will make a full-timer appear as a part-timer whereas even with under-reporting a part-timer may still appear as a part-timer. We also believe that Assumption (iv) is likely to hold. Its rationale is that work intensity obviously impacts the distribution since the possible extent of under-reporting is a function of x^* , and a worker’s future is related to current work intensity. But beyond the effect that works through work intensity, there should be no relation between the future and current patterns of reporting. Assumption (v) is also credible. Together, the conditions a.-g. are chosen so as to isolate full-time workers and we believe that the conditions do that quite well in the sense that not many part-timers will remain after conditioning on a.-g. Obviously many full-timers will not be contained in this set, but there should be no reason to believe that the full-timers identified with the conditions a.-g. are systematically different from the full-timers that are left out in terms of the probability of over-reporting. Whether Assumption (ii) holds in practice is unknown, but it cannot be too far off, and in any case it is the best we can do.

Observables.

Assumption (i) guarantees that x_2^*, x_3^*, T, T' , and T^c are observable. In addition, we observe $P(x|T^c)$ and $P(x|T)$, the distributions of observed u_{min}/u -ratios, and the spell length SL .

Step 1.

With our notation and assumptions, our interest lies in $P(x^* = 1|x = a)$ for $a < 1$. (When $a = 1$, u_{min}/u equals u_{top}/u .) We do not point-identify the conditional probabilities since it would require stronger assumptions than we are willing to make. Instead we stay cautious and obtain a lower bound of the conditional probability. In this way, we stick to the conservative approach and refrain from making changes when we do not feel we can justify them. Rewrite

$$\begin{aligned}
P(x^* = 1|x = a) &= P(x^* = 1|x = a, T^c)P(T^c|x = a) + P(x^* = 1|x = a, T)P(T|x = a) \\
&= P(x = a|x^* = 1, T^c) \frac{P(x^* = 1, T^c)}{P(x = a, T^c)} P(T^c|x = a) + P(x^* = 1|x = a, T)P(T|x = a) \\
&= P(x = a|x^* = 1, T) \frac{P(x^* = 1, T^c)}{P(x = a, T^c)} P(T^c|x = a) + P(x^* = 1|x = a, T)P(T|x = a) \\
&= P(x = a, x^* = 1|T) \frac{P(T)}{P(x^* = 1, T)} \frac{P(x^* = 1, T^c)}{P(x = a, T^c)} P(T^c|x = a) \\
&\quad + P(x = a, x^* = 1|T) \frac{P(T)}{P(x = a, T)} P(T|x = a) \\
&\geq P(x = a, x^* = 1|T) \left(\frac{P(x = 1, T^c)}{P(x^* = 1|T)P(x = a)} + \frac{P(T)}{P(x = a)} \right) \\
&= P(x = a|(obs. char. = 1), T)P(x_2 = 1|T') \cdot \\
&\quad \left(\frac{P(x = 1, T^c)}{P(x_2 = 1|T')P(x = a)} + \frac{P(T)}{P(x = a)} \right).
\end{aligned}$$

We estimate all the probabilities by industry. The reason is that different industries might have different mobility patterns, meaning that in some industries the tendency for workers to separate already after the first year could be higher than in other industries. Such mobility would for instance reduce the probability $P(T|x = a)$. We estimate the probability $P(x = a|(obs. char. = 1), T)$ only for the years 1990-2007 since it takes five years to identify four years of work history and our sample begins in 1985.

Step 2.

For all workers in their first year, the obtained values for the probabilities $P(x^* = 1|x = a)$ are used to quantify f in the following way:

$$f\left(\frac{u_{min}}{u}\right) = \left(P(x^* = 1|x = a)\frac{u_{min}}{u} + (1 - P(x^* = 1|x = a))\frac{u_{top}}{u}\right) 1[G],$$

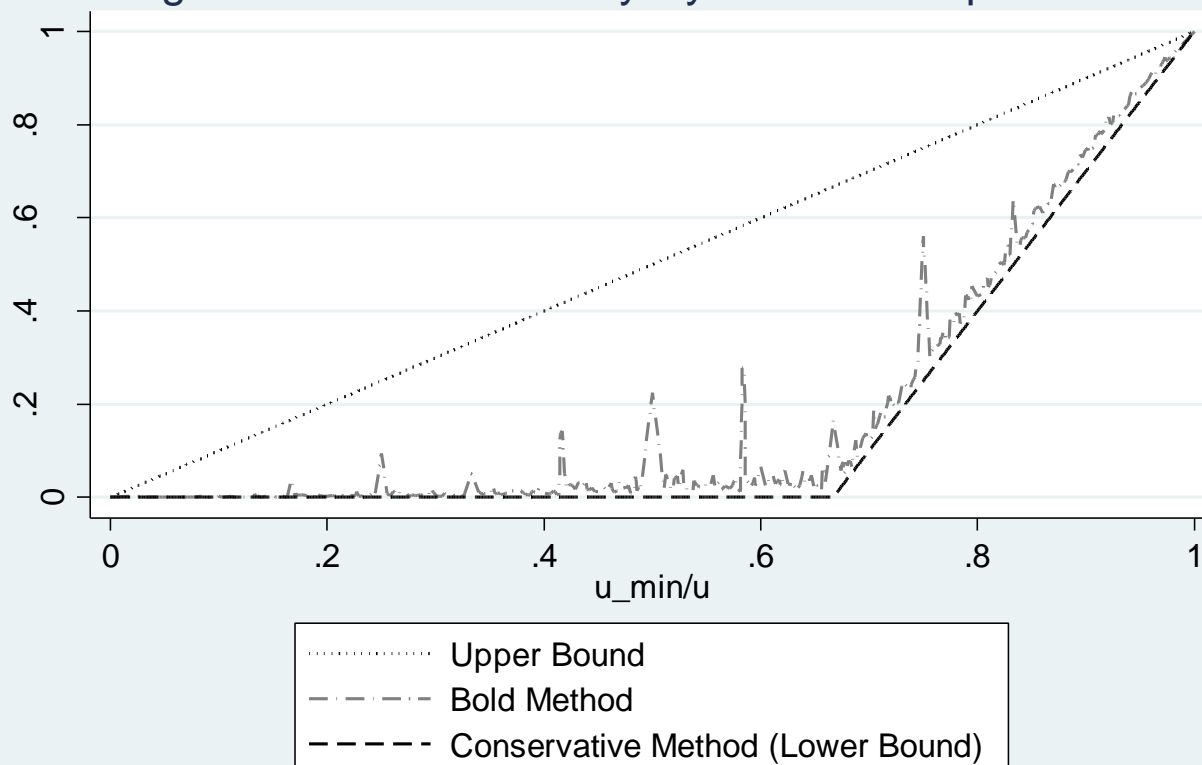
$$t_1^* = t_{min} + \frac{1}{2} \left(\left(P(x^* = 1|x = a)\frac{u_{min}}{u} + (1 - P(x^* = 1|x = a))\frac{u_{top}}{u}\right) 1[G] + 1 \right) ku.$$

For workers in their second year and above,

$$f = \frac{u_{top}}{u} 1[G].$$

With the assumptions made, this way of estimating the hours is only a bit less cautious than the Conservative method. While it cannot be completely ruled out that it over-corrects the IDA hour estimates, it is still very unlikely. Firstly because the obtained weight on minimal smoothing u_{min}/u is just a lower bound on the “true” weight, and secondly because the residual weight is put on maximal smoothing u_{top}/u when in fact in the residual cases smoothing could be anything between minimal and maximal. Therefore, if the assumptions are deemed acceptable, the Bold method (as we call it to distinguish it from the Conservative) is potentially much better than the Conservative method. The difference between the Conservative and Bold methods is nicely summed up for new hires in the following Figure 3.7.

Figure 3.7: Fraction of Pay Cycles in the Top-Bracket



Source: Authors' calculations of wages in the Danish IDA database.

The figure shows the span of the possible fractions of pay cycles in the top-bracket, bounded by the Conservative method and the upper bound u_{min}/u . The estimated fraction of pay cycles in the top-bracket turns out to be very close to the Conservative method, except in nine roughly equidistant spikes. At these spikes, there is a high probability of being an over-reported full-timer. One might wonder why exactly the over-reporting is concentrated at 9 equidistant points. The explanation is simple. It is time consuming for employers to keep track of the exact commencement dates or to look them up at the time of reporting if they were noted down. It is much easier to just tick off the “entire year”-option. Consider then a full-timer who began work on March 1st and was reported to have begun on January 1st. Instead of $x^* = 1$ we will observe $x = u_{min}/u = u_{min}^*/u = u^*/u = 10/12 = 0.833$. There is a spike in the figure right at this point. Similarly, if the worker began on April 1st, the observed x is 0.75 where there again is a spike. In

fact, all the spikes are right at the points $2/12$, $3/12$, ... $9/12$, $10/12$. There is no spike at $1/12$ or $11/12$, possibly because for these values of x the upper and lower bounds are not that far from each other and probably also because it is easy to remember for employers if an employee was hired just one month ago ($1/12$). Other reasons could play a role as well. On a related note, one might wonder why the spike at 0.5 is so big. The reason could be that the workers who are hired in the middle of the summer (and who stay for the rest of the year) are more likely to be full-timers compared to those hired at other times. Not many workers are hired in the summer, but those who are also skip their vacation and could very well be intent on getting some work done. In addition, part-time work is often done by students who either are away for the summer or quit their jobs when the summer ends. Either way, they are not both hired in the summer and still in the job come November.

Apart from the spikes, the Conservative and Bold methods are very similar. With both methods, the share of the pay cycles with changes to the IDA hours rises rapidly after $u_{min}/u = 2/3$ until it reaches all pay cycles for full-timers. Thus, in the limit both yield hours that are identical to those of the full-timers with the current method. Similarly, to the left on the continuum both yield hours equal to what is obtained with the current method for workers who are “enough” part-timers. Given that the Conservative and Bold methods are so similar, we should not expect their tenure profiles to exhibit big differences, cf. Figure 3.8 below.

Using Standard Variables to Identify and Estimate the Bounds

As was the case with the corrections to the bounds on hours proposed in Chapter 2, also the changes in this chapter to how the bounds are used are easily implemented with access to only standard variables. The identification is exactly as Section 5 of Chapter 2. But compared with the corrections in that chapter, the task is harder here since we not only would like to exclude part-timers with reduced hours and workers in second jobs but also the part-timers who are part-time

insured or part-time unemployed at some point during the year. These were the worker groups with $1[G] = 0$. Unfortunately it is not possible to distinguish these groups precisely using standard variables, so mistakes will invariably be made. Yet the number of mistakes is reducible. For instance, by definition all of the part-timers with $u_{min} = u$ are known to have either reduced hours, a spell of part-time unemployment or are part-time insured. In turn, many of the workers who are always in the top-bracket can be identified with the standard variable *jobatp*. If $jobatp \geq 1166$, then $ATP \geq A$ and for sure $u_{min} = u$. Moreover, some of the remaining observations with reduced hours, part-time unemployment, or part-time insurance can be identified with *tlonkval*, the variable proxying for the relative uncertainty. The part-timers in main jobs have their relative uncertainty computed in two ways, depending on whether their hours are reduced or not. There is a big overlap between the ranges of values that each relative uncertainty produces (see Section 2 of Chapter 1), but some of the lowest values obtained with reduced hours are not possible outcomes without reduced hours. It is easy to exclude these observations by computing the lower bound on the range of values without reduced hours. Furthermore, we know that the relative uncertainty decreases in u_{min}/u . We also know the rough percentages of part-timers who are always in the top-bracket (Table 1 of Chapter 1). Most of them do not have reduced hours. If it is the case that there is a discrete jump of roughly that percentage in the distribution of *tlonkval* values and this jump happens around the lower *tlonkval*-bound of the part-timers without reduced hours, then we are quite confident that all of these observations are part-time insured or part-time unemployed at some point in the year. In this way, a good portion of the part-timers always in the top-bracket and of the workers with reduced hours can be identified and exempted from the correction. In practice, in order to select the proper levels of *tlonkval* in each year, we do not stick firmly to the computed values of the relative uncertainty since *tlonval* deviates from them because of rounding. Instead we also use the cumulative distributions of *tlonkval* in each year to judge. The exact thresholds and the SAS code can be

found in the Appendix. Note that the SAS code also incorporates the corrections to the bounds in Chapter 2.

4 Performance of the New Methods of Estimating Hours

We now turn to investigating the performances of the suggested solutions to Puzzle 4. The first figure below shows a replication of Figure 1.5 in Chapter 1. As in that figure, we include the tenure profile of the Survey wage for illustrative purposes. Again it should be kept in mind that the sample of the Survey wage is very different. For example, government-run firms and big firms are over-represented.

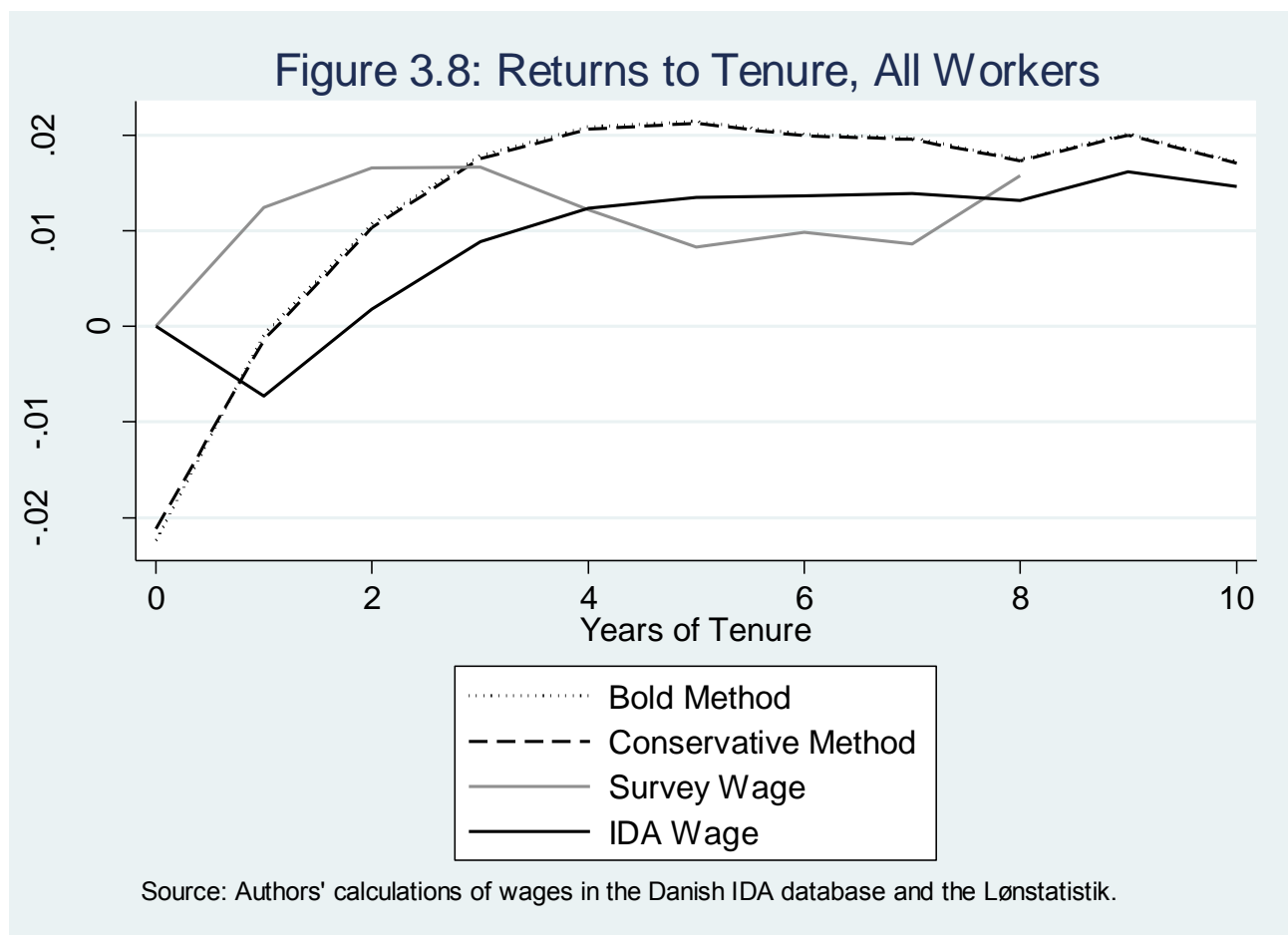
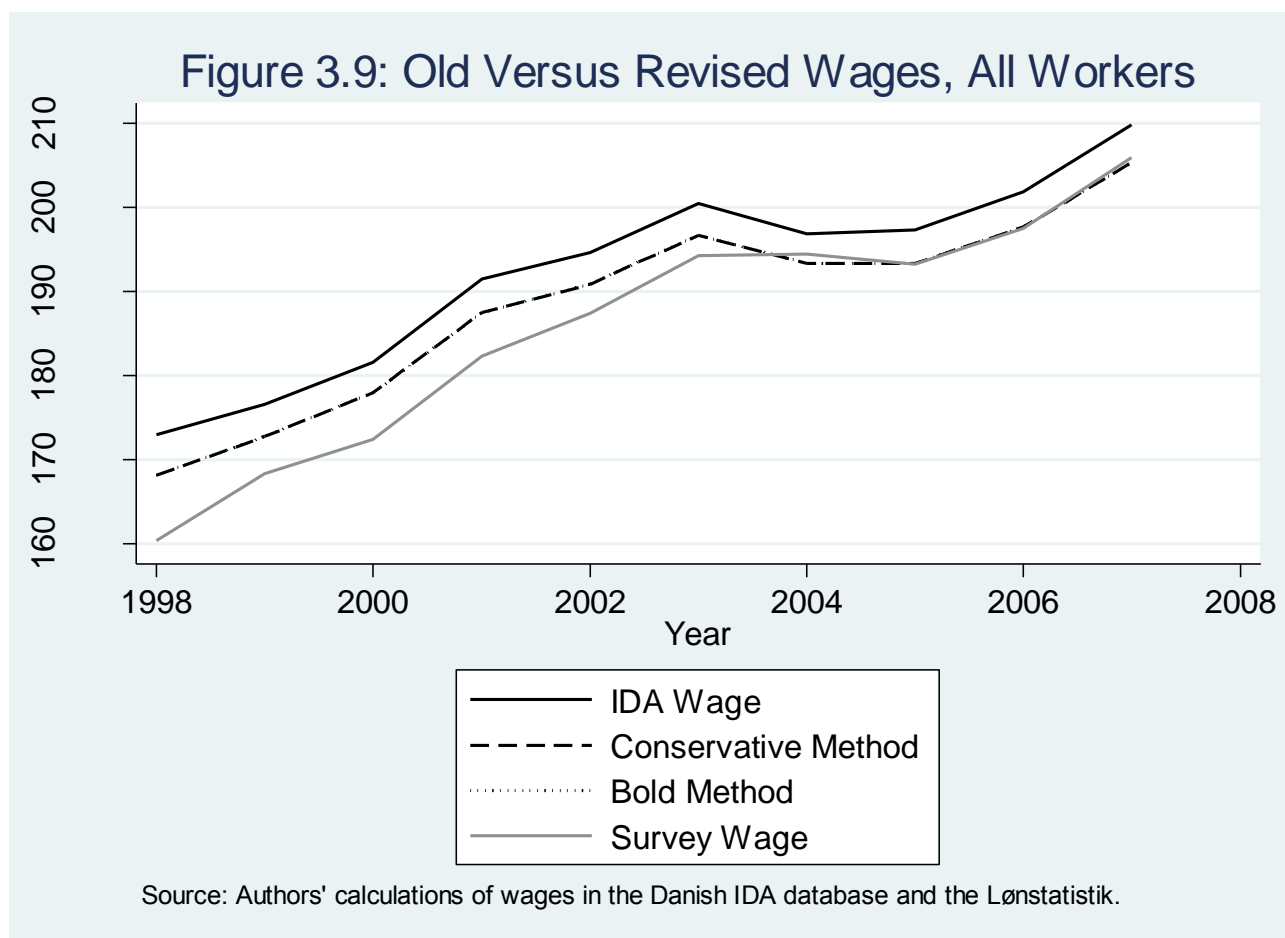


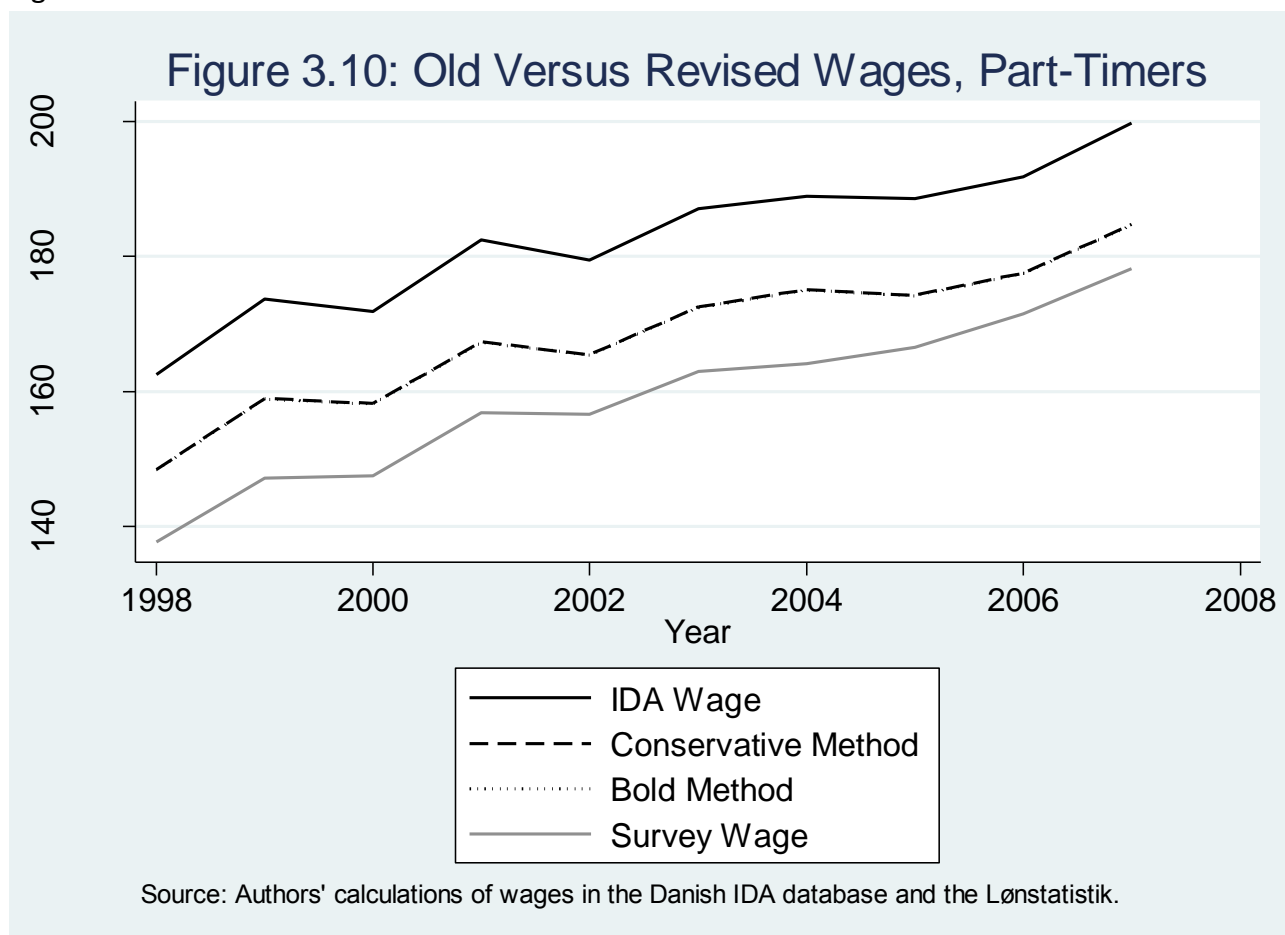
Figure 3.8 shows that our proposed solutions completely solve Puzzle 4. There is nothing left to be desired. The Bold method produces a slightly steeper path from the first year (0) to the second (1), but the difference is very small.

In Figure 3.9 we consider Puzzle 3 and show how much closer the wages generated by the Conservative and Bold methods are to the Survey wage compared to in Figure 1.4a. Recall that the sample used in the figure is the matched sample as in Figures 1.4a-c.



The Bold and Conservative methods are here identical, and both offer a significant improvement over the IDA wage. In effect, after 2003 the difference between the Survey wage and the IDA wage has completely disappeared. Given that the sample improves over the years due to better firm identifiers (see Chapter 1), perhaps we should not even expect a zero difference in the early years.

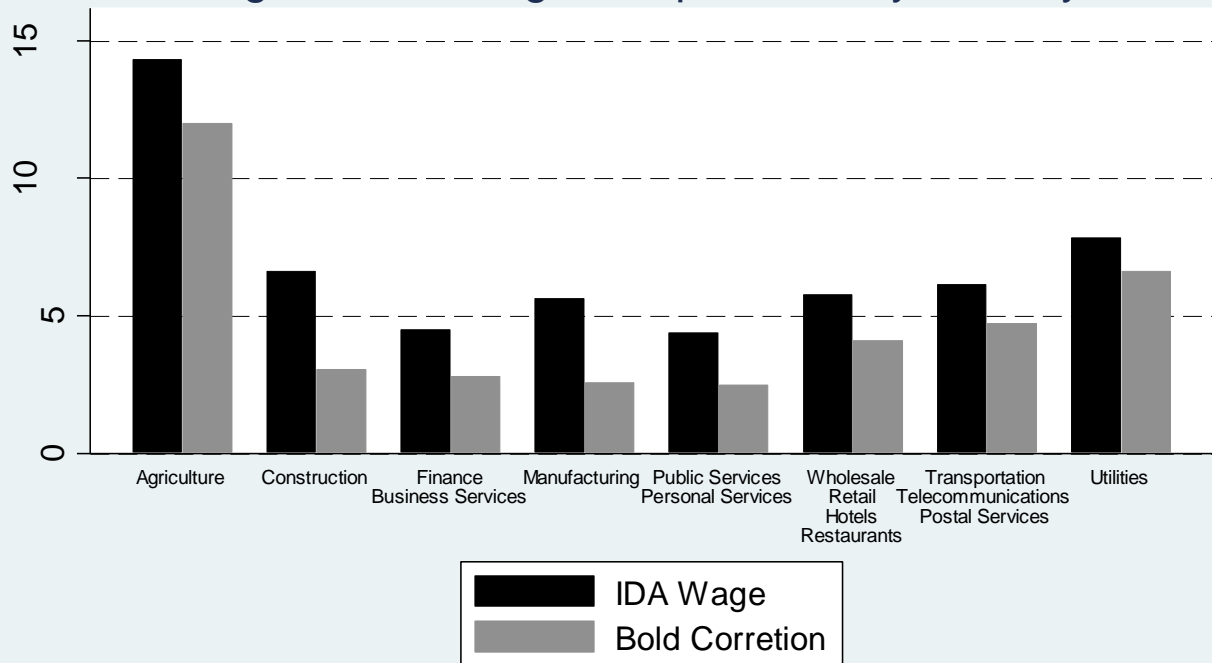
The improvement must stem from the part-timers since the wages of the full-timers are un-changed. In Figure 3.10 we show the improvement in the part-timer wages by replicating Figure 1.4c.



The improvement is significant, even if it only goes half the way. It does not matter what method we consider since they perform identically in practice, as they did in Figures 3.7-3.9. In terms of which of the methods is best, it follows from the similarity that the assumptions behind the Bold method cannot be too strong since otherwise the methods would not yield identical wages. As a result, we prefer the Bold method since it addresses over-reporting conceptually even if the magnitudes are small in practice.

In Figure 3.11 we replicate Figure 3.6 and show how the wage biases change across industries. We use the Bold method.

Figure 3.11: Wage Comparisons, by Industry



Note: Each bar shows the numerical difference between the IDA and Survey wages or between the Bold Correction and the Survey wage.
 Personal Services cover: Entertainment, Culture, Sports, and Other Services.
 Source: Authors' calculations of wages in the Danish IDA database and the Lønstatistik, Dansk Branchekode 2003.

As is seen, the biggest bias reductions are found in Manufacturing and Construction, entirely as expected if the biases were due to some workers being attributed hours in the middle of the possible range in all pay cycles instead of sometimes in the middle and sometimes at the right end-point of the range. However, against our expectations the bias in Agriculture by and large seems to persist. This fact indicates that the IDA and Survey wage concepts are not identical, with some component differing systematically in some industries. A similar finding was reported after Figure 3.6.

5 Combining With the Corrections from Chapter 2

Having solved Puzzle 1 in Chapter 2 and Puzzle 4 and (some of) 3 in the present chapter, we turn to combining all the proposed changes. Rewriting the expressions for t_0^* and t_1^* in a way that bridges the gap to Chapter 2, we obtain the following Conservative hours measure *for all workers*:

$$t_0^* = t_{min} + \frac{1}{2} \left(\frac{u_{top}}{u} (1 + 1[G]) + \left(1 - \frac{u_{top}}{u} \right) \cdot 1 \right) ku.$$

The Bold hours measure *for new hires* is the following:

$$t_1^* = t_{min} + \frac{1}{2} \left(\left(P(x^* = 1|x = a) \frac{u_{min}}{u} + (1 - P(x^* = 1|x = a)) \frac{u_{top}}{u} \right) (1 + 1[G]) \right. \\ \left. + \left(1 - \left(P(x^* = 1|x = a) \frac{u_{min}}{u} + (1 - P(x^* = 1|x = a)) \frac{u_{top}}{u} \right) \right) \cdot 1 \right) ku.$$

For workers in their second year or above, $t_1^* = t_0^*$.

Combining with Chapter 2, we obtain with $t_{min}^* = t_{min} \cdot \tau_3 / (h - k)$:

- *Conservative, all workers:*

$$\hat{t}_0^* = t_{min}^* + \frac{1}{2} \left(\frac{u_{top}}{u} (1 + 1[G]) \frac{h - \tau_3}{k} + \left(1 - \frac{u_{min}}{u} \right) \frac{\tau_3 - \tau_2}{k} + \frac{u_{min} - u_{top}}{u} \cdot 1 \right) ku.$$

- *Bold, new hires:*

$$\hat{t}_1^* = t_{min}^* + \frac{1}{2} \left(\left(P(x^* = 1|x = a) \frac{u_{min}}{u} + (1 - P(x^* = 1|x = a)) \frac{u_{top}}{u} \right) (1 + 1[G]) \frac{h - \tau_3}{k} \right. \\ \left. + \left(1 - \frac{u_{min}}{u} \right) \frac{\tau_3 - \tau_2}{k} + \left(\frac{u_{min}}{u} - \left(P(x^* = 1|x = a) \frac{u_{min}}{u} + (1 - P(x^* = 1|x = a)) \frac{u_{top}}{u} \right) \right) \cdot 1 \right) ku.$$

- *Bold, workers in their second year or above:*

$$\hat{t}_1^* = \hat{t}_0^*.$$

These formulas solve both Puzzles 1, 4, and some of 3. \hat{t}_1^* is the recommended one, to be further enhanced in Chapter 4.

Appendix C3

Code Used to Correct the Wages (Including Changes From Chapter 2)

The code is for the Conservative measure, but it can be easily modified using the formula for \hat{t}_1^* in

Section 5 of this chapter.

```
data w_corrections;
set sample;
  if type='H' & (tilknyt not in('01','11','21','31')) then do; /*Main jobs and
part-timers*/
  if jobatp<1166 then do;
    if aar in (1984,1985,1986) & tlonkval>=30 then cor_1=1;
    if aar=1987 & tlonkval>=31 then cor_1=1;
    if aar in(1988,1989,1993,1994,1995,1997,2000,2002,2003,2005)
      & tlonkval>=32 then cor_1=1;
    if aar in(1990,1991,1992,1996,1998,1999,2001,2004,2006) &
      tlonkval>=33 then cor_1=1;
  end;
  if cor_1=. then cor_1=0;
  cor1=1+cor_1;
  drop cor_1;
  r=tlonkval/100-1+sqrt(1+(tlonkval/100)^2);
  hours=joblon/timelon;
  t_min=hours / (1+r/2);
  alpha_u=hours*r/(1+r/2);
  if aar<=1986 then do;
    u_min_share=1/(3*r); /*k/(h-k)=1/3, and r=alpha_u/t_min*/
    indi=(u_min_share>=2/3);
    u_full_share=(3*u_min_share-2)*indi;
    u_small_share=1-u_min_share;
    h_cor_factor=cor1*u_full_share+1*u_small_share+1*(1-u_full_share-
      u_small_share);
    new_alpha_u=h_cor_factor*alpha_u;
    new_hours=t_min+1/2*new_alpha_u;
  end;
  if aar=1987 then do;
    new_t_min=(30/29)*t_min;
    u_min_share=1/r*10/29;
    indi=(u_min_share>=2/3);
    u_full_share=(3*u_min_share-2)*indi;
    u_small_share=1-u_min_share;
    h_cor_factor= cor1*9/10*u_full_share+1*u_small_share+1*(1-
      u_full_share-u_small_share);
    new_alpha_u=h_cor_factor*alpha_u;
    new_hours=new_t_min+1/2*new_alpha_u;
  end;
  if aar=1988 then do;
    new_t_min=(30/28.5)*t_min;
    u_min_share=1/r*10/28.5;
    indi=(u_min_share>=2/3);
    u_full_share=(3*u_min_share-2)*indi;
    u_small_share=1-u_min_share;
    h_cor_factor= cor1*8.5/10*u_full_share+1*u_small_share+1*(1-
      u_full_share-u_small_share);
    new_alpha_u=h_cor_factor*alpha_u;
```



```

    new_hours=new_t_min+1/2*new_alpha_u;
end;
if aar=1989 then do;
    new_t_min=(30/28)*t_min;
    u_min_share=1/r*10/28;
    indi=(u_min_share>=2/3);
    u_full_share=(3*u_min_share-2)*indi;
    u_small_share=1-u_min_share;
    h_cor_factor= cor1*8/10*u_full_share+1*u_small_share+1*(1-
        u_full_share-u_small_share);
    new_alpha_u=h_cor_factor*alpha_u;
    new_hours=new_t_min+1/2*new_alpha_u;
end;
if aar=1990 then do;
    new_t_min=(30/27.5)*t_min;
    u_min_share=1/r*10/27.5;
    indi=(u_min_share>=2/3);
    u_full_share=(3*u_min_share-2)*indi;
    u_small_share=1-u_min_share;
    h_cor_factor= cor1*7.5/10*u_full_share+1*u_small_share+1*(1-
        u_full_share-u_small_share);
    new_alpha_u=h_cor_factor*alpha_u;
    new_hours=new_t_min+1/2*new_alpha_u;
end;
if aar in(1991,1992) then do;
    new_t_min=(30/27)*t_min;
    u_min_share=1/r*10/27;
    indi=(u_min_share>=2/3);
    u_full_share=(3*u_min_share-2)*indi;
    u_small_share=1-u_min_share;
    h_cor_factor= cor1*7/10*u_full_share+1*u_small_share+1*(1-
        u_full_share-u_small_share);
    new_alpha_u=h_cor_factor*alpha_u;
    new_hours=new_t_min+1/2*new_alpha_u;
end;
if aar>=1993 then do;
    u_min_share=1/r*10/27;
    indi=(u_min_share>=2/3);
    u_full_share=(3*u_min_share-2)*indi;
    u_small_share=1-u_min_share;
    h_cor_factor= cor1*u_full_share+9/10*u_small_share+1*(1-
        u_full_share-u_small_share);
    new_alpha_u=h_cor_factor*alpha_u;
    new_hours=t_min+1/2*new_alpha_u;
end;
rename timelon=old_timelon;
new_timelon=joblon/new_hours;
drop hours new_hours;
end;
/*Full-timers: No correction necessary as the corrections cancel out.*/
/*Type='B': I'm not able to correct type Bs;*/
if type^='H' | (tilknyt in('01','11','21','31')) then do;
    new_timelon=timelon;
    rename timelon=old_timelon;
end;
run;

```

Chapter 4: A New Wage Measure

1 Introduction

In the previous three chapters, we put forward several changes to the hourly wage measure in the IDA database. Each change solves a specific puzzle. In the same way, in this chapter, we tackle Puzzle 2 by recommending a change that removes much or all of the fall in the wages of full-timers from 2003 to 2004 that was documented in Chapter 1. However, the main purpose of this chapter is not only to propose new changes but also to combine all the changes made in this and previous chapters into a new wage measure. The hope is that it will emerge as the next default wage measure used by researchers, statisticians, think tanks, perhaps the Danish Economic Councils, and, ultimately, anybody with an interest in wage statistics. In order to minimize the risk of a need for future changes to our new measure and to increase the accuracy of it, we also take an extra careful look at the programming code of the IDA wage. In doing so, we not only identify and correct one error but also implement a few, minor improvements. In this way, we reduce the likelihood of revisions.

In total we propose four changes in this chapter. The first is to get rid of the premature rounding documented in Chapter 1. It really serves no purpose nowadays to round the numbers before arriving at the final result since we have access to sufficient computing power. The second is to scale periods of unemployment and illness with the number of work weeks in the year divided by the number of weeks in the year. The effect of the scaling is to avoid a double-subtraction of vacation when computing the upper bound on calendar weeks with work (see Chapter 1 for more on this concept). The rationale is that part of the unemployment spell would have been spent on vacation if the employee had had a job, and that vacation has already been subtracted and should not be subtracted again. Third, instead of using January hours in the years 1986-1990 (see Chapter 1), we use the more accurate weighted average of the weekly hours.

Finally, and this is the error that we identify, many of the employees who qualify for reduced hours (see Chapter 1) are not assigned the lower bound on hours, and many of the employees who do not qualify in fact are assigned the lower bound.

The remedy to Puzzle 2 that we identify is one of these four changes. Indeed, while the scaling has no effect whatsoever, the imposed rounding in the IDA wage changes the slope from 2003 to 2004 from slightly positive to the negative slope that we see. The incorrect criterion for reduced hours only changes the levels of the average wages, not the time series patterns.

This chapter begins in Section 2 with a description of the final changes that we make and their rationales, then shows in Section 3 the effects of the changes, also on Puzzle 2. In Section 4, we combine all the changes in this and previous chapters into a new wage measure and show how all the puzzles identified in Chapter 1 have disappeared or been much improved upon. Finally, in Section 5 we briefly touch on areas where we think there is room for future improvements that we have not had the resources to undertake.

2 The Final Changes

In order to increase the likelihood of the new wage measure standing the test of time, we scrutinize the programming code for possible improvements and outright mistakes. Here are the results of the examination. Firstly, we prefer not rounding the numbers that are used in intermediate steps in the estimation of the IDA wage (see Chapter 1). The effects are likely to be small but visible for two reasons. Firstly, the rounding is applied to numerators and denominators separately before they are divided, not after. Secondly, the parameter values that enter the numerators and denominators are identical for all workers, so part of the rounding error is not limited to a subgroup of workers but applies to all.

Secondly, we prefer scaling periods of unemployment or sickness with a factor that takes into account that a fraction of the period would have been spent on vacation if the

employee had instead been working. In the IDA database, the information on part-time and full-time unemployment and periods with sickness is based on the actual period of absence, not on how much work could have been carried out in that period. The unemployment and sickness information is used to compute an upper bound on the number of calendar weeks of employment in which there is actual work (see Chapter 1). If z denotes the period of absence in the data, the upper bound should be estimated as

$$(52 - z) \cdot \frac{52 - V}{52} = 52 - V - z \cdot \frac{52 - V}{52},$$

where $V = 5$ is the number of weeks of vacation. However, Statistics Denmark estimates the upper bound as $52 - V - z$, thus subtracting part of the vacation twice.

Thirdly, since the weekly hours fell from 1986-1990 but not on 1st of January, instead of relying on the hours as they were on 1st of January as Statistics Denmark does, we prefer computing a weighted average of the weekly hours with the weights reflecting for how long each value of the weekly hours was in place. For instance, if hours were 39 for $\frac{3}{4}$ of the year and 38.5 for $\frac{1}{4}$, we use $39 \cdot \frac{3}{4} + 38.5 \cdot \frac{1}{4}$. This approach is more accurate.

Finally, we find that the current condition defining when and how the reduced hours are applied is incorrect and damaging. Recall from Chapter 1 that employees with two or more jobs may be assigned reduced hours in their main job if they put in sufficiently many hours in their second job. Reduced hours mean that the lower bound on the hours in the main job is used as the estimate. In the code though, reduced hours are tied to a meaningless comparison between the earned pensions and the maximum pensions one can earn in one job, instead of whether an employee earns pensions from a second job or not. In addition, the reduced hours are undefined in many cases and missing values are generated. In the SAS-language, a missing value is the lowest value there is, and for this reason they satisfy a key inequality that really should not have been satisfied. The end result is a criterion that does quite a bit of damage to the wages. For

completeness, in the Appendix we use counterfactual exercises to verify that the programming code we have seen and analysed really is the one that Statistics Denmark uses.

3 Puzzle 2 and the Effects of the Final Changes

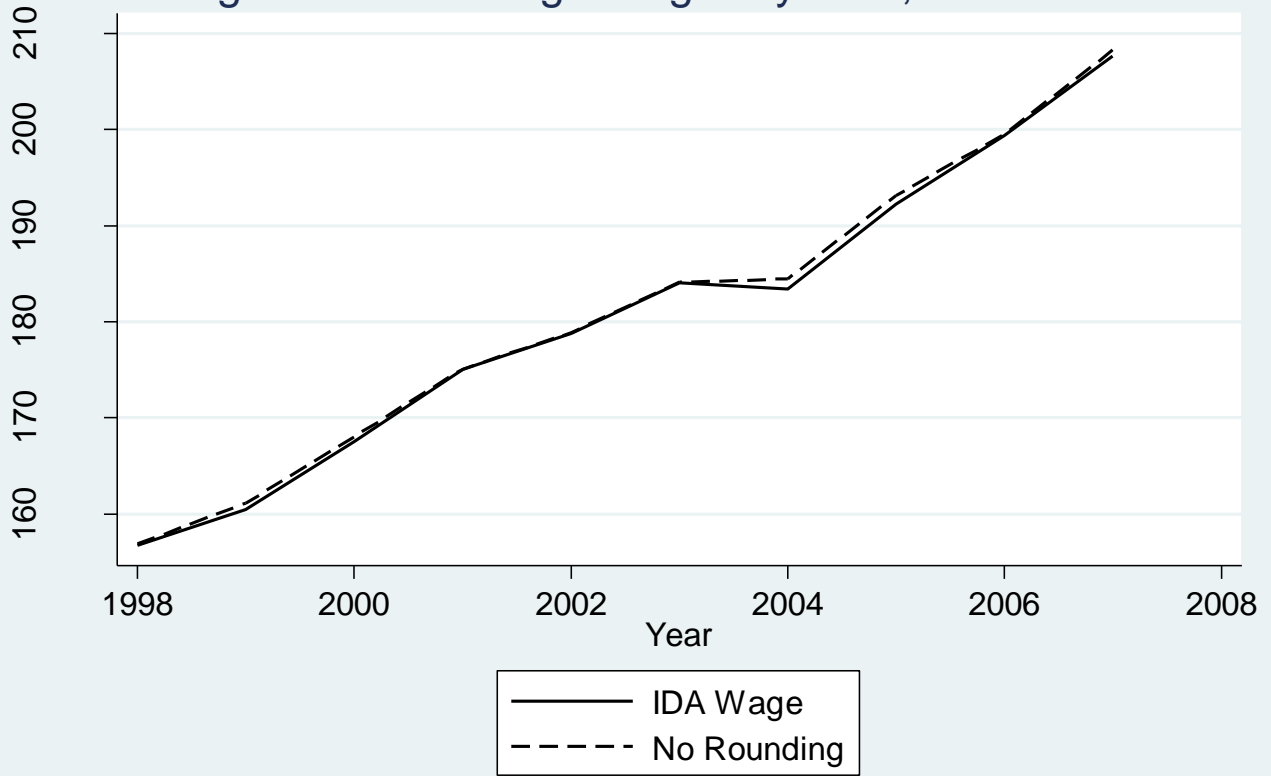
Chapter 1 documented how the average wage of full-timers fell from 2003 to 2004 while the average wage of part-timers increased slightly. Such a pattern is unusual in the Danish labour market since widespread wage cuts are rare and the wages of part-timers and full-timers most often track each other. For this reason we checked in Figures 1.4a-c whether the pattern reflected an issue with the estimation or a real development in the labour market. We found that the estimation was the culprit and referred to the pattern as “Puzzle 2”. Here we approach the puzzle head on and conjecture that the solution is to be found among the four recommended changes from the previous section. If we are right there will be no need for a more comprehensive investigation.

The first candidate explanation for the fall in the wages is the incorrect criterion for reduced hours that Statistics Denmark uses. For the 54.43% of the observations that should have reduced hours but do not (cf. Appendix), the effect is too high hours and too low wages among the group of full-timers. Similarly, among those who should not have reduced hours, 1.59% (Appendix) are allocated to the part-timer group and given reduced hours, thus increasing the wage in that group. This logic suggests that the incorrect reduced hours-criterion could have an impact on the wage levels, but it is not clear why the effect of using the wrong criterion should materialize itself so strongly in 2004 but not in other years. The second candidate explanation is rounding which can potentially change the time series properties since the estimated hours depend on the yearly full-time hours T that change every year. The rounding might go in opposite directions even if T only changes slightly, potentially causing visible effects that vary with time. On the other hand, it is not clear what effects the scaling of the unemployment and sickness periods might mean for the

wages. We can rule out any impact of using the weighted average of the weekly hours since that change concerned the years 1986-1991 and not 2003-2004.

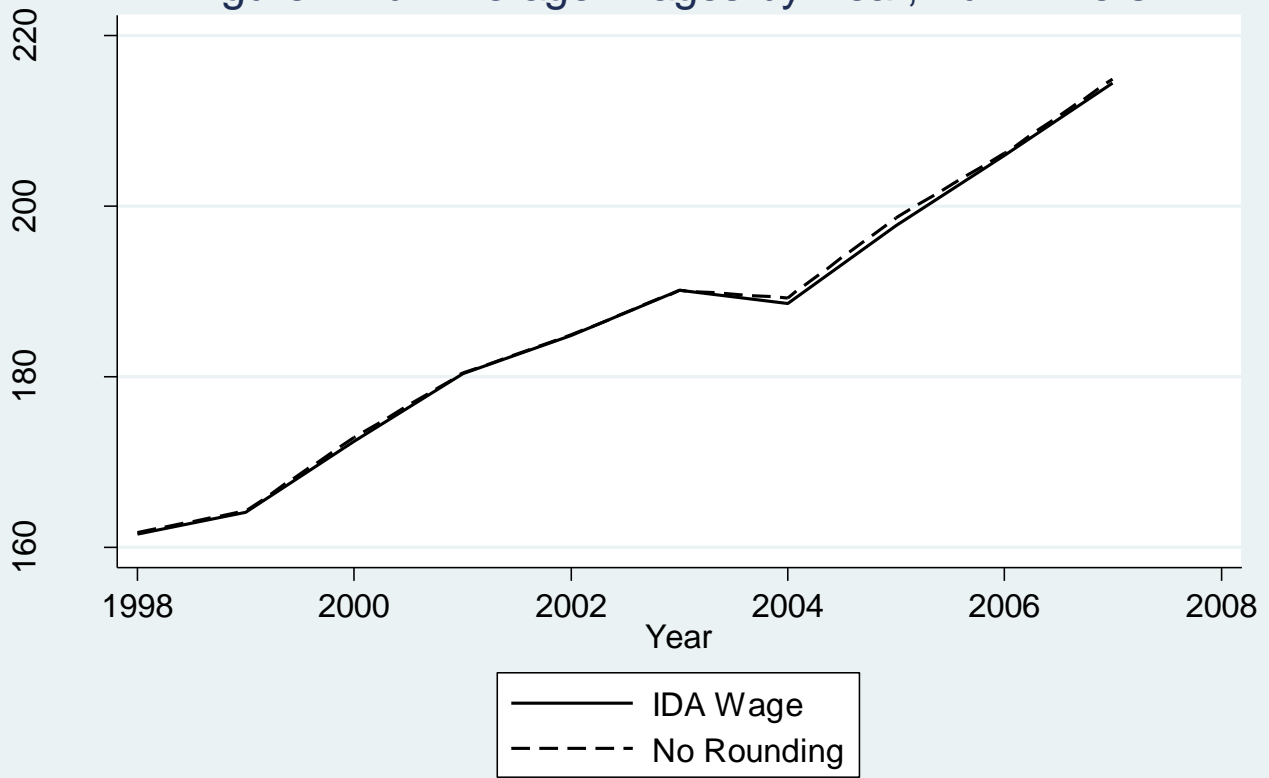
We now turn to evaluating the effects of the three relevant changes. We consider each explanation in isolation and find that removing the premature rounding fully accounts for the negative average wage growth from 2003 to 2004, but only when the average includes part-timers and full-timers alike. Indeed, we find that the average wage growth becomes flat or slightly positive when the rounding is removed, matching the growth rate of the Survey wage in Figure 1.4a in Chapter 1. However, if we split on part-timers and full-timers, then we can no longer match the growth rates of the Survey wage. Indeed, removing the rounding should remove a fall in the wage of the full-timers, whereas it increases the wage growth of the part-timers. Thus, we are not sure how much emphasis to put on the Survey wage as a benchmark in Figures 1.4a-c. Still, the fact remains that removing the rounding increases the slope from negative to positive. We conclude that the premature rounding is part of the problem but perhaps not the whole story. The effects of removing the rounding are shown in Figures 4.1a-4.1c.

Figure 4.1a: Average Wages by Year, All Workers



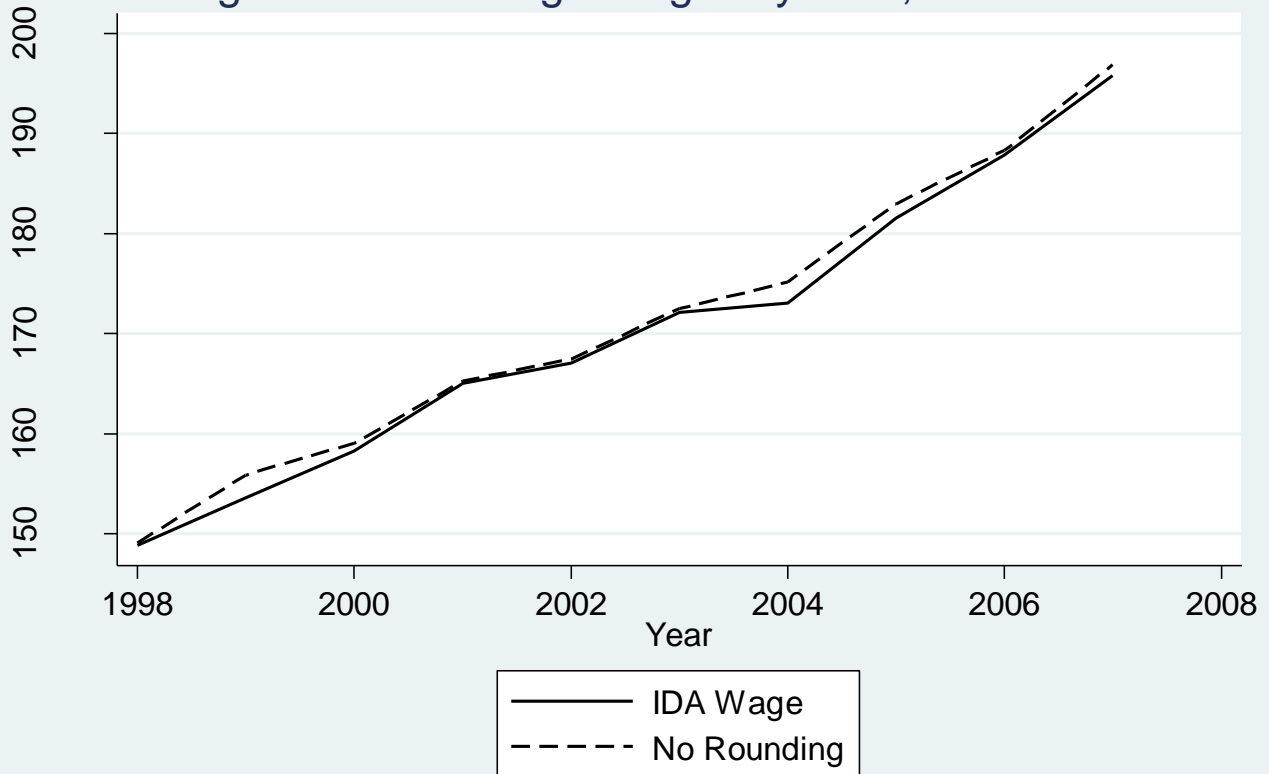
Source: Authors' calculations of wages in the Danish IDA database.

Figure 4.1b: Average Wages by Year, Full-Timers



Source: Authors' calculations of wages in the Danish IDA database.

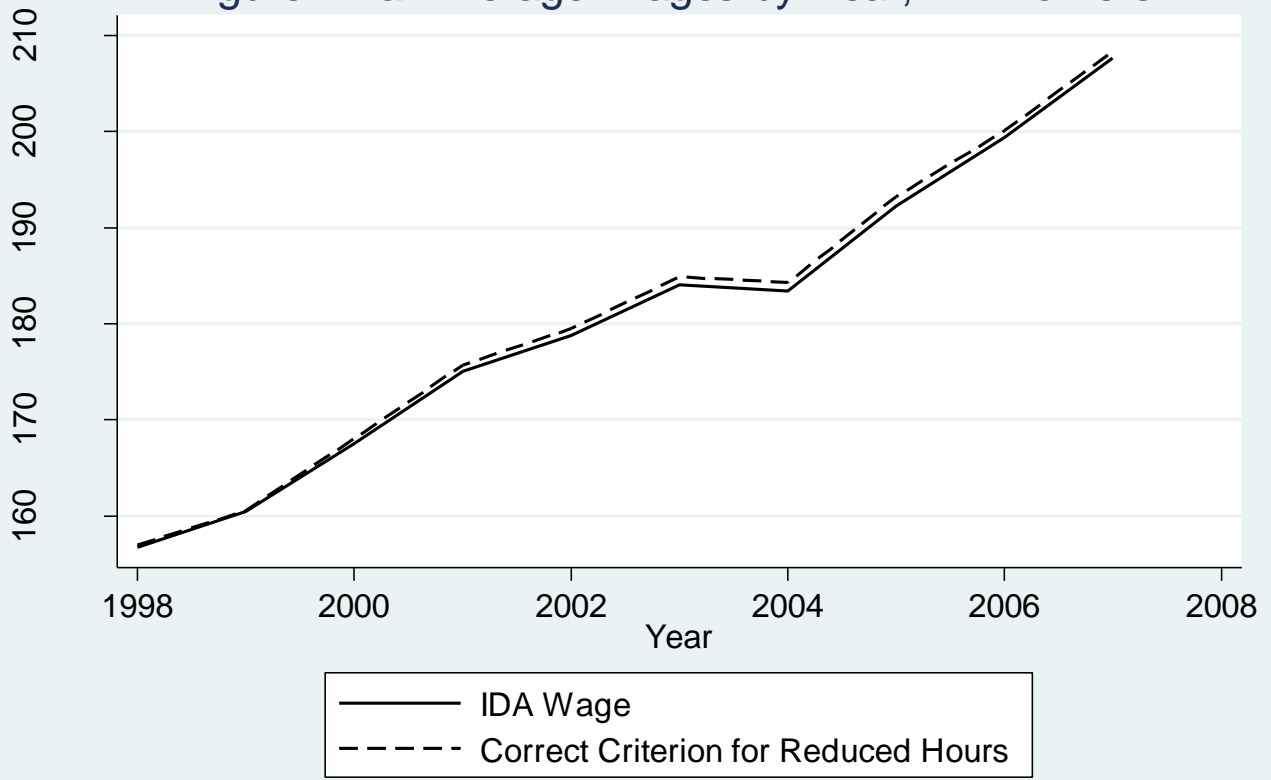
Figure 4.1c: Average Wages by Year, Part-Timers



Source: Authors' calculations of wages in the Danish IDA database.

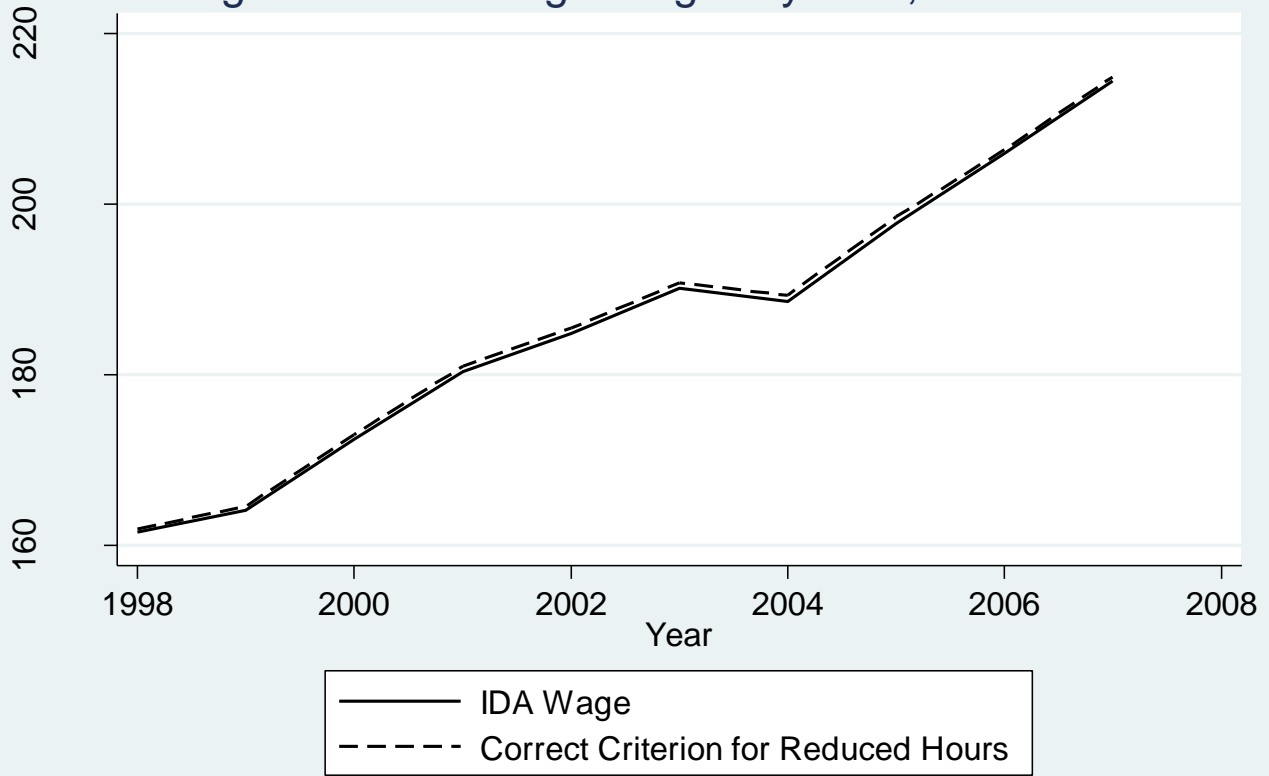
Turning to the effects of correcting the criterion that is used for identifying workers with reduced hours, we find that the growth rates of the wages do not change much. Instead, the overall wage levels rise, especially for the part-timers, cf. Figures 4.2a, 4.2b, and 4.2c.

Figure 4.2a: Average Wages by Year, All Workers



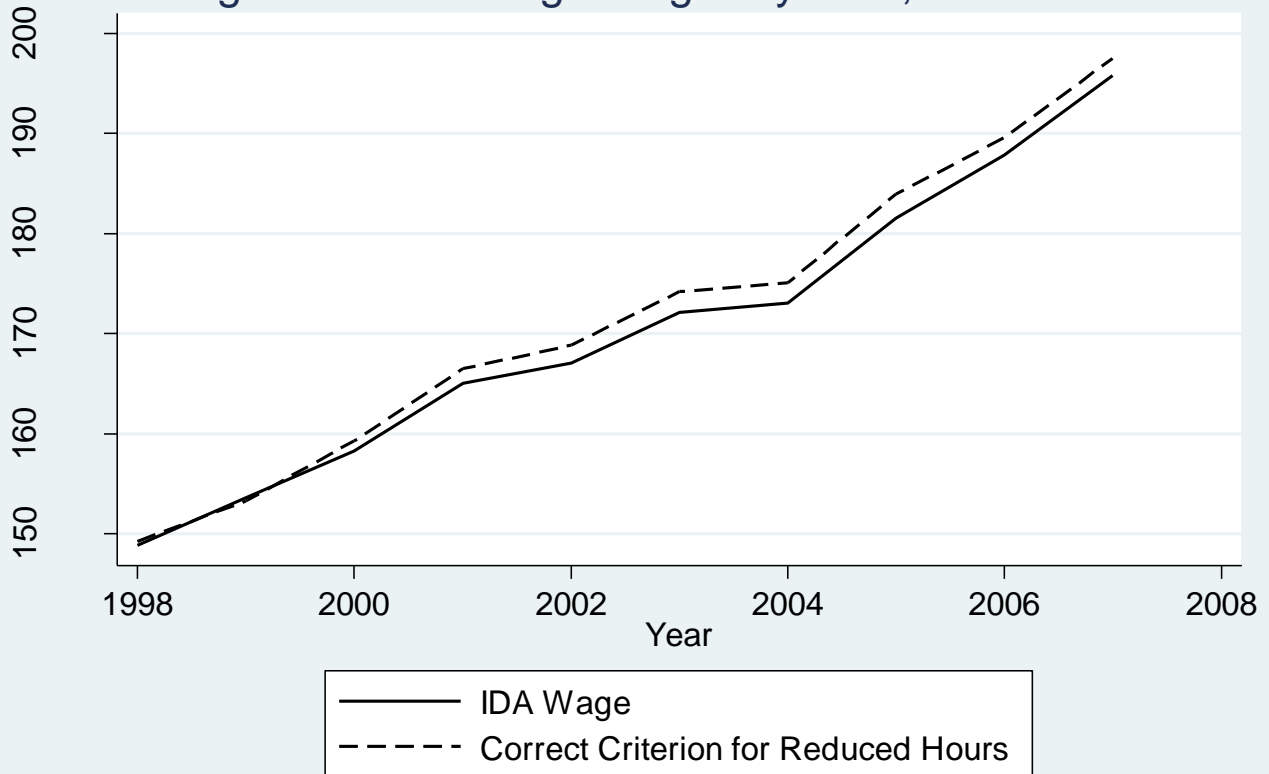
Source: Authors' calculations of wages in the Danish IDA database.

Figure 4.2b: Average Wages by Year, Full-Timers



Source: Authors' calculations of wages in the Danish IDA database.

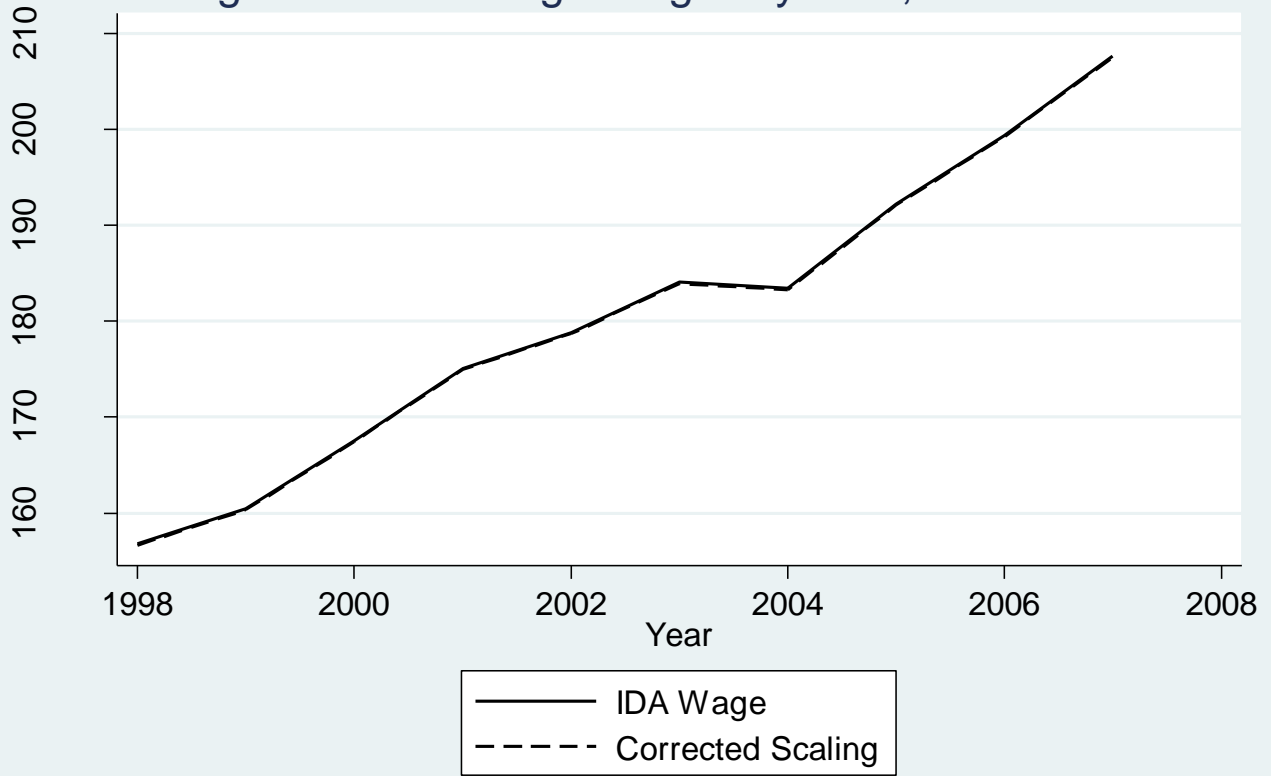
Figure 4.2c: Average Wages by Year, Part-Timers



Source: Authors' calculations of wages in the Danish IDA database.

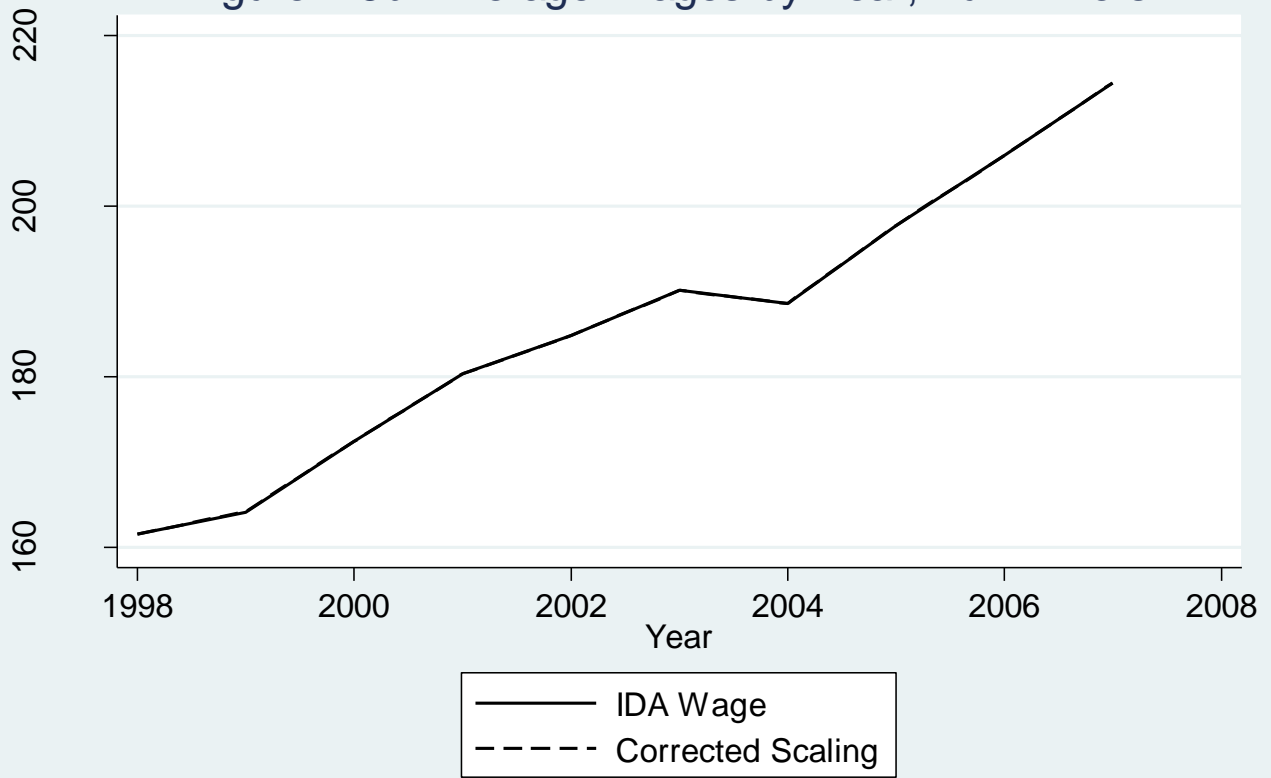
Finally, the changed scaling factor that is applied to periods with illness and unemployment has no effect whatsoever, as shown in Figures 4.3a-4.3c.

Figure 4.3a: Average Wages by Year, All Workers



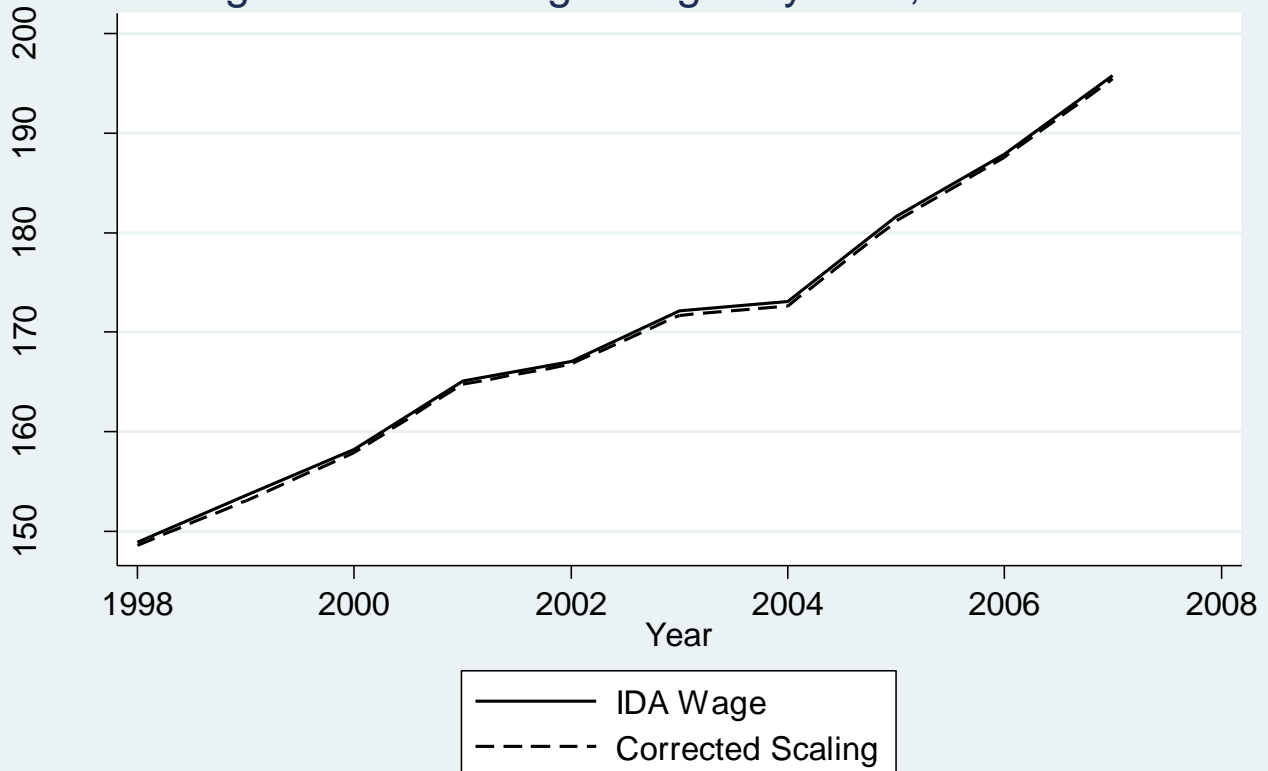
Source: Authors' calculations of wages in the Danish IDA database.

Figure 4.3b: Average Wages by Year, Full-Timers



Source: Authors' calculations of wages in the Danish IDA database.

Figure 4.3c: Average Wages by Year, Part-timers



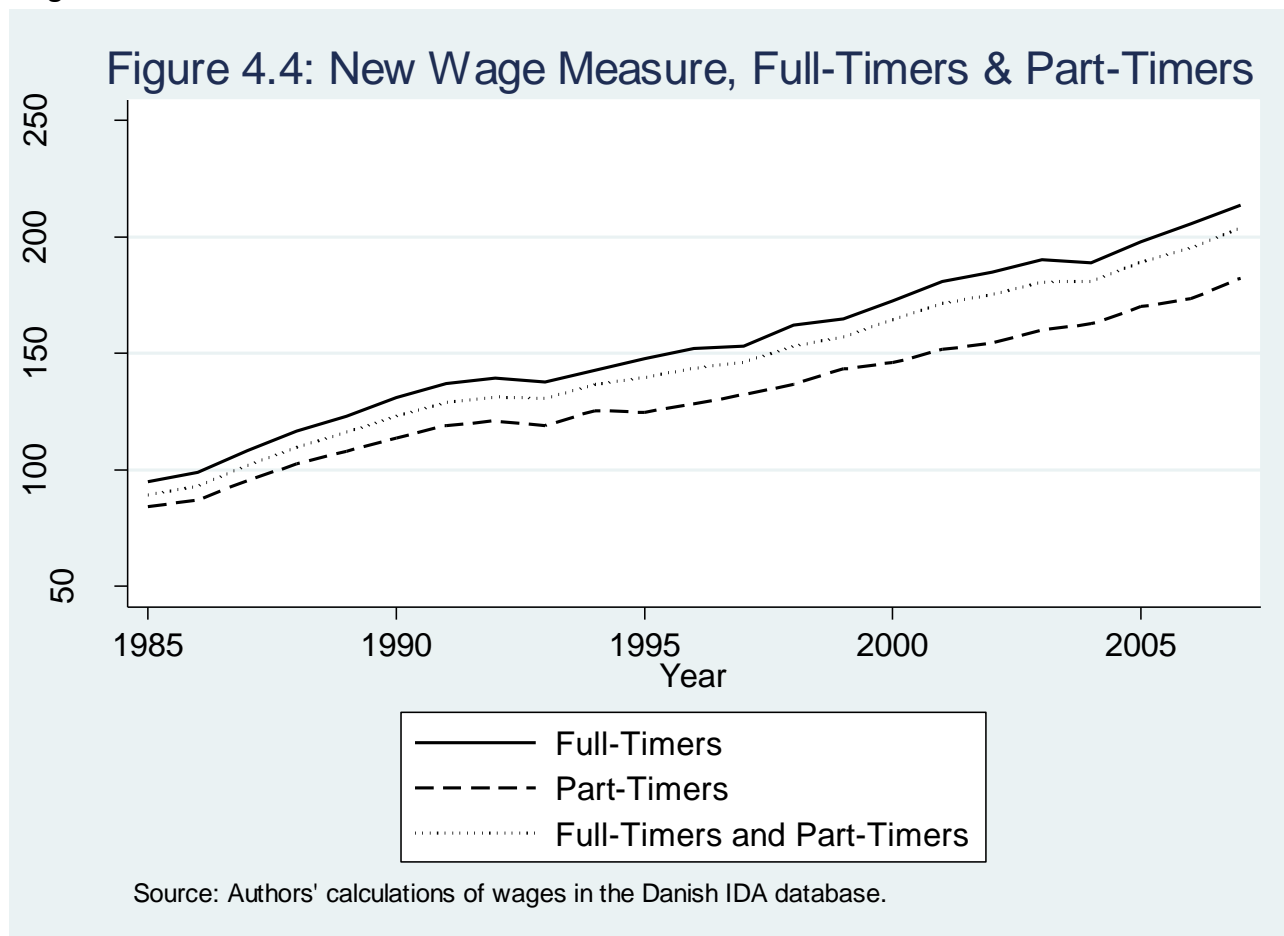
Source: Authors' calculations of wages in the Danish IDA database.

In sum, the rounding has no pronounced effect on the levels of the average wages but removing it does increase the growth rate from 2003 to 2004, especially for the part-timers. In contrast, changing the criterion for reduced hours raises the level of the average wages.

4 The New Wage Measure And the Puzzles

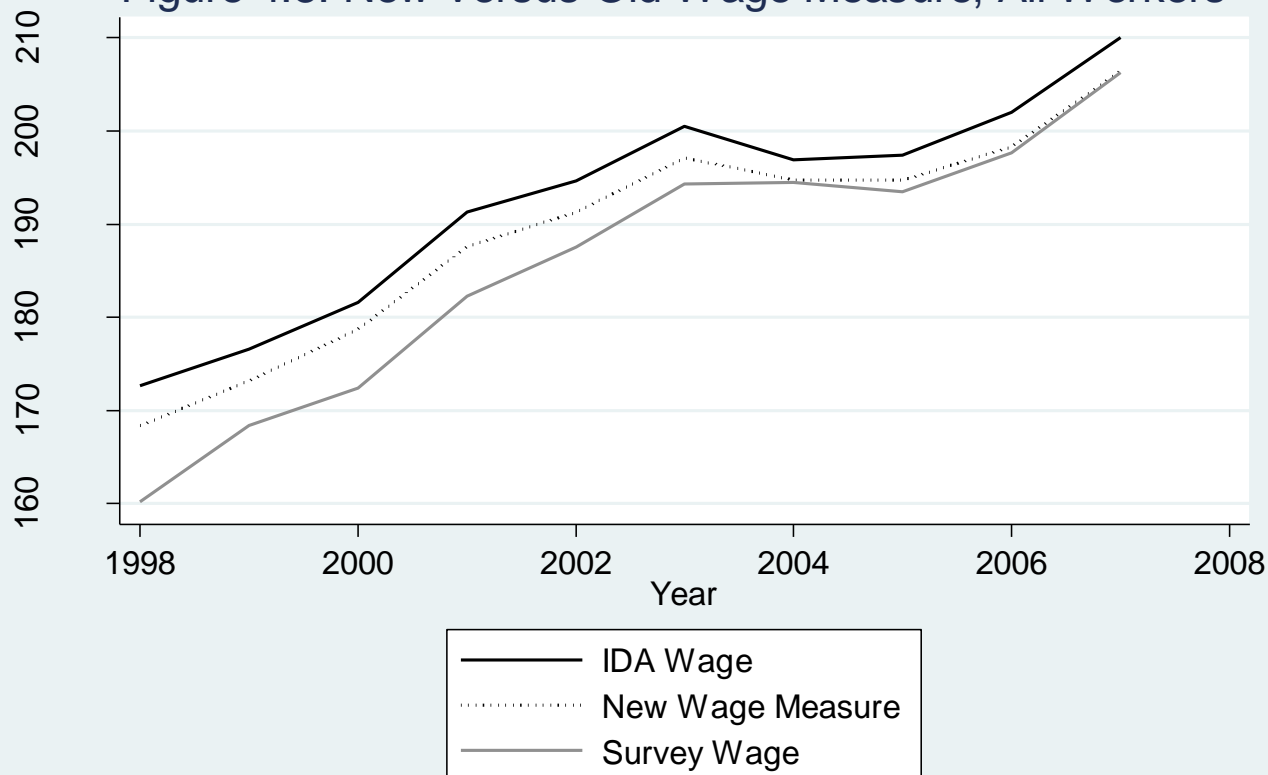
We now combine all the changes made in this and previous chapters into a new hourly wage measure. In Chapter 3 there were two changes to choose from, and we pick what we called the Bold method. Here we demonstrate the performance of our new measure in terms of how well it solves the puzzles identified in Chapter 1. The first figure replicates Figure 1.3 of Chapter 1 that showed the time series puzzles 1 and 2. The fall in the wages from 1992 to 1993 has disappeared, the part-timer wages exhibit the same relative difference to the full-timer wages before and after

1993, there is no drop in the aggregate wage from 2003 to 2004, and the wages of the full-timers stagnate instead of fall.



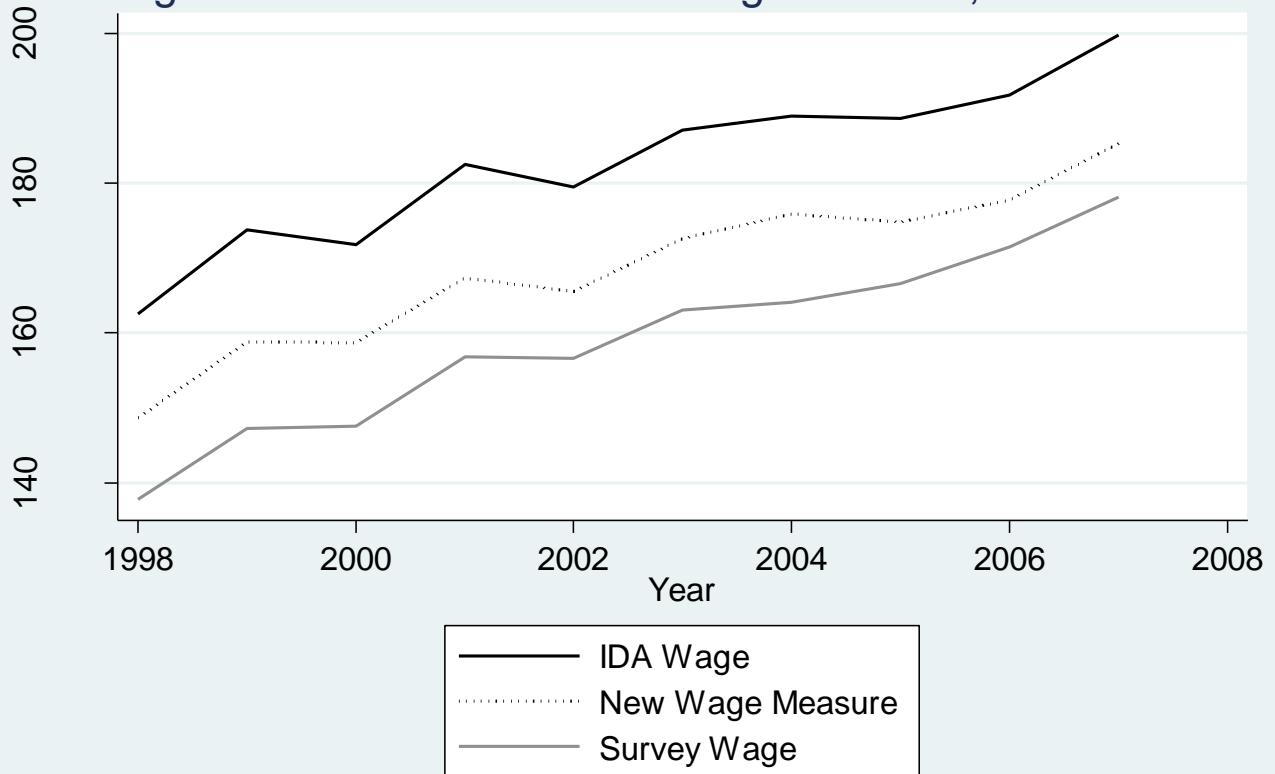
Next, we replicate Figures 3.9 and 3.10 and show how the levels of the part-timer and aggregate wages are lowered, improving considerably on Puzzle 3. We also note that seemingly there is still room for improvement.

Figure 4.5: New Versus Old Wage Measure, All Workers



Source: Authors' calculations of wages in the Danish IDA database and the Lønstatistik.

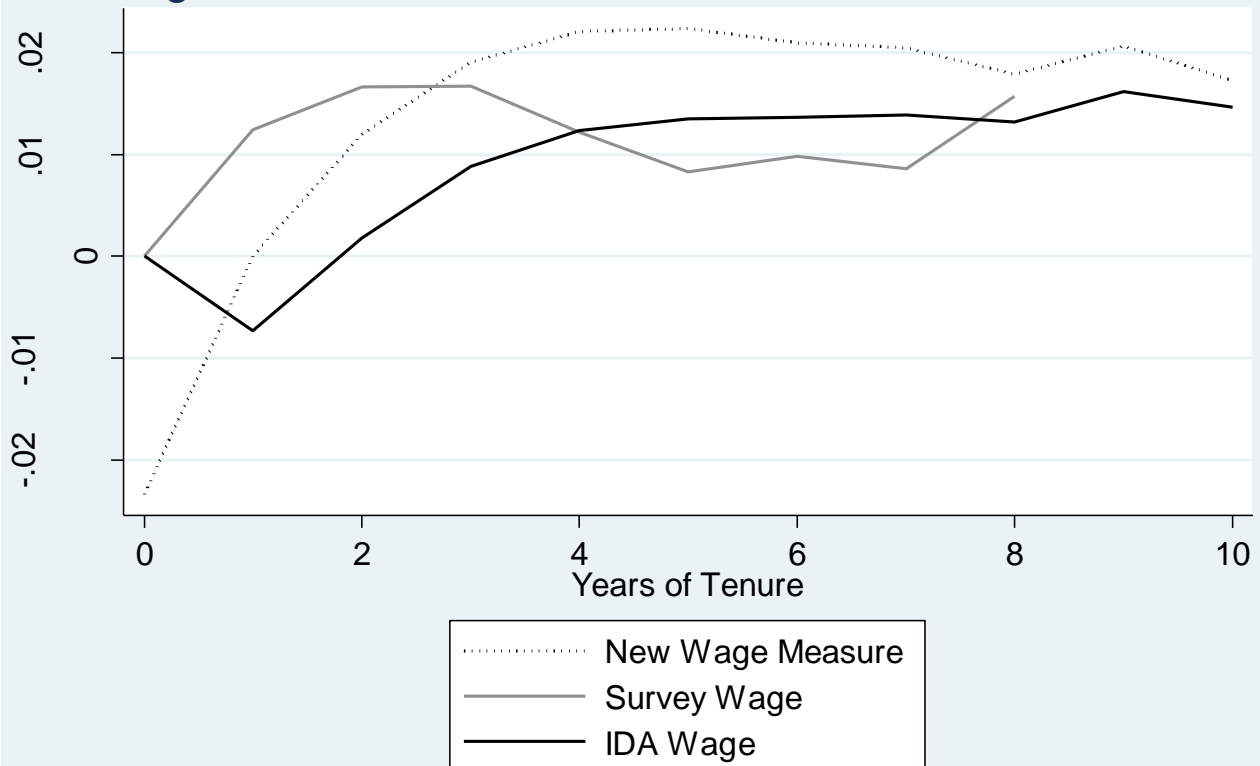
Figure 4.6: New Versus Old Wage Measure, Part-Timers



Source: Authors' calculations of wages in the Danish IDA database and the Lønstatistik.

Finally, we show how Puzzle 4, the tenure profile puzzle, has been completely solved.

Figure 4.7: New and Old Returns to Tenure, All Workers



Source: Authors' calculations of wages in the Danish IDA database and the Lønstatistik.

5 Other Improvements of the IDA Wage Measure?

Even if we have conducted an expansive study on how the IDA wage is estimated, what parameters are used, what short-comings the wage measure has, and how to solve them, we still have not investigated everything. As in all projects, we have not had unlimited resources and there are some corners that we have not been able to shed light on. Perhaps other researchers and statisticians will have an opportunity in the future to think about further improvements of the IDA wage. To that effect, we list four areas where we think the returns are highest.

- *The level of part-timer wages.* The comparison with our constructed Survey wage shows that it is probably still too high. Indeed, the discrepancy has only shrunk by a half. However, it is difficult to assess what the target discrepancy should be since the wage concepts embodied in the IDA and Survey

wages need not be completely identical, as argued in the text around Figures 3.6 and 3.11. Yet, as argued in Chapter 1, the wage concepts are not too different so we should probably not expect the difference to be as large as we see. If that is indeed the case, one possible solution is to correct the IDA wages less conservatively. As explained in Chapter 3, if the hours of part-timers vary a lot, we under-correct by assuming a minimum of variation. Instead, we could use an (ad hoc) average of the upper and lower bounds on the share of pay cycles in the top-bracket and in that way assume an intermediate amount of variation.

- *The wage comparison by industry (Figures 3.6 and 3.11 in Chapter 3).* Why are there so big differences between the Survey and IDA wages in Agriculture, Utilities, and Construction? Do the wage concepts differ more in these industries compared to in the others? Is the corrected IDA wage level still wrong for these industries?
- *The time series puzzle of the full-timers.* Figures 1.3 and 1.4a-c in Chapter 1 and Figure 4.4 of this chapter suggest that there is variation in the decline of the full-timer wage across firm size and public versus non-public sector firms. This could be explored further in order to possibly explain Puzzle 2 better.
- *Second jobs (type B-jobs).* We think that there is a mistake in how the tlonkval (relative uncertainty) variable is coded for second jobs. We have not had access to the variables needed to prove this conjecture, but it should be very easy for Statistics Denmark to verify. Indeed, we think that the agency computes tlonkval as

$$\frac{w}{t_{min}} - \frac{w}{t_{max}}$$

instead of

$$\frac{t}{t_{min}} - \frac{t}{t_{max}}.$$

Here w and t are the estimated wage and hours. As in Chapter 1, the latter expression would make sense, but not the former.

Appendix C4

We verify here that Statistics Denmark really estimates reduced hours as in the programming code that we have had access to. To this end, we need to replicate the bounds on hours perfectly. If we can do that, we can divide the earnings by the IDA hourly wage and recover the hours that were used, then compare to our bounds and see which one was used or if an average was used instead. Given that we consider only main jobs, we can infer which observations qualified for reduced hours, namely those with the lower bound. In order to test what the underlying criterion was, we can compute reduced hours according to how it is done in the programming codes that we have seen, and according to how we think they should have been calculated. Then a comparison can be made with how it is done in practice in the data.

Unfortunately we cannot replicate the bounds perfectly, at least not in all years. Recall from Chapter 1 and its Appendix that Statistics Denmark has used a variable capturing sickness allowances without documenting its values over time, which has made it necessary for us to measure sickness and unemployment differently. Fortunately, through conversations with former employees we have recovered the values of the variable for the years 1985-1991. As a result, we can replicate the estimation of Statistics Denmark almost perfectly in these years. Accordingly, we perform the verification exercise for these years only. The results are presented in Table A1.

Table A1. Comparison of the Incorrect Criterion for Reduced Hours with the Criterion Used In the Data

| Years 1985-1991 Frequencies | | Incorrect criterion | |
|--------------------------------|-------------------|---------------------|---------------|
| | | Non-reduced Hours | Reduced Hours |
| Data | Non-reduced Hours | $1.534 \cdot 10^7$ | 51917 |
| | Reduced Hours | 1 | 288939 |

There is a clear correspondence between reduced hours in the data and the incorrect criterion for reduced hours. In effect, 100% of those with reduced hours in the data would also receive reduced

hours with the incorrect criterion. Similarly, 99.7% of those with non-reduced hours in the data would also have non-reduced hours if the incorrect criterion were used. From the perspective of the incorrect criterion, 100% of those with non-reduced hours also do not have reduced hours in the data, also pointing to the incorrect criterion as the one used in the data. However, 15.23% of the observations with reduced hours according to the incorrect criterion do not have reduced hours in the data, indicating that perhaps a stricter criterion than the incorrect one was actually used. However, that impression disappears once we compare the criteria by year.

Table A2. Comparison of the Incorrect Criterion With the Criterion Used In the Data, by Year

| | | Incorrect Criterion, Reduced Hours | | | | | | |
|------|-------------------|------------------------------------|--------|--------|--------|--------|--------|--------|
| | | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 |
| Data | Non-reduced Hours | 0.65% | 0.44% | 0.42% | 90.31% | 0.43% | 0.46% | 0.48% |
| | Reduced Hours | 99.35% | 99.56% | 99.58% | 9.69% | 99.57% | 99.54% | 99.52% |

The table shows why the incorrect criterion appears as too permissive in Table A1. The 15.23% of the observations with reduced hours according to the incorrect criterion that did not have reduced hours in the data can be traced to the year 1988 in which 90.31% or 50554 observations did not have reduced hours in the data. In all other years, only insignificantly small proportions of those with reduced hours according to the incorrect criterion do not have reduced hours in the data. Moreover, a tabulation of the actual, estimated wage and the wage that would have been obtained from the lower bound on hours in 1988 reveals that 49827 of the 50554 observations had a difference within 2 DKK. This difference is so small that it must be due to the premature rounding being done differently in 1988 for a subset of the observations. In other words, the lower bound on hours must have been used also for the 49827 observations. Accordingly, the hours are reduced in the data also and there is a virtually perfect correspondence between the incorrect criterion and the data.

On the other hand, the correct criterion does not track the reduced hours in the data at all. In effect, only 14.28% of the observations that have reduced hours in the data would have had reduced hours with the correct criterion, and only 45.57% of the observations that would have had reduced hours with the correct criterion have reduced hours in the data. This is a significant misallocation of workers that has a huge effect on the wages of some individuals. Table A3 shows the numbers.

Table A3. Comparison of the Correct Criterion for Reduced Hours with the Criterion Used In the Data

| Years 1985-1991 Frequencies | | Correct criterion | |
|--------------------------------|-------------------|--------------------|---------------|
| | | Non-reduced Hours | Reduced Hours |
| Data | Non-reduced Hours | $1.535 \cdot 10^7$ | 49303 |
| | Reduced Hours | 247667 | 41273 |

References

Bagger, J., Fontaine, F., Postel-Vinay, F., and Robin, J. M.: "Tenure, Experience, Human Capital and Wages: A Tractable Equilibrium Search Model of Wage Dynamics", *American Economic Review*, forthcoming 2014.

Buhai, I. S., Portela, M. A., Teulings, C. N., Van Vuuren, A.: "Returns to Tenure or Seniority?", *Econometrica*, forthcoming 2014.

Taber, C., Vejlin, R.: "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market", Unpublished Manuscript 2014.

DØRS (2003): "Timeløn og arbejdsomfang i den enkelte ansættelse", Arbejdsnotat nr. 31, Det Økonomiske Råd, Sekretariatet 2003

Online Resources

Online (1): Statistics Denmark's online documentation of the hourly wage in the IDA database.
<http://www.dst.dk/da/Statistik/dokumentation/Times/ida-databasen/ida-ansættelser/timelon.aspx>

Online (2): Pension rates tabulated based on type of rate and compensation mode.
<http://www.virk.dk/home/myndigheder/atp-livslang-pension/offentlig-virksomhed/bidragssatser.html>

Online (3): Historical weekly, monthly, and yearly pension rates.
http://www.virk.dk/files/live/sites/virk/files/PDF-Filer_og%20Word-dokumenter/ATP_LP/Alle%20ATP-satser%20private%20virksomheder%20til%20www%20-%20opdateres%20p%C3%A5%20excell.pdf

Online (4): Weeks of vacation in 1980:
<http://www.aff.dk/Om%20fonden/Historie.aspx>

Online (5): Overview of the labour market agreements 1985-1990:
http://faos.ku.dk/pdf/temasider/ok/Overenskomstforhandlinger_paa_LO-DA_omraadet_1985-2007_OK2010.pdf

Online (6): Weekly hours changed on October 1st for some workers:
<http://www.mejerierne.dk/default.aspx?page=GetFile&name=/ProtectedFiles/Cirkulare/c-426-1987.pdf>

Online (7): Weekly hours changed on September 1st for many workers:
<http://www.mejerierne.dk/default.aspx?page=GetFile&name=/ProtectedFiles/Cirkulare/c-410-1987.pdf>

Online (8): Weekly hours changed on September 1st for many workers:

<http://www.dors.dk/graphics/Synkron-Library/Publikationer/Rapporter/Gamle%20rapporter%201962-1996/1987%20DECEMBER.pdf>

Online (9): Extra days off ("Feriefridage"):

<https://www.danskerhverv.dk/Raadgivning/Overenskomst/feriefridage/Sider/Feriefridage.aspx>

Online (10): Extra days off ("Feriefridage"):

<http://di.dk/personale/personalejura/ferieogfridage/feriefridage/pages/feriefridage.aspx>

Online (11): Extra days off ("Feriefridage"):

<https://www.hk.dk/RaadOgStoette/Emner/Ferie/Handel/5-ting-du-skal-vid-om-feriefridage-den-sjette-ferieuge>

Online (12): Extra days off ("Feriefridage"):

<http://www.3f.dk/overenskomster/pshr/horesta/kapitler/kapitel-6---feriedage-og-ferie/24-feriefridage>

Online (13): Extra days off ("Feriefridage"):

<http://www.danskmatal.dk/Nyheder%20og%20presse/OK2014/Historiske%20OK-resultater.aspx>

Online (14): Old calendars with legal holidays indicated:

<http://www.rmadsen.dk/kal>

Sources from Statistics Denmark

Statistics Denmark: "Hourly Wages and Working Hours in each Job", Arbejdsnotat 31, 1991.

Statistics Denmark: "Udviklingen i IDA's timeløn mellem 1992 og 1993", un-dated and anonymous memo.

Statistics Denmark: A sample of old program codes, un-dated and confidential.

Statistics Denmark: An Excel working file, un-dated and confidential.

Laws and Promulgations

| Law No. | Type of Law | Related Promulgations | Topic |
|-------------------|-------------|-----------------------|------------------------|
| 1964-03-07 nr 46 | Lov, Main | BKG 1964-03-07 nr 47 | Labour market pensions |
| 1965-05-26 nr 180 | Lov, Change | | Labour market pensions |
| 1972-02-09 nr 34 | Lov, Change | | Labour market pensions |
| 1977-06-02 nr 231 | Lov, Change | | Labour market pensions |
| 1978-04-12 nr 159 | Lov, Change | | Labour market pensions |
| 1980-05-30 nr 216 | Lov, Change | BKG 1981-??-?? nr 492 | Labour market pensions |
| 1982-04-28 nr 169 | Lov, Change | | Labour market pensions |
| 1982-05-28 nr 225 | Lov, Change | | Labour market pensions |

| | | | |
|--------------------|------------------------|-----------------------|------------------------|
| 1985-03-31 nr 123 | Lov, Main | | Working hours and pay |
| 1987-06-10 nr 367 | Lov, Change | BKG 1987-??-?? nr 695 | Labour market pensions |
| 1987-12-23 nr 878 | Lov, Change | | Labour market pensions |
| 1987-12-23 nr 882 | Lov, Change | | Labour market pensions |
| 1990-05-02 nr 267 | Lov, Change | | Labour market pensions |
| 1991-12-21 nr 891 | Lov, Change | | Labour market pensions |
| 1992-05-20 nr 372 | Lov, Change | BKG 1992-06-23 nr 559 | Labour market pensions |
| 1992-12-23 nr 1090 | Lov, Change | | Labour market pensions |
| 1992-08-06 nr 693 | LBKG, Law-Promulgation | BKG 1992-09-29 nr 822 | Labour market pensions |
| 1993-06-30 nr 432 | Lov, Change | | Labour market pensions |
| 2013-02-22 nr 202 | LBKG, Law-Promulgation | | Vacation |

Economics Working Papers

- 2014-17: Kristine Vasiljeva: On the importance of macroeconomic factors for the foreign student's decision to stay in the host country
- 2014-18: Ritwik Banerjee: On the Interpretation of Bribery in a Laboratory Corruption Game: Moral Frames and Social Norms
- 2014-19: Ritwik Banerjee and Nabanita Datta Gupta: Awareness programs and change in taste-based caste prejudice
- 2014-20: Jos Jansen and Andreas Pollak: Strategic Disclosure of Demand Information by Duopolists: Theory and Experiment
- 2014-21: Wenjing Wang: Do specialists exit the firm outsourcing its R&D?
- 2014-22: Jannie H. G. Kristoffersen, Morten Visby Krægpøth, Helena Skyt Nielsen and Marianne Simonsen: Disruptive School Peers and Student Outcomes
- 2014-23: Erik Strøjer Madsen and Yanqing Wu: Globalization of Brewing and Economies of Scale
- 2014-24: Niels-Hugo Blunch and Nabanita Datta Gupta: Social Networks and Health Knowledge in India: Who You Know or Who You Are?
- 2014-25: Louise Voldby Beuchert and Anne Brink Nandrup: The Danish National Tests - A Practical Guide
- 2015-01: Ritwik Banerjee, Tushi Baul, and Tanya Rosenblat: On Self Selection of the Corrupt into the Public Sector
- 2015-02: Torben M. Andersen: The Nordic welfare model and welfare services - Can we maintain acceptable standards?
- 2015-03: Nina Neubecker, Marcel Smolka and Anne Steinbacher: Networks and Selection in International Migration to Spain
- 2015-04: Sylvanus Kwaku Afesorgbor and Kaleb Girma Abreha: Preferential Market Access, Foreign Aid and Economic Development
- 2015-05: Ritwik Banerjee: Corruption, Norm Violation and Decay in Social Capital
- 2015-06: Christian Gjødesen Lund and Rune Vejlin: Documenting and Improving the Hourly Wage Measure in the Danish IDA Database

Economics Working Papers

- 2014-17: Kristine Vasiljeva: On the importance of macroeconomic factors for the foreign student's decision to stay in the host country
- 2014-18: Ritwik Banerjee: On the Interpretation of Bribery in a Laboratory Corruption Game: Moral Frames and Social Norms
- 2014-19: Ritwik Banerjee and Nabanita Datta Gupta: Awareness programs and change in taste-based caste prejudice
- 2014-20: Jos Jansen and Andreas Pollak: Strategic Disclosure of Demand Information by Duopolists: Theory and Experiment
- 2014-21: Wenjing Wang: Do specialists exit the firm outsourcing its R&D?
- 2014-22: Jannie H. G. Kristoffersen, Morten Visby Krægpøth, Helena Skyt Nielsen and Marianne Simonsen: Disruptive School Peers and Student Outcomes
- 2014-23: Erik Strøjer Madsen and Yanqing Wu: Globalization of Brewing and Economies of Scale
- 2014-24: Niels-Hugo Blunch and Nabanita Datta Gupta: Social Networks and Health Knowledge in India: Who You Know or Who You Are?
- 2014-25: Louise Voldby Beuchert and Anne Brink Nandrup: The Danish National Tests - A Practical Guide
- 2015-01: Ritwik Banerjee, Tushi Baul, and Tanya Rosenblat: On Self Selection of the Corrupt into the Public Sector
- 2015-02: Torben M. Andersen: The Nordic welfare model and welfare services - Can we maintain acceptable standards?
- 2015-03: Nina Neubecker, Marcel Smolka and Anne Steinbacher: Networks and Selection in International Migration to Spain
- 2015-04: Sylvanus Kwaku Afesorgbor and Kaleb Girma Abreha: Preferential Market Access, Foreign Aid and Economic Development
- 2015-05: Ritwik Banerjee: Corruption, Norm Violation and Decay in Social Capital
- 2015-06: Christian Gjødesen Lund and Rune Vejlin: Documenting and Improving the Hourly Wage Measure in the Danish IDA Database